Journal of Probability and Statistics

Actuarial and Financial Risks: Models, Statistical Inference, and Case Studies

Guest Editors: Ričardas Zitikis, Edward Furman, Abdelhakim Necir, Johanna Nešlehová, and Madan L. Puri

Actuarial and Financial Risks: Models, Statistical Inference, and Case Studies

Journal of Probability and Statistics

Actuarial and Financial Risks: Models, Statistical Inference, and Case Studies

Guest Editors: Ričardas Zitikis, Edward Furman, Abdelhakim Necir, Johanna Nešlehová, and Madan L. Puri

Copyright © 2010 Hindawi Publishing Corporation. All rights reserved.

This is an issue published in volume 2010 of "Journal of Probability and Statistics." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Mohammad F. Al-Saleh, Jordan
Vo V. Anh, Australia
Zhidong Bai, China
Ishwar Basawa, USA
Shein-chung Chow, USA
Murray Clayton, USA
Dennis Dean Cox, USA
Junbin B. Gao, Australia
Arjun K. Gupta, USA
Tomasz J. Kozubowski, USA
Debasis Kundu, India
Michael Lavine, USA
Nikolaos E. Limnios, France
Chunsheng Ma, USA
Eric Moulines, France
Hung T. Nguyen, USA
Madan L. Puri, USA
José María Sarabia, Spain
Lawrence A. Shepp, USA
H. P. Singh, India
Man Lai Tang, Hong Kong
Robert J. Tempelman, USA
A. Thavaneswaran, Canada
Peter van der Heijden, The Netherlands
Rongling Wu, USA
Kelvin K. W. Yau, Hong Kong
Philip L. H. Yu, Hong Kong
Daniel Zelterman, USA
Ricardas Zitikis, Canada

Contents

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 392498, 3 pages
doi:10.1155/2010/392498

Editorial

Actuarial and Financial Risks: Models, Statistical Inference, and Case Studies

Ričardas Zitikis,[1] Edward Furman,[2] Abdelhakim Necir,[3] Johanna Nešlehová,[4] and Madan L. Puri[5, 6]

[1] *Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON, Canada N6A 3K7*
[2] *Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3*
[3] *Laboratory of Applied Mathematics, Mohamed Khider University, Biskra 07000, Algeria*
[4] *Department of Mathematics and Statistics, McGill University, Motreal, QC, Canada H3A 2K6*
[5] *Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia*
[6] *Department of Mathematics, Indiana University, Bloomington, Indiana 47405, USA*

Correspondence should be addressed to Ričardas Zitikis, zitikis@stats.uwo.ca

Received 10 June 2010; Accepted 10 June 2010

Copyright © 2010 Ričardas Zitikis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding actuarial and financial risks poses major challenges. The need for reliable approaches to risk assessment is particularly acute in the present context of highly uncertain financial markets. New regulatory guidelines such as the Basel II Accord for banking and Solvency II for insurance are being implemented in many parts of the world. Regulators in various countries are adopting risk-based approaches to the supervision of financial institutions.

Many researchers in a variety of areas have been dealing with nontrivial and highly multifaceted problems in an attempt to answer seemingly plain questions such as how to assess and quantify risks (Crouhy, Galai, and Mark [1]). The present issue of the Journal of Probability and Statistics provides a glimpse of how challenging such problems are, both philosophically and mathematically, through a collection of papers that cover a large spectrum of applied and theoretical problems.

A number of ideas concerning measuring risks stem from the economic theory and in particular from the classical utility theory (Neumann and Morgenstern [2]) as well as from the prospect theory (Kahneman and Tversky [3]), which were subsequently developed into the anticipated, also known as rank-dependent or generalized expected, utility theory (Quiggin [4]), and most recently into a ground-breaking theory of choice under uncertainty that allows for the presence of catastrophic risks (Chichilnisky [5]).

This special Journal of Probability and Statistics issue offers many articles based on such economic theories and their extensions. G. Chichilnisky develops foundations for dealing with catastrophic risks, called "black swans", which require tools beyond the classical σ-additive probability theory. M. Finke, E. Belasco, and S. Huston review household property risk management theory in order to compare optimal risk retention to conventional practice. Aided with ideas of behavioral economics and finance, H. Takahashi investigates the forecast accuracy of fundamental values in financial markets and clarifies issues such as price fluctuations. F. Greselin, L. Pasquazzi, and R. Zitikis develop statistical inference for Zenga's index of economic inequality, whose construction brings to mind the relative nature of notions such as small and large, poor and rich.

To make us aware of the scope and complexity of the problem, several authors have contributed papers tackling risks within and beyond the financial sector. G. Chichilnisky and P. Eisenberger have written a far-reaching article on asteroid risks. They convince us about the critical importance of these risks, for which very little research has been carried out, and they provide an interesting methodology for comparing asteroid risks with the risks of climate change to make better decisions about research and development. For further information on this and other related topics, we refer to the web site http://www.chichilnisky.com/.

W. J. Braun, B. L. Jones, J. S. W. Lee, D. G. Woolford, and M. Wotton examine the risk assessment of forest fires in order to generate burn probability risk maps, concentrating on the district municipality of Muskoka in Ontario, Canada, as an illustrative example. D. L. K. Hoag and J. Parsons discuss their new program, Risk Navigator SRM, which greatly expands the number of managers that can address risks in agriculture. The program lays down solid foundations and provides state-of-the-art practical support tools, including a web site (http://www.risknavigatorsrm.com/), a risk management simulator (http://www.agsurvivor.com/), and a book that puts it all together (Hoag [6]).

Specific insurance risk-related problems are tackled by a number of authors. P. Gaillardetz develops an elaborate evaluation approach to equity-linked insurance products under stochastic interest rates. O. Furman and E. Furman propose layer-based counterparts of a number of risk measures and investigate their properties and analytic tractability, especially within the framework of exponential dispersion models.

One of the basic measures of risk is the so-called value-at-risk, which amounts to a high quantile from the statistical point of view. It is arguably one of the most challenging measures to estimate and to work with in practice. This risk measure was advocated by the Basel Committee on Banking Supervision (http://www.bis.org/bcbs/) and implemented worldwide. C. Gouriéroux and J. Jasiak suggest and develop a novel nonparametric method for estimating the conditional value-at-risk and illustrate its performance on real-life portfolio returns. G. Dmitrasinovic-Vidovic, A. Lari-Lavassani, X. Li, and A. Ware explore the conditional capital-at-risk measure in the context of portfolio optimization and offer optimal strategies.

This special Journal of Probability and Statistics issue also includes papers that develop statistical inference for distortion risk measures and related quantities in the case of heavy-tailed distributions. A. Necir and D. Meraghni deal with the estimation of L-functionals, which are generalizations of the distortion risk measure, and which naturally arise in the aforementioned anticipated utility theory. A. Necir, A. Rassoul, and D. Meraghni develop a theory for estimating the renewal function of interoccurrence times of heavy-tailed risks. A. Necir, A. Rassoul, and R. Zitikis introduce a new estimator of the conditional tail expectation, which is one of the most important examples of the distortion risk measure and demonstrate the performance of the new estimator within the framework of heavy-tailed risks.

We, the editors of this special issue, most sincerely thank three groups of people, without whom this special issue would not have reached its fruition: first and foremost, the authors who have shared with us their research achievements; second, the helpful and efficient Hindawi Publishing Corporation staff; third, the professional and diligent referees whose efforts resulted in useful feedback incorporated into often several rounds of revisions of the herein published papers. We are also grateful to various granting agencies (the Actuarial Foundation, the Actuarial Education and Research Fund, the Society of Actuaries Committee on Knowledge Extension Research, and the Natural Sciences and Engineering Research Council of Canada) for their support of our research.

Ričardas Zitikis
Edward Furman
Abdelhakim Necir
Johanna Nešlehová
Madan L. Puri

References

[1] M. Crouhy, D. Galai, and R. Mark, *The Essentials of Risk Management*, McGraw-Hill, NY, USA, 2006.
[2] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, USA, 1944.
[3] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, pp. 263–291, 1979.
[4] J. Quiggin, *Generalized Expected Utility Theory*, Kluwer Academic, Dodrecht, The Netherlands, 1993.
[5] G. Chichilnisky, "An axiomatic approach to choice under uncertainty with catastrophic risks," *Resource and Energy Economics*, vol. 22, no. 3, pp. 221–231, 2000.
[6] D. L. Hoag, *Applied Risk Management in Agriculture*, CRC Press, Boca Raton, Fla, USA, 2009.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 823018, 26 pages
doi:10.1155/2010/823018

Research Article

Forest Fire Risk Assessment: An Illustrative Example from Ontario, Canada

W. John Braun,[1] Bruce L. Jones,[1] Jonathan S. W. Lee,[1] Douglas G. Woolford,[2] and B. Mike Wotton[3]

[1] *Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario, ON, Canada N6A 5B7*
[2] *Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada N2L 3C5*
[3] *Faculty of Forestry, The University of Toronto, Toronto, ON, Canada M5S 3B3*

Correspondence should be addressed to W. John Braun, braun@stats.uwo.ca

Received 2 October 2009; Revised 6 March 2010; Accepted 27 April 2010

Academic Editor: Ricardas Zitikis

Copyright © 2010 W. John Braun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents an analysis of ignition and burn risk due to wildfire in a region of Ontario, Canada using a methodology which is applicable to the entire boreal forest region. A generalized additive model was employed to obtain ignition risk probabilities and a burn probability map using only historic ignition and fire area data. Constructing fire shapes according to an accurate physical model for fire spread, using a fuel map and realistic weather scenarios is possible with the Prometheus fire growth simulation model. Thus, we applied the Burn-P3 implementation of Prometheus to construct a more accurate burn probability map. The fuel map for the study region was verified and corrected. Burn-P3 simulations were run under the settings (related to weather) recommended in the software documentation and were found to be fairly robust to errors in the fuel map, but simulated fire sizes were substantially larger than those observed in the historic record. By adjusting the input parameters to reflect suppression effects, we obtained a model which gives more appropriate fire sizes. The resulting burn probability map suggests that risk of fire in the study area is much lower than what is predicted by Burn-P3 under its recommended settings.

1. Introduction

Fire is a naturally occurring phenomenon on the forested landscape. In Canada's boreal forest region, it plays an important ecological role. However, it also poses threats to human safety and can cause tremendous damage to timber resources and other economic assets.

Wildfires have recently devastated parts of British Columbia, California, and several other locations in North America, Europe, and Australia. The economic losses in terms of suppression costs and property damage have been staggering, not to mention the tragic loss of human life. Many of these fires have taken place at the wildland-urban interface—predominantly natural areas which are increasingly being encroached upon by human habitation. As the population increases in these areas, there would appear to be potential for increased risk of economic and human loss.

A wildland-urban interface is defined as "any area where industrial or agricultural installations, recreational developments, or homes are mingled with natural, flammable vegetation" [1]. The Province of Ontario has several areas which could be classified as wildland-urban interface. These areas include the Lake of the Woods region, the Thunder Bay region, the region surrounding Sault St. Marie, and North Bay among others. One of the most significant of these is the District of Muskoka which is a popular recreational area. This district, located in Southern Ontario (Figure 1), is commonly referred to as "cottage country". It spans 6,475 square kilometers and contains over 100,000 seasonal properties or cottages. Many of these properties are nestled in forested areas, which make up most of the region. This concentration of values is of particular interest to the Canadian insurance industry due to the risk of claims from damage caused by wildfire.

Unlike British Columbia and California where topography plays a major role in the rate of spread of wildfire, Ontario is relatively flat but is dominated geographically by the Boreal and Taiga forests, where some of the largest fires in Canada have burned [2]. The Boreal forest has a large percentage of coniferous trees which are susceptible to high intensity crown wildfires. The Muskoka region is on the southern edge of the Boreal forest, and thus there is potential for substantial property damage from fires originating further north.

We are focusing on the Muskoka region to provide an illustration of how the tools that have been developed by the forest management community can be applied to assess fire risk. The methods described here can be adapted easily to other wildland-urban interface locations. The Muskoka area presents some technical challenges which do not exist to the same degree in most other wildland-urban interface settings.

Although there have not yet been substantial losses due to wildfire in the Muskoka area, it is important to assess the risk because of what is being observed elsewhere (e.g., British Columbia and California) and because of possible climate change effects which could ultimately lead to increased fire activity across Canada.

Wildfires usually start from point ignitions, either by people or by lightning, and if not detected immediately, they can spread rapidly under appropriate weather conditions. Approximately half of the forest fires in Canada are ignited by lightning. Such fires account for approximately 80 percent of area burned [3].

The spread of a wildfire in a particular area depends on many factors but most importantly, it is influenced by local weather, vegetation, and geography [2]. Of these three factors, the geographical features remain static, while vegetation changes gradually over time. In addition, changes in human land use patterns, such as industrial forestry, or urban expansion can lead to changes in vegetation. Weather is the most dynamic factor affecting fire risk. The unpredictable nature of weather makes modelling forest fire spread a difficult task. Nonetheless, the risk of wildfire in a region can be estimated using the methodology described in this paper.

In Canada, the fire season can last from early April through October each year. During this period, the probability of fire ignition and fire spread potential changes depending on the time of year, primarily influenced by seasonal weather patterns. Each year an average of 2.5 million hectares are burned by 8,500 individual wildfires.

Most regions which are within the Boreal and Taiga zones have very accurate and up-to-date fuel information because the provincial fire management agencies maintain these records rigorously. The forest resource inventory information in our study area, and hence the fuel map which is based upon it, is not updated as frequently by the Ontario Ministry of

Figure 1: Location of the District of Muskoka within the Province of Ontario.

Natural Resources in this region because there is a higher proportion of private land under municipal fire protection agreements with the province and relatively little area under forest management planning. Thus, it was necessary for us to validate the fuel map by doing a field survey. To apply the methodology in other instances would be straightforward, not requiring this kind of fieldwork.

The remainder of this paper will proceed as follows. The next section provides a description of the study area and the fire data for that region. Section 3 contains results of an ignition risk assessment which uses historic fire data only. This section also contains a crude burn risk assessment.

In Section 4, we briefly describe the Prometheus fire growth model [4] and how it is used in the Burn-P3 simulator [5] to generate a burn probability map. This section also provides a description of the required data inputs and the procedure that was used to obtain and verify this data. In Section 5, the results of the analysis are presented along with a summary of the limitations of this study.

2. The Data and Study Area

2.1. Study Area

Of the properties in the Muskoka District, the most expensive are concentrated along the shores of the three major lakes: Lake Joseph, Lake Muskoka, and Lake Rosseau. A 25×35 km rectangular study area that encompasses a large portion of these lakes was selected (Figure 2) for the our study. In order to reduce possible biases near the boundaries of this region, we also considered a 5-km wide "buffer" zone which surrounds the study area. Fires originating in this zone could spread into the study area, and this possibility needs to be accounted for.

2.2. Description of Historic Fire Data

Fire data for over 12,200 fires from 1980 through 2007 were obtained for a region encompassing the study area. For each fire, a number of covariates were recorded including

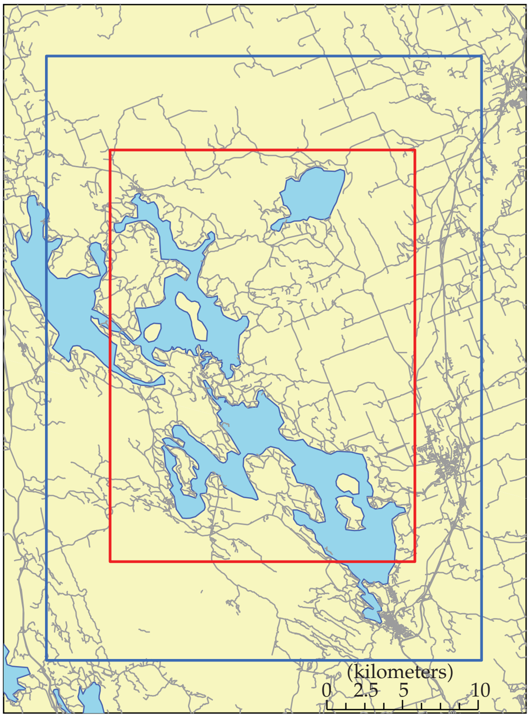


Figure 2: Map illustrating the 25 × 35 km study area which is enclosed in the red box as well as the buffer zone used in the Burn-P3 modelling denoted by the blue box.

the date, ignition location, and final area. Figure 3 shows an estimate of the density of the natural log-transformed fire sizes of escaped fires from this dataset. Here, we use the Ontario Ministry of Natural Resources definition of an escaped fire: any fire where final area exceeds 4 hectares. The fuel composition, weather conditions, and fire suppression capabilities for this region are relatively homogeneous and hence are representative of our smaller study area. Within this dataset, 319 fires were located in the study area. Figures 4 and 5 show locations of human-caused and lightning-caused ignitions

3. Ignition and Burn Probability Modelling Using Generalized Additive Models

3.1. Ignition Modelling

Brillinger et al. [6] provide a method for assessing fire risk in a region using generalized additive models. Their technique uses pixellated data on a fine scale where each pixel is assigned a 1 or a 0 depending on whether or not a fire was ignited at that location. (To be precise, they considered temporal effects as well, while our focus will be to produce only a spatial risk map.) The resulting data set is very large with an overwhelming proportion of 1×1 km pixels (sites) without a fire ignition between 1980 and 2007, indicated by a value of 0. However, a simple random sample of these 0-sites can be analyzed in the same way as the full dataset with the addition of an offset of the form $\log(1/\pi_s)$. Here π_s denotes the (constant) inclusion probability for site $s = (s_1, s_2)$, where s_1 and s_2 refer to the easting and northing geographic coordinates, respectively.

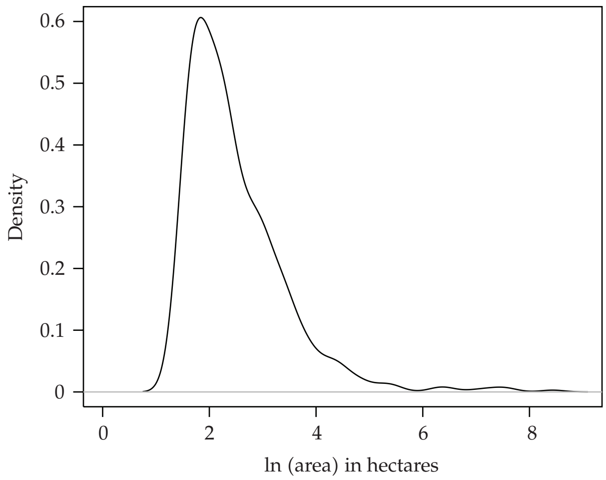


Figure 3: Estimated density of natural log-transformed fire sizes of historic escaped fires (1980–2007).

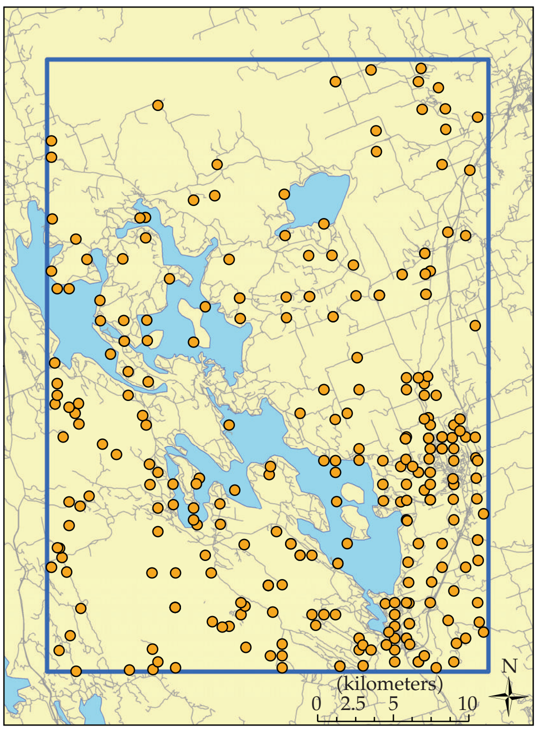


Figure 4: Map of human ignition locations in study area (1980–2007).

We have explored our data set with a simple model from within this family of models:

$$\operatorname{logit}(p_s) = f(s_1, s_2) + \log\left(\frac{1}{\pi_s}\right), \tag{3.1}$$

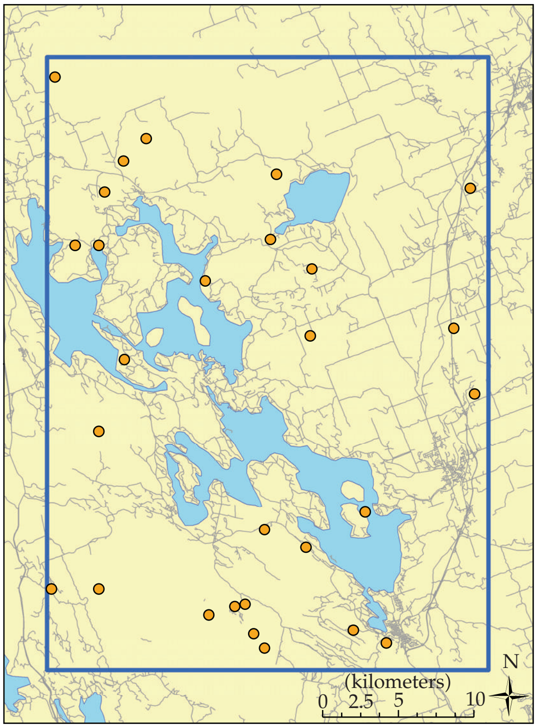


Figure 5: Map of lightning ignition locations in study area (1980–2007).

where p_s = probability of ignition at site s and $f(s_1, s_2)$ is a penalized tensor product spline smoother using the cubic B-spline basis in each dimension [see 19, Chapter 4]. We have taken

$$\pi_s = \begin{cases} 0.01 & \text{if sites do not contain an ignition (i.e., a “0”)} \\ 1 & \text{if sites contain an ignition (i.e., a “1”).} \end{cases} \tag{3.2}$$

We chose this value for π_s in order to have a manageable data set which has sufficient covariate information for inference. The resulting ignition risk map is shown in Figure 6. We note that there is a relatively high risk of ignition in the southeast region. This is the region closest to the town of Gravenhurst. The rest of the region is less heavily populated, and thus less likely to be subject to human-caused ignitions.

3.2. *Simple Burn Probability Map*

We also used the above modelling approach to assess the probability of burning by applying the same methodology but instead of assigning a value of 1 to a pixel that had an ignition, we assign a value of 1 to pixels that have burned, either directly by an ignition or spread from an ignition point. Unfortunately, actual final fire shape was not available in the database, so we made a crude approximation based on the observed final burned areas. The resulting burn probability map is pictured in Figure 7. Notice the decreased fire risk near the town and the increase in fire risk to the north and to the west. Because of the proximity to town, it may be that the ignited fires in the southeast may be suppressed relatively quickly, leading to smaller burned area. This phenomenon has been well documented (e.g., [2]).

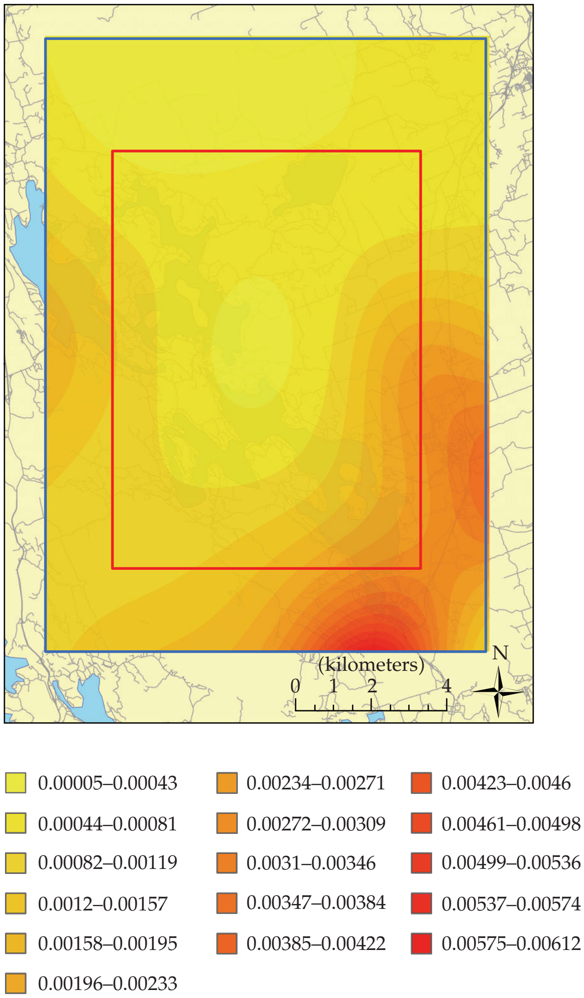


Figure 6: Model of ignition risk using generalized additive models with historic ignition data.

In addition to the loss of accuracy due to incorrect fire shape, the presence of relatively large lakes in the study area causes some difficulties for the smoother; essentially, boundary-like effects are introduced into the interior of the region. Furthermore, vegetation type and presence of other nonfuel fire barriers is not accounted for in this model.

For these reasons, we are motivated to consider a different modelling approach which is based partially on a physical model for wildfire growth and which incorporates fuel and fuel breaks. However, this map, based on historic records, can serve as a partial check on the reasonableness of the model we will propose next.

4. Burn Probability Modelling Using a Fire Growth Model

4.1. The Prometheus Fire Growth Model

Another method of forest fire risk assessment is based on computer simulation of fires, taking account of fuel information and local weather patterns. To model fire growth, we will employ the Prometheus Fire Growth Model [4].

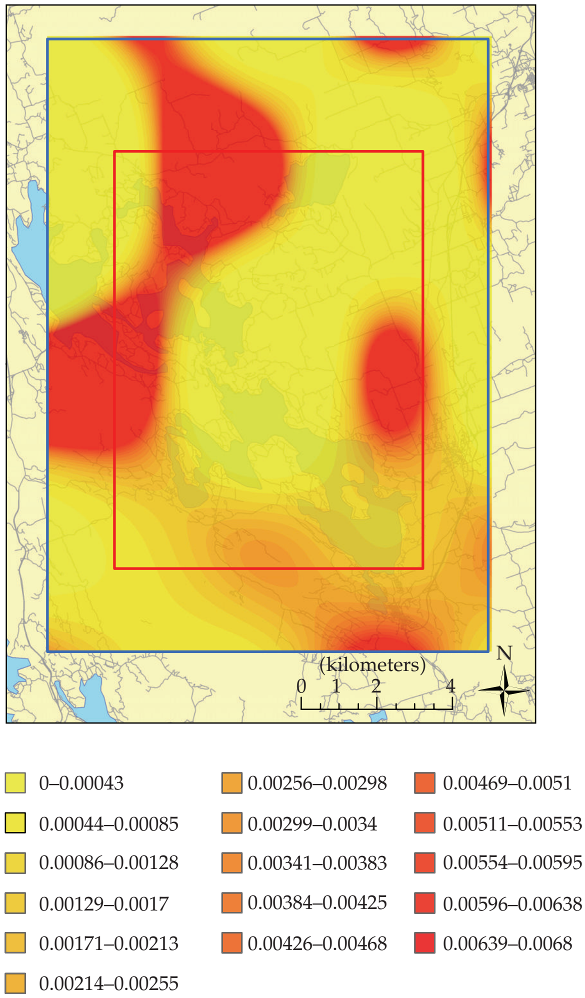


Figure 7: A simple burn probability map using generalized additive models with historic ignition data.

The evolution of a fire front simulated by Prometheus relies on the theory developed by Huygens for wave propagation: each point of a fire front at a given time acts as an ignition point for a small fire which grows in the shape of an ellipse based at that point. The size and shape of each ellipse depend on fuel composition information, weather, and various fire growth parameters as well as the time duration. The envelope containing all of the ellipses is taken to be the fire perimeter at the next time step (Figure 8).

In the absence of topographic variation, the orientation of each ellipse is aligned with the direction of the wind. The shapes of the ellipses at each time step are calculated from empirical models based on the Canadian Fire Behaviour Prediction (FBP) system which is described in the next subsection. The length of each ellipse is related to a local estimate of the forward rate of spread plus an analogous estimate of the back rate of spread, while the width of an ellipse is related to a local estimate of the flank rate of spread. These local rates of spread are, in turn, inferred from the empirical FBP models which relate spread rate to wind speed, fuel moisture, and fuel type. The measurements required for this calculation are based on local estimates of the weather conditions which have been extrapolated from the nearest reliable weather station. Diurnal changes in fuel moisture (as it is affected by temperature and relative humidity) and wind speed are also incorporated into the model.

4.2. *Canadian Fire Behaviour Prediction System and Fire Weather Index System*

In Canada, forest fire danger is assessed via the Canadian Forest Fire Danger Rating System (CFFDRS). As described by Natural Resources Canada [7], the current form of this system has been in development since 1968. The structure of the CFFDRS is modular and currently consists of four subsystems. Two of these subsystems are of interest in our study: the Canadian Fire Weather Index (FWI) System and the Canadian Forest Fire Behaviour Prediction (FBP) System, both of which are fully documented and are used operationally across Canada.

Many parts of the CFFDRS rely on information obtained using the FWI System. This system is comprised of six components which summarize aspects of the relative fire danger at its midafternoon peak [8]. All calculations are based on locally observed weather readings recorded at local noon: temperature, relative humidity, wind speed (usually a 10-minute average), and rainfall (over the last 24 hours). Three fuel moisture codes are calculated, each representing the dryness in a different layer of the forest floor. Three fire behaviour indices, estimating the risk of fire spread, the fuel available for combustion, and the potential intensity of a fire, are also calculated. For a recent exposition on the CFFDRS, see the account by Wotton [9] which is a review designed for modellers who require an understanding of this system and how it is to be interpreted.

The FWI System is used to estimate forest fire potential. Its outputs are unitless indicators of aspects of fire potential and are used for guiding fire managers in their decisions about resource movements, presuppression planning, and so forth. However, this is only a part of fire management. There is also the need, once fires have begun, to estimate characteristics of fire behaviour at a point on the landscape; this is done with the FBP System.

Given inputs that fall into one of five categories—fuels, weather, topography, foliar moisture content, and type and duration of prediction—the FBP System can be used to estimate fire behaviour quantitatively [10]. The FBP System calculations yield four primary and eleven secondary outputs as fire behaviour indices. It gives estimates which can be used as the basis for predictions.

The primary outputs are Rate of Spread, Fuel Consumption (as either the surface or crown consumption, or total), Head Fire Intensity, and a fire description code (crown fraction burned and fire type). The secondary outputs are Flank and Back Fire Rate of Spread; Flank and Back Fire Intensity; Head, Flank, and Back Fire Spread Distances; Elliptical Fire Area; Fire Perimeter; Rate of Perimeter Growth; and Length-to-Breadth Ratio. The primary outputs are based on a fire intensity equation and the secondary outputs are determined by assuming elliptical fire growth. All underlying models and calculations are based on an extensive 30-year field experimental burning program and are fully documented [10].

4.3. *Burn-P3 Simulation Model*

A burn risk probability map can be generated using the Burn-P3 simulation model software developed by Marc Parisien of the Canadian Forest Service [5]. P3 stands for Probability, Prediction, and Planning. Burn-P3 runs repeated simulations of the Prometheus Fire Growth

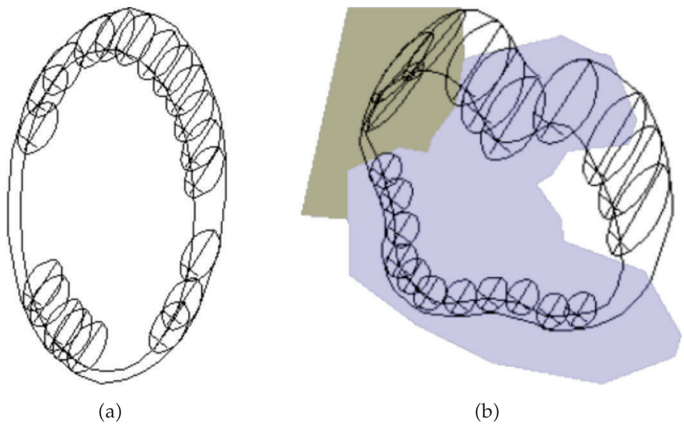


Figure 8: Illustration of fire perimeter growth under uniform burning conditions for homogenous fuels (a) and nonhomogenous fuels (b) [11].

Model, under different weather scenarios, to give estimates of the probability distribution of locations being burned during a single fire season.

In each iteration of a Burn-P3 simulation, a pseudorandom number is generated and used to sample a number from the empirical distribution of the annual number of escaped fires in the region. This empirical distribution is based on historic data. This number represents the number of simulated fires for one realization of one fire season.

For each of these fires, a random cause, season, and ignition location combination is selected from an ignition table. Burn-P3 creates an ignition table by combining ignition grids for each cause/season combination. Ignition grids partition the study area into coarse cells and represent the relative likelihood of a fire occurrence of an ignition in each cell. This spatial distribution can be empirically based on historic ignition patterns or it can be a uniform distribution, for example. The probability of selecting a certain row in the ignition table is proportional to the ignition probability of that particular cell specified in the matching ignition grid.

The duration of each simulated fire is also randomly drawn from an empirical fire duration distribution based on historic data. Given the location and fuel conditions, the Prometheus program is then used to simulate the growth of each fire individually given a random weather stream consisting of conditions conducive to fire growth from the appropriate season. All simulated fires in a single iteration are collectively used as an independent realization of a fire season.

Repeatedly simulating such fire seasons allows for construction of a burn probability map. Specifically, dividing the number of times each cell in the rasterized map of the study region has been burned by the number of simulations run gives an estimate of the probability that the particular cell will burn in a single fire season. See Figure 9 for a step-by-step illustration of this process.

The version of Burn-P3 used in this paper is not programmed to handle vectorized fuel breaks, that is, features in the landscape which tend to prevent fire from spreading. All fuel breaks such as roads are rasterized in Burn-P3 which sometimes leads to anomalous behaviour where a simulated fire passes between grid cells connected only at a single vertex. By using a small grid cell size, we can avoid this problem.

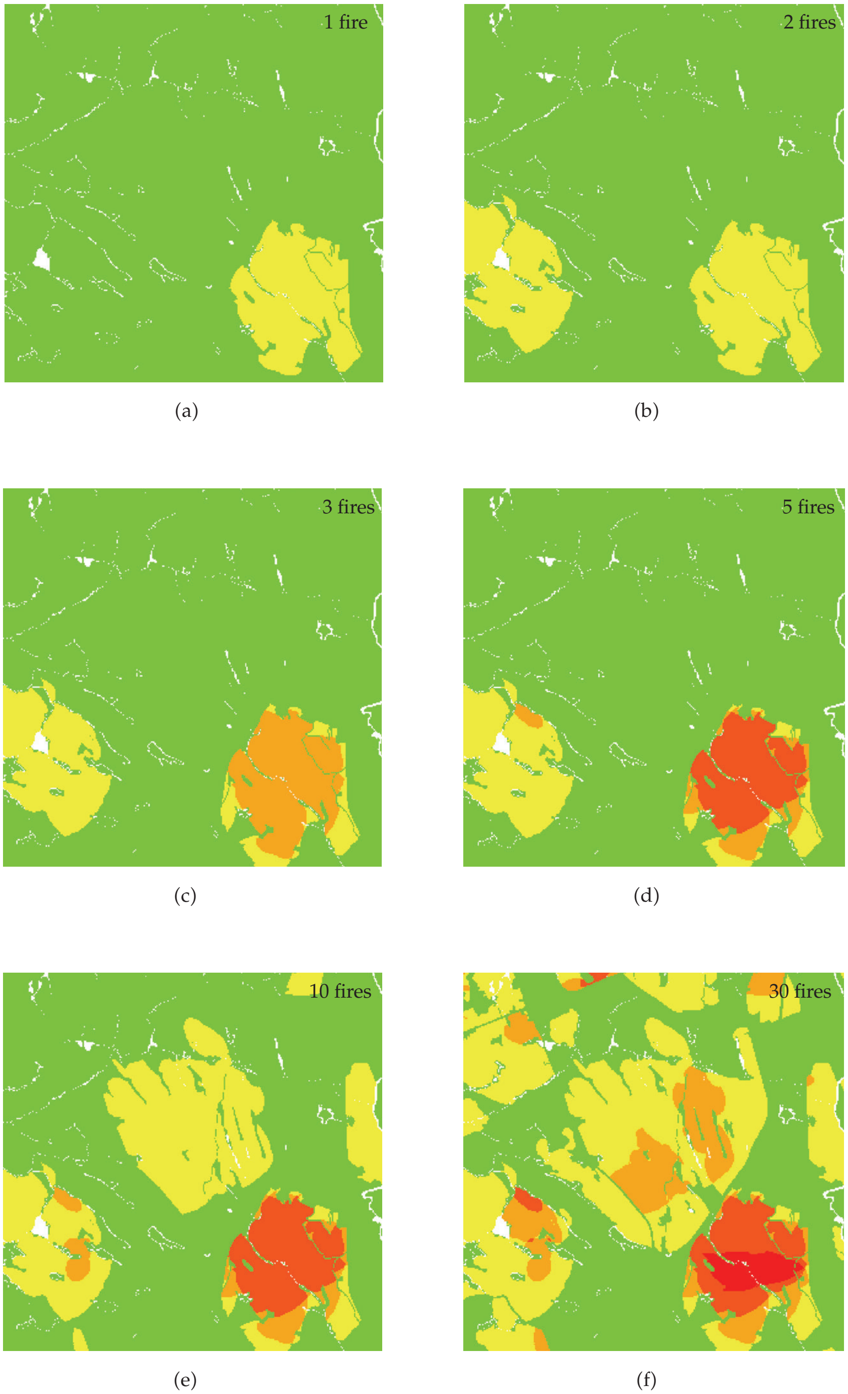


Figure 9: Step-by-step illustration of 30 iterations of the Burn-P3 simulation model. Darker colours indicate areas that have been burned more often. Green areas are unburned fuels. White areas represent nonfuel. (a) The yellow patch represents a single fire. (b) The two yellow patches represent two fires occurring in two different years. (c) Yellow patches denote areas burned by one of 3 fires occurring in different years. The orange patch represents an overlap of 2 of these fires. (d) The red patch represents an area burned by fires in 3 or more different years.

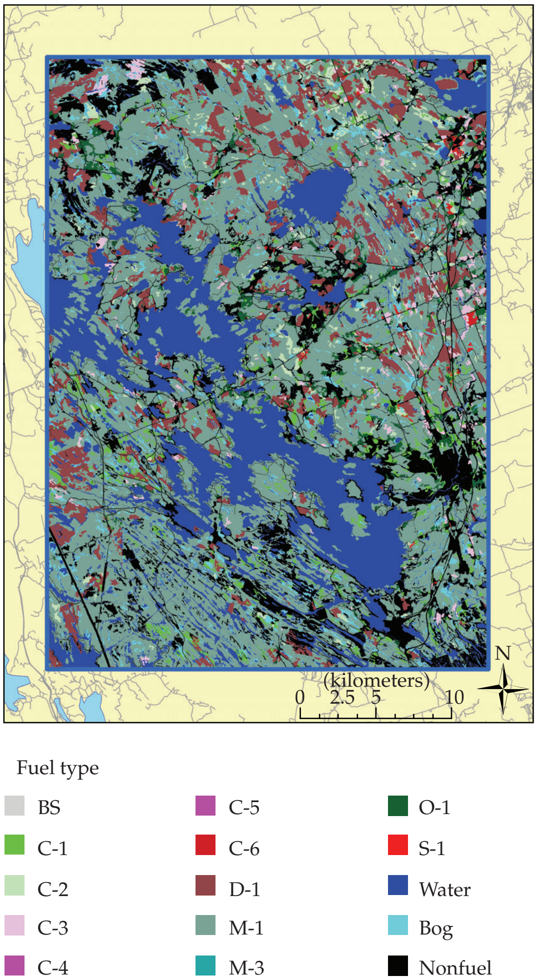


Figure 10: Corrected vector fuel map.

4.4. Inputs

4.4.1. Fuel Map

The FBP System classifies vegetation into 16 distinct fuel types (Table 1), that can further be grouped into the five categories: coniferous, deciduous, mixed wood, slash, and open [10]. A map of the fuel types in the District of Muskoka was obtained from the Ontario Ministry of Natural Resources. The fuel map was created manually from aerial photography in 1994. Fieldwork was conducted over the course of 7 days to verify and correct a subsample of the fuel map which appears in Figure 10.

4.4.2. Verification in the Field

Regardless of the accuracy of the fuel map at the time of its creation, fuel types and extents change over time due to land use changes, urban expansion, and natural causes such as forest succession. For example, in the study area, a large area of fuel mapped as C-6 (Conifer

Table 1: Fire Behavior Prediction System fuel types [10].

Group/Identifier	Descriptive name
Coniferous	
C-1	Spruce-lichen woodland
C-2	Boreal spruce
C-3	Mature jack or lodgepole pine
C-4	Immature jack or lodgepole pine
C-5	Red and white pine
C-6	Conifer plantation
C-7	Ponderosa pine-Douglas-fire
Deciduous	
D-1	Leafless aspen
Mixedwood	
M-1	Boreal mixedwood-leafless
M-2	Boreal mixedwood-green
M-3	Dead balsam fir mixedwood-leafless
M-4	Dead balsam fir mixedwood-green
Slash	
S-1	Jack or lodgepole pine slash
S-2	White spruce-balsam slash
S-3	Coastal cedar-hemlock-Douglas-fir slash
Open	
O-1	Grass

Plantation) was found to be harvested and hence, was reclassified (Figure 11). Not all areas were accessible by public roads and thus could not be verified by fieldwork. Consequently, satellite imagery was used to further supplement our fieldwork to help confirm such areas.

To get an estimate of the accuracy of the fuel map, multistage cluster sampling procedure was carried out in the field. First, 20 roads were selected at random with probability proportional to the length of the road (Figure 12). For each of these roads, observations were taken at the beginning, the end, and at various points along the road. The number of observations taken was randomly generated from a Poisson distribution with rate equal to the length of the road in kilometers (Table 2). The exact locations of these observations were randomly selected from a uniform distribution from the beginning to the end of the road.

At each observation location, the width of the road (including shoulder) was measured and recorded. One person without prior knowledge of the given fuel classification gave his best assessment of fuel classification of the fuels on either side of the road, making sure to look beyond the immediate vegetation at the tree line. Both the assessed fuel classification and original fuel classification were recorded.

Three of these selected roads were privately owned or not maintained enough to be traversable; these were not included in the sample. A summary of results is given in Table 3. The subjective nature of fuel classification can be seen in the 81.1% misclassification rate

Figure 11: Area mapped as C-6 that has since been harvested and required reclassification.

assuming no tolerance for classification error. However, not all differences in observations can be considered practically important. For example, an area assessed as 10% mixed wood originally classified as 20% mixed wood was not updated because the resulting change in fire behaviour is very slight; the rate of spread changes marginally and the direction of spread would not be affected at all. On the other hand, if a nonfuel was incorrectly classified as some form of fuel in the original map, the correction was made because the difference in fire behaviour could be substantial. Using this criterion, the misclassification rate was found to be to 22.7% in our sample: most of the fuel types were close to what we assessed them to be.

4.4.3. Unmapped Private Properties

The main properties of interest are those located along the waterfront because they represent the highest concentration of values at risk. Unfortunately, on the fuel map obtained from the Ontario Ministry of Natural Resources, such areas are almost always mapped as nonfuels because this is a private land not included in the Forest Resource Inventories on which the fuels classification is based. From what was observed in the field, most of these properties are located very close to fuels and could in fact be treated as a separately defined fuel type (Figure 13). The fuel types in these properties are almost always similar to what is located on the opposite side of the road further inland. For the purpose of this paper, we will assume that forest stands are continuous across roads into waterfront properties mapped as nonfuel.

Some properties have isolated buildings separated from surrounding fields (e.g., by a well-kept lawn or driveway). Although these buildings are unlikely to be damaged directly by a wildfire, they are still at risk to ignitions caused by spotting. Fire spotting is the situation where firebrands are transported long distances by the wind to start new fires. On waterfront properties, isolated buildings are not common. Thus, we further assume that all such properties have the same fuel type as the surrounding area.

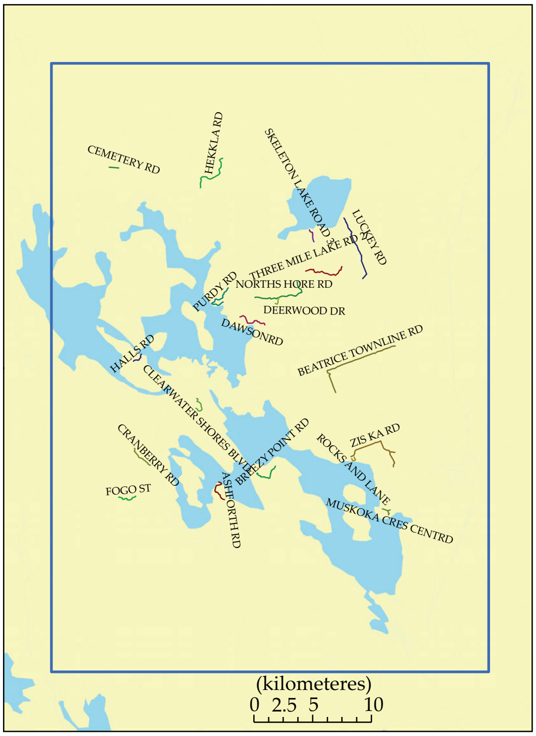


Figure 12: Map of the 20 randomly selected roads to be sampled.

Table 2: Names of the 20 randomly selected roads to be sampled, the number of observations taken on each road, and the location of each observation.

Name of Road	Number of observations	Locations along road (km)
Cemetery Road	3	0, 0.5, 0.8
Hekkla Road	5	0, 1.0, 2.3, 2.9, 4.3
Deerwood Drive	2	0, 0.7
Purdy Road	4	0, 0.1, 1.4, 2.4
North Shore Road	8	0, 0.4, 0.6, 3.0, 3.7, 3.9, 4.9, 5.3
Skeleton Lake Road 3	2	0, 1.2
Three Mile Lake Road 2	3	0, 2.5, 3.3
Walkers Road*	10	0, 1.9, 2.7, 3.6, 3.7, 4.4, 5.1, 5.5, 5.6, 6.5
Luckey Road*	7	0, 1.0, 1.4, 2.4, 2.5, 3.1, 4.0
Clearwater Shores Blvd	2	0, 1.5
Halls Road	2	0, 0.8
Dawson Road	8	0, 0.1, 1.7, 2.0, 2.1, 2.5, 2.6, 2.9
Beatrice Townline	9	0, 1.0, 2.1, 5.2, 7.1, 7.5, 7.6, 8.0, 8.4
Cranberry Road	4	0, 0.6, 1.9, 2.1
Fogo Street	3	0, 1.3, 1.8
Breezy Point Road	5	0, 0.5, 0.9, 1.3, 2.6
Ashworth Road	4	0, 0.3, 1.2, 1.7
Ziska Road	8	0, 0.9, 1.9, 3.3, 4.6, 4.9, 5.2, 5.3
Rock Sand Lane*	3	0, 0.2, 0.7
Muskoka Crescent Road	2	0, 6.1

*Inaccessible roads.

Table 3: Summary of fuel map accuracy sampling procedure.

Tolerance Margin	% Misclassified
Zero tolerance	81.1%
Little practical difference	22.7%

Figure 13: A structure embedded in a region of continuous fuel.

4.4.4. Fuel Breaks

As we have seen, fuel breaks are wide regions of what are essentially nonfuels that have the potential to prevent a fire from spreading across. The most common fuel breaks are roads, water bodies, and power lines.

The fuel map classifies major roads as nonfuels. Smaller private roads have not been included on the map. In some cases, they are located within larger private property regions mapped as nonfuel. We traced such roads from road data provided by the National Road Network. These data contain the locations of almost all roads as well as the number of lanes per road. We took a random sample of roads of varying numbers of lanes to measure their effective widths. Not only are lane widths not uniform on smaller roads, the size of the shoulder and distance to the tree line are highly variable. We used an average of sampled road widths for roads for which the widths are unknown or cannot be reasonably estimated.

Bodies of water such as lakes and rivers are clearly and accurately identified on the fuel map. However, bogs and swamps are problematic, since they are occasionally classified as water. Although a sufficiently dry bog could potentially become a fuel source in the heat of summer, the current fire growth model does not account for such a phenomenon.

Power lines introduce a further difficulty, since they are not identified on the fuel map. When power lines are built in a forested area, a path is cleared and growth underneath the line is regularly maintained. The width of this clearing and the amount of growth directly underneath the power line vary depending on how regular such maintenance occurs. Without a map of the smaller power lines, power lines are assumed to be negligible as fuel breaks.

4.4.5. Rasterization

The corrected vector fuel map must be converted to a raster map before it can be used by the Burn-P3 program. In doing so, detail at resolutions smaller than the grid cell resolution of the raster fuel map may be lost. However, refining the resolution of the raster fuel map

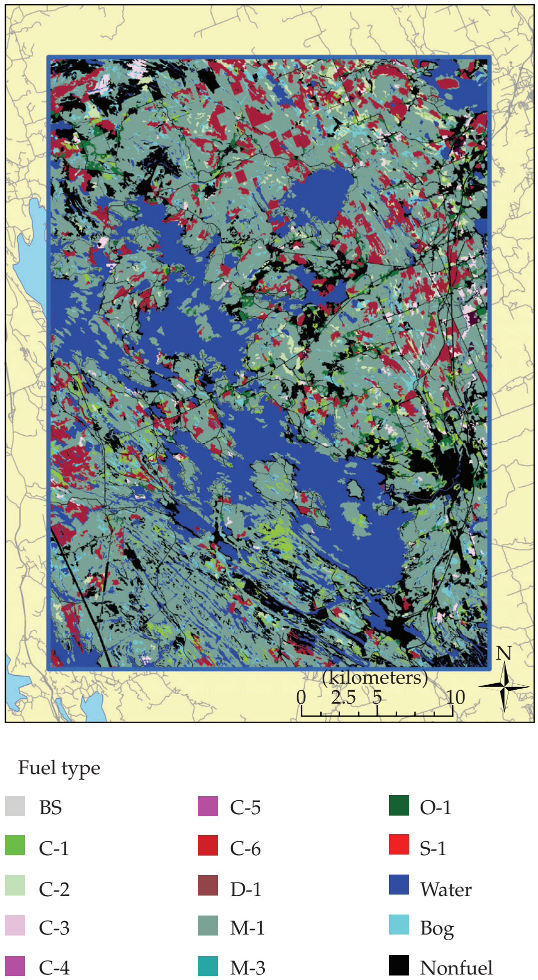


Figure 14: Rasterized fuel map at 25 m resolution.

directly increases computation time. In this assessment, a 25 m resolution (Figure 14) was used. A coarser resolution of 150 m was also tested, but we have not included the results of this rasterization, because important features such as fuel breaks were not respected.

4.4.6. Historic Weather

We obtained weather data from surrounding weather stations. The most complete weather record also happens to be the station closest to the study region so only weather data from this station was used in our analysis. This weather record begins with the 1980 fire season.

Only the extreme weather days are used for input to the Burn-P3 program. These days, in which there is potential for substantial fire growth, are referred to as *spread event days*. Days when the initial spread index (ISI) is less than 7.5 are considered to be nonextreme and are deleted from the weather stream. The resulting data set consists of 232 cases, each representing to a single day. Given the large number of simulations to be run, these weather conditions are sampled frequently.

4.4.7. Seasons and Causes

To properly simulate the growth of fires for an entire year, fires need to be classified by cause and by season in which they occur.

In regions with a mixture of deciduous and coniferous vegetation, there are often two distinct fire subseasons each year, the first immediately following snowmelt, and the second in summer. The spring fire subseason is a period of increased fire risk because leaves have not appeared on the deciduous trees leading to drier surface conditions. Since there is limited lightning activity during this period, ignitions are primarily due to people.

After the leaves appear, there is often a brief interval with few fires. Fire occurrence increases with temperature increase and as lightning activity increases. Thus, during the summer fire subseason, ignitions are due to people and lightning. This results in different spatial ignition patterns depending on time of year.

Operationally, early to mid June is typically taken as the transition date between the spring and summer seasons when classifying fires. However, this date is inferred from observations taken in the northern boreal forest. The District of Muskoka is further south and experiences a slightly warmer climate. Consequently, it is reasonable to assume that the actual transition date occurs approximately a week earlier. Looking at the distribution of human caused fires (Figure 15), we can see a dip on June 11th in human caused ignitions which is used as an estimate of the transition date and is highlighted with a vertical line in the figure.

All fire ignitions can be classified into either human- or lightning-caused fires. Human-caused fires can be further subdivided into eight specific causes: recreational, residential, railway, forestry industrial, nonforestry industrial, incendiary, miscellaneous, and unknown. Each of these causes is associated more strongly with either the spring or summer season (Figure 16). Between 1996 and 2005 inclusive, across the province, there were a total of 12,974 wildfires resulting in over 1.5 million hectares burned. Of these ignitions, nearly 7,000 can be attributed to lightning.

4.4.8. Ignition Grids

For each season and cause combination, ignition grids were created to represent the relative likelihood of an ignition in a certain cell. To create the ignition grids for human caused fires, a grid twenty times coarser than the fuel grid was created and assigned a value of 1. For each historic fire, the cell in which the ignition occurred had its value incremented by 1. Lightning ignitions in the region appear to be uniformly random and so a uniform ignition grid was used for lightning-caused ignitions.

4.4.9. Estimation of Spread Event Days

Burn-P3 models fire growth only on spread event days. The number of spread event days is not necessarily equal to the total duration of a fire; there may be days for which a fire does not spread (either due to suppression or nonconducive weather). The fire data do not contain information on fire spread days so a best estimate is made by counting the number of days from the fire start date to the date in which the fire was reported as “held”. A held fire is one that has been completely surrounded by fire line (i.e., fuel breaks constructed by suppression efforts as well as naturally occurring fuel breaks). The historic data on the distributions of spread event days per fire, as defined previously, is displayed in Figure 17.

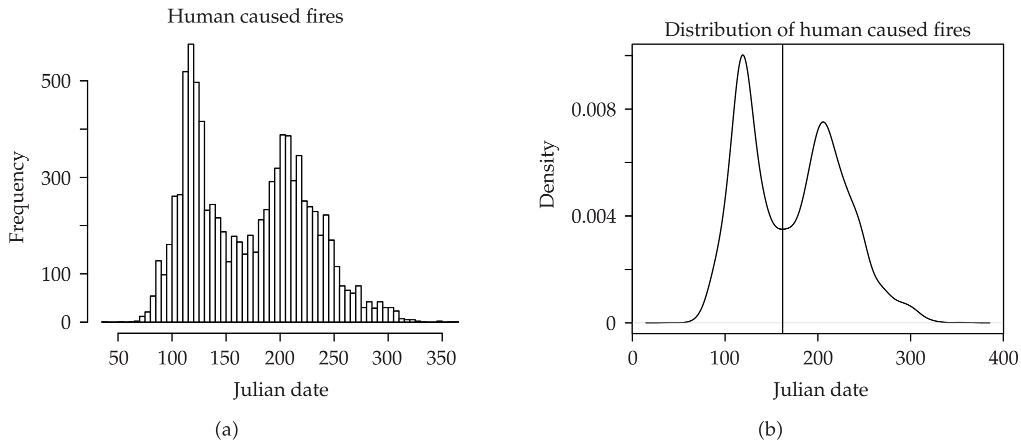


Figure 15: Distribution of human caused forest fires during the year. The vertical line indicates the minimum of the dip, corresponding to June 11th.

The FBP System's rate of spread is based on peak burning conditions, which are assumed to occur in the late afternoon, generally specified as 1600 hours [10]. Consequently, any fire growth models based on this system are effectively simulating spread event days [12].

5. Analysis and Results

We will begin this section by discussing the application of the Burn-P3 simulator to the corrected fuel map when the recommended settings are used. We will then discuss the results of a sensitivity analysis in which the effects of fuel break misclassification on the burn probabilities are studied. This will provide an indication of the uncertainty induced by possible inaccuracies in the fuel map. We next compare the burn probability map obtained from Burn-P3 with the map obtained using generalized additive models, and finally compare simulated fire size distributions with the historical record; this ultimately guides us to what we believe is a more accurate burn probability map.

A uniform ignition grid for lightning caused fires is normally recommended for use in Burn-P3. For human-caused fires, the ignition grid is also uniform, but with increased probability at locations where previous ignitions occurred. The distribution of spread event days is based on historic weather data where the weather stream has been adjusted so that it only contains extreme weather, conducive to fire growth. Using these recommended settings, we obtain the burn probability map and fire size distribution as shown in the left panel of Figure 18.

In that figure, it can be seen that the fire risk is higher in the north. This result seems plausible, since there are larger forest stands in that region, and we have already conjectured on the possibility of large fires spreading into the study region from further north.

In order to assess the uncertainty induced by possible inaccuracies in the fuel map, we conducted the same simulation but with 20% of randomly selected nonfuel grid cells converted to the M-1 20% mixed wood fuel type. By making this kind of change, we should observe the largest range of realistic fire behaviour in the study region, since much of the forest in the area is of M-1 type, and changes within this categorization have minimal effect on fire behaviour. By contrast, changes from nonfuel to any kind of of fuel can have relatively

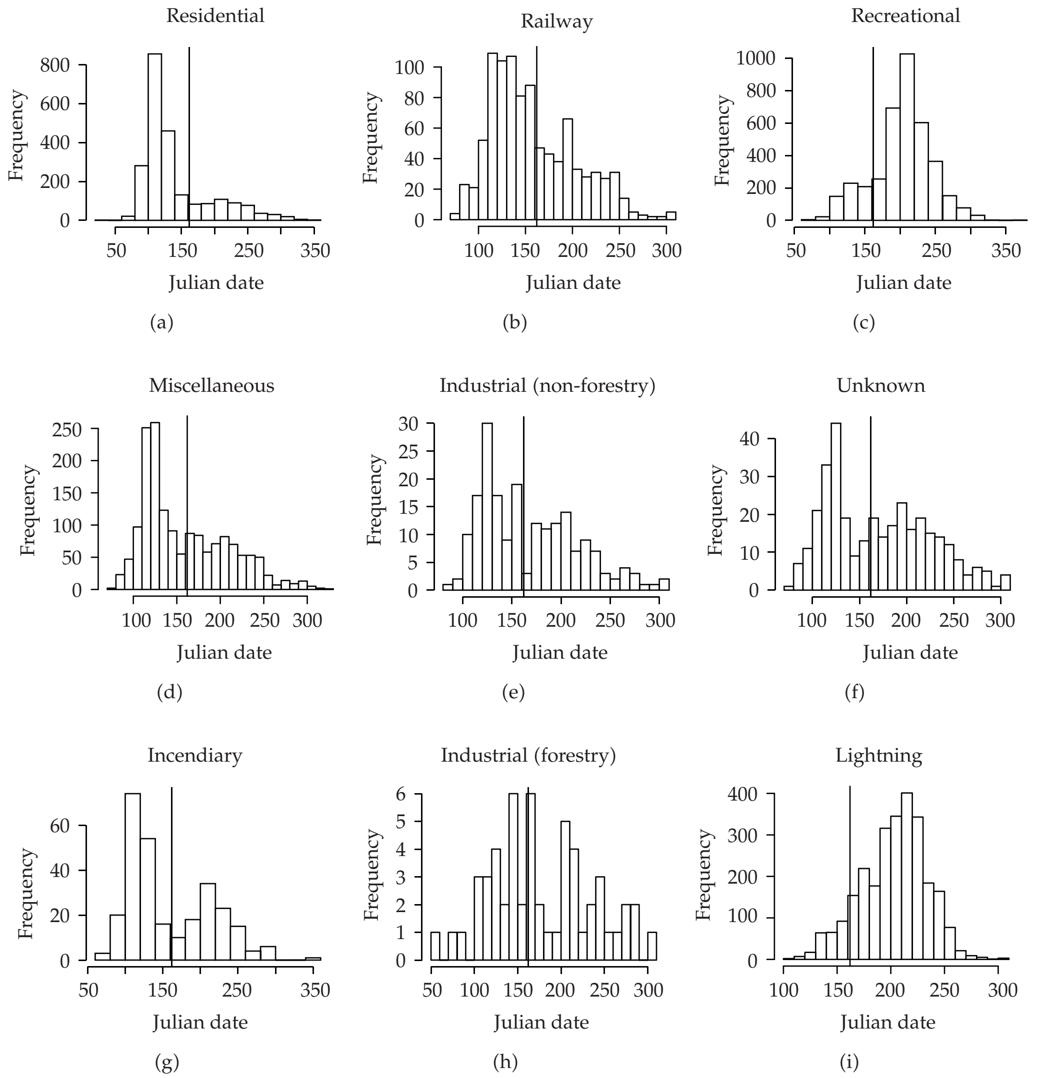


Figure 16: Histograms of forest fires during the year by cause. Vertical line indicates June 11th, the estimated transition date used to separate "spring" and "summer" subseasons.

dramatic effects on fire behaviour, since nonfuel regions often serve as fuel breaks; replacing parts of such regions with fuel allows for the possibility of a fire breach where it would not otherwise have been possible. As expected, the burn probability map (Figure 18(b)) exhibits larger regions of relatively high probability than in the original map, especially in the eastern region as well as in the north. Note that regions where the burn probability was already relatively high do not see a substantial gain in burn probability when the nonfuels are perturbed. Rather, we see somewhat more substantial increases in burn probability in those areas where the probability was much less. An additional simulation was run with only 10% of the nonfuel randomly converted to fuel; we have not shown the resulting map because of its similarity to the map in the right panel of Figure 18. We conclude that gross misclassification of fuel as nonfuel could lead to a moderate underestimate of the area at elevated risk.

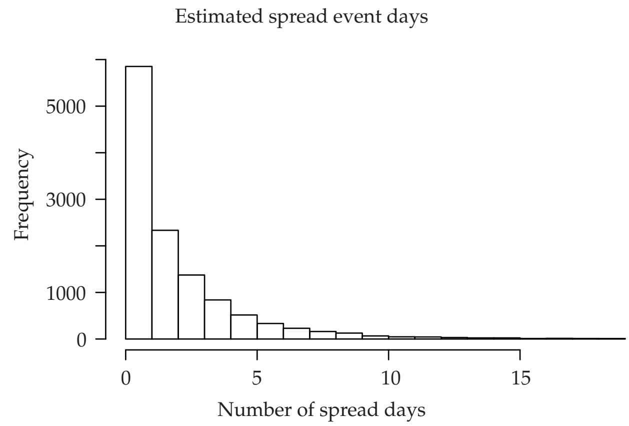


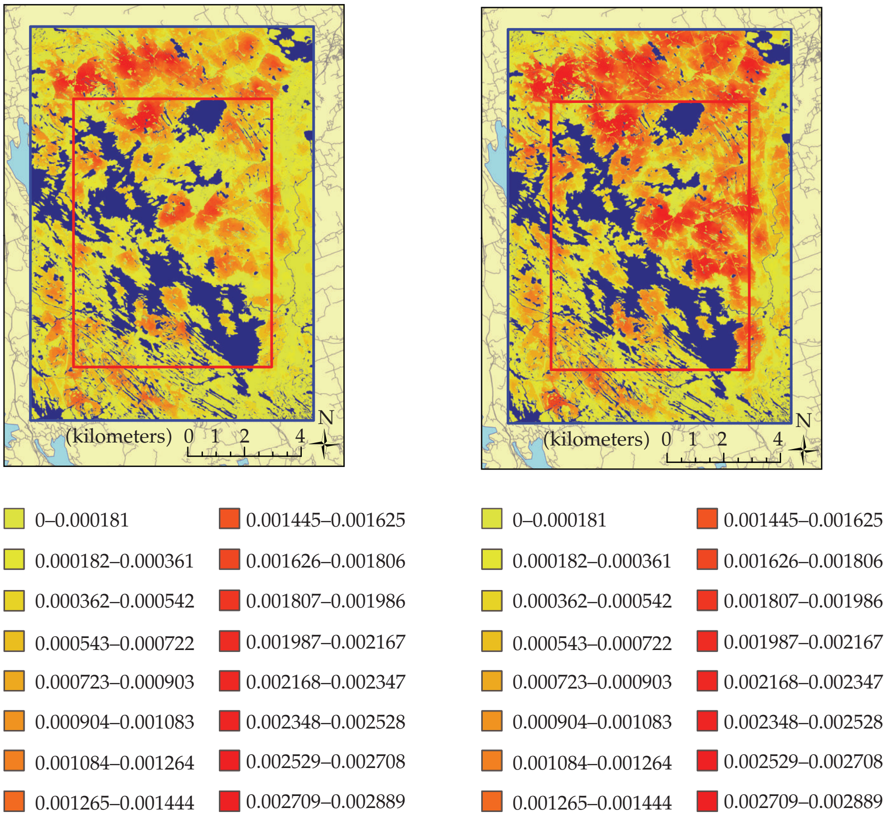
Figure 17: Histogram of estimated spread event days.

Comparing the map obtained from Burn-P3 using the corrected and unperturbed fuel map (Figure 18) with the burn probability map obtained using the generalized additive model (Figure 7), we see some similarities. Both maps exhibit elevated burn risk in the north, but the prevalence of ignitions in the southeast seems to figure more prominently in the map obtained from the generalized additive model. It should be noted that the latter map is based on accurate fire sizes but incorrect fire shapes, while the former map is based on what are possibly more realistic fire shapes, but with a fire shape and size distribution that is determined by the weather and fuels.

We can then use the historic fire size distribution as a check on the accuracy of the Burn-P3 output. Figure 19 shows the estimated density of fire sizes in the study region (solid black curve) on the natural logarithmic scale. The dashed curve represents the estimated density of the simulated log fire sizes under the recommended settings, and the dotted curve corresponds to the perturbed nonfuel simulation. Both densities fail to match that of the historic record. Modal log fire sizes are close to 2 in the historic record, while the simulations give modes exceeding 5. Note that, in accordance with our earlier observations regarding the nonfuels, the fire sizes indeed increase when fuel breaks are removed.

In order to find a model which matches the historic record more closely, we could introduce additional fuel breaks, but we have no way of determining where they should be located without additional (substantial) fieldwork, and the earlier sensitivity study indicates that even fairly substantial errors in the fuel map will lead to only modest discrepancies in the fire size distribution. Instead, it may be more important to consider the effects due to weather. To investigate this, we have run four additional Burn-P3 simulations under different settings. The resulting burn probability maps appear in Figure 20. We now proceed to describe these simulations and their resulting fire size distributions.

First, we replaced the spread event day distribution with a point mass at 1 day. In other words, we made the assumption that even if fires in the area burn for several days, there would only be one day in which the fire would burn a nonnegligible amount. All other simulation settings remain as before. The fire size distribution for this situation is pictured in Figure 19 as the long-dashed curve, having a mode near 4. This is closer to the historic record, but still unsatisfactory.



(a) (b)

Figure 18: (a) Burn probability map of simulated fires using recommended Burn-P3 settings. (b) Burn probability map of simulated fires using recommended settings and 20% of nonfuels randomly converted to M-1 fuel type.

The next simulation made use of the entire weather record, dispensing with the notion of spread event days completely. Fire durations were sampled from historic fire duration distribution. Again, all other simulation settings were the same as before. The resulting fire size distribution is displayed in Figure 19 as the dashed-dotted curve, having a mode near 3—a substantial improvement, but still not satisfactory. The difference between this result and the earlier simulations which depend only on extreme weather calls such practice into question.

In the succeeding simulation run, the duration of the fires was reduced to a single day, again sampling from the full weather stream. The resulting density estimate is displayed in Figure 19 as the long-dashed-dotted curve, having a mode near the historic mode, although its peak is not nearly as pronounced. An additional simulation was conducted using the same settings but with an ignition grid based on the generalized additive model for ignitions obtained in Section 3.1. The resulting fire size distribution is also pictured in Figure 19 and is very similar to the result of the preceding simulation.

We conclude that the use of the full weather stream and that limiting the duration of the fires to one day give more accurate fire size distributions. Use of the uniform ignition grid is slightly less accurate than the use of the modelled grid based on historic ignitions.

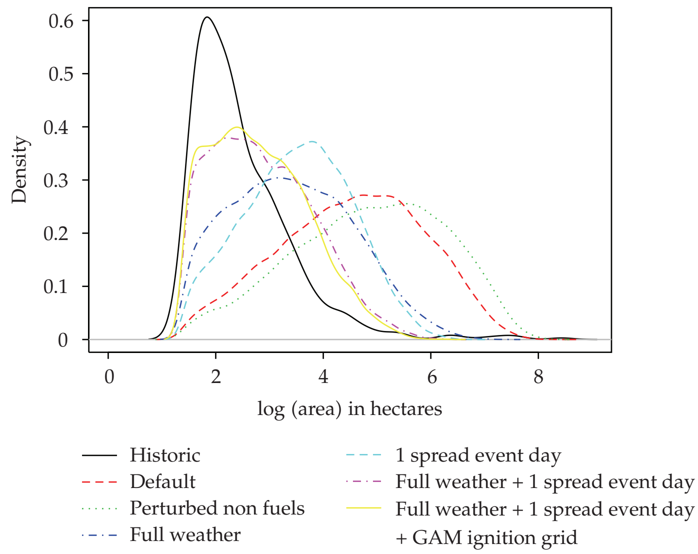


Figure 19: Estimated density functions of log-transformed simulated and observed fire sizes under various scenarios. The historic fire size density is based on data from an area encompassing the study region (1980–2007). The other fire curves are based on Burn-P3 simulations, using the default setting with the corrected fuel map as well as a fuel map with 20% of the nonfuel randomly selected to be changed to mixed wood-type fuel. The Full weather curve is based on a simulation using all weather data but with the same spread event day distribution as before.

6. Discussion

We have shown how to estimate a burn probability map which could be used by insurers to estimate expected losses due to wildfire risk in the region under study. We found that substantial perturbation of the fuel map, converting nonfuels to fuels, gives rise to moderate changes in the fire risk.

We have also used historic fire size distribution information as a check on its accuracy and found that the recommendation to use a spread event day distribution for fire duration overestimates the fire size distribution. The use of spread event days in the Burn-P3 model could be degrading the probability estimates. The spread event day distribution may be biased since it is based on the time between when a fire was first reported and when it was declared as being successfully held by suppression activities. A fire would not necessarily be spreading rapidly during this entire period.

Note that fire suppression is not accounted for directly in Burn-P3. This could account for the difference between the simulated and observed fire size distributions. By using the full weather stream and a one day fire duration, the simulated fire size distribution comes closer to matching the historic record. In fact, using the 1 day fire duration may be realistic because of suppression effects; it is unlikely that most fires are allowed to burn for more than 1 day in this region without being attacked. If allowed to burn longer, it would not be under extreme weather conditions, and such fires would not be spreading fast.

Thus, there is some justification for our approach. We note, however, that there is still a discrepancy between the simulated fire size distribution and the historic record. As we saw, fuel/nonfuel misclassification can have a modest effect on the fire size distribution estimates. It is possible that some of the small roads that are not classified as fuel breaks may in fact

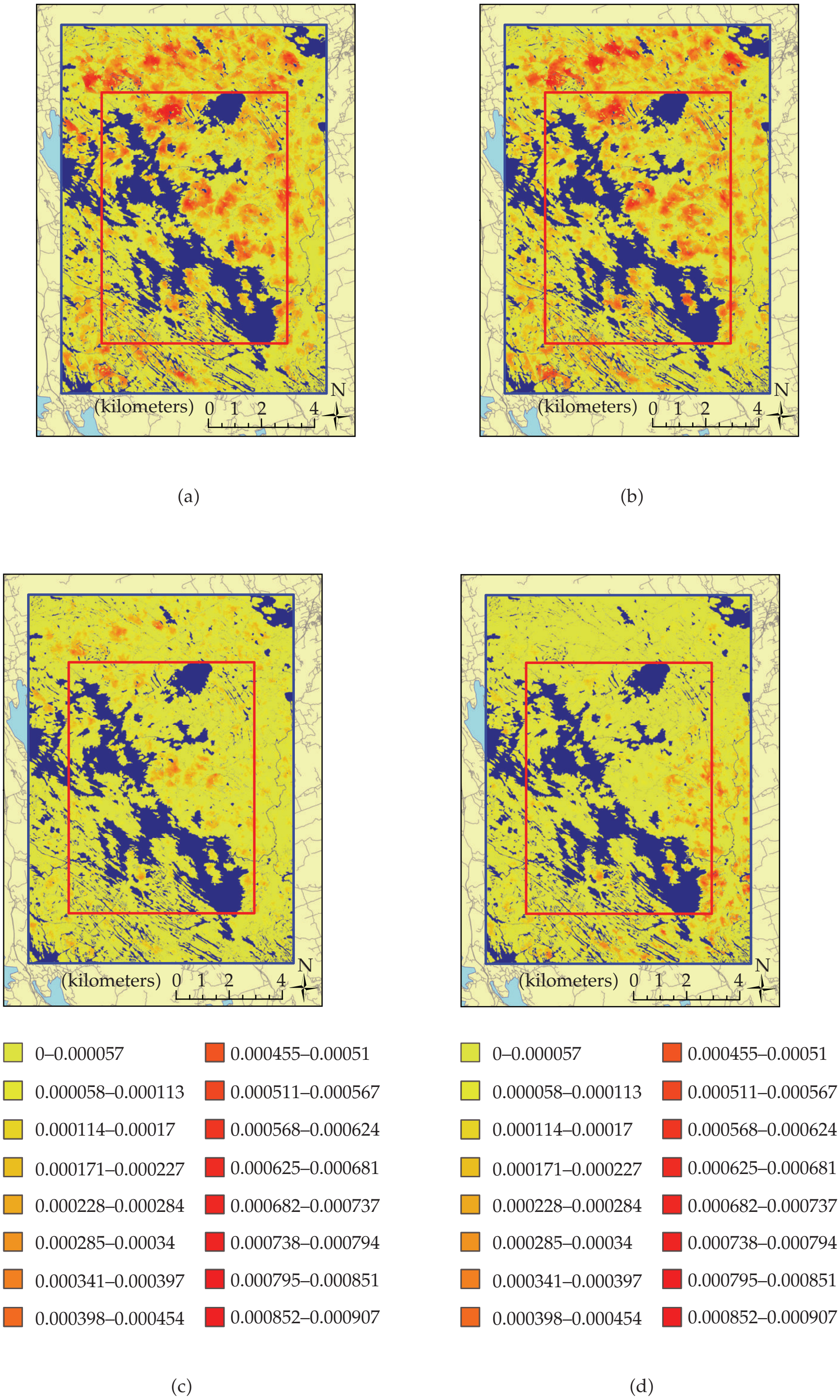


Figure 20: Burn probability maps under various simulation scenarios. (a) Using full weather stream. (b) Using one spread event day. (c) Using full weather stream and one spread event day. (d) Using full weather stream and one spread event day with ignition grid from GAM model.

be serving as fuel breaks; this kind of error could explain the discrepancy in the fire size distributions.

Our final model indicates a relatively low burn probability across the region. Depending on the ignition grid used, the higher risk areas are either in the north (uniform grid) or to the east and southeast (historic ignition grid). The latter indicates a somewhat more serious risk for the more heavily populated area, and may be the more realistic scenario, since there is no reason to believe that the pattern of human-caused ignitions (which are largely responsible for the fires here) will change in the future without some form of intervention.

However, there are several limitations to this approach. The results obtained in this assessment need to be interpreted with some care. We should note that by using the fire size distribution as our standard for accuracy checking, we are assuming that the fuel distribution and composition is similar to how it was in the past and that the climate has not changed substantially. Future research in which this model is run under various realistic climate change scenarios will be important. Any changes in fire management in the area, either past or future, have not been factored into our estimates of fire risk.

The Prometheus Fire Growth Model has been used extensively in Canadian fire management operations. It has been found to be reasonably accurate under many circumstances, especially under peak burning conditions, which is where the FBP system is at its most accurate. Indeed, predictions from the FBP System, which forms the foundation of fire growth in the Prometheus model and consequently in Burn-P3, are used as a regular and important part of forecasting the active growth of fires and the consequent planning of fire suppression resource needs. While the FBP has limitations (see discussion below), it constitutes the best available system for predicting fire ignition and growth in the forests of Canada. Consequently, the Prometheus fire growth model as well as Burn-P3 have been used in a number of research studies in a wide variety of locations in Canada [5, 13–16].

However, the size of fires may be overstated under moderate weather conditions [12]. Since Prometheus is based on the FBP system which is, in turn, based on empirical observations, the process under which these empirical observations were collected influences model performance. Some of these observations were from controlled burns, so spread rates of wildfires in different fuel types may be quite different, at least during the acceleration phase. The reason for this is that the prescribed fires were started with a line ignition under somewhat variable weather conditions. Because of the line ignition, the estimated spread rate may be biased upwards, since point fire ignitions are more common in naturally occurring fires. Spread rates for mixed wood fuel types were not empirically developed from observed fire behaviour; instead, they were calculated as weighted averages of spread rates of coniferous and deciduous fuel types.

The Burn-P3 simulation model is also limited in that it is not programmed to handle vectorized fuel breaks so any fuel breaks smaller than the chosen cell resolution do not prevent a fire from spreading. Furthermore, inputs for Burn-P3 are based on empirical observations which makes an assumption that what will be observed in future fire seasons is similar to what has happened in the past.

Acknowledgments

Funding was provided by the Institute for Catastrophic Loss Reduction, the MITACS Accelerate internship program, and GEOIDE. The authors would also like to thank Marc Parisien for assistance in the use of the Burn-P3 simulation model, Jen Beverly and Cordy

Tymstra for a number of helpful discussions, and the Ontario Ministry of Natural Resources for providing the fuel map of the area as well as the fire and weather data. Comments from two anonymous referees are also gratefully acknowledged.

References

[1] S. McCaffrey, "Thinking of wildfire as a natural hazard," *Society and Natural Resources*, vol. 17, no. 6, pp. 509–516, 2004.
[2] B. J. Stocks, J. A. Mason, J. B. Todd et al., "Large forest fires in Canada, 1959–1997," *Journal of Geophysical Research D*, vol. 108, no. 1, pp. 5.1–5.12, 2003.
[3] D. G. Woolford and W. J. Braun, "Convergent data sharpening for the identification and tracking of spatial temporal centers of lightning activity," *Environmetrics*, vol. 18, no. 5, pp. 461–479, 2007.
[4] C. Tymstra, R. W. Bryce, B. M. Wotton, and O. B. Armitage, "Development and structure of Prometheus: the Canadian wildland fire growth simulation model," Information Report NOR-X-417, Natural Resources Canada, Canadian Forestry Service, Northern Forestry Centre, Edmonton, Canada, 2009.
[5] M. A. Parisien, V. G. Kafka, K. G. Hirsch, J. B. Todd, S. G. Lavoie, and P. D. Maczek, "Using the Burn-P3 simulation model to map wildfire susceptibility," Information Report NOR-X-405, Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Edmonton, Canada, 2005.
[6] D. R. Brillinger, H. K. Preisler, and J. W. Benoit, "Risk assessment: a forest fire example," in *Statistics and Science: A Festschrift for Terry Speed*, D. R. Goldstein, Ed., vol. 40 of *IMS Lecture Notes Monograph Series*, pp. 177–196, Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2003.
[7] Natural Resources Canada, "Canadian Forest Fire Danger Rating System," NRC, August 2009, http://fire.nofc.cfs.nrcan.gc.ca/en/background/bi_FDR_summary_e.php .
[8] C. E. van Wagner, "Development and structure of the Canadian forest fire weather index system," Forest Technical Report 35, Canadian Forest Service, Ottawa, Canada, 1987.
[9] B. M. Wotton, "Interpreting and using outputs from the Canadian Forest Fire Danger Rating System in research applications," *Environmental and Ecological Statistics*, vol. 16, no. 2, pp. 107–131, 2009.
[10] Forestry Canada Fire Danger Group, "Development and structure of the Canadian forest fire behavior prediction system," Information Report ST-X-3, Forestry Canada, Science and Sustainable Development Directorate, Ottawa, Canada, 1992.
[11] FARSITE, 2008, http://www.firemodels.org/.
[12] J. J. Podur, *Weather, forest vegetation, and fire suppression influences on area burned by forest fires in Ontario*, Ph.D. dissertation, Graduate Department of Forestry, University of Toronto, Toronto, Canada, 2006.
[13] C. Tymstra, M. D. Flannigan, O. B. Armitage, and K. Logan, "Impact of climate change on area burned in Alberta's boreal forest," *International Journal of Wildland Fire*, vol. 16, no. 2, pp. 153–160, 2007.
[14] R. Suffling, A. Grant, and R. Feick, "Modeling prescribed burns to serve as regional firebreaks to allow wildfire activity in protected areas," *Forest Ecology and Management*, vol. 256, no. 11, pp. 1815–1824, 2008.
[15] M. A. Parisien, C. Miller, A. Ager, and M. Finney, "Use of artificial landscapes to isolate controls on burn probability," *Landscape Ecology*, vol. 25, no. 1, pp. 79–93, 2010.
[16] J. L. Beverly, E. P. K. Herd, and J. C. R. Conner, "Modeling fire susceptibility in west central Alberta, Canada," *Forest Ecology and Management*, vol. 258, no. 7, pp. 1465–1478, 2009.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 838240, 11 pages
doi:10.1155/2010/838240

Research Article

The Foundations of Probability with Black Swans

Graciela Chichilnisky

Departments of Economics and Mathematical Statistics, Columbia University, 335 Riverside Drive, New York, NY 10027, USA

Correspondence should be addressed to Graciela Chichilnisky, chichilnisky1@gmail.com

Received 8 September 2009; Accepted 25 November 2009

Academic Editor: Ričardas Zitikis

Copyright © 2010 Graciela Chichilnisky. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We extend the foundation of probability in samples with rare events that are potentially catastrophic, called *black swans*, such as natural hazards, market crashes, catastrophic climate change, and species extinction. Such events are generally treated as "outliers" and disregarded. We propose a new axiomatization of probability requiring equal treatment in the measurement of rare and frequent events—the Swan Axiom—and characterize the subjective probabilities that the axioms imply: these are neither finitely additive nor countably additive but a combination of both. They exclude countably additive probabilities as in De Groot (1970) and Arrow (1971) and are a strict subset of Savage (1954) probabilities that are finitely additive measures. Our subjective probabilities are standard distributions when the sample has no black swans. The finitely additive part assigns however more weight to rare events than do standard distributions and in that sense explains the persistent observation of "power laws" and "heavy tails" that eludes classic theory. The axioms extend earlier work by Chichilnisky (1996, 2000, 2002, 2009) to encompass the foundation of subjective probability and axiomatic treatments of subjective probability by Villegas (1964), De Groot (1963), Dubins and Savage (1965), Dubins (1975) Purves and Sudderth (1976) and of choice under uncertainty by Arrow (1971).

1. Introduction

Black swans are rare events with important consequences, such as market crashes, natural hazards, global warming, and major episodes of extinction. This article is about the foundations of probability when catastrophic events are at stake. It provides a new axiomatic foundation for probability requiring sensitivity both to rare and frequent events. The study culminates in Theorem 6.1, that proves existence and representation of a probability satisfying three axioms. The last of these axioms requires sensitivity to rare events, a property that is desirable but not respected by standard probabilities. The article shows the connection between those axioms and the Axiom of Choice at the foundation of Mathematics. It defines a new type of probabilities that coincide with standard distributions when the sample is populated only by relatively frequent events. Generally, however, they are a mixture of countable and finitely additive measures, assigning more weight to black swans than do normal distributions, and predicting more realistically the incidence of "outliers," "power laws," and "heavy tails" [1, 2].

The article refines and extends the formulation of probability in an uncertain world. It provides an argument, and formalization, that probabilities must be additive functionals on $L_\infty(\mathcal{U})$ (where $\mathcal{U}$ is a σ-field of "events" represented by their indicator bounded and real valued functions), that are neither countably additive nor finitely additive. The contribution is to provide an axiomatization showing that subjective probabilities must lie in the full space L_∞^* rather than L_1 as the usual formalization (Arrow, [3]) forcing countable additivity implies. The new axioms refine both Savage's [4] axiomatization of finitely additive measures, and Villegas' [5] and Arrow's [3] that are based on countably additive measures, and extend both to deal more realistically with catastrophic events.

Savage [4] axiomatized subjective probabilities as finitely additive measures representing the decision makers' beliefs, an approach that can ignore frequent events as shown in the appendix. To overcome this, Villegas [5] and Arrow [3] introduced an additional continuity axiom (called "*Monotone Continuity*") that yields countably additivity of the measures. However Monotone Continuity has unusual implications when the subject is confronted with rare events, for example, it predicts that in exchange for a couple of cents, one should be willing to accept a small risk of death (measured by a countably additive probability), a possibility that Arrow called "outrageous" [3, Pages 48–49]. This article defines a realistic solution: for some, very large, payoffs and in certain situations, one may be willing to accept a small risk of death—but not in others. This means that Monotone Continuity holds in some cases but not in others, a possibility that leads to the axiomatization proposed in this article and is consistent with the experimental observations reported by (Chanel and Chichilnisky [6, 7]). The results are as follows. We show that countably additive measures are insensitive to *black swans:* they assign negligible weight to rare events, no matter how important these may be, treating catastrophes as outliers. Finitely additive measures, on the other hand, may assign no weight to frequent events, which is equally troubling. Our new axiomatization balances the two approaches and extends both, requiring sensitivity in the measurement of rare as well as frequent events. We provide an existence theorem for probabilities that satisfy our axioms, and a characterization of all that do.

The results are based on an axiomatic approach to choice under uncertainty and sustainable development introduced by Chichilnisky [8–10] and illuminate the classic issue of continuity that has always been at the core of "subjective probability" axioms (Villegas, [5], Arrow [3]). To define continuity, we use a topology that tallies with the experimental evidence of how people react to rare events that cause fear (Le Doux [11], Chichilnisky [12]), previously used by Debreu [13] to formalize a market's Invisible Hand, and by Chichilnisky [9, 12, 14] to axiomatize choice under uncertainty with rare events that inspire fear. The new results provided here show that the standard axiom of decision theory, Monotone Continuity, is equivalent to De Groot's Axiom SP_4 that lies at the foundation of classic likelihood theory (Proposition 2.1) and that both of these axioms underestimate rare events no matter how catastrophic they may be. We introduce here a new Swan Axiom (Section 3) that logically negates them both, show it is a combination of two axioms defined by Chichilnisky [9, 14] and prove that any subjective probability satisfying the Swan Axiom is neither countably additive nor finitely additive: it has elements of both (Theorem 4.1). Theorem 6.1 provides a complete characterization of all subjective probabilities that satisfy linearity and the Swan Axiom, thus extending earlier results of Chichilnisky [1, 2, 9, 12, 14].

There are other approaches to subjective probability such as Choquet Expected Utility Model (CEU, Schmeidler, [15]) and Prospect Theory (Kahneman and Tversky, [16, 17]). They use a nonlinear treatment of probabilities of likelihoods (see, e.g., Dreze, [18], or Bernstein, [19]), while we retain linear probabilities. Both have a tendency to give higher weight to small probabilities, and are theoretical answers to experimental paradoxes found by Allais in 1953 and Ellsberg in 1961, among others refuting the *Independence Axiom* of the Subjective Expected Utility (SEU) model. Our work focuses instead directly on the foundations of probability by taking the logical negation of the *Monotone Continuity Axiom*. It is striking that weakening or rejecting this axiom—respectively, in decision theory and in probability theory—ends up in probability models that are more in tune with observed attitudes when facing catastrophic events. Presumably each approach has advantages and shortcomings. It seems that the approach offered here may be superior on four counts: (i) it retains linearity of probabilities, (ii) it identifies Monotone Continuity as the reason for underestimating the measurement of catastrophic events, an axiom that depends on a technical definition of continuity and has no other compelling feature, (iii) it seems easier to explain and to grasp, and therefore (iv) it may be easier to use in applications.

2. The Mathematics of Uncertainty

Uncertainty

Uncertainty is described by a set of distinctive and exhaustive possible *events* represented by a family of sets $\{U_\alpha\}$, $\alpha \in N$, whose union describes a universe $\mathcal{U} = \bigcup_\alpha U_\alpha$. An event $U \in \mathcal{U}$ is identified with its *characteristic function* $\phi_U : \mathcal{U} \to R$ where $\phi_U(x) = 1$ when $x \in U$ and $\phi_U(x) = 0$ when $x \notin U$. The subjective probability of an event U is a real number $W(U)$ that measures how likely it is to occur according to the subject. Generally we assume that the probability of the universe is 1 and that of the empty set is zero $W(\emptyset) = 0$. In this article we make no difference between subjective probabilities and likelihoods, using both terms intercheangeably. Classic axioms for subjective probability (resp. likelihoods) are provided by Savage [4] and De Groot [20]. The likelihood of two disjoint events is the sum of their likelihoods: $W(U_1 \cup U_2) = W(U_1) + W(U_2)$ when $U_1 \cap U_2 = \emptyset$; a property called *additivity*. These properties correspond to the definition of a probability or likelihood as a *finite additive measure* on a family (σ-algebra) of measurable sets of $\mathcal{U}$, which is Savage's [4] definition of subjective probability. W is countably additive when $W(\bigcup_{\mathrm{i}=1}^{\infty} U_\mathrm{i}) = \sum_{\mathrm{i}=1}^{\infty} W(U_\mathrm{i})$ whenever $U_\mathrm{i} \cap U_\mathrm{j}$ if $\mathrm{i} \neq \mathrm{j}$. A *purely finitely additive probability* is one that is additive but not countably additive. Savage's subjective probabilities can be purely finitely additive or countably additive. In that sense they include all the probabilities in this article. However as seen below, this article excludes probabilities that are either purely finitely additive, or countably additive, and therefore our characterization of a subjective probability is strictly finer than that of Savage's [4], and different from the view of a measure as a countably additive set function (e.g. De Groot , [21]) The following Axioms were introduced by Villegas [5]; and others for the purpose of obtaining countable additivity.

Monotone Continuity Axiom (MC) (Arrow [3])

For every two events f and g with $W(f) > W(g)$, and every *vanishing* sequence of events $\{E_\alpha\}_{=1,2\ldots}$ (defined as follows: for all α, $E_{\alpha+1} \subset E_\alpha$ and $\bigcap_{\alpha=1}^{\infty} E_\alpha = \emptyset$) there exists N such that altering arbitrarily the events f and g on the set E_i, where $i > N$, does not alter the subjective probability ranking of the events, namely, $W(f') > W(g')$, where f' and g' are the altered events.

This axiom is equivalent to requiring that the probability of the sets along a vanishing sequence goes to zero. Observe that the decreasing sequence could consist of infinite intervals of the form (n, ∞) for $n = 1, 2 \ldots$. Monotone continuity therefore implies that the likelihood of this sequence of events goes to zero, even though all its sets are unbounded. A similar example can be constructed with a decreasing sequence of bounded sets, $(-1/n, 1/n)$ for $n = 1, 2 \ldots,$ which is also a vanishing sequence as it is decreasing and their intersection is empty.

De Groot's Axiom SP_4 (De Groot, [20], Chapter 6, page 71)

If $A_1 \supset A_2 \supset \cdots$ is a decreasing sequence of events and B is some fixed event that is less likely than A_i for all i, then the probability of the intersection $\bigcap_i^\infty A_i$ is larger than that of B.

The following proposition establishes that the two axioms presented above are one and the same; both imply countable additivity.

Proposition 2.1. *A relative likelihood (subjective probability) satisfies the Monotone Continuity Axiom if and only if it satisfies Axiom SP_4. Each of the two axioms implies countable additivity.*

Proof. Assume that De Groot's axiom SP_4 is satisfied. When the intersection of a decreasing sequence of events is empty $\bigcap_i A_i = \emptyset$ and the set B is less likely to occur than every set A_i, then the subset B must be as likely as the empty set; namely, its probability must be zero. In other words, if B is more likely than the empty set, then regardless of how small is the set B, it is impossible for every set A_i to be as likely as B. Equivalently, the probability of the sets that are far away in the vanishing sequence must go to zero. Therefore SP_4 implies Monotone Continuity. Reciprocally, assume that MC is satisfied. Consider a decreasing sequence of events A_i and define a new sequence by substracting from each set the intersection of the family, namely, $A_1 - \bigcap_i^\infty A_i$, $A_2 - \bigcap_i^\infty A_i, \ldots$. Let B be a set that is more likely than the empty set but less likely than every A_i. Observe that the intersection of the new sequence is empty, $\bigcap_i^\infty (A_i - \bigcap_i^\infty A_i) = \emptyset$ and since $A_i \supset A_{i+1}$ the new sequence is, by definition, a vanishing sequence. Therefore by MC $\lim_i W(A_i - \bigcap_i^\infty A_i) = 0$. Since $W(B) > 0$, B must be more likely than $A_i - \bigcap_i^\infty A_i$ for some i onwards. Furthermore, $A_i = (A_i - \bigcap_i^\infty A_i) \cup (\bigcap_i^\infty A_i)$ and $(A_i - \bigcap_i^\infty A_i) \cap (\bigcap_i^\infty A_i) = \emptyset$, so that $W(A_i) > W(B)$ is equivalent to $W(A_i - \bigcap_i^\infty A_i) + W(\bigcap_i^\infty A_i) > W(B)$. Observe that $W(\bigcap_i^\infty A_i) < W(B)$ would contradict the inequality $W(A_i) = W(A_i - \bigcap_i^\infty A_i) + W(\bigcap_i^\infty A_i) > W(B)$, since as we saw above, by MC, $\lim_i W(A_i - \bigcap_i^\infty A_i) = 0$, and $W(A_i - \bigcap_i^\infty A_i) + W(\bigcap_i^\infty A_i) > W(B)$. It follows that $W(\bigcap_i^\infty A_i) > W(B)$, which establishes De Groots's Axiom SP_4. Therefore Monotone Continuity is equivalent to De Groot's Axiom SP_4. A proof that each of the axioms implies countable additivity is in Villegas [5], Arrow [3] and De Groot [20]. □

The next section shows that the two axioms, Monotone Continuity and SP_4, are biased against rare events no matter how catastrophic these may be.

3. The Value of Life

The best way to explain the role of *Monotone Continuity* is by means of an example provided by Arrow [3, Pages 48–49]. He explains that if a is an action that involves receiving one cent, b is another that involves receiving zero cents, and c is a third action involving receiving one cent and facing a small probability of death, then *Monotone Continuity* requires that the third action involving death and one cent should be preferred to the action with zero cents if the probability of death is small enough. Even Arrow says of his requirement "this may sound outrageous at first blush…" (Arrow [3, Pages 48–49]). Outrageous or not, Monotone Continuity (MC) leads to neglect rare events with major consequences, like death. Death is a black swan.

To overcome the bias we introduce an axiom that is the logical negation of MC: this means that sometimes MC holds and others it does not. We call this the *Swan Axiom*, and it is stated formally below. To illustrate this, consider an experiment where subjects are offered a certain amount of money to choose a pill at random from a pile, which is known to contain one pill that causes death. It was shown experimentally (Chanel and Chichilnisky [7]) that in some cases people accept a sum of money and choose a pill provided that the pile is large enough—namely, when the probability of death is small enough—thus satisfying the Monotone Continuity axiom and determining the statistical value of their lives. But there are also cases where the subjects will not accept to choose any pill, no matter how large is the pile. Some people refuse the payment of one cent if it involves a small probability of death, no matter how small the probability may be (Chanel and Chichilnisky, [6, 7]). This conflicts with the Monotone Continuity axiom, as explicitly presented by Arrow [3].

Our Axiom provides a reasonable resolution to this dilemma that is realistic and consistent with the experimental evidence. It implies that there exist catastrophic outcomes such as the risk of death, so terrible that one is unwilling to face a small probability of death to obtain one cent versus nothing, no matter how small the probability may be. According to our Axiom, no probability of death may be acceptable when one cent is involved. Our Axiom also implies that in other cases there may be a small enough probability that the lottery involving death may be acceptable, for example if the payoff is large enough to justify the small risk. This is a possibility discussed by Arrow [3]. In other words: sometimes one is willing to take a risk with a small enough probability of a catastrophe, in other cases one is not. This is the content of our Axiom, which is formally stated as follows.

The Swan Axiom

This axiom is the logical negation of Monotone Continuity: There exist events f and g with $W(f) > W(g)$, and for every vanishing sequence of events $\{E_i\}_{\mathrm{i}=1,2\ldots}$ an $N > 0$ such that altering arbitrarily the events f and g on the set E_i, where $i > N$, does not alter the probability ranking of the events, namely, $W(f') > W(g')$, where f' and g' are the altered events. For other events f and g with $W(f) > W(g)$, there exist vanishing sequence of events $\{E_i\}_{\mathrm{i}=1,2\ldots}$ where for every N, altering arbitrarily the events f and g on the set E_i, where $i > N$, does alter the probability ranking of the events, namely $W(f') < W(g')$, where f' and g' are the altered events.

Definition 3.1. A probability W is said to be *biased against rare events* or *insensitive to rare events* when it neglects events that are small according to Villegas and Arrow; as stated in Arrow [3, page 48]: "An event that is far out on a *vanishing sequence* is "*small*" by any reasonable standards" (Arrow [3, page 48]). Formally, a probability is insensitive to rare events when given two events f and g and any vanishing sequence of events $\{E_j\}$, $\exists N = N(f, g) > 0$, such that $W(f) > W(g) \Leftrightarrow W(f') > W(g')$ for all f', g' satisfying $f' = f$ and $g' = g$ a.e. on $E_j^c \subset R$ when $j > N$, and E^c denotes the complement of the set E.

Proposition 3.2. *A subjective probability satisfies Monotone Continuity if and only if it is biased against rare events.*

Proof. This is immediate from the definitions of both [3, 12]. □

Corollary 3.3. *Countably additive probabilities are biased against rare events.*

Proof. It follows from Propositions 2.1 and 3.2 [9, 12]. □

Proposition 3.4. *Purely finitely additive probabilities are biased against frequent events.*

Proof. See example in the appendix. □

Proposition 3.5. *A subjective probability that satisfies the Swan Axiom is neither biased against rare events, nor biased against frequent events.*

Proof. This is immediate from the definition. □

4. An Axiomatic Approach to Probability with Black Swans

This section proposes an axiomatic foundation for subjective probability that is unbiased against rare and frequent events. The axioms are as follows:

Axiom 1. Subjective probabilities are continuous and additive.

Axiom 2. Subjective probabilities are unbiased against rare events.

Axiom 3. Subjective probabilities are unbiased against frequent events.

Additivity is a natural condition and *continuity* captures the notion that "nearby" events are thought as being similarly likely to occur; this property is important to ensure that "sufficient statistics" exist. "Nearby" has been defined by Villegas [5] and Arrow [3] as follows: two events are *close* or *nearby* when they differ on a *small set* as defined in Arrow [3], see previous section. We saw in Proposition 3.2 that the notion of continuity defined by Villegas and Arrow—namely, monotone continuity—conflicts with the Swan Axiom. Indeed Proposition 3.2 shows that countably additive measures are biased against rare events. On the other hand, Proposition 3.4 and the Example in the appendix show that purely finitely additive measures can be biased against frequent events. A natural question is whether there is anything left after one eliminates both biases. The following proposition addresses this issue.

Theorem 4.1. *A subjective probability that satisfies the Swan Axiom is neither finitely additive nor countably additive; it is a strict convex combination of both.*

Proof. This follows from Propositions 3.2, 3.4 and 3.5, Corollary 3.3 above, and the fact that convex combinations of measures are measures. It extends Theorem 6.1 of Section 6 below, which applies to the special case where the events are Borel sets in R or in an interval $(a, b) \subset R$. □

Theorem 4.1 establishes that neither Savage's approach nor Villegas' and Arrow's satisfy the three axioms stated above. These three axioms require more than the additive subjective probabilities of Savage, since purely finitely additive probabilities are finitely additive and yet they are excluded here. At the same time the axioms require less than the countably subjective additivity of Villegas and Arrow, since countably additive probabilities are biased against rare events. Theorem 4.1 above shows that a strict combination of both does the job.

Theorem 4.1 does not however prove the existence of likelihoods that satisfy all three axioms. What is missing is an appropriate definition of continuity that does not conflict with the Swan Axiom. The following section shows that this can be achieved by identifying an event with its characteristic function, so that events are contained in the space of bounded real-valued functions on the universe space $\mathcal{U}$, $L_\infty(\mathcal{U})$, and endowing this space with the sup norm.

5. Axioms for Probability with Black Swans, in R or (a, b)

From here on events are the Borel sets of the real line R or the interval (a, b). This is a widely used case that make the results concrete and allows to compare the results with the earlier axioms on choice under uncertainty of Chichilnisky [9, 12, 14]. We use a concept of "continuity" based on a topology that was used earlier by Debreu [13] and by Chichilnisky [1, 2, 9, 10, 12, 14]: observable events are in the space of measurable and essentially bounded functions $L = L_\infty(R)$ with the sup norm $\|f\| = \operatorname{ess\,sup}_{x\in R}|f(x)|$. This is a sharper and more stringent definition of closeness than the one used by Villegas and Arrow, since two events can be close under the Villegas-Arrow definition but not under ours, see the appendix.

A subjective probabiliy satisfying the classic axioms by De Groot [20] is called a *standard probability*, and is countably additive. A classic result is that for any event $f \in L_\infty$ a standard probability has the form $W(f) = \int_R f(x) \cdot \phi(x) d\mu$, where $\phi \in L_1(R)$ is an integrable function in R.

The next step is to introduce the new axioms, show existence and characterize all the distributions that satisfy the axioms. We need more definitions. A subjective probability $W : L_\infty \to R$ is called *biased against rare events*, or *insensitive to rare events* when it neglects events that are small according to a probability measure μ on R that is absolutely continuous with respect to the Lebesgue measure. Formally, a probability is insensitive to rare events when given two events f and g $\exists \varepsilon = \varepsilon(f, g) > 0$, such that $W(f) > W(g) \Leftrightarrow W(f') > W(g')$ for all f', g' satisfying $f' = f$ and $g' = g$ a.e. on $A \subset R$ and $\mu(A^c) < \varepsilon$. Here A^c denotes the complement of the set A. $W : L_\infty \to R$ is said to be *insensitive to frequent events* when given any two events f, g $\exists \varepsilon(f, g) > 0$ that $W(f) > W(g) \Leftrightarrow W(f') > W(g')$ for all f', g' satisfying $f' = f$ and $g' = g$ a.e. on $A \subset R$ and $\mu(A^c) > 1 - \varepsilon$. W is called *sensitive* to rare (respectively frequent) events when it is *not insensitive* to rare (respectively frequent) events.

The following three axioms are identical to the axioms in last section, specialized to the case at hand.

Axiom 1. $W : L_\infty \to R$ is linear and continuous.

Axiom 2. $W : L_\infty \to R$ is sensitive to frequent events.

Axiom 3. $W : L_\infty \to R$ is sensitive to rare events.

The first and the second axiom agree with classic theory and standard likelihoods satisfy them. The third axiom is new.

Lemma 5.1. *A standard probability satisfies Axioms 1 and 2, but it is biased against rare events and therefore does not satisfy Axiom 3.*

Proof. Consider $W(f) = \int_R f(x)\phi(x)dx, \int_R \phi(x)dx = K < \infty$. Then

$$\begin{aligned} W(f) + W(g) &= \int_R f(x)\phi(x)dx + \int_R g(x)\phi(x)dx \\ &= \int_R f(x) + g(x) \cdot \phi(x)dx = W(f+g), \end{aligned} \tag{5.1}$$

since f and g are characteristic functions and thus positive. Therefore W is linear. W is continuous with respect to the L_1 norm $\|f\|_1 = \int_R |f(x)|\phi(x)d\mu$ because $\|f\|_\infty < \varepsilon$ implies

$$W(f) = \int_R f(x) \cdot \phi(x)dx = \int_R |f(x)| \cdot \phi(x)dx \leq \varepsilon \int \phi(x)dx = \varepsilon K. \tag{5.2}$$

Since the sup norm is finer than the L_1 norm, continuity in L_1 implies continuity with respect to the sup norm (Dunford and Schwartz, [22]). Thus a standard subjective probability satisfies Axiom 1. It is obvious that for every two events f, g, with $W(f) > W(g)$, the inequality is reversed namely $W(g') > W(f')$ when f' and g' are appropriate variations of f and g that differ from f and g on sets of sufficiently large Lebesgue measure. Therefore Axiom 2 is satisfied. A standard subjective probability is however not sensitive to rare events, as shown in Chichilnisky [1, 2, 9, 10, 12, 14, 23]. □

6. Existence and Representation

Theorem 6.1. *There exists a subjective probability $W : L_\infty \to R$ satisfying Axioms 1, 2, and 3. A probability satisfies Axioms 1, 2 and 3 if and only if there exist two continuous linear functions on L_∞, denoted ϕ_1 and ϕ_2 and a real number λ, $0 < \lambda < 1$, such that for any observable event $f \in L_\infty$*

$$W(f) = \lambda \int_{x \epsilon R} f(x)\phi_1(x)dx + (1-\lambda)\phi_2(f) \tag{6.1}$$

where $\phi_1 \in L_1(R, \mu)$ defines a countably additive measure on R and ϕ_2 is a purely finitely additive measure.

Proof. This result follows from the representation theorem by Chichilnisky [9, 12]. □

Example 6.2 ("Heavy" Tails). The following illustrates the additional weight that the new axioms assign to rare events; in this example in a form suggesting "heavy tails." The finitely additive measure ϕ_2 appearing in the second term in (6.1) can be illustrated as follows. On the subspace of events with limiting values at infinity, $L'_\infty = \{f \epsilon L_\infty : \lim_{x\to\infty}(x) < \infty\}$, define $\phi_2(f) = \lim_{x\to\infty} f(x)$ and extend this to a function on all of L_∞ using Hahn Banach's theorem. The difference between a standard probability and the likelihood defined in (6.1) is the second term ϕ_2, which focuses all the weight at infinity. This can be interpreted as a “heavy tail,” a part of the distribution that is not part of the standard density function ϕ_1 and gives more weight to the sets that contain *terminal* events, namely sets of the form (x, ∞).

Corollary 6.3. *In samples without rare events, a subjective probability that satisfies Axioms 1, 2, and 3 is consistent with classic axioms and yields a countably additive measure.*

Proof. Axiom 3 is an empty requirement when there are no rare events while, as shown above, Axioms 1 and 2 are consistent with standard relative likelihood. □

7. The Axiom of Choice

There is a connection between the new axioms presented here and the Axiom of Choice that is at the foundation of mathematics (Godel, [24]), which postulates that there exists a universal and consistent fashion to select an element from every set. The best way to describe the situation is by means of an example, see also Dunford and Schwartz [22], Yosida [25, 26], Chichilnisky and Heal [27], and Kadane and O'Hagan [28].

Example 7.1 (illustration of a purely finitely additive measure). Consider a possible measure ρ satisfying the following: for every interval $A \subset R$, $\rho(A) = 1$ if $A \supset \{x : x > a$, for some $a \in R\}$, and otherwise $\rho(A) = 0$. Such a measure would not be countably additive, because the family of countably many disjoint sets $\{V_i\}_{i=0,1,\ldots}$ defined as $V_i = (i, i+1] \bigcup (-i-1, -i]$, satisfies $V_i \bigcap V_i = \emptyset$ when $i \neq j$, and $\bigcup_{i=0}^{\infty} V_i = \bigcup_{i=0}^{\infty} (i, i+1] \bigcup (-i-1, -i] = R$, so that $\rho(\bigcup_{i=0}^{\infty} V_i) = 1$, while $\sum_{i=0}^{\infty} \rho(V_i) = 0$, which contradicts countable additivity. Since the contradiction arises from assuming that ρ is countably additive, such a measure could only be purely finitely additive.

One can illustrate a function on L_∞ that represents a purely finitely additive measure ρ if we restrict our attention to the closed subspace L'_∞ of L_∞ consisting of those functions $f(x)$ in L_∞ that have a limit when $x \to \infty$, by the formula $\rho(f) = \lim_{x \to \infty} f(x)$, as in Example 6.2 of the previous section. The function $\rho(\cdot)$ can be illustrated as a limit of a sequence of delta functions whose supports increase without bound. The problem however is to extend the function ρ to another defined on the entire space L_∞. This could be achieved in various ways but as we will see, each of them requires the Axiom of Choice.

One can use Hahn—Banach's theorem to extend the function ρ from the closed subspace $L'_\infty \subset L_\infty$ to the entire space L_∞ preserving its norm. However, in its general form Hahn—Banach's theorem requires the Axiom of Choice (Dunford and Schwartz, [22]). Alternatively, one can extend the notion of a *limit* to encompass all functions in L_∞ including those with no standard limit. This can be achieved by using the notion of convergence along a *free ultrafilter* arising from compactifying the real line R as by Chichilnisky and Heal [27]. However the existence of a *free ultrafilter* also requires the Axiom of Choice.

This illustrates why any attempts to construct *purely finitely additive measures*, requires using the Axiom of Choice. Since our criteria include purely finitely additive measures, this provides a connection between the Axiom of Choice and our axioms for relative likelihood. It is somewhat surprising that the consideration of rare events that are neglected in standard statistical theory conjures up the Axiom of Choice, which is independent from the rest of mathematics (Godel, [24]).

Appendix

Example A.1 (Illustration of a probability that is biased against frequent events). Consider the function $W(f) = \liminf_{x \epsilon R}(f(x))$. This is insensitive to frequent events of arbitrarily large Lebesgue measure (Dunford and Schwartz, [22]) and therefore does not satisfy Axiom 2. In addition it is not linear, failing Axiom 1.

Example A.2 (two approaches to "closeness"). Consider the family $\{E_i\}$ where $E_i = [i, \infty)$, $i = 1, 2, \ldots$. This is a vanishing family because for all $i, E_i \supset E_{i+1}$ and $\bigcap_{i=1}^{\infty} E_i = \emptyset$. Consider now the events $f^i(t) = K$ when $t \in E_i$ and $f^i(t) = 0$ otherwise, and $g^i(t) = 2K$ when $t \in E_i$ and $g^i(t) = 0$ otherwise. Then for all $i, \sup_{E^i}|f^i(t) - g^i(t)| = K$. In the sup norm topology this implies that f^i and g^i are *not* "close" to each other, as the difference $f^i - g^i$ does not converge to zero. No matter how far along the vanishing sequence E^i the two events f^i, g^i differ by K. Yet since the events f^i, g^i differ from $f \equiv 0$ and $g \equiv 0$ respectively only in the set E_i, and $\{E_i\}$ is a vanishing sequence, for large enough i they are as "close" as desired according to Villegas-Arrow's definition of "nearby" events.

The Dual Space L_∞^*: Countably Additive and Finitely Additive Measures

The space of continuous linear functions on L_∞ with the sup norm is the "dual" of L_∞, and is denoted L_∞^*. It has been characterized, for example, in Yosida [25, 26]. L_∞^* consists of the sum of two subspaces (i) L_1 functions g that define countably additive measures ν on R by the rule $\nu(A) = \int_A g(x)dx$ where $\int_R |g(x)|dx < \infty$ so that υ is *absolutely continuous* with respect to the Lebesgue measure, and (ii) a subspace consisting of purely finitely additive measures. A countable measure can be identified with an L_1 function, called its "density," but purely finitely additive measures cannot be identified by such functions.

Example A.3. Illustration of a Finitely Additive Measure that is not Countably Additive
See Example 7.1 in Section 7.

Acknowledgments

This research was conducted at Columbia University's Program on Information and Resources and its Columbia Consortium for Risk Management (CCRM). The author acknowledges support from Grant no 5222-72 of the US Air Force Office of Research directed by Professor Jun Zheng, Arlington VA. The initial results (Chichilnisky [8]) were presented as invited addresses at Stanford University's 1993 Seminar on Reconsideration of Values, organized by Professor Kenneth Arrow, at a 1996 Workshop on Catastrophic Risks organized at the Fields Institute for Mathematical Sciences of the University of Toronto, Canada, at the NBER Conference *Mathematical Economics: The Legacy of Gerard Debreu* at UC Berkeley, October 21, 2005, the Department of Economics of the University of Kansas National Bureau of Economic Research General Equilibrium Conference, September 2006, at the Departments of Statistics of the University of Oslo, Norway, Fall 2007, at a seminar organized by the late Professor Chris Heyde at the Department of Statistics of Columbia University, Fall 2007, at seminars organized by Drs. Charles Figuieres and Mabel Tidball at LAMETA Universite de Montpellier, France December 19 and 20, 2008, and by and at a seminar organized by Professor Alan Kirman at GREQAM Universite de Marseille, December 18 2008. We are grateful to the above institutions and individuals for supporting the research, and for helpful comments and suggestions. An anonymous referee provided insightful comments that improved this article.

References

[1] G. Chichilnisky, "The Limits of Econometrics: Non Parametric Estimation in Hilbert Spaces," *Econometric Theory*, vol. 25, no. 4, pp. 1070–1086, 2009.
[2] G. Chichilnisky, "The Work and Legacy of Kenneth Arrow," in *Encyclopedia of Quantitative Finance*, R. Cont, Ed., pp. 76–82, John Wiley & Sons, Chichester, 2010.
[3] K. J. Arrow, *Essays in the Theory of Risk-Bearing*, North-Holland, Amsterdam, The Netherlands, 1970.
[4] L. J. Savage, *The Foundations of Statistics*, Dover, New York, USA, revised edition, 1972.
[5] C. Villegas, "On quantitative probability σ—algebras," *The Annals of Mathematical Statistics*, vol. 35, pp. 1789–1800, 1964.
[6] O. Chanel and G. Chichilnisky, " The influence of fear in decisions: experimental evidence," *Journal of Risk and Uncertainty*, vol. 39, no. 3, pp. 271–298, 2009.
[7] O. Chanel and G. Chichilnisky, "The value of life: theory and experiments," Working Paper, GREQE, Universite de Marseille and Columbia University, New York, NY, USA, 2009.
[8] G. Chichilnisky, "An axiomatic approach to sustainable development," *Social Choice and Welfare*, vol. 13, no. 2, pp. 231–257, 1996.
[9] G. Chichilnisky, "An axiomatic approach to choice under uncertainty with catastrophic risks," *Resource and Energy Economics*, vol. 22, no. 3, pp. 221–231, 2000.
[10] G. Chichilnisky, "Updating von Neumann Morgenstern axioms for choice under uncertainty," in *Proceedings of a Conference on Catastrophic Risks, the Fields Institute for Mathematical Sciences*, Toronto, Canada, 1996.
[11] J. Le Doux, *The Emotional Brain*, Simon and Schuster, New York, NY, USA, 1996.
[12] G. Chichilnisky, "The topology of fear," *Journal of Mathematical Economics*, vol. 45, no. 12, pp. 807–816, 2009.
[13] G. Debreu, "Valuation equilibrium and Pareto optimum," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 40, pp. 588–592, 1954.
[14] G. Chichilnisky, "Catastrophic risk," in *Encyclopedia of Environmetrics*, A. H. El-Shaarawi and W. W. Piegorsch, Eds., vol. 1, pp. 274–279, John Wiley & Sons, Chichester, UK, 2002.
[15] D. Schmeidler, "Subjective probability and expected utility without additivity," *Econometrica*, vol. 57, no. 3, pp. 571–587, 1989.
[16] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decisions under risk," *Econometrica*, vol. 47, pp. 263–291, 1979.
[17] A. Tversky and D. Kahneman, "Advances in prospect theory: cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.
[18] J. H. Dreze, *Essays on Economic Decisions under Uncertainty*, Cambridge University Press, Cambridge, UK, 1987.
[19] P. L. Bernstein, *Against the Gods: A Remarkable Story of Risk*, John Wiley & Sons, New York, NY, USA, 1996.
[20] M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, NY USA, 1970.
[21] M. H. DeGroot, *Optimal Statistical Decisions*, Wiley Classics Library, John Wiley & Sons, Hoboken, NJ, USA, 2004.
[22] N. Dunford and J. T. Schwartz, *Linear Operators, Part I*, Interscience, New York, NY, USA, 1958.
[23] G. Chichilnisky and H.-M. Wu, "General equilibrium with endogenous uncertainty and default," *Journal of Mathematical Economics*, vol. 42, no. 4-5, pp. 499–524, 2006.
[24] K. Godel, *The Consistency of the Continuum Hypothesis*, Annals of Mathematics Studies, No 3, Princeton University Press, Princeton, NJ, USA, 1940.
[25] K. Yosida and E. Hewitt, "Finitely additive measures," *Transactions of the American Mathematical Society*, vol. 72, pp. 46–66, 1952.
[26] K. Yosida, *Functional Analysis*, vol. 12 of *Die Grundlehren der mathematischen Wissenschaften*, Springer, New York, NY, USA, 4th edition, 1974.
[27] G. Chichilnisky and G. Heal, "Social choice with infinite populations: construction of a rule and impossibility results," *Social Choice and Welfare*, vol. 14, no. 2, pp. 303–318, 1997.
[28] J. B. Kadane and A. O'Hagan, "Using finitely additive probability: uniform distributions on the natural numbers," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 626–631, 1995.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 954750, 15 pages
doi:10.1155/2010/954750

Research Article

Asteroids: Assessing Catastrophic Risks

Graciela Chichilnisky and Peter Eisenberger

Columbia University, New York, NY 10027, USA

Correspondence should be addressed to Graciela Chichilnisky, chichilnisky1@gmail.com

Received 28 January 2010; Accepted 9 May 2010

Academic Editor: Ricardas Zitikis

Copyright © 2010 G. Chichilnisky and P. Eisenberger. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We evaluate two risk profiles: (i) global warming risks and (ii) collisions with asteroids that can cause the extinction of our species. The expected values computed for these two risks suggest that no action will be taken to avoid extinction. The result is somewhat counterintuitive, but it is typical of the results of using classic decision theory to evaluate catastrophic risks in the distant future, see the study by Posner (2004). We establish why expected value is insensitive to catastrophic risks see the study by Chichilnisky (1996), and use another criterion to evaluate risk based on axioms for choice under uncertainty that update the classic Von Neumann theory and require equal treatment for rare and frequent events. Optimizing according to the new criterion is shown to be equivalent to optimizing expected utility with a restriction on the worst outcome in the case of a catastrophe. The evaluation obtained from the new criterion seems more intuitively plausible, and suggests a more practical and realistic approach to catastrophic risks: optimizing expected value while minimizing losses in the case of a catastrophe.

1. Asteroids

Sixty five million years ago, an asteroid crashed into earth. Global winds distributed the dust throughout the atmosphere, blocking sunlight, and many life forms that relied on the sun eventually perished. In a short period of time, experts believe, the mighty dinosaurs that dominated our planet went extinct. Realistically the same fate awaits us. Over 99.99% of the species that have ever existed are now extinct [1, 2]. If our species survives long enough, we will be exposed to an asteroid and could suffer the same fate as the dinosaurs. The data suggests that asteroids of that caliber will hit our planet on average once every 100 million years [2]. The last one was 65 million years ago. Under current conditions, when the next one hits the earth, humans and many other species could go extinct.

What should we do about this threat to our survival and others like it? And if the issue is serious, why is this issue getting so little attention whereas the less catastrophic threat of global warming is in the news almost daily?

The purpose of this paper is to provide answers to these questions. We examine systematically how to deal with *catastrophic risks* such as asteroid impacts, which are small-probability events with enormous consequences, events that could threaten the survival of our species, and compare their treatment with risks like global warming that are more imminent and familiar but possibly less catastrophic.

The task is not easy. Classic tools for risk management are notoriously poor for managing catastrophic risks, (see Posner [2] and Chichilnisky [3, 4]). There is an understandable tendency to ignore rare events, such as an asteroid impact, which are unlikely to occur in our lifetimes or those of our families [2, 5]. Yes this is a questionable instinct at this stage of human evolution where our knowledge enables to identify such risks. Standard decision tools make this task difficult. We show using the existing data that a major disturbance caused by global warming of less than 1% of GDP overwhelms in expected value the costs associated with an asteroid impact that can plausibly lead to the extinction of the human species. We show that the expected value of the loss caused by an asteroid that leads to extinction—is between \$500 million and \$92 billion. A loss of this magnitude is smaller than that of a failure of a single atomic plant—the Russians lost more than \$140 billion with the accident at Chernobyl—or with the potential risks involved in global warming that is between \$890 billion and \$9.7 trillion [2]. Using expected values therefore we are led to believe that preventing asteroid impacts should not rank high in our policy priorities. Common sense rebels against the computation we just provided. The ability to anticipate and plan for threats that have never been experienced by any current or past member of the species and are unlikely to happen in our lifespans, appears to be unique to our species. We need to use a risk management approach that enables us to deal more effectively with such threats [2]. To overcome this problem this paper summarizes a new axiomatic approach to catastrophic risks that updates current methods developed initially by John Von Neumann, see Chichilnisky [3, 4, 6–9], and offers practical figures to evaluate possible policies that would protect us from asteroid impacts. Our conclusion is that we are underinvesting in preventing the risk of asteroid like threats. Much can and should be done at a relatively small cost; this paper suggests a methodology and a range of dollar values that should be spent to protect against such risks to help prevent the extinction of our species.

2. Catastrophes and the Survival of the Species

A catastrophe is a rare event with enormous consequences. In a recent book, Posner [2] classifies catastrophes into various types, each of which threats the survival of our species. He uses a classic approach to value the importance of a risk by quantifying its *expected value*, namely, the product of the probability times the loss. For example, the expected value of an event that occurs with ten percent probability and involves \$1 billion loss is $\$10^9 \times 10^{-1} = \$100$ million. This approach is used by actuaries to price the cost of life insurance policies, and is also by law the measure used in US Congress when evaluating budget plans with uncertain outcomes. The notion of *expected value* started with Von Neumann and Morgenstern about 60 years ago [10], and it is based on their axioms or principles for decision making under uncertainty formalized in [11, 12]. Posner [2] uses the concept of *expected value* to evaluate risks but warns the reader about its weaknesses for evaluating catastrophic risks (see Posner [2, Chapter 3, pages 150–154]). This weakness is exposed in the case of asteroids, when we ask how much we should invest in preventing the impact of an asteroid that can destroy all of the earth's economic value forever. Posner [2] argues that *expected value* does not capture the true impact of such a catastrophe; that something else is at stake. Because of his loyalty to the concept of expected value, which does not work well in these cases, Posner appears to be arguing that rationality does not work in the case of catastrophes, that we cannot deal rationally with small probabilities events that cause such large and irreversible damage.

Perhaps the problem is not one of rationality. There may be a *different* rationality needed when considering the long-range future of the species. It could be that *expected value* is a good measure for evaluating risks that have a good chance to occur in our lifetime, but not for evaluating risks that are important but have essentially a zero chance to occur while we are alive. For such risks we may need another approach overall, for both the present and the future. In our current state of evolution it would seem useful to oppose a human tendency based on our hunter-gatherer origins to give preference to immediate outcomes as opposed to more distant ones; see the study by McClure et al. [5]. When using *expected value* the response we obtain seems to clash with our intuition because the probabilities involved are so small that they render the computation almost meaningless, as seen numerically in the examples provided below. The experimental evidence summarized below provides further support for this view.

3. Experimental Evidence and Alternative Approaches

Expected utility optimization derives from Von Neumann's (NM) axioms, but it is well known for sometime that it conflicts with the experimental evidence on how humans choose under uncertainty; for example, see the studies by Allais [13], Machina, [14, 15], Tversky and Wakker [16]. Problems arise when there are infrequent events involved; examples are weather catastrophes like hurricanes or mass extinctions. Similar types of conflicts appear when using the standard criterion of *present value optimization* for choosing among projects that evolve through time. Other problems arise when the plans involve very long time horizons as in the disposal of nuclear waste (Heal [17]). While the problem areas mentioned above are quite different, they all share a common mathematical root: the relative insensitivity of the classic axioms of choice towards (1) small-probability events (2) the far away future [3, 4, 6, 17]. The mathematical structure of the problem is the same in all cases: it arises from the use of "normal" distributions; the "bell curves" to describe the frequency with which we expect everything to occur from weather events to returns on investments or corporate profits. Normal distributions arise when many independent events contribute to some outcome. However when there are unexpected interconnections or catastrophic events, normal distribution can understate (1) the role of small-probability events (2) the role of events that are very far into the future. We formalize this problem below and provide a solution.

Taking a leaf from Von Neumann and Morgenstern's book, Chichilnisky [3, 4, 6] reconsidered the foundations of the expected value approach, which are the VNM axioms for choice under uncertainty, see also Arrow [11] and Hernstein and Milnor [12]. A first step is to show that classic axioms can be "biased" against small-probability events, as was established by Chichilnisky in [6]. She introduced new axioms for choice under uncertainty that require more symmetry in the treatment of small and large probability events [3, 4]. The new axioms coincide with those of Von Neumann and Morgenstern when the events involved have "normal" probabilities, for example, when they are likely to occur in our lifetime. But the new axioms give rise to a rather different decision making criterion when the probabilities involved are extremely small, or equivalently when the events are only likely to occur in a very long future. The two sets of axioms are consistent with each other for “normal” events while they are quite different on catastrophic events. How can this be?

A somewhat far-fetched analogy is the relationship between classical mechanics and general relativity. The former applies to “normal scales” that are closer to own reality on earth, while the latter applies to large-scale phenomena involving astral bodies. Both are correct in their respective spheres, and neither contradicts the other. The same could be the case with the Von Neumann-Morgenstern and the Chichilnisky’s axioms. The next section presents the new axioms. It has been shown empirically and theoretically (Posner [2] and Chichilnisky [3]) that standard tools of decision making under uncertainty are ill suited to evaluate such risks, more on this below.

In sum: the standard approaches do not provide a satisfactory answer and we provide here an alternative approach to risk management that seems better suited to the management of catastrophic risks and risks that are most likely to occur in the very distant future. This approach has an axiomatic treatment that parallels Von Neumann’s theory of choice under uncertainty, but extends it requiring equal treatment for frequent and rare events.

The next section provides empirical motivation for the new approach by comparing it with expected utility in two risk profiles: asteroids and global warming risks.

4. Two Risk Profiles: Asteroids and Global Warming

In September 16, 2000, the British Task Force on Potentially Hazardous Near Earth Objects (NEOs) produced a report classifying asteroids risks by their size, energy yield, and average interval between impacts. Large-mass extinctions—for example, the Cretaceous Terciary Geological boundary—follow from the impact of asteroids of 16 km in diameter, which occur on the average once every 100 million years, and threaten the survival of all advanced life forms on the planet of which 99.9% have already gone extinct [18]. Below we compare the risk profile presented by asteroids with global warming risks.

(i) An asteroid impact of this magnitude occurs on average once every 100 million years.

(ii) It produces damage of about $120 trillion [2], obliterating all human-produced value in the planet.

(iii) The damage is permanent—it continues annually for about 1 billion years, the expected lifetime of our planet before it is destroyed by our sun becoming a red star.

(iv) Existing observations indicate that such an impact will not take place in the next 30 years.

Below we compare this risk with the risk of “global warming” with the following simplified profile.

(i) The probability of global warming is 1, namely, it is happening.

(ii) The best estimate is that it will produce damage that is calculated for the catastrophic case to bring in a permanent loss of about $2 trillion a year loss in the US and globally about $8 trillion a year; see, for example, the study by (Posner in [2]). There is no consensus on whether the gradual or the catastrophic case for global warming is more likely.

Before examining the two risk profiles, we explain the connection between rare events and events in the distant future.

5. Decisions over Time and under Uncertainty

We adopt a simple approach that views “choices over time” and “choices under uncertainty” as two aspects of one and the same phenomenon.

The *probability* of an event is viewed as the *frequency* with which this event is likely to occur through time. (A word of warning is in order. This is not the only approach to defining probabilities—indeed many people object to it because of it views reality as an experiment that can repeat itself. Yet for our purposes here the approach has an important advantage in that it simplifies matters and at the same time generalizes the results.) For example, drawing ”heads” with a fair coin is an event with probability 0.50 because in repeated flipping of the coin, “tails” tend to occur 50% of the time. The appearance of ”heads” is thus a relatively frequent event, one that will on average occur one out of every two trials. If we flip a coin every year, for example, heads will occur almost surely in our lifetime. In this context, high frequency over time translates into high-probability and vice versa. Low-frequency events translate into low probability and vice versa. This way we treat “time” and “uncertainty”. as two aspects of the same phenomenon.

We saw that high-probability events are those that occur frequently in time. In the case of asteroids, we know with certainty that at some point in time one will hit the earth and destroy most life on the planet unless we take action. The probability of such destructive event sometime in the future is one, although the event is so infrequent that the probability is essentially zero in any person’s lifetime.

A *catastrophe* has been defined in [2, 6] as an event with enormous negative consequences—such as the extinction of the species—an event that may occur with probability one sometime in the distant future, but has nearly zero probability of occurring during the lifetime of any one individual. There is basically zero risk that the “catastrophe” will occur in our lifetime, although we know for sure it will occur at some point in the future. It is possible although very unlikely that we will pick up the papers tomorrow and read that some astronomer has discovered a massive comet taking deadly aim at our planet. With this definition, it becomes clear that dealing with catastrophes is the same as dealing with events that take place in the distant future. It is well known that the losses from events that take place in the very long-run future are “discounted” in an exponential manner to obtain a “present discounted value”, and that this exponential discount renders long-term losses almost meaningless. Having connected “choices over time” and “choices under uncertainty”, this explains why expected value, which is the parallel to “present discounted value”, leads to counterintuitive evaluation of catastrophic losses or for that matter distort outcomes generally. Next we provide a formal framework.

6. The Mathematics of Risk

A system is in one of several possible states that can be described by real numbers. To each state $s \in R$ there is an associated outcome, so that one has $x(s) \in R^N,\ N \geq 1$. A description of outcomes across all states is called a *lottery* $x : R \to R^N$. The space of all lotteries is therefore a function space L (identified in the following by the space of all essentially bounded real-valued functions with the "sup norm", denoted L_∞). Under conditions of uncertainty one makes decisions by ranking lotteries in L.

Von Neumann-Morgenstern (NM) axioms provided a mathematical formalization of how to rank or order lotteries. Optimization according to such an order is called *expected utility* (EU) and defines standard decision making under uncertainty. The main result from the NM axioms is that the decision procedure is obtained by optimizing a function of the following form:

$$W(x) = \int_{s\in R} u(x(s))d\mu(x), \tag{6.1}$$

where the real line R is the state space, $x : R \to R^N$ is a "lottery", $u : R^N \to R$ is a *utility function* describing the utility provided by the outcome of the lottery in each state s, $u(s)$, and where $d\mu(x)$ describes a countably additive measure over sets of states (or events) in R that determines their relative frequency. Arrow [11] explains why utility functions must be essentially bounded so as to overcome the St. Petersburg paradox. This implies that the space L of utility values provided by lotteries is L_∞, namely, it is the space of measurable and (essentially) bounded functions (An essentially bounded function is a measurable function that is bounded except perhaps on a set of measure zero.) on the line R. Using the EU criterion, a lottery x is ranked above another y if and only if W assigns to x a larger real number:

$$x \succ y \Longleftrightarrow W(x) > W(y), \tag{6.2}$$

where W satisfies (6.1). The optimization of expected utility (EU) is a widely used procedure for evaluating choices under uncertainty. (The Euler-Lagrange equations are typically used to characterize optimal solutions.)

In the following examples we consider the space of lotteries to be the space of all continuous linear real-valued functions, L_∞ with the sup norm, and the *dual* of L_∞, denoted L_∞^*, consists of all continuous real-valued functions on L_∞. L_∞^* includes integrable functions on R as well as purely finite additive measures [19–21] that are not representable by functions, for example a measure that assigns measure zero to all bounded measurable subsets of the line. Other examples are provided below. (An example of a purely finitely additive measure is the continuous linear real-valued function $\phi : L_\infty \to R$ defined by $\phi(f) = \lim_{s\to\infty} f(s)$ on all functions that have such a limit, and extended by using Hahn-Banach theorem to the rest.)

6.1. Catastrophic Risks

A *catastrophic risk* is a rare event leading to major widespread losses. Expected utility undervalues such risks: Chichilnisky [3, 4, 6] showed formally that using expected utility criteria underestimates catastrophic risks, and by doing so conflicts with the observed evidence of how humans evaluate such risks if they are likely to occur in their lifetimes. In order to formalize the problem we need some definitions.

Definition 6.1. A ranking $W : L_\infty \to R$ is said to be *insensitive to rare events* when $W(x) > W(y) \Leftrightarrow W(x') > W(y')$ for any two lotteries x' and y' that are obtained by modifying arbitrarily x and y on any set of states $S \subset R$ with an arbitrarily small-probability $\varepsilon = \varepsilon(x, y)$. Formally,

W is *insensitive to rare* events if $\forall x, y\ \exists \varepsilon = \varepsilon(x, y) : W(x) > W(y) \Leftrightarrow W(x') > W(y')$ for all x', y' satisfying

$$y' = y \quad \text{a.e. on } S^c \subset R, \qquad x = x' \quad \text{a.e. on } S^c \subset R, \quad \text{where } \mu(S) < \varepsilon. \tag{6.3}$$

Otherwise, W is said to be *sensitive* to rare events.

Definition 6.2. W is *insensitive to frequent events* if

$$\begin{gathered} \forall x, y\ \exists \varepsilon = \varepsilon(x, y) : W(x) > W(y) \Longleftrightarrow W(x') > W(y') \quad \forall x', y' \text{ s.t.} \\ x' = x, \quad y' = y \quad \text{a.e. on } S^c \subset R : \mu(S) > 1 - \varepsilon. \end{gathered} \tag{6.4}$$

Otherwise, W is called *sensitive* to frequent events.

Mathematically, the problem with expected utility is that it is insensitive to rare events no matter how catastrophic these may be.

Proposition 6.3. *Any expected utility function $\int_{s \in R} u(c(s))\gamma(s)ds$, where $\gamma(s) \in L_1(R)$ is insensitive to rare events.*

For a proof see the theorem in Chichilnisky [6].

6.2. New Axioms of Choice Under Uncertainty

Introduced in [3, 6] the following axioms contrast with Von-Neumann Morgenstern's axioms (NM) in that they treat symmetrically rare and frequent risks. They postulate that the ranking of lotteries $W : L_\infty \to R$ must satisfy the following.

Axiom 1. Sensitivity to rare events.

Axiom 2. Sensitivity to frequent events.

Axiom 3. Linearity and continuity of the ranking W.

Axioms 2 and 3 are standard; they satisfied, for example, by expected utility EU. However Axiom 1 is not satisfied by EU (Proposition 6.3 above).

To clarify the meaning of these axioms, the following are examples of rankings W that do *not* satisfy our axioms.

Example 6.4. Consider a *criterion of choice* $W : L \to R$ that ranks lotteries assigning measure zero to any bounded set in R [19–21]. Such functionals are ruled out by Axiom 2 which requires sensitivity to frequent events.

Example 6.5. Expected utility maximization is ruled out, as is shown in Proposition 6.3 above (see also [6]), because it does not satisfy Axiom 1.

Like the NM axioms, the new Axioms 1, 2, and 3 lead to a representation theorem.

Theorem 6.6. *There exist ranking criteria $\Psi : L_\infty \rightarrow R$ that satisfy all three axioms. Any criterion that satisfies the axioms is a convex combination of an expected utility plus a purely finitely additive measure focused on catastrophic events, for example:*

$$\Psi = \int_R u(c(s))\lambda(s) + \lambda(\Phi(c(s))), \tag{6.5}$$

where $c(s)$ describes the value of the lottery in state $s \in R$, $\gamma(s)$ is an integrable real-valued function on the line R, for example, $\lambda(s) = e^{-\delta s}$, and $u : R \rightarrow R$ is a bounded utility function.

The first term is thus an expected utility with an $L_1(R)$ density function $\gamma(s)$, and the second term is a purely finitely additive measure such as $\Phi(s) = \lim_{c\to\infty} c(s)$ for lotteries that have such a limit and extended otherwise to all lotteries by Hahn-Banach's theorem. For a proof see the study by Chichilnisky in [6]. (The optimization of functionals such as Ψ is not amenable to standard tools of calculus of variations, which must be developed in new directions; see, for example, the studies by Chichilnisky [3, 6] and Heal [17].)

A recent result established that the new criterion in (6.5) is a way to formalize the notion of optimizing expected utility while bounding the worst outcome in the case of a catastrophe:

Theorem 6.7. *Optimizing the ranking criterion in* (6.5) *is equivalent to optimizing an expected utility function* $\int_R u(c(s))\lambda(s)$ *subject to a constraint on the possible loss in case of a catastrophe.*

For a proof see [4].

An interpretation of his theorem is as follows: the first term $\sum_{s=1}^{\infty} \lambda^{-s} u(x(s))$ is an integral operator with a countably additive kernel $\{\lambda^{-s}\}_{s\in Z}$ which emphasizes the weight of frequent events in the ranking of a lottery $x \in l_\infty$. The second purely finitely additive part $\Phi u(x(s))$ assigns positive weight to rare events, which have small-probability according to μ. Both parts are present, so Ψ is sensitive to small and to large probability events. Catastrophic risks are therefore ranked more realistically by such functionals.

Next section applies this framework for evaluating risks to the two risk profiles described above, asteroids and global warming.

7. Evaluating Asteroid Impacts

We first use standard methods for calculating the value of the damage caused by a risk calculated by its present value, which brings future value into the present (Posner [2] for both, and in particular page 150 for the latter). The reader should refer to the risk profile of an asteroid impact presented above.

(i) In the case of an asteroid as described above, the expected cost of the loss in any given year is obtained by multiplying \$120 trillion or $\$120 \times 10^{12}$—by the probability of the loss, which occurs on average once every 100 million years, and is therefore 10^{-8}. Once the loss occurs, however, it is assumed to be permanent. Therefore the expected value of the loss in year N is

$$\begin{aligned}\sum_{t=1}^{\infty}\left(120\times 10^{12}\times 10^{-8}\right)\cdot\delta^{N+t} &= \left(120\times 10^{12}\times 10^{-8}\times\delta^{N}\right)\sum_{t=1}^{\infty}\delta^{t}\\ &= \left(120\times 10^{12}\times 10^{-8}\times\delta^{N}\right)\left(\frac{\delta}{1-\delta}\right),\end{aligned}\tag{7.1}$$

where δ is the time "discount factor", $0<\delta<1$, and $1-\delta$ is the "discount rate".

If the risk does not occur in year N, then it can occur in year $N+1$, and if not in year $N+2$, $N+3$, and so forth, and each time it occurs, it lasts permanently. Therefore the total risk is the sum of the risk of it occurring in year 30, plus the risk of it occurring in year 31, plus the risk of it occurring in year 32, and so forth, namely,

$$\left(120\times 10^{12}\times 10^{-8}\times\delta^{N}\right)\left(\frac{\delta}{1-\delta}\right)\sum_{j=1}^{\infty}\delta^{j} = \left(120\times 10^{12}\times 10^{-8}\times\delta^{N}\right)\left(\frac{\delta}{1-\delta}\right)^{2}.\tag{7.2}$$

At a 5% discount rate $\delta = 0.95$, and the total expected discounted value of the loss from such as asteroid is

$$120\times 10^{12}\times 10^{-8}\times\frac{95}{100}^{30}\times\left(\frac{95/100}{1-95/100}\right)^{2} = 9.2982\times 10^{7}\quad\text{or about \$92 million.}\tag{7.3}$$

At a 10% discount rate, that value is

$$120\times 10^{12}\times 10^{-8}\times\frac{90}{100}^{30}\times 81 = 4.1204\times 10^{6}\quad\text{or about \$4 million.}\tag{7.4}$$

At a 3% discount rate, the value is

$$120\times 10^{12}\times 10^{-8}\times\frac{97}{100}^{30}\times 32^{2} =: 4.9276\times 10^{8}\quad\text{or about \$500 million.}\tag{7.5}$$

These values pale by comparison with the estimated value of other losses such as global warming, which are estimated to be in the tens of trillions, as shown below. In all cases, therefore, it appears to makes sense to allocate more funding to the global warming problem than to the problem of preventing asteroid impacts; more on this below.

7.1. Evaluating Global Warming

The United Nations Intergovernmental Panel on Climate Change (IPCC) found in 1996 that human-induced global warming is already occurring. There are two main scenarios for the damages that global warming will cause: (1) catastrophic global warming effects, and (2) a more gradual buildup of damages

Scenario 1 (Catastrophic Global Warming). The catastrophic scenario is described by Posner in [2, page 181], as follows. There is rapid increase in temperature, which produces a damage that is calculated to bring in a permanent loss of about \$2 trillion a year in the US and globally about \$8 trillion a year. Considering the risk profile already established above (cf. also [Posner, 1992, page 181]), the present discounted value of such a disaster at a 3% discount rate is

$$2 \times 10^{12} \times \frac{\delta}{1-\delta} = 2 \times 10^{12} \times \frac{97}{3} = 6.4667 \times 10^{13} \quad \text{or about \$65 trillion.} \tag{7.6}$$

At a 5% discount rate the number is

$$2 \times 10^{12} \times \frac{95}{5} =: 3.8 \times 10^{13} \quad \text{or about \$38 trillion,} \tag{7.7}$$

and at a 10% discount rate the number is

$$2 \times 10^{12} \times 9 =: 1.8 \times 10^{13} \quad \text{or about \$18 trillion.} \tag{7.8}$$

Scenario 2 (Gradual Buildup of Damages). In the second scenario global warming is also here today, but temperature increases slowly and its damages increase for about 100 years to reach 1% of the planet's GDP. Global GDP is calculated to be about \$120 trillion then (the same number used in the asteroid risk). After we reach maximum damage, we consider various possibilities going forward: (1) the annual damage remains the same *a perpetuity*, and (2) damages decrease slowly and disappear 100 years later. Let us compute using standard technique the present discounted value of the losses.

In the first case using a 3% discount rate we obtain

$$\sum_{i=1}^{\infty} \frac{10^{12}}{100} \times i \times \frac{97}{100}^{\,i} = 1.077\,8 \times 10^{13}, \quad \text{which is about \$10 trillion} \tag{7.9}$$

using a 5% discount rate

$$\sum_{i=1}^{\infty} \frac{10^{12}}{100} \times i \times \frac{95}{100}^{\,i} = 3.8 \times 10^{12} \quad \text{which is about \$3.8 trillion,} \tag{7.10}$$

and using a 10% discount rate,

$$\sum_{i=1}^{\infty} \frac{10^{12}}{100} \times i \times \frac{9}{10}^{\,i} =: 9.0 \times 10^{11}, \quad \text{which is about \$900 billion.} \tag{7.11}$$

In the second case, when the damage gradually decreases until it vanishes after 100 years following its maximum impact we have using a 3% discount rate

$$\left[\sum_{i=1}^{100}\frac{10^{12}}{100}\times i\times\frac{97}{100}^{i}\right]+\left[\sum_{i=2}^{100}\frac{10^{12}}{100}\times(100-(i-1))\times\frac{97}{100}^{100+i}\right]=9.7456\times10^{12}\sim\$9.7\text{ trillion},\tag{7.12}$$

using a 5% discount rate

$$\left[\sum_{i=1}^{100}\frac{10^{12}}{100}\times i\times\frac{95}{100}^{i}\right]+\left[\sum_{i=2}^{100}\frac{10^{12}}{100}\times(100-(i-1))\times\frac{95}{100}^{100+i}\right]=3.7506\times10^{12}\sim\$3.7\text{ trillion}\tag{7.13}$$

and using a 10% discount rate

$$\left[\sum_{i=1}^{100}\frac{10^{12}}{100}\times i\times\frac{9}{10}^{i}\right]+\left[\sum_{i=2}^{100}\frac{10^{12}}{100}\times(100-(i-1))\times\frac{9}{10}^{100+i}\right]=8.9993\times10^{11}\sim\$890\text{ billion}.\tag{7.14}$$

As was indicated above, in all cases, and with all three discount rates, 3%, 5%, and 10%, the global warming problem overwhelms in terms of present discounted values the costs involved with asteroid impacts. This is despite the fact that even in the noncatastrophic case global warming decreased GDP by a small fraction, only 1% and only after 100 years.

8. Comparing Global Warming and Asteroid Impacts

Using expected values we are led to believe that preventing asteroid impacts should not rank high in our policy priorities. The results from the numerical computations provided above to evaluate the risks of asteroids and global warming seem counterintuituve. How can it be that a major disturbance caused by global warming—even when we take very conservative estimated losses of less than 1% of GDP building up slowly over 100 years—overwhelm the costs associated with an asteroid impact that can plausibly lead to the extinction of the human species? The expected value of the loss caused by an asteroid that leads to extinction—is between \$500 million and \$92 billion, as seen above. A loss of this magnitude is smaller that of a failure of a single atomic plant—the Russians lost more than \$140 billion with the accident at Chernobyl—or with the potential risks involved in global warming that is between \$890 billion and \$9.7 trillion. Common sense rebels against the computation we just provided. Let us use other examples for comparison. In year 2004, the profits of the 10 biggest oil companies were about \$100 billion. It seems therefore unreasonable to think of losses from asteroid impacts as valued between \$500 million and \$92 billion. At this rate, it seems difficult to believe that we will ever do anything about averting human extinction, since current priorities will always outweighed such infrequent events, no matter how important they may be. Is there anything wrong with this argument? The alternative is to use the evaluation criterion arising from Axioms 1, 2, and 3 above. In view of the representation theorem, the next section utilizes a ranking function Φ as defined in (6.5), and to make the computation explicit, provides a plausible number for the parameter μ that appears in the definition of Φ above.

9. Catastrophes and the Survival of the Species

The axioms proposed here lead us to evaluate catastrophic risks by a formula that adds to the present expected value a second term that focuses on rare events. To focus on catastrophes that involve extinction of the human species, we evaluate the cost of an event using a sum of the presented expected value plus the cost of extinction of the species. The most conservative scenario for the cost of extinction of the species (Posner [2]) is when everyone alive today dies without warning, and at probabilities that are so small that the value of a human life is computed (according to experimental evidence) at about \$50, 000. Recall (Posner [2]) that the value of life decreases with the probability of death, so this number corresponds to events with probabilities lower that 1 in 100 million. At such small probabilities, with the current population, the species extinction event amounts to \$600 trillion. We may therefore assume that, in the following mathematical expression for our criterion,

$$\Psi(x) = \sum_{s=1}^{\infty} \lambda^{-s} u(x(s)) + \mu(\Phi u(x(s))), \tag{9.1}$$

the expression $\Phi u(x(s) = \$600$ trillion.

To use this approach, we need to specify now a value for μ, which is the "weight" one gives to extremely small-probability events in one's decision making. It is clear that the weight to give to the second term—which addresses the value of the low probability and thus likely distant event—is somewhat subjective. Those concerned about the long-term fate of our species may argue that we should consider it equally to the more familiar current—and likely more frequent threat like global warming. Others may argue that our capability to deal with such a threat will improve with time and therefore we should not spend too much on a threat that will most likely only occur in the distant future. For this discussion we will just note that many people would say that it is a mistake to spend money on something that will occur millions of years from now devalue the threat for an incorrect reason. They incorrectly conclude that something that occurs only every 100 million years on average will not occur in their lifetime. But it can, in fact the likelihood is the same every year.

Example 9.1. Valuating the Parameter μ: An Illustration.

Consider the criterion Φ defined in (6.5) above. In the studies by Chichilnisky [4, 7–9], we assume that we give the catastrophic event a weight of only 1 in 100,000, namely,

$$1 - \mu = 10^{-5}, \quad \text{or equivalently} \quad \mu = 1 - 10^{-5}. \tag{9.2}$$

On this basis we can compare the global warming scenario with the asteroid collision scenario. One takes into consideration the following.

(1) Neither of the two cases of global warming—abrupt or gradual—involve human extinction.

(2) The asteroid impact considered here does involve extinction of the human species.

Under those conditions, the total cost involved in global warming is (approximately) \$66 trillion at 3% discount rates, as shown above, while the total cost involved in an asteroid impact (neglecting the presented discounted value which is no larger than \$500 million as shown above), is about

$$\$600 \times 10^{12} \times 10^{-5} = 6 \times 10^{9} = 6 \text{ billion}. \tag{9.3}$$

Under the conditions, therefore, the yearly investment in prevention of asteroid impacts should be about 1/10 of the yearly investment in prevention of global warming, which is currently \$1.7 billion (Posner [2, page 182]) leading to \$170 million, while the current expenditures are instead \$3.9 million, requiring therefore to be increased by a factor of about 60.

Our rational decision maker who values the future of the species and understands what probabilities really mean, could go through the following simple analysis. For any value of μ even close to one-half the expected value we have calculated makes asteroids more threatening than global warming that is attracting all the attention of policy makers and the public today. In one sense this is satisfying since we would like to believe that we would give great value to prevent our extinction. However, we used the number of US\$300 trillion ($\mu = 1/2$) for the expected value and argued that it is what we should spend to defend against extinction. This does not seem intuitively correct for many reasons, not the least of which is that we would have no resources left to do anything else. The answer to this dilemma is to recognize that what we are really interested in is utility loss from extinction rather than expected value for the dollars we allocate. This view can help us achieve an intuitively pleasing answer that we should spend as much money today on defenses against extinction as can be usefully transferred into improved protection. In the case of asteroids based on current estimates many experts believe this might be only about 10 times what we are now spending which is about US\$30 million dollars. This is a small number and the corrected valuation of the risk is high enough that we should need no further analysis to decide to increase our efforts now and when new opportunities become available in the future.

10. Conclusions

We believe that the above analysis is the beginning of a much more extensive assessment and research about our response to all kinds of catastrophic risks. Recent results provide ways to enhance our subjective judgments about the value of μ, which is approximated by the marginal utility of avoiding extinction near the catastrophe, see the study by Chichilnisky in [4].

Other methods could include the application of Bayesian analysis involving experts who understand the nature of the threats as well as the correct meaning of low probability events. A Bayesian approach can be helpful to determine both the true risk profile and the most plausible utility function for the use of resources to combat a given threat. Such evaluations identify not only high expected value but also high utility. If there are very expensive things we can do to prevent the risk the the allocations of a large amount of resources may be warranted and the problem becomes more complicated. Our political leaders will need to make the more difficult choices between meeting todays' needs compared with the need to defend against distant catastrophic threats. This is not a new challenge since we and other nations spend a significant part of our resources to defend against the threat of nuclear war or the nuclear winter that would follow it. What is new is that now we recognize that many serious threats like those arising from glaciation, asteroid impact, and biodiversity loss are unlikely to occur within our lifetimes, yet we do not want to wake up one day and find that we are facing the impact of what was an avoidable catastrophic risk. Furthermore the same type of deficiency in our approach also exists for very rare events like tsunamis and earthquakes also leading to a poor allocation of resources, as was likely the case for the 2005 Asian tsunami. This work provides a framework to address these threats in a way that agrees with our intuition. We would like to allocate resources in a way that can be useful in reducing the catastrophic threats we face.

In conclusion we offer another perspective that might also be useful for understanding why it is now that we are confronting the dilemmas. An analogy might help. Early on nobody spent a lot of money on personal insurance to protect him/herself. As we gained more knowledge of the risks we face and as we became affluent enough we decided to spend increasing amounts of money on insurance. In a similar way our species only recently has obtained the knowledge of some of the catastrophic risks we face and developed ways to cope with them. For the moment we are seriously underinsured so any way that we can do useful things to reduce our risk we should do so. Someday in the future we may be challenged as we were doing the cold war to decide between present risks and future ones.

Acknowledgments

This paper reports the research performed at the Columbia Consortium for Risk Management (CCRM), Columbia University, New York; see Columbia University Program on Information and Resources (PIR), http://columbiariskmanagement.org/leadership.php. Some of the references are available at http://www.chichilnisky.com/. The authors are indebted to Duncan Foley for valuable comments and suggestions. The authors acknowledge support from the US Air Force, research grant FA9550-09-1-0467 "Optimal Statistical Decisions with Catastrophic Risks" from AFOSR, Arlington VA to Columbia Consortium for Risk Management (CCRM) directed by Graciela Chichilnisky and Columbia University's Department of Economics, New York, NY 10027.

References

[1] D. M. Raup, *Extinction: Bad Genes or Bad Luck?* W. W. Norton & Company, New York, NY, USA, 1st edition, 1992.
[2] R. Posner, *Catastrophes, Risk and Resolution*, Oxford University Press, Oxford, UK, 2004.
[3] G. Chichilnisky, "An axiomatic approach to choice under uncertainty with catastrophic risks," *Resource and Energy Economics*, vol. 22, no. 3, pp. 221–231, 2000.
[4] G. Chichilnisky, "Avoiding extinction: equal treatment of the present and the future," *Economics*, vol. 3, no. 2009-32, 2009.
[5] S. M. McClure, D. I. Laibson, G. Loewenstein, and J. D. Cohen, "Separate neural systems value immediate and delayed monetary rewards," *Science*, vol. 306, no. 5695, pp. 503–507, 2004.
[6] G. Chichilnisky, "Updating von Neumann Morgenstern axioms for choice under uncertainty with catastrophic risks," in *Proceedings of the Conference on Catastrophic Risks*, The Fields Institute for Mathematical Sciences, Toronto, Canada, June 1996.
[7] G. Chichilnisky, "Avoiding extinction: the future of economics," *International Journal of Green Economics*, vol. 3, no. 1, pp. 1–18, 2009.
[8] G. Chichilnisky, "The foundations of probability with black swans," *Journal of Probability and Statistics*, vol. 2010, Article ID 838240, 11 pages, 2010.

[9] G. Chichilnisky, "The foundations of statistics with black swans," *Mathematical Social Sciences*, vol. 59, no. 2, pp. 184–192, 2010.

[10] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, USA, 3rd edition, 1980.

[11] K. J. Arrow, *Essays in the Theory of Risk Bearing*, North-Holland, Amsterdam, The Netherlands, 1970.

[12] I. N. Herstein and J. Milnor, "An axiomatic approach to measurable utility," *Econometrica*, vol. 21, pp. 291–297, 1953.

[13] M. Allais, "The general theory of random choices in relation to the invariant cardinal utility function and the specific probability function," in *Risk Decision and Rationality*, B. R. Munier, Ed., pp. 233–289, Reidel, Dordrecht, The Netherlands, 1988.

[14] M. J. Machina, "Expected utility analysis without the independence axiom," *Econometrica*, vol. 50, no. 2, pp. 277–323, 1982.

[15] M. Machina, "Dynamic consistency and non-expected utility models of choice under uncertainty," *Journal of Economic Literature*, vol. 22, pp. 1622–1668, 1989.

[16] A. Tversky and P. Wakker, "Risk attitudes and decision weights," *Econometrica*, vol. 63, no. 6, pp. 1255–1280, 1995.

[17] G. M. Heal, *Valuing the Future*, Columbia University Press, New York, NY, USA, 2002.

[18] A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen, "The neural basis of economic decision-making in the Ultimatum Game," *Science*, vol. 300, no. 5626, pp. 1755–1758, 2003.

[19] K. Yosida and E. Hewitt, "Finitely additive measures," *Transactions of the American Mathematical Society*, vol. 72, pp. 46–66, 1952.

[20] K. Yosida, *FunctionalAnalysis*, Classics in Mathematics, Springer, New York, NY, USA, 4th edition, 1995.

[21] A. Tversky and P. Wakkler, "Risk attitudes and decision weights," *Econometrica*, vol. 63, pp. 1225–1280, 1995.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 976371, 26 pages
doi:10.1155/2010/976371

Research Article

Continuous Time Portfolio Selection under Conditional Capital at Risk

Gordana Dmitrasinovic-Vidovic, Ali Lari-Lavassani, Xun Li, and Antony Ware

The Mathematical and Computational Finance Laboratory, University of Calgary, Calgary, AB, Canada T2N 1N4

Correspondence should be addressed to Gordana Dmitrasinovic-Vidovic, gdmitras@ucalgary.ca

Received 29 October 2009; Revised 4 February 2010; Accepted 26 March 2010

Academic Editor: Ričardas Zitikis

Copyright © 2010 Gordana Dmitrasinovic-Vidovic et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Portfolio optimization with respect to different risk measures is of interest to both practitioners and academics. For there to be a well-defined optimal portfolio, it is important that the risk measure be coherent and quasiconvex with respect to the proportion invested in risky assets. In this paper we investigate one such measure—conditional capital at risk—and find the optimal strategies under this measure, in the Black-Scholes continuous time setting, with time dependent coefficients.

1. Introduction

The choice of risk measure has a significant effect on portfolio investment decisions. Downside risk measures—that focus attention on the downside tail of the distribution of portfolio returns—have received considerable attention in the financial world. Value at risk (VaR) is probably the most famous among these measures, having featured heavily in various regulatory frameworks. It can be defined for a random variable X and a confidence level α by $\mathrm{VaR}(X) = E[X] - q_\alpha$, where q_α is the α-quantile of X (see e.g., [1, Equation (1.2)]) (Another common definition is that the VaR of a loss distribution L is the smallest number x_α such that $P[L > x_\alpha] = \alpha$. This is equivalent to the definition given here if we define the loss of the portfolio X to be given by $L = E[X] - X$, and identify $x_\alpha = \mathrm{VaR}(X)$.). A closely-related downside risk measure is capital at risk (CaR), defined in [2] (see also [3, 4]) as the difference between the riskless investment and the quantile q_α.

Quantile-based risk measures such as VaR and CaR suffer from several shortcomings. First, while they measure the best of the worst outcomes at the $100(1-\alpha)\%$ confidence level, they do not answer the question of how severe the loss can be. Also, one of the most important concerns is that these measures are not in general subadditive; that is, when used to measure risk, they do not always satisfy the notion that the diversification should not create more risk [5]. Finally, as illustrated in [6], VaR can exhibit multiple local extrema.

These issues were addressed in the widely cited article by Artzner at al. [7], where the authors define coherent risk measures by four conditions that such measures should satisfy. The article motivated a number of authors [5, 8–13] to propose and investigate different types of coherent risk measures, all of which are tail mean-based risk measures.

One such measure, that does not suffer from the critical shortcomings of VaR and CaR, is conditional capital at risk (CCaR). This is defined (in [14]) as the difference between the riskless investment and the conditional expected wealth, under the condition that the wealth is smaller than the corresponding quantile, for a given risk level. As such, this measure provides an indication of the likely severity of the loss in the event that the loss exceeds a given quantile. In this paper we prove that CCaR is strongly quasiconvex as a function of the portfolio, which is an essential property for optimization. We investigate conditional capital at risk in a multiasset Black-Scholes setting, in continuous time, and with time-dependent coefficients. We generalize and extend the optimization approach of Emmer at al. (see [2, 14]) to the continuous-time setting.

The outline of this paper is as follows. In Section 2, we give the notation and define the portfolio process and CCaR. Section 3 provides the proof that CCaR is a coherent risk measure and that it satisfies the property of strong quasiconvexity. In Section 4, we derive an analytical solution for the minimal CCaR problem, up to a scalar constant which has to be evaluated numerically. Section 5 is devoted to the derivation of an analytical strategy for the maximal expected wealth, subject to constrained CCaR. Section 6 provides some numerical examples, and Section 7 concludes this paper.

2. Preliminaries

We introduce the following notation. The m-dimensional column vector with each component equal to 1 is denoted by e, the Euclidean norm of a matrix or vector by $\|\cdot\|$, and the space of $\mathbb{R}^n$-valued, square-integrable functions defined on $[0,t]$ by $L^2([0,t],\mathbb{R}^n)$, or just L^2. The natural inner product of this space is denoted by $\langle\cdot,\cdot\rangle_t$, and the corresponding norm by $\|\cdot\|_t$.

We work under the following assumptions.

Assumption 2.1. (i) The securities are perfectly divisible.

(ii) Negative positions in securities are possible.

(iii) Rebalancing of the holdings does not lead to transaction costs.

Assumption 2.2. (i) $m+1$ assets are traded continuously over a finite horizon $[0,T]$.

(ii) m of these assets are stocks that follow the generalized Black-Scholes dynamics:

$$dS_i(t) = S_i(t)\left(b_i(t)dt + \sum_{j=1}^{m}\sigma_{ij}(t)dW^j(\mathrm{t})\right), \quad t\in[0,T],\ S_i(0)>0,\ i=1,\dots,m, \tag{2.1}$$

where $W^j(t)$, $j=1,\dots,m$, are independent standard Brownian motions.

(iii) One of the assets is a bond, whose price $S_0(t)$, $t \geq 0$, evolves according to the differential equation:

$$dS_0(t) = r(t)S_0(t)dt, \quad t \in [0,T],\ S_0(0) = S_0 > 0, \tag{2.2}$$

where $r(t)(> 0)$ is the interest rate of the bond. Throughout this work, we assume that borrowing in the bond is unconstrained.

(iv) The volatility matrix $\sigma(t)$ (with ijth element $\sigma_{ij}(t)$), its inverse $\sigma^{-1}(t)$, the drift vector $b(t) := (b_1(t), \ldots, b_m(t))'$, and the interest rate $r(t)$ are deterministic, Borel measurable, bounded functions over $[0,T]$, so that they belong to the appropriate L^2 spaces.

(v) $\sigma(t)$ satisfies the nondegeneracy condition:

$$x'\sigma(t)\sigma(t)'x \geq \delta x'x, \quad \forall t \in [0,T],\ \forall x \in \mathbb{R}^m, \tag{2.3}$$

where $\delta > 0$ is a given constant.

We note that, under the above assumptions, the market is complete.

At any time t, $N_i(t)$ shares are held in the asset $S_i(t)$, leading to the wealth $X^{\pi}(t) = \sum N_i(t)S_i(t)$. The $m+1$-dimensional vector-valued function $N(t) = (N_0(t), \ldots, N_m(t))'$ is called the *trading strategy*. We denote the fraction of the wealth $X^{\pi}(t)$ invested into the risky asset $S_i(t)$ by

$$\pi_i(t) = \frac{N_i(t)S_i(t)}{X^{\pi}(t)}, \quad i = 1, \ldots, m, \tag{2.4}$$

and call $\pi(t) := (\pi_1(t), \ldots, \pi_m(t))' \in \mathbb{R}^m$ the *portfolio*. The fraction held in the bond is $\pi_0(t) = 1 - \pi'(t)e$. Under the assumption that the trading strategy is *self-financing,* the wealth process follows the dynamics

$$dX^{\pi}(t) = X^{\pi}(t)\big(\big(r(t) + B(t)'\pi(t)\big)dt + \pi'(t)\sigma(t)dW(t)\big), \quad X(0) = X_0, \tag{2.5}$$

where X_0 is the initial wealth, and the risk premium vector $B(t)$ is defined by

$$B(t) := b(t) - r(t)e, \quad t \in [0,T]. \tag{2.6}$$

Proceeding as in [3], to ensure a minimal tractability of the optimization problems which we solve through the following sections, we restrict our attention in this work to the class Q of portfolios $\pi(\cdot)$ which are Borel measurable, deterministic, and bounded over $[0,T]$. Such portfolios are called *admissible*. Note that for an admissible portfolio $\pi(\cdot)$, (2.5) is guaranteed to have a strong solution $X^{\pi}(\cdot)$ (see [15, Theorem 5.2.9]). Note that, by allowing for time-dependent coefficients, this generalises the class of portfolios considered in [14].

Under condition (2.3) the *market price of risk* is uniquely defined by

$$\theta(t) = \sigma(t)^{-1}B(t). \tag{2.7}$$

It will be shown throughout this work that the magnitude of the L^2-norm of the market price of risk is the determining criterion for optimal investment strategies, which turn out to be the weighted averages of the bond and *Merton's portfolio* defined by

$$\pi_M(t) := (\sigma(t)\sigma(t))^{-1} B(t). \tag{2.8}$$

This result is an illustration of the mutual fund theorem in complete markets.

We now recall the definitions of the quantiles and quantile-based risk measures for $X^{\pi}(t)$.

Definition 2.3. Suppose that $F_t(x)$ is the cumulative distribution function of the wealth $X^{\pi}(t)$, at time $t \in [0, T]$. For a risk level $\alpha \in (0, 0.5)$, the *α-quantile* of $X^{\pi}(t)$ is defined as

$$q_\alpha(X_0, \pi, t) := \inf\{x \in \mathbb{R} \mid F(x) \geq \alpha\}. \tag{2.9}$$

The tail mean or expected shortfall of the wealth process $X^{\pi}(t)$, which we denote by $\mathrm{TM}_\alpha(X^{\pi}(t))$, is the expected value of $X^{\pi}(t)$ conditional on $X^{\pi}(t) \leq q_\alpha(X_0, \pi, t)$, that is,

$$\mathrm{TM}_\alpha(X^{\pi}(t)) := E\left[X^{\pi}(t) \mid X^{\pi}(t) \leq q_\alpha(X_0, \pi, t)\right]. \tag{2.10}$$

The conditional capital at risk, which we denote by $\mathrm{CCaR}(X_0, \pi, t)$, is defined to be the difference between the riskless investment and the tail mean, that is,

$$\mathrm{CCaR}(X_0, \pi, t) := X_0 R(t) - \mathrm{TM}_\alpha(X^{\pi}(t)), \tag{2.11}$$

where $R(t)$ is defined as

$$R(t) = \exp\left(\int_0^t r(s)ds\right). \tag{2.12}$$

Remark 2.4. Note that, with CCaR defined in this way, we get the following.

(i) An investment in a riskless asset corresponds to zero CCaR.

(ii) An increase in the tail mean corresponds to a decrease in CCaR. Thus, negative CCaR corresponds to the case when the tail mean is above the riskless return, that is, a desired result, while positive CCaR corresponds to the case when the tail mean is below the riskless return which we try to reverse by minimizing CCaR.

It was shown in [3] that the α-quantile of the wealth process $X^{\pi}(t)$ can be written as

$$q_\alpha(X_0, \pi, t) = X_0 R(t) \exp\left(\langle B, \pi\rangle_t - \frac{1}{2}\|\sigma'\pi\|_t^2 - |z_\alpha| \|\sigma'\pi\|_t\right), \tag{2.13}$$

where, for a given risk level α, z_α denotes the corresponding α-quantile of the standard normal distribution (note that this is negative when $\alpha < 0.5$). In the following, we let φ and Φ denote the density and cumulative distribution functions of the standard normal random variable. We have the following proposition.

Proposition 2.5. *The tail mean of the wealth process* $X^\pi(t)$ *solving* (2.5) *can be expressed as*

$$TM_\alpha(X^\pi(t)) = X_0 R(t)\frac{1}{\alpha}\exp(\langle B,\pi\rangle_t)\Phi\left(-|z_\alpha| - \left\|\sigma'\pi\right\|_t\right), \tag{2.14}$$

and the expected value is given by

$$E[X^\pi(t)] = X_0 R(t)\exp(\langle B,\pi\rangle_t). \tag{2.15}$$

The proof is given in the appendix. From Proposition 2.5 we get the following corollary.

Corollary 2.6. *The conditional capital at risk of the solution to* (2.5) *can be written as*

$$CCaR(X_0,\pi,t) = X_0 R(t)\left(1 - \frac{1}{\alpha}\exp(\langle B,\pi\rangle_t)\Phi\left(-|z_\alpha| - \left\|\sigma'\pi\right\|_t\right)\right). \tag{2.16}$$

Throughout the paper we will use the following function:

$$g_\alpha(\pi,t) := \langle B,\pi\rangle_t + \ln\left(\Phi\left(-|z_\alpha| - \left\|\sigma'\pi\right\|_t\right)\right) - \ln\alpha, \tag{2.17}$$

which transforms the expression for conditional capital at risk into

$$\mathrm{CCaR}(X_0,\pi,t) = X_0 R(t)\left(1 - \exp\left(g_\alpha(\pi,t)\right)\right). \tag{2.18}$$

We now turn our attention to investigating the properties of CCaR.

3. Coherency and Quasiconvexity of CCaR

The notion of coherent risk measures has been widely discussed in the recent literature (see [5, 7–13] e. g.). Below we recall the definition of coherency, confirm that CCaR is a coherent risk measure, and prove that it is quasiconvex.

Definition 3.1. Let V be the set of real-valued random processes on a complete probability space $(\Omega,\mathcal{F},P)$, with its natural filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$, satisfying $E[S(t)] < \infty$ for all $S(t) \in V$, and for $t \in [0,T]$. Then $\rho : V \to \mathbb{R}$ is a coherent risk measure if it satisfies the following properties.

(1) Subadditivity: ρ is subadditive if, for all random processes $S(t), Y(t) \in V$,

$$\rho(S(t) + Y(t)) \leq \rho(S(t)) + \rho(Y(t)) \quad \text{for } t \in [0, T]. \tag{3.1}$$

(2) Positive Homogeneity: ρ is positive homogeneous if, for all $S(t) \in V$ and constant $c > 0$,

$$\rho(cS(t)) = c\rho(S(t)) \quad \text{for } t \in [0, T]. \tag{3.2}$$

(3) Monotonicity: ρ is monotone if, for all $S(t), Y(t) \in V$, such that $S(t) \geq Y(t)$ almost everywhere, and $S_0 = Y_0$,

$$\rho(S(t)) \leq \rho(Y(t)) \quad \text{for } t \in [0, T]. \tag{3.3}$$

(4) Translation Invariance: ρ is translation invariant if, for all $S(t) \in V$ and $c \in \mathbb{R}$,

$$\rho(S(t) + c) = \rho(S(t)) \quad \text{for } t \in [0, T]. \tag{3.4}$$

To prove that $\mathrm{CCaR}(X_0, \pi, t)$ of a wealth process $X^{\pi}(t) \in V$ is a coherent risk measure we refer to [8, Definition 2.6], where the expected shortfall of a random process $X(t)$, with the corresponding α-quantile q_α, is defined as

$$\mathrm{ES}_\alpha(X(t)) = -\alpha^{-1}\left(E\left[X(t)I_{X(t)\leq q_x}\right] + q_x\left(\alpha - P\left(X(t) \leq q_x\right)\right)\right) \tag{3.5}$$

and is shown to be coherent [8, Proposition 3.1]. We should note that the definition of coherency used in [8] involves translation invariance in the sense that $\rho(X(t)+c) = \rho(X(t)) - c$, for all $X(t) \in V$ and $c \in \mathbb{R}$.

In order to relate the result in [8] to the coherency of CCaR in the sense used here, we first note that, if $X(t)$ is a random process with a continuous probability distribution, we have $\mathrm{ES}_\alpha(X(t)) = -\mathrm{TM}_\alpha(X(t))$, so that (with a slight abuse of notation) $\mathrm{CCaR}(X(t)) = X_0 R(t) + \mathrm{ES}_\alpha(X(t))$. If we consider the shifting of the portfolio value by an amount c, we have

$$\begin{aligned}
\mathrm{CCaR}(X(t) + c) &= X_0 R(t) + c + \mathrm{ES}_\alpha(X(t) + c) \\
&= X_0 R(t) + c + \mathrm{ES}_\alpha(X(t)) - c \\
&= X_0 R(t) + \mathrm{ES}_\alpha(X(t)) = \mathrm{CCaR}(X(t)),
\end{aligned} \tag{3.6}$$

and so we have the following result.

Corollary 3.2. *Conditional capital at risk $\mathrm{CCaR}(X_0, \pi, t)$ of a wealth process $X^{\pi}(t) \in V$, at time $t \in [0, T]$, for a risk level $\alpha \in (0, 0.5)$, is a coherent risk measure.*

Remark 3.3. (i) While properties 2 and 3 are quite natural, subadditivity has the obvious, yet important consequence that diversification does not create more risk.

(ii) Properties 1 and 2 together are very important as they guarantee the convexity of the risk measure which is essential for optimization.

(iii) In our definition of a coherent risk measure we say that a risk measure is translation invariant, if $\rho(X(t)+c) = \rho(X(t))$, for all $X(t) \in V$, and $c \in \mathbb{R}$. While in article [8], It is said that arisk measure is translation invariant, if $\rho(X(t)+c) = \rho(X(t)) - c$, for all $X(t) \in V$ and $c \in \mathbb{R}$ It is our belief that the definition of the translation invariance, as given in this paper, corresponds more with the intuition behind the notion of translation invariance than the definition given in [8].

Remark 3.4. In the above remarks we investigated the impact of diversification to the portfolio risk, as measured by CCaR. We saw that the axiom of subadditivity requires that the risk of the sum of two risky processes, say two stock price processes, be less than or equal to the sum of the individual risks associated with these stock price processes. This means that diversification does not create more risk, as measured by CCaR, and that, as long as we diversify, we expect risk reduction.

However, if the portfolio consists of all market assets, the diversification is completed. Then the risk can be further reduced by optimization, that is, by rebalancing the positions across these assets. We then look at the risk measure CCaR as a function of the portfolio π, and prove that it is strongly quasiconvex in π which further implies the uniqueness of the corresponding optimization problems' solutions.

We now turn to the notion of strong quasiconvexity, which we note has not been discussed in the context of portfolio optimization in any of the quoted references except in [3, 4]. Its usefulness lies in its role in establishing the existence of unique solutions to portfolio optimization problems.

We first recall (see [16, Definition 3.5.8]) that a function $f : U \subset \mathbb{R}^m \to \mathbb{R}$ is said to be *strongly quasiconvex* if

$$f(\lambda\pi + (1-\lambda)\xi) < \max\{f(\pi), f(\xi)\} \quad \forall \pi,\ \xi \in U, \pi \neq \xi,\ \lambda \in (0,1). \tag{3.7}$$

In the following theorem we prove that CCaR has this important property, when viewed as a function of the portfolio weights.

Theorem 3.5. *For all distinct $\pi(\cdot),\xi(\cdot) \in Q$ (where the set $\{t \in [0,T] \mid \pi(t) \neq \xi(t)\}$ has a positive Lebesgue measure), and for all $\lambda \in (0,1)$*

$$CCaR(X_0, \lambda\pi + (1-\lambda)\xi, T) < \max\{CCaR(X_0,\pi,T), CCaR(X_0,\xi,T)\}. \tag{3.8}$$

Proof of Theorem 3.5. We suppose, without loss of generality, that

$$\mathrm{CCaR}(X_0,\pi,T) > \mathrm{CCaR}(X_0,\xi,T), \tag{3.9}$$

in which case, from (2.18), we have

$$g_\alpha(\pi,T) < g_\alpha(\xi,T). \tag{3.10}$$

Let $\lambda \in (0,1)$. We claim that

$$\mathrm{CCaR}(X_0, \lambda\pi + (1-\lambda)\xi, T) < \mathrm{CCaR}(X_0, \pi, T), \tag{3.11}$$

which is equivalent to the statement

$$g_\alpha(\lambda\pi + (1-\lambda)\xi, T) > g_\alpha(\pi, T). \tag{3.12}$$

For ease of notation, we define the function

$$\gamma(x) := \ln(\Phi(-|z_\alpha| - x)). \tag{3.13}$$

Since $g_\alpha(\pi, T) = \langle B, \pi\rangle_T - \ln\alpha + \gamma(\|\sigma'\pi\|_T)$, (3.10) is equivalent to

$$\langle B, \xi - \pi\rangle_T + \gamma\left(\|\sigma'\xi\|_T\right) - \gamma\left(\|\sigma'\pi\|_T\right) > 0, \tag{3.14}$$

while (3.12) becomes

$$(1-\lambda)\langle B, \xi - \pi\rangle_T + \gamma\left(\|\sigma'(\lambda\pi + (1-\lambda)\xi)\|_T\right) - \gamma\left(\|\sigma'\pi\|_T\right) > 0. \tag{3.15}$$

Clearly, from (3.14), the left-hand side of (3.15) is greater than

$$\gamma\left(\|\sigma'(\lambda\pi + (1-\lambda)\xi)\|_T\right) - \lambda\gamma\left(\|\sigma'\pi\|_T\right) - (1-\lambda)\gamma\left(\|\sigma'\xi\|_T\right), \tag{3.16}$$

and the theorem will be proved if we can establish that this is nonnegative. In order to do this, we make use of the following lemma, the proof of which is given in the appendix.

Lemma 3.6. *The function $k(x) = \ln\Phi(-x)$ is decreasing and strictly concave for all $x > 0$.*

Since $\gamma(x) = k(|z_\alpha| + |x|)$, one can make use of Lemma 3.6 to establish the following:

$$\begin{aligned}\gamma\left(\|\sigma'(\lambda\pi + (1-\lambda)\xi)\|_T\right) &\geq \gamma\left(\lambda\|\sigma'\pi\|_T + (1-\lambda)\|\sigma'\xi\|_T\right)\\ &\geq \lambda\gamma\left(\|\sigma'\pi\|_T\right) + (1-\lambda)\gamma\left(\|\sigma'\xi\|_T\right).\end{aligned} \tag{3.17}$$

The proof is complete. □

This theorem has an immediate, important consequence. Namely, from [16, Theorem 3.5.9], if a function $f : U \subset \mathbb{R}^m \to \mathbb{R}$ is strongly quasiconvex, then its local minimum is its unique global minimum. Therefore, the following corollary is true.

Corollary 3.7. *If $\mathrm{CCaR}(X_0, \pi, T)$ has a local minimum at $\pi^*(\cdot) \in Q$, then $\pi^*(\cdot)$ is its unique global minimum.*

This guarantees that minimization of $\mathrm{CCaR}(X_0, \pi, T)$ with respect to π gives the true, global minimum. To illustrate strong quasiconvexity, we give the plot of CCaR for

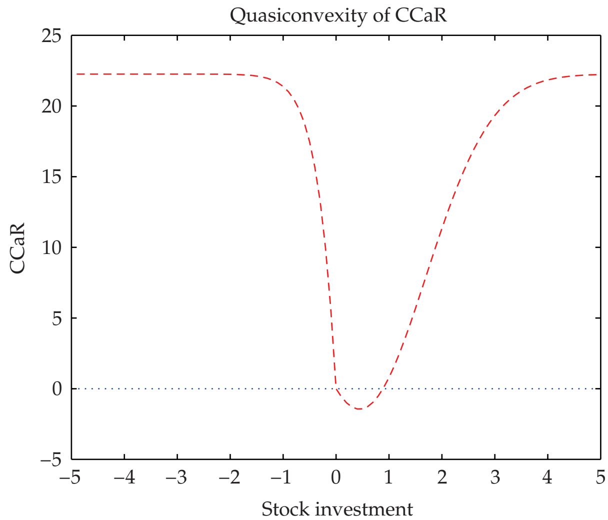


Figure 1: The graph of CCaR as a function of the fraction of wealth π invested in the stock, for $\pi \in [-5, 5]$, $T = 16$, $S(0) = 10$, $r = 0.05$, $\sigma = 0.2$, $b = 0.15$. Observe that, in this case (with a long-time horizon and constantly high-performing stocks), the global minimum π_ε lies in the interval $[0, 1]$ and satisfies $\mathrm{CCaR}(\pi_\varepsilon) < 0$.

the following example (where the parameters were chosen to represent rather extreme conditions, for the purposes of illustration).

Example 3.8. We consider a market consisting of one stock following the SDE

$$dS(t) = S(t)(0.15dt + 0.2dW(t)), \quad t \in [0, 16], \ S(0) = 10, \tag{3.18}$$

and the bond with the constant interest rate $r = 0.05$. The graph of $\mathrm{CCaR}(\pi)$ is given in Figure 1.

4. Minimal Conditional Capital at Risk

The first portfolio selection problem we consider is to minimize risk as measured by CCaR, that is, to determine its minimal value over all $\pi \in Q$. The problem can be stated as

$$\min_{\pi \in Q} \mathrm{CCaR}(X_0, \pi, T). \tag{4.1}$$

Using (2.18), problem (4.1) transforms into

$$\min_{\pi \in Q} X_0 R(T)\left(1 - \exp\left(g_\alpha(\pi, T)\right)\right). \tag{4.2}$$

Since $R(T)$ is a constant, problem (4.2) is equivalent to

$$\max_{\pi \in Q} g_\alpha(\pi, T). \tag{4.3}$$

We now introduce the fundamental *dimension reduction procedure*, used throughout this work. Following [2, 14], we project the optimization problems considered in this paper onto the family of surfaces $Q_\varepsilon = \{\pi(\cdot) \in L^2 : \|\sigma'\pi\|_T^2 = \varepsilon^2\}$, and note that $Q = \bigcup_{\varepsilon \geq 0} Q_\varepsilon$.

We denote by g_α^ε the restriction of g_α to Q_ε, so that

$$g_\alpha^\varepsilon(\pi, T) = \langle B, \pi \rangle_T + \ln(\Phi(-|z_\alpha| - \varepsilon)) - \ln \alpha. \tag{4.4}$$

Taking into account the definition of Q_ε, problem (4.3) can be stated as

$$\max_{\varepsilon \geq 0} \max_{\pi \in Q_\varepsilon} g_\alpha^\varepsilon(\pi, T). \tag{4.5}$$

We deal with this problem in two stages. First, fixing ε reduces the problem to

$$\max_{\pi \in Q_\varepsilon} g_\alpha^\varepsilon(\pi, T). \tag{4.6}$$

If π_ε denotes the unique maximising portfolio for this problem, then (4.5) can be solved through the one-dimensional problem:

$$\max_{\varepsilon \geq 0} g_\alpha^\varepsilon(\pi_\varepsilon, T). \tag{4.7}$$

It remains to solve the subproblem (4.6). Since ε is fixed, we see from (4.4) that (4.6) is equivalent to the problem

$$\max_{\pi \in Q_\varepsilon} \langle B, \pi \rangle_T, \tag{4.8}$$

the solution of which is given by the following proposition (the proof of which is given in [6, Proposition 2.1]).

Proposition 4.1. *For a fixed ε, the solution of problem* (4.6) *is attained by*

$$\pi_\varepsilon(t) = \frac{\varepsilon}{\|\theta\|_T} \left(\sigma(t)\sigma(t)'\right)^{-1} B(t), \tag{4.9}$$

where θ denotes the market price of risk, defined in (2.7).

We are now ready to solve problem (4.3), and we have the following theorem.

Theorem 4.2. *Let $\theta(\cdot)$ be the market price of risk* (2.7). *$g_\alpha(\pi, T)$ attains its maximum when*

$$\pi(t) := \pi_{\varepsilon^*}(t) = \varepsilon^* \|\theta\|_T^{-1} \left(\sigma(t)\sigma(t)'\right)^{-1} B(t), \tag{4.10}$$

where, if $\|\theta\|_T \geq \varphi(|z_\alpha|)/\alpha$, ε^* is defined to be the unique solution of the equation

$$\frac{\varphi(|z_\alpha| + \varepsilon)}{\Phi(-|z_\alpha| - \varepsilon)} = \|\theta\|_T, \tag{4.11}$$

and, if $\|\theta\|_T < \varphi(|z_\alpha|)/\alpha$, $\varepsilon^* = 0$. *The corresponding minimum conditional capital at risk is*

$$CCaR(X_0, \pi_{\varepsilon^*}, T) = X_0 R(T)\left(1 - \frac{1}{\alpha}\exp(\varepsilon^*\|\theta\|_T)\Phi(-|z_\alpha| - \varepsilon^*)\right) \tag{4.12}$$

(which is 0 *if* $\varepsilon^* = 0$*), and the expected wealth is*

$$E[X^{\pi_{\varepsilon^*}}(T)] = X_0 R(T)\exp(\varepsilon^*\|\theta\|_T). \tag{4.13}$$

Proof of Theorem 4.2. Using the definition of $\theta(t)$ and substituting (4.9) into (4.5) allows us to rewrite (4.5) as

$$\max_{\varepsilon \geq 0} g_\alpha^\varepsilon(\pi_\varepsilon, T) := \varepsilon\|\theta\|_T + \ln(\Phi(-|z_\alpha| - \varepsilon)) - \ln\alpha. \tag{4.14}$$

If we define the function

$$f(\varepsilon) := g_\alpha^\varepsilon(\pi_\varepsilon, T) = \varepsilon\|\theta\|_T + \ln(\Phi(-|z_\alpha| - \varepsilon)) - \ln\alpha, \tag{4.15}$$

we get

$$\begin{aligned}
f'(\varepsilon) &= \|\theta\|_T - \frac{\varphi(-|z_\alpha| - \varepsilon)}{\Phi(-|z_\alpha| - \varepsilon)} = \|\theta\|_T - \frac{\varphi(|z_\alpha| + \varepsilon)}{\Phi(-|z_\alpha| - \varepsilon)}, \\
f'(0) &= \|\theta\|_T - \frac{\varphi(|z_\alpha|)}{\Phi(-|z_\alpha|)} = \|\theta\|_T - \frac{\varphi(|z_\alpha|)}{\alpha}, \\
f''(\varepsilon) &= \frac{\varphi(|z_\alpha| + \varepsilon)}{(\Phi(-|z_\alpha| - \varepsilon))^2}\big((1 - \Phi(|z_\alpha| + \varepsilon))(|z_\alpha| + \varepsilon) - \varphi(|z_\alpha| + \varepsilon)\big).
\end{aligned} \tag{4.16}$$

We see that $f''(\varepsilon)$ has the same form as $k''(x)$, where $k(x)$ is defined in Lemma 3.6, with $|z_\alpha| + \varepsilon$, instead of x, so that $f''(\varepsilon) \leq 0$. Since $f'(\varepsilon)$ is a decreasing function of ε for $\varepsilon > 0$, we have two cases.

(i) If $f'(0) \geq 0$, that is, $\|\theta\|_T \geq \varphi(|z_\alpha|)/\alpha$, then the equation

$$f'(\varepsilon) = \|\theta\|_T - \frac{\varphi(|z_\alpha| + \varepsilon)}{\Phi(-|z_\alpha| - \varepsilon)} = 0 \tag{4.17}$$

has a unique solution which we denote by ε^*. Since $f(\varepsilon)$ is a concave function, it reaches its maximum at ε^*.

(ii) If $f'(0) < 0$, that is, $\|\theta\|_T < \varphi(|z_\alpha|)/\alpha$, then $f'(\varepsilon) < 0$ for all $\varepsilon \geq 0$, and so (4.17) has no solution. This implies that $f(\varepsilon)$ is decreasing for all $\varepsilon \geq 0$, so that the optimal solution to problem (4.14) is $\varepsilon = 0$.

To complete the proof of Theorem 4.2, note that (4.9), with $\varepsilon = \varepsilon^*$, is the optimal solution of problem (4.3). One can then write

$$\langle B, \pi_{\varepsilon^*} \rangle_T = \frac{\varepsilon^*}{\|\theta\|_T} \|\sigma^{-1} B\|_T^2 = \varepsilon^* \|\theta\|_T, \qquad \|\sigma' \pi\|_t^2 = \varepsilon^{*2}, \tag{4.18}$$

leading to

$$\begin{gathered} \mathrm{CCaR}(X_0, \pi_{\varepsilon^*}, T) = X_0 R(T) \left(1 - \frac{1}{\alpha} \exp(\varepsilon^* \|\theta\|_T) \Phi(-|z_\alpha| - \varepsilon^*) \right), \\ E[X^{\pi_{\varepsilon^*}}(t)] = X_0 R(T) \exp(\varepsilon^* \|\theta\|_T). \end{gathered} \tag{4.19}$$

□

Remark 4.3. (i) Note that, if $\|\theta\|_T \geq \varphi(|z_\alpha|)/\alpha$, we can deduce by using (A.12) that

$$\|\theta\|_T \geq \frac{\varphi(|z_\alpha|)}{\alpha} = \frac{\varphi(|z_\alpha|)}{1 - \Phi(|z_\alpha|)} \geq |z_\alpha|. \tag{4.20}$$

Therefore, the condition $\|\theta\|_T \geq \varphi(|z_\alpha|)/\alpha$, which has to be satisfied for investing into stocks under conditional capital at risk, is stronger than the condition $\|\theta\|_T \geq |z_\alpha|$. The latter condition is sufficient in order to include stocks into the optimal strategy using capital at risk as a risk measure. Otherwise stated, conditional capital at risk is a more conservative risk measure than capital at risk, which is consistent with its definition.

(ii) Increasing the time horizon T leads to increasing the L^2 norm of the market price of risk, so that, in case (ii), the optimal strategy changes from a pure bond strategy to a mixed bond-stocks strategy. In case (i), increasing the L^2 norm $\|\theta\|_T$ leads to increasing the expected wealth $E[X^\pi(T)]$ and decreasing conditional capital at risk $\mathrm{CCaR}(X_0, \pi, T)$.

(iii) As was noted in the preliminary remarks, the optimal portfolio is a weighted average of Merton's portfolio and the bond, which is an illustration of the two-fund separation theorem.

We note that the solution provided in Theorem 4.2 is not an explicit analytical solution, but it is expressed in terms of the solution of the one-dimensional equation (4.11), whose solution ε^* can be easily computed. Analytical upper and lower bounds for ε^* are given in the following lemma, whose proof is given in the appendix.

Lemma 4.4. *The unique solution ε^* of* (4.17) *satisfies the inequality*

$$\left(\|\theta\|_T \left(1 - \frac{1}{|z_\alpha|^2} \right) - |z_\alpha| \right)^+ \leq \varepsilon^* \leq \|\theta\|_T - |z_\alpha|, \tag{4.21}$$

where $a^+ := \max\{a, 0\}$.

Remark 4.5. Note that, for $\alpha = 0.05$, that is, at the 95% confidence level, $(1 - 1/|z_\alpha|^2) \approx 0.63$, (4.21) approximates the result from [14], which states

$$\left(\frac{2}{3}\|\theta\|_T - |z_\alpha|\right)^+ \le \varepsilon^* \le \|\theta\|_T - |z_\alpha| \quad \text{for } \alpha < 0.15. \tag{4.22}$$

However, at a higher confidence level, that is, $\alpha < 0.05$, (4.21) gives a better approximation for ε^*, that is, a smaller interval to which ε^* belongs.

5. Portfolio Optimization with Respect to Conditional Capital at Risk

We now turn to the problem of maximizing wealth subject to constrained CCaR, that is,

$$\max_{\pi \in Q} E[X^\pi(T)] \quad \text{subject to } \mathrm{CCaR}(X_0, \pi, T) \le C. \tag{5.1}$$

Using (2.18), the above problem can be written in the form

$$\max_{\pi \in Q} X_0 R(T) \exp(\langle B, \pi \rangle_T) \quad \text{subject to } X_0 R(T)\left(1 - \exp\left(g_\alpha(\pi, T)\right)\right) \le C, \tag{5.2}$$

which is equivalent to

$$\max_{\pi \in Q} \langle B, \pi \rangle_T \quad \text{subject to } \exp\left(g_\alpha(\pi, T)\right) \ge 1 - \frac{C}{X_0 R(T)}. \tag{5.3}$$

Since CCaR, from its definition, is smaller than total wealth, to avoid trivial cases we only consider C such that

$$C < X_0 R(T) \tag{5.4}$$

Under condition (5.4), problem (5.3) can be written as

$$\max_{\pi \in Q} \langle B, \pi \rangle_T \quad \text{subject to } g_\alpha(\pi, T) \ge c, \tag{5.5}$$

where

$$c = \ln\left(1 - \frac{C}{X_0 R(T)}\right). \tag{5.6}$$

Note that condition (5.4) guarantees that c is well defined. Using the dimension reduction procedure of [2, 14] (see Section 4), we can write problem (5.5) as a one-parameter optimization problem:

$$\max_{\varepsilon \ge 0} \max_{\pi \in Q_\varepsilon} \langle B, \pi \rangle_T \quad \text{subject to } g_\alpha(\pi, T) \ge c. \tag{5.7}$$

The solution to (5.7) is given in the following theorem.

Theorem 5.1. *Suppose that the constant risk level C satisfies the following condition:*

$$\begin{aligned} X_0R(T)\left(1-\frac{1}{\alpha}\exp(\varepsilon^*\|\theta\|_T)\Phi(-|z_\alpha|-\varepsilon^*)\right) \le C < X_0R(T), \quad \textit{if } \|\theta\|_T \ge \frac{\varphi(|z_\alpha|)}{\alpha} \\ 0 \le C < X_0R(T), \quad \textit{if } \|\theta\|_T < \frac{\varphi(|z_\alpha|)}{\alpha}, \end{aligned} \tag{5.8}$$

where ε^* is defined in Theorem 4.2. Then the optimal solution to problem (5.7) *is*

$$\pi_{\varepsilon^{**}}(t) = \frac{\varepsilon^{**}}{\|\theta\|_T}\left(\sigma(t)\sigma(t)'\right)^{-1}B(t), \tag{5.9}$$

where c is defined in (5.6)*, and $\varepsilon^{**} \in [\varepsilon^*, \infty)$ is the unique solution of the equation*

$$h(\varepsilon) := \varepsilon\|\theta\|_T + \ln(\Phi(-|z_\alpha|-\varepsilon)) - \ln\alpha - c = 0. \tag{5.10}$$

The corresponding expected wealth is equal to

$$E[X^{\pi_{\varepsilon^{**}}}(T)] = X_0R(T)\exp(\varepsilon^{**}\|\theta\|_T), \tag{5.11}$$

with the corresponding conditional capital at risk:

$$CCaR(X_0, \pi_{\varepsilon^{**}}, T) = C. \tag{5.12}$$

The efficient frontier is given by the following curve, whose first component is an increasing function of C defined implicitly through (5.6)*,* (5.10)*, and* (5.11)*:*

$$\{(C, E[X^{\pi_{\varepsilon^{**}}}(T)]) \mid C \textit{ satisfies } (5.8)\}. \tag{5.13}$$

Proof of Theorem 5.1. Proposition 4.1 states that the problem

$$\max_{\pi\in Q_\varepsilon}\langle B,\pi\rangle_T \tag{5.14}$$

has the optimal solution

$$\pi_\varepsilon(t) = \frac{\varepsilon}{\|\theta\|_T}\left(\sigma(t)\sigma(t)'\right)^{-1}B(t). \tag{5.15}$$

Substituting (5.15) into (5.7) transforms it into the problem

$$\max_{\varepsilon\ge 0}\varepsilon\|\theta\|_T \quad \text{subject to } g_\alpha(\pi_\varepsilon, T) \ge c. \tag{5.16}$$

Using (5.10), the above problem can be written as

$$\max_{\varepsilon\geq 0} \varepsilon\|\theta\|_T \quad \text{subject to } h(\varepsilon) = \varepsilon\|\theta\|_T + \ln(\Phi(-|z_\alpha| - \varepsilon)) - \ln\alpha - c \geq 0. \tag{5.17}$$

Clearly, the problem achieves its optimal solution for maximal ε for which the constraint is satisfied. The solution to this problem is given in the following lemma, whose proof is given in the appendix.

Lemma 5.2. *Under condition* (5.8)*, the equation $h(\varepsilon) = 0$ has a maximal solution $\varepsilon^{**} \in [\varepsilon^*, \infty)$. Furthermore, $h'(\varepsilon) < 0$, for $\varepsilon > \varepsilon^*$.*

We substitute ε^{**} into (5.15) to get the optimal portfolio

$$\pi_{\varepsilon^{**}}(t) = \frac{\varepsilon^{**}}{\|\theta\|_T}\left(\sigma(t)\sigma(t)'\right)^{-1}B(t), \tag{5.18}$$

which yields

$$E[X^{\pi_{\varepsilon^{**}}}(T)] = X_0R(T)\exp(\varepsilon^{**}\|\theta\|_T), \qquad \mathrm{CCaR}(X_0, \pi_{\varepsilon^{**}}, T) = C. \tag{5.19}$$

Finally, to prove that the expected wealth is an increasing function of the risk constant C, we rewrite (5.10) in the form

$$h(\varepsilon^{**}) = \varepsilon^{**}\|\theta\|_T + \ln(\Phi(-|z_\alpha| - \varepsilon^{**})) - \ln\alpha - \ln\left(1 - \frac{C}{X_0R(T)}\right) = 0, \tag{5.20}$$

which defines ε^{**} as an implicit function of C. Differentiating the above, we get

$$\frac{\partial h}{\partial\varepsilon^{**}}\frac{d\varepsilon^{**}}{dC} + \frac{\partial h}{\partial C} = 0, \tag{5.21}$$

so that

$$\frac{d\varepsilon^{**}}{dC} = -\frac{\partial h/\partial C}{\partial h/\partial\varepsilon^{**}}. \tag{5.22}$$

Since $(\partial h/\partial C) = (1/X_0R(T) - C)$, and since $C < X_0R(T)$, it follows that $\partial h/\partial C > 0$. From Lemma 5.2 it follows that $dh/d\varepsilon < 0$, for all $\varepsilon > \varepsilon^*$, and hence for ε^{**}, so that $d\varepsilon^{**}/dC > 0$. Thus, ε^{**} is an increasing function of C. From (5.11), we see that the expected wealth is an increasing function of ε^{**}, and consequently of C. This completes the proof of the theorem. □

Remark 5.3. If a risk-averse investor decides to take the minimal risk which, in the first case, means taking for the risk constant

$$C = X_0R(T)\left(1 - \frac{1}{\alpha}\exp(\varepsilon^*\|\theta\|_T)\Phi(-|z_\alpha| - \varepsilon^*)\right), \tag{5.23}$$

implying

$$c = -\ln\alpha + \varepsilon^{*}\|\theta\|_{T} + \ln(\Phi(-|z_{\alpha}| - \varepsilon^{*})), \tag{5.24}$$

then the unique solution of the equation

$$\varepsilon\|\theta\|_{T} + \ln(\Phi(-|z_{\alpha}| - \varepsilon)) - \ln\,\alpha - c = 0 \tag{5.25}$$

is $\varepsilon^{**} = \varepsilon^{*}$.

In the second case, for a risk averse investor, the minimal risk constant is $C = 0$, which implies $c = 0$, leading to the equation

$$h(\varepsilon) = \varepsilon\|\theta\|_{T} + \ln(\Phi(-|z_{\alpha}| - \varepsilon)) - \ln\alpha = 0, \quad \text{with } h(0) = 0, \qquad h'(\varepsilon) \leq 0, \tag{5.26}$$

so that the unique solution of the equation $h(\varepsilon) = 0$ is $\varepsilon^{**} = 0$.

Therefore, in both cases, choosing the minimal risk constant leads to the optimal strategies coinciding with the strategies from the previous chapter, in which minimal $\mathrm{CCaR}(X_0, \pi, T)$ was determined.

6. Applications

To illustrate the results from previous sections, we give some numerical examples. We recall that the system of SDEs that models the stocks' prices is

$$dS_i(t) = S_i(t)\left(b_i(t)dt + \sum_{j=1}^{m}\sigma_{ij}(t)dW^{j}(t)\right), \quad S_i(0) > 0,\ i = 1, \ldots, m, \tag{6.1}$$

and that the stocks' returns variance-covariance matrix, which we denote by $\Gamma(t)$, is equal to $\sigma(t)\sigma(t)'$. We also recall that $\Gamma(t)$ can be decomposed as

$$\Gamma(t) = \nu(t)\Delta(t)\nu(t), \tag{6.2}$$

where $\Delta(t)$ is the stocks' returns correlation matrix, and $\nu(t)$ is a diagonal matrix with the entries equal to the stocks' returns standard deviations. Therefore, from (6.2) we get

$$\Gamma(t) = \sigma(t)\sigma(t)' = \nu(t)\Delta(t)\nu(t). \tag{6.3}$$

Although, theoretically, we assume that the vector of independent Brownian motions $W(t)$ is $\{\mathcal{F}_t\}_{t\in[0,T]}$-adapted, that is, known at time $t \in [0, T]$, it is a common practice that we only observe $\Gamma(t)$ or, equivalently, $\nu(t)$ and $\Delta(t)$, but not $\sigma(t)$. From (6.3) we see that this leads to a nonunique decomposition of $\Gamma(t)$ into the product $\sigma(t)\sigma(t)'$. Despite that fact, the Euclidean norm, and consequently, the L^2-norm of the market price of risk, is uniquely determined by

$$\|\theta(t)\|^2 = \left\|\sigma(t)^{-1}B(t)\right\|^2 = B(t)'(\sigma(t)\sigma(t))^{-1}B(t), \tag{6.4}$$

or, in the terms of the standard deviation and the correlation matrix, as

$$\|\theta(t)\|^2 = B(t)'\nu(t)^{-1}\Delta(t)^{-1}\nu(t)^{-1}B(t). \tag{6.5}$$

To keep the exposition simple we consider the cases where the interest rate $r(t)$, as well as the volatility matrix, is constant, and the number of stocks is $m = 3$. In all numerical computations and the corresponding plots we use an annual time scale for the drifts, standard deviations, and the correlation matrix $\Delta(t)$ of the 3 stocks. We model time dependency in the drift $b_i(t)$ by assuming that it oscillates periodically around a constant level μ_i. In order to capture cycles in the economy or in the dynamics of the stocks we model the drifts as

$$b_i(t) = \mu_i + \beta_i \cos(\varphi_i t), \quad i = 1, 2, 3. \tag{6.6}$$

We note that the above model for stocks' drifts was already used in [17], as the amplitude and frequency coefficients β_i and φ_i allow a high degree of flexibility in adjusting the shape of this time dependency. We also note that, when modeling real market data, it is quite easy to deal with the above functional form and estimate these two parameters by maximum likelihood techniques, rather than detrending the data.

We now look at four special cases with the following characteristics.

(i) We let $\varphi_1 = \varphi_2 = \varphi_3 = \varphi$, with $\varphi = 0.75$; that is, the economic cycles of all three stocks are the same. We consider $\beta_1 = 0.75\beta$, $\beta_2 = 0.5\beta$, $\beta_3 = 0.25\beta$, with $\beta = 0.015$, which corresponds to a 1.5% deviation around the constant values μ_i.

(ii) We assume that the interest rate is $r = 0.05$ and numerically explore the sensitivity of the optimal strategies with respect to μ_i and ρ.

(iii) We assume that the stocks' returns have constant standard deviations given as follows:

$$\nu_1 = 20\%, \quad \nu_2 = 25\%, \quad \nu_3 = 30\%. \tag{6.7}$$

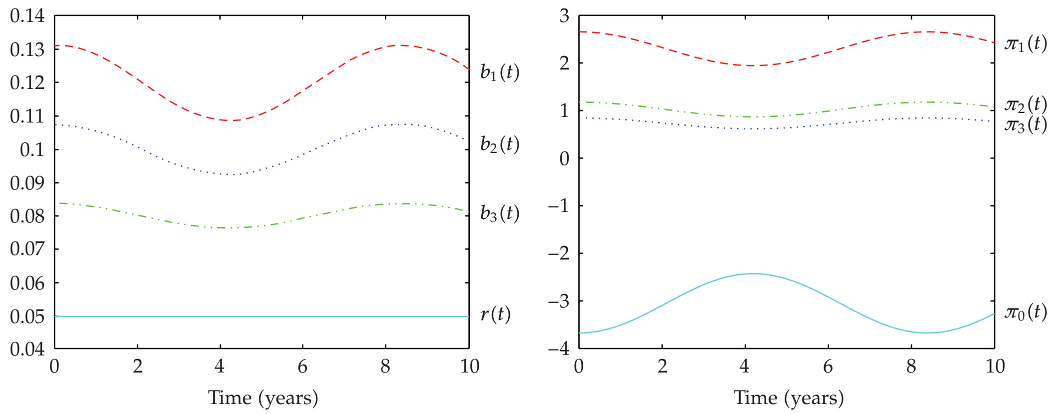
To emphasize the importance of the diversification effect and, consequently, of the market price of risk, for the optimal strategies, in Examples 6.2, 6.3, and 6.4 we keep the same drift and volatility coefficients and change the correlation matrix.

Example 6.1. We assume that $\mu_1 = 0.12$, $\mu_2 = 0.10$, and $\mu_3 = 0.08$, and that the correlation matrix is

$$\Delta = \begin{bmatrix} 1.0 & -0.6 & -0.8 \\ -0.6 & 1.0 & 0.5 \\ -0.8 & 0.5 & 1 \end{bmatrix}. \tag{6.8}$$

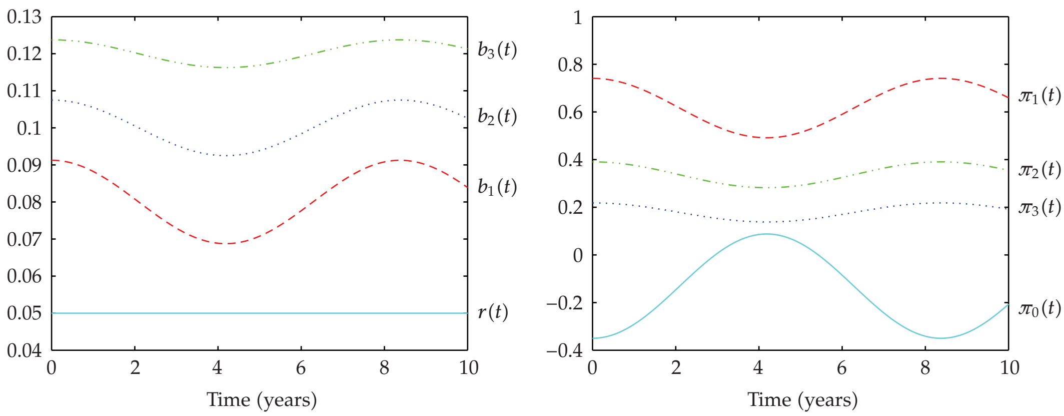
In Figure 2, we show (a) the stocks' drifts over a ten year period, with daily granularity, and (b) the optimal strategy, under minimal CCaR.

In this example, we see the expected result, that is, stock 1, which has the largest constant part in the drift, and the smallest volatility is present in the optimal portfolio in



(a) (b)

Figure 2: Plot of the Stock Drifts (a) and portfolio weights (b) for Example 6.1.



(a) (b)

Figure 3: Plot of the Stock Drifts (a) and portfolio weights (b) for Example 6.2.

the highest percentage. However, to assume that just drifts and volatilities determine the optimal strategy would be misleading, as the following examples show.

Example 6.2. We assume that $\mu_1 = 0.08$, $\mu_2 = 0.10$, $\mu_3 = 0.12$, and that the correlation matrix is the same as in Example 6.1.

The stocks' drifts and portfolio weights are given in Figure 3.

We see that, despite the fact that stock 1 has the smallest constant part of the drift, it is still present in the optimal portfolio in the highest percentage, due to its high negative correlations with both stocks 2 and 3.

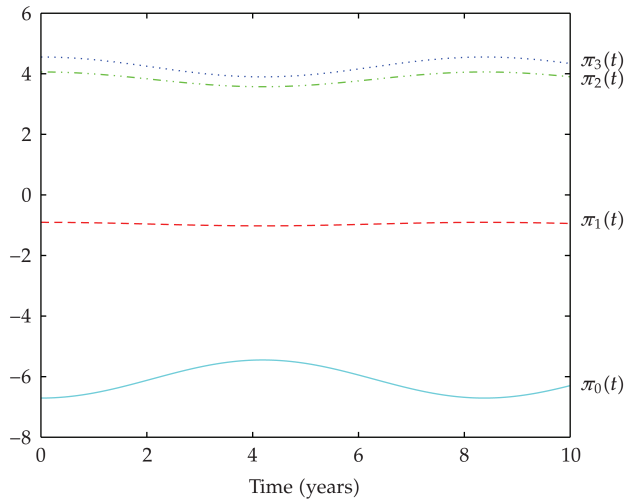


Figure 4: Portfolio Weights under Minimal CCaR for Example 6.3.

Example 6.3. In this example we assume that $\mu_1 = 0.08$, $\mu_2 = 0.10$, and $\mu_3 = 0.12$; that is, the stocks have the same constant parts of the drifts as in Example 6.2 but a different correlation matrix given by

$$\Delta = \begin{bmatrix} 1.0 & -0.3 & 0.5 \\ -0.3 & 1.0 & -0.9 \\ 0.5 & -0.9 & 1 \end{bmatrix}. \tag{6.9}$$

The optimal strategy, under minimal CCaR, is given in Figure 4.

In this example, stocks 2 and 3, which are highly negatively correlated, are present in the optimal portfolio in percentages above 400%, while stock 1 and the bond are being borrowed.

Example 6.4. In this example, we assume again that the stocks have the same constant parts of the drifts as in Examples 6.2 and 6.3, while the correlation matrix is

$$\Delta = \begin{bmatrix} 1.0 & 0.2 & -0.3 \\ 0.2 & 1.0 & 0.1 \\ -0.3 & 0.1 & 1 \end{bmatrix}. \tag{6.10}$$

The criterion for investing into stocks, under the minimal CCaR, is $\|\theta\|_T \geq \varphi(|z_\alpha|)/\alpha$. For $\alpha = 0.05$, $|z_\alpha| = 1.645$, so that we must have $\|\theta\|_T \geq 2.0620$, in order to start investing into stocks. In the case of low correlation coefficients, such as in the above example, the market price of risk increases very slowly, so that it takes the time horizon of 34 years to invest into stocks with the risk level of 0.05. For $\alpha = 0.01$, $|z_\alpha| = 2.33$, so that we must have $\|\theta\|_T \geq 2.6424$ to start investing into stocks. If we choose the risk level to be 0.01, we have to assume the very

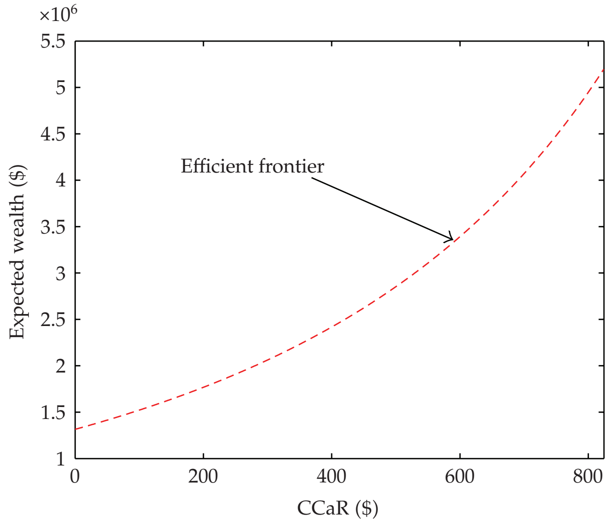


Figure 5: Efficient Frontier for Example 6.1.

Table 1

T	10	20	30	33	34	40	50	56	57
$\|\theta\|_T$	1.1420	1.5944	1.9344	2.0302	2.0705	2.2293	2.5030	2. 6424	2.6655

long time horizon, of 57 years, in order to include stocks into the optimal strategy, under minimal CCaR. The relation between $\|\theta\|_T$ and T can be found in Table 1.

Efficient Frontiers

Figure 5 shows the efficient frontier for Example 6.1 created using Theorem 5.1. The theorem states that the expected wealth is an increasing function of C, which is bounded above by

$$C_{\max} = X_0 R(T). \tag{6.11}$$

In order to avoid extremely risky strategies as $C \to C_{\max}$, and for the sake of more transparency of the graphs, we restrict C to the interval $[0, 0.5C_{\max}]$. We note that the efficient frontiers for the other three examples are of the same exponential form, so that we omit their graphs.

The graph given in Figure 5 illustrates the fact that the efficient frontiers are an increasing function of the risk constant C which bounds CCaR.

Time-CCaR-Expected Wealth Surfaces

In Figure 6 we give a three-dimensional plot of the expected wealth as a function of time and the risk constant C. From Theorem 5.1 we know that the expected wealth increases when the risk constant C increases. We recall that the expected wealth is also an increasing function of the time horizon T. The following figure illustrates these facts for Example 6.1. In order

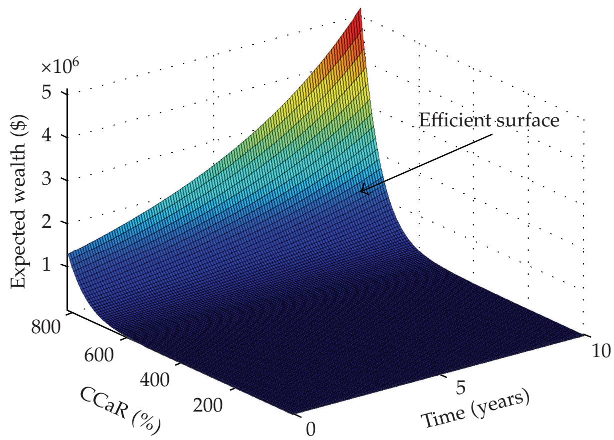


Figure 6: Time-CCaR-Expected Wealth Surface for Example 6.1.

to avoid extremely risky strategies, and to get a better representation, we restrict the upper bound for CCaR to the interval $[0, 0.5C_{\max}]$.

7. Conclusion

In this work we investigated continuous time portfolio selection under the notion of conditional capital at risk, within the Black-Scholes asset pricing paradigm, with time dependent coefficients. We showed that conditional capital at risk is a coherent risk measure and proved that it satisfies an important property of strong quasiconvexity. Based on an idea from [14], generalized in [3], we introduced the fundamental dimension reduction procedure which transforms m-dimensional optimization problems into one-dimensional problems, within the class of admissible portfolios, which are Borel measurable, bounded, and deterministic. We further developed optimal strategies for portfolio optimization under constrained conditional capital at risk. It is important to emphasize that we considered time dependent portfolios, where the methods developed in [14] no longer work. We illustrated the two-fund separation theorem, by showing that all optimal strategies are the weighted averages of Merton's portfolio and the bond, and the weights depend on the investor's risk tolerance only. Finding optimal strategies requires solving nonlinear equations which include the cumulative distribution function of the normal random variable, so that the weights can be only found numerically. We provide several numerical examples which illustrate the importance of diversification, given by the correlation matrix. The correlation matrix significantly impacts the magnitude of the L^2 norm of the market price of risk, which, in turn, is the determining criterion for investment strategies.

Appendix

Proof of Proposition 2.5. Let

$$\mu(t) = r(t) + B(t)'\pi(t), \qquad \eta(t) = \sigma(t)'\pi(t). \tag{A.1}$$

Then the differential equation of the wealth process given by (2.5) can be written as

$$dX^{\pi}(t) = X^{\pi}(t)\big(\mu(t)dt + \eta(t)dW(t)\big), \quad X^{\pi}(0) = X_0. \tag{A.2}$$

For convenience, set $Y(t) = \ln X^{\pi}(t)$. Applying the multidimensional version of Itô's Lemma, it can be shown (see the proof of Proposition 2.1 of [3]) that $Y(t)$ follows the dynamics

$$dY(t) = \left(\mu(t) - \frac{1}{2}\|\eta(t)\|^2\right)dt + \eta(t)dW(t), \qquad Y(0) = \ln(X_0) = Y_0. \tag{A.3}$$

The α-quantile of $Y(t)$ is equal to

$$q_{\alpha}(Y_0, \pi, t) = Y_0 + \int_0^t \left(\mu(s) - \frac{1}{2}\|\eta(s)\|^2\right)ds - |z_{\alpha}|\sqrt{\int_0^t \|\eta(s)\|^2 ds}, \tag{A.4}$$

so that the α-quantile of $X(t)$ is equal to $q_{\alpha}(X_0, \pi, t) = \exp(q_{\alpha}(Y_0, \pi, t))$. We will further simplify the notation by introducing

$$q_x = q_{\alpha}(X_0, \pi, t), \qquad q_y = q_{\alpha}(Y_0, \pi, t). \tag{A.5}$$

From

$$X^{\pi}(t)I_{X^{\pi}\le q_x} = \exp(Y(t))I_{Y\le q_y}, \tag{A.6}$$

where $I_{(\cdot)}$ is the corresponding indicator function, and from Bayes' theorem, we obtain

$$\mathrm{TM}_{\alpha}(X^{\pi}(t)) = \alpha^{-1}E\left[\exp(Y(t))I_{Y\le q_y}\right]. \tag{A.7}$$

Choose $t \in [0, T]$. We will evaluate (A.7) using the characteristics of the distribution of $Y(t)$, for fixed t. From (A.3) we have that $Y(t)$ is a normal random variable with the parameters

$$\mu_y = Y_0 + \int_0^t \left(\mu(s) - \frac{1}{2}\|\eta(s)\|^2\right)ds, \qquad \sigma_y = \sqrt{\int_0^t \|\eta(s)\|^2 ds}. \tag{A.8}$$

Taking into account (A.7) and (A.8), we can write

$$\mathrm{TM}_{\alpha}(X^{\pi}(t)) = \alpha^{-1}E\left[\exp(Y(t))I_{Y\le q_y}\right] = \alpha^{-1}\int_{-\infty}^{q_y} \exp(y)\varphi(y)dy. \tag{A.9}$$

Using the standard integration techniques, and taking into account that $z_{\alpha} < 0$, we get

$$\mathrm{TM}_{\alpha}(X^{\pi}(t)) = \alpha^{-1}\exp\left(\mu_y + \frac{\sigma_y^2}{2}\right)\Phi(-|z_{\alpha}| - \sigma_y). \tag{A.10}$$

If we substitute the original notation from (A.8) and (A.1), we get (2.10). This completes the proof of Proposition 2.5. □

Proof of Lemma 3.6. Let $x > 0$. The function $k(x) = \ln \Phi(-x)$ is clearly decreasing. To prove that it is strictly concave, we differentiate twice to get

$$k''(x) = \frac{\varphi(x)}{(\Phi(-x))^2}\big((1-\Phi(x))x - \varphi(x)\big). \tag{A.11}$$

To evaluate $k''(x)$, we use the following standard inequality. For $x > 0$,

$$\frac{1}{x} - \frac{1}{x^3} \le \frac{1-\Phi(x)}{\varphi(x)} \le \frac{1}{x}. \tag{A.12}$$

Applying the above, we get that $k''(x) \le 0$ which means that $k(x)$ is concave. To prove that $k(x)$ is strictly concave we need to show that

$$x(1-\Phi(x)) < \varphi(x) \quad \text{for } x > 0. \tag{A.13}$$

If we define the function

$$w(x) := x(1-\Phi(x)) - \varphi(x), \tag{A.14}$$

it is an easy exercise to prove that $w(x) < 0$, for all $x > 0$; that is, the function $k(x)$ is strictly concave for all $x > 0$, which ends the proof of Lemma 3.6. □

Proof of Lemma 4.4. We prove that the optimal solution ε^* of (4.17) satisfies (4.21). From (4.17) we have that

$$\frac{(1-\Phi(|z_\alpha| + \varepsilon^*))}{\varphi(|z_\alpha| + \varepsilon^*)} = \frac{1}{\|\theta\|_T}. \tag{A.15}$$

We use again the standard inequality (A.12). From inequality (A.12), with $x = |z_\alpha| + \varepsilon^*$, we have

$$\frac{1}{|z_\alpha| + \varepsilon^*} - \frac{1}{(|z_\alpha| + \varepsilon^*)^3} \le \frac{1}{\|\theta\|_T} = \frac{(1-\Phi(|z_\alpha| + \varepsilon^*))}{\varphi(|z_\alpha| + \varepsilon^*)} \le \frac{1}{|z_\alpha| + \varepsilon^*}. \tag{A.16}$$

The right-hand side of the inequality implies

$$\varepsilon^* \le \|\theta\|_T - |z_\alpha|. \tag{A.17}$$

From the left-hand side of the inequality we have

$$\|\theta\|_T\left(1 - \frac{1}{(|z_\alpha| + \varepsilon^*)^2}\right) \le |z_\alpha| + \varepsilon^*. \tag{A.18}$$

Since

$$-\frac{1}{(|z_\alpha|+\varepsilon^*)^2} \geq -\frac{1}{|z_\alpha|^2}, \tag{A.19}$$

inequality (A.18) implies

$$\|\theta\|_T\left(1-\frac{1}{|z_\alpha|^2}\right)-|z_\alpha| \leq \varepsilon^*. \tag{A.20}$$

The fact that $\varepsilon^* \geq 0$ and (A.20) give the desired result, so that Lemma 4.4 is proved. □

Proof of Lemma 5.2. In this proof we find a maximal solution of the equation

$$h(\varepsilon) = \varepsilon\|\theta\|_T + \ln(\Phi(-|z_\alpha|-\varepsilon)) - \ln\,\alpha - c = 0. \tag{A.21}$$

Since $h'(\varepsilon) = f'(\varepsilon)$, defined by (4.15), we will apply some arguments from the proof of Theorem 4.2.

(i) Suppose that $\|\theta\|_T \geq \varphi(|z_\alpha|)/\alpha$, which means that $h'(0) = f'(0) \geq 0$. Then the unique maximum of $h(\varepsilon)$ is achieved at ε^*. From condition (5.8) (i) we have that

$$\begin{aligned}
c &= \ln\left(1-\frac{C}{X_0R(T)}\right) \\
&\leq \ln\left(1-\left(1-\frac{1}{\alpha}\exp(\varepsilon^*\|\theta\|_T)\right)\Phi(-|z_\alpha|-\varepsilon^*)\right) \\
&= -\ln\alpha + \varepsilon^*\|\theta\|_T + \ln(\Phi(-|z_\alpha|-\varepsilon^*)).
\end{aligned} \tag{A.22}$$

Hence

$$\begin{aligned}
h(\varepsilon^*) &= \varepsilon^*\|\theta\|_T + \ln(\Phi(-|z_\alpha|-\varepsilon^*)) - \ln\alpha - c \\
&\geq \varepsilon^*\|\theta\|_T + \ln(\Phi(-|z_\alpha|-\varepsilon^*)) - \ln\alpha \\
&\quad + \ln\alpha - \varepsilon^*\|\theta\|_T - \ln(\Phi(-|z_\alpha|-\varepsilon^*)) = 0.
\end{aligned} \tag{A.23}$$

Therefore, $h(\varepsilon^*) \geq 0$. We further distinguish the following two subcases.

(a) Suppose that $C < 0$. Then (A.21) implies that $h(0) = -c < 0$. From the proof of Theorem 4.2, for $\varepsilon > \varepsilon^*$, $h'(\varepsilon) \leq 0$, $h(\varepsilon)$ is concave, and since $h(\varepsilon^*) \geq 0$, it follows that the equation $h(\varepsilon) = 0$ has at least one solution, with the bigger solution $\varepsilon^{**} \in [\varepsilon^*, \infty)$.

(b) For $C \geq 0$, $h(0) = -c \geq 0$. Using the same arguments as in case (a), we get that (A.21) has a unique solution $\varepsilon^{**} \in [\varepsilon^*, \infty)$.

(ii) Suppose $\|\theta\|_T < \varphi(|z_\alpha|)/\alpha$, that is, $h'(0) < 0$. We recall that in this case $\varepsilon^* = 0$. The definition of $h(\varepsilon)$ yields

$$h(0) = \ln(\Phi(-|z_\alpha|)) - \ln \alpha - c = -c. \tag{A.24}$$

From condition (5.8) (ii), we have that $c \le 0$, that is, $h(0) \ge 0$. Using $h'(\varepsilon) = f'(\varepsilon)$, and the proof of Theorem 4.2, we have

$$h''(\varepsilon) \le 0; \qquad h'(0) = \|\theta\|_T - \frac{\varphi(|z_\alpha|)}{\alpha} < 0. \tag{A.25}$$

Therefore, $h'(\varepsilon) < 0$ for all $\varepsilon \ge 0$, with $h(0) \ge 0$, so that $h(\varepsilon) = 0$ has a unique solution $\varepsilon^{**} \in [\varepsilon^*, \infty)$. This completes the proof of Lemma 5.2. □

Acknowledgments

This research was partially supported by the National Science and Engineering Research Council of Canada and the Network Centre of Excellence, Mathematics of Information Technology and Complex Systems.

References

[1] P. Jorion, *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, New York, NY, USA, 2nd edition, 2000.
[2] S. Emmer, C. Klüppelberg, and R. Korn, "Optimal portfolios with bounded capital at risk," *Mathematical Finance*, vol. 11, no. 4, pp. 365–384, 2001.
[3] G. Dmitrašinović-Vidović, A. Lari-Lavassani, X. Li, and T. Ware, *Dynamic Portfolio Selection Under Capital at Risk*, University of Calgary Yellow Series, University of Calgary, Alberta, Canada, 2003, Report 833.
[4] G. Dmitrašinović-Vidović, *Portfolio selection under downside risk measures*, Ph.D. thesis, University of Calgary, Alberta, Canada, 2004.
[5] C. Acerbi, "Spectral measures of risk: a coherent representation of subjective risk aversion," *Journal of Banking and Finance*, vol. 26, no. 7, pp. 1505–1518, 2002.
[6] G. Dmitrašinović-Vidović and T. Ware, "Asymptotic behaviour of mean-quantile efficient portfolios," *Finance and Stochastics*, vol. 10, no. 4, pp. 529–551, 2006.
[7] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.
[8] C. Acerbi and D. Tasche, "On the coherence of expected shortfall," *Journal of Banking and Finance*, vol. 26, no. 7, pp. 1487–1503, 2002.
[9] H. Föllmer and A. Schied, "Convex measures of risk and trading constraints," *Finance and Stochastics*, vol. 6, no. 4, pp. 429–447, 2002.
[10] G. Pflug, "Some remarks on the value-at-risk and the conditional value-at-risk," in *Probabilistic Constrained Optimization*, vol. 49 of *Methodology and Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
[11] F. Riedel, "Dynamic coherent risk measures," *Stochastic Processes and Their Applications*, vol. 112, no. 2, pp. 185–200, 2004.
[12] T. Rockafellar and S. Uryasev, "Conditional value at risk: optimization algorithms and applications," *Financial Engineering News*, vol. 14, 2000.
[13] D. Tasche and D. Bundesbank, "Expected shortfall and beyond," in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, Statistics for Industry and Technology, Springer, Berlin, Germany, 2002.

[14] S. Emmer, C. Klüppelberg, and R. Korn., *Optimal Portfolios with Bounded Downside Risks*, Working Paper, Center of Mathematical Sciences, Munich University of Technology, Munich, Germany, 2000.
[15] I. Karatzas and S. E. Shreve, *Methods of Mathematical Finance*, Applications of Mathematics, Springer, New York, NY, USA, 1999.
[16] M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming*, John Wiley & Sons, New York, NY, USA, 1979.
[17] A. Lari-Lavassani, A. A. Sadeghi, and T. Ware, *Modeling and Implementing Mean Reverting Price Processes in Energy Markets*, Electronic Publications of the International Energy Credit Association, 2001, http://www.ieca.net/.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 805309, 11 pages
doi:10.1155/2010/805309

Research Article

Individual Property Risk Management

Michael S. Finke,[1] Eric Belasco,[2] and Sandra J. Huston[1]

[1] *Division of Personal Financial Planning, Texas Tech University, Lubbock, TX 79409, USA*
[2] *Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX 79409, USA*

Correspondence should be addressed to Michael S. Finke, michael.finke@ttu.edu

Received 2 October 2009; Accepted 21 January 2010

Academic Editor: Ričardas Zitikis

Copyright © 2010 Michael S. Finke et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper reviews household property risk management and estimates normatively optimal choice under theoretical assumptions. Although risk retention limits are common in the financial planning industry, estimates of optimal risk retention that include both financial and human wealth far exceed limits commonly recommended. Households appear to frame property losses differently from other wealth losses leading to wealth-reducing, excess risk transfer. Possible theoretical explanations for excess sensitivity to loss are reviewed. Differences between observed and optimal risk management imply a large potential gain from improved choice.

1. Introduction

Property risk management, a fundamental aspect of individual financial planning, has perhaps been subject to the least amount of rigor. While investment management draws directly from a theoretical structure of modern portfolio theory, risk management often involves only the identification of risk exposures and products available to eliminate these exposures. A common method of ensuring consistency in choice among insurance products is to retain all risks beneath a risk retention limit; however the practice literature offers little insight into how much retention is appropriate. This paper uses expected utility theory to estimate optimal risk retention limits for households given reasonable assumptions about risk aversion, human and financial wealth, and cost of insurance. Estimated retention limits are generally much larger than limits chosen by individuals or recommended by professionals. This suggests that households are either overweighting losses in a manner consistent with Kahneman and Tversky's [1] prospect theory or unaware of normatively efficient insurance decision making.

Risky decision making involves consideration of the likelihood of expected outcomes and the consequences of each outcome on expected well being. While the profit motive of firms suggests a preference for risky decisions that have a positive net expected value, households are willing to pay a greater premium to mitigate risk. More formally, consumers make decisions that maximize expected utility (U) by sacrificing expected wealth to reduce the variance of possible outcomes.

For example, a household can face m possible states of nature where the likelihood of an event is generated from a $\mathrm{Ber}(\rho_j)$ distribution where ρ_j is the likelihood of event j occurring. Along with the probabilities associated with each state is the payout or return (R) associated with each event. In this case, we can compute the expected payout associated as

$$E(\text{Payout}) = \sum_{j=1}^{m} \rho_j * R_j. \tag{1.1}$$

If a household is risk neutral, they will make insurance decisions that maximize their expected payout. Since the insurance product is costly, this assumption typically leads to no new purchases of insurance. A more realistic scenario is one where households have some aversion to risk. For example, if a household is asked whether they prefer an annual salary of \$50,000 with full certainty or either \$30,000 with a 95% likelihood or \$450,000 with a 5% likelihood, more households are likely to take the certain salary even though the expected income from the uncertain scenario is higher. This is an illustration of the Von Neumann-Morgenstern utility function, which is generally assumed to be strictly concave with respect to wealth (W). If W were a random variable and utility ($u(W)$) is strictly concave, then Jensen's inequality results in the following relationship where

$$\mathrm{EU}[W] < u(E[W]). \tag{1.2}$$

This implies that when facing large positive payouts, the utility associated with expected wealth, $u(E[W])$, is greater than the expected utility associated wealth, $\mathrm{EU}[W]$. The difference between these two points represents the welfare gain to the household as well as the profit opportunity for the insurer. In other words, a household can achieve greater utility when faced with uncertain outcomes that are less extreme and will be willing to give up expected wealth in order to forego these potential losses. The amount that a household is willing to pay to mitigate risk is dependent on their degree of risk aversion, which is known empirically as the risk aversion parameter. Risk aversion parameters are embedded into utility functions, where an individual with $U(W)'' < 0$ is risk averse, $U(W)'' > 0$ is risk seeking, and $U(W)'' = 0$ is risk neutral.

An actuarially fair premium rate is exactly equal to the expected loss or the product of the expected loss and probability of loss:

$$\text{rate} = E(\text{loss}) = \rho_{\text{loss}} E(\text{payout} \mid \text{loss}). \tag{1.3}$$

Risk-averse agents have incentive to purchase insurance in order to mitigate their risk and increase expected utility. Insurance firms are able to take on new risk due to their ability to spread risks among a diversified insurance pool while taking advantage of their risk neutrality. At the same time, consumers are willing to pay a fee for this protection. If insurance firms charged exactly the actuarial fair premium rate, they would make zero profits if they effectively diversified their insurance pool and have no moral hazard or adverse selection. Insurance firms use load fees to pay for administrative costs and generate profits. The cost of insurance to individuals is

$$\text{cost} = (1 + \text{load}) * E(\text{loss}), \tag{1.4}$$

where *load* is some percentage greater than zero and adds to the cost. A rational, risk-averse agent might still have incentive to purchase this insurance product with a negative expected payout.

Insurance is also subject to other costs that are included in the load. Employees need to be hired to verify the claim that a negative state has occurred. Employees also need to be hired to estimate the likelihood that an uncertain state will occur in order to calculate the cost of the contingent claim (actuaries). Additional employees need to be hired to sell the contingent claims and manage the collection of fees charged for the claims. These costs, incurred to provide insurance to households, impact appropriate use of contingent claims to maximize welfare. They further increase how much wealth is sacrificed to decrease risk.

Household investment choice assumes that there is an optimal amount of portfolio risk for each investor at the point where an additional unit of risk no longer provides greater expected utility despite greater expected wealth. Insurance involves this same tradeoff of expected wealth for reduced risk. The next section focuses on calculating the point at which an additional unit of insurance no longer provides an increase in expected utility.

2. Estimating Optimal Insurance

Estimating insurance needs is similar to computing an optimal investment portfolio. In a simple model, we need only know the wealth of the household, its risk aversion, and the cost of the contingent claim.

2.1. Human Wealth

Total household wealth consists of both net worth and human wealth (discounted expected future household earnings). A formula to estimate human wealth is provided by Ibbotson et al. [2], where human capital *HC* is equal to the expected future earnings *E* from next year until retirement, discounted each year by the discount rate *r* and a risk premium *v*:

$$\text{HC}(x) = \sum_{t=x+1}^{n} \frac{E[h_t]}{(1 + r + v)^{t-x}}. \tag{2.1}$$

A good proxy for the discount rate *r* is the rate on taxable, low-risk corporate bonds, since income streams are similar to a bond and fully taxable. The risk premium *v* further discounts expected future income streams that may be more volatile. To illustrate the importance of human wealth in a household portfolio, consider that a 25-year-old with an income of $75,000 at a 6% discount rate has an estimated present value wealth of $1.1 million, assuming no income growth and no volatility if the individual expects to work until age 65. Alternatively, a 55-year-old with an income of $100,000 has a human wealth of $640,000, given the same assumptions. As we age, we transform our human wealth into income until it is exhausted at retirement. Generally, the wealth of younger households will consist primarily of human wealth.

Risks to net worth include investment risk and negative events which will reduce the value of assets (destruction of property, a lawsuit, etc.). Risks to human wealth include uncertain events that decrease the expected value of future earnings, including disability, death, illness, or a labor market shock. Some risks are insurable through contingent claims, for example, disability and property destruction, while others must either be retained or insured through the public sector.

2.2. Cost of the Contingent Claim

Investors are often induced to accept greater risk by the opportunity to realize greater expected returns. Similarly, the decision to retain or to transfer risk is influenced by the cost of the insurance product. More expensive insurance will provide a disincentive to transfer risk and conversely a greater incentive to retain risk. Unfortunately, it is often difficult to estimate with any degree of accuracy the actual cost of the contingent claim.

In general, insurance products with heavy sales, underwriting, claims, moral, and morale hazard expenses will be more costly. These types of costs are endemic to many property and casualty policies. Loss ratios, or claims paid to policyholders, range from roughly 55% for homeowners insurance to 60% for automobile insurance to 75% for group health insurance. (Based on 2007 national insurance loss ratio statistics provided by the Missouri Department of Insurance, Financial Institutions & Professional Registration, available at http://insurance.mo.gov/reports/lossratio/.)

A loss ratio of 60% for property insurance policies implies that the expected return on the premium paid for the policies is −40%. Of every dollar paid for an insurance premium the household can expect to receive 60 cents back in claims on average. A $2,000 autopolicy will thus yield an expected annual loss in wealth of $800. This is the cost of risk transfer. However, the policy prevents a wealth loss if a negative state occurs. By choosing to buy the insurance policy the individual is revealing that the expected utility from preventing the uncertain negative state is greater than the certain loss of $800 in expected wealth, assuming full information.

It is also important to note that the ownership of any asset that may decline in value due to a negative state (peril) implies an expected annual loss in wealth that is a function of the magnitude and likelihood of this loss. It is the cost of insurance (1-loss ratio) that represents the loss in wealth above the expected loss inherent in asset ownership. For example, the ownership of a $100,000 home with a 1 in 100 chance of complete loss from a fire involves an implied cost of $1,000 per year in risk on average. A policy with a loss ratio of 50% would cost $2,000 (expected payout/loss ratio) implying an additional expected wealth loss of $1,000 per year.

2.3. Calculation of Insurance Needs

The estimate of insurance needs analysis relies on the following assumptions.

(1) Life is risky. It is possible that negative states may occur that will reduce wealth.

(2) Individuals are risk averse and are willing to reduce expected wealth in order to avoid a risk.

(3) Insurance reduces expected wealth to the extent that premiums exceed expected payouts.

(4) Risk aversion and wealth determine optimal risk retention and transfer.

Optimal risk retention will occur at the point where the expected utility from retaining an additional dollar of possible loss is equal to the expected utility of transferring risk to prevent the loss. According to Hanna [3], if we assume that this is a pure risk with two potential states (loss or no loss), the decision to insure will involve a comparison of utility over four possible outcomes. The amount of utility gained or lost is a function of wealth and risk aversion.

Suppose that individuals are assumed to exhibit constant relative risk aversion (CRRA) which can be written as

$$U = \frac{W^{1-r}}{1-r}, \tag{2.2}$$

where r is the coefficient of relative risk aversion. (While CRRA utility functions are commonly used, decreasing relative risk aversion (DRRA) can also be assumed where the relative risk coefficient decreases for higher levels of wealth. Because a DRRA utility function assumes that individuals are relatively less risk averse at higher levels of wealth, the optimal premium rate may be lower relative to CRRA as individuals are more willing to take on additional relative risk for higher levels of wealth. Other utility functions include more flexible forms as well as a class of absolute relative risk aversion functions (CARA, IARA, and DARA).) Notice that the Arrow-Pratt coefficient ($\overline{r}$) of relative risk aversion has the following relationship in this scenario:

$$\overline{r} = -\frac{U''(w)}{U'(w)}x = -\left(\frac{\partial U'}{\partial x}\right)\frac{x}{U'} = r \tag{2.3}$$

implying that the negative of the elasticity of utility with respect to wealth is constant. Also, notice that a higher r implies greater risk aversion and an r closer to 1 implies greater risk tolerance. A household with a coefficient of relative risk aversion equal to 1 is indifferent between a 50/50 chance of a 100% increase in expected lifetime income or a 50% reduction in income. A coefficient of 4 implies indifference between a gamble whose outcome is either a 100% increase in income or a 20% decline in income. Empirical estimates from Kaplow [4] suggest that most fall near a relative risk aversion coefficient of 3 to 5.

As an illustration, if we assume a wealth of W and assume one's house is worth H where ρ is the probability of fire damage destroying the house where $\rho \in [0, 1]$, the choice to insure requires a premium payment of π and a deductible payment (d). The loss ratio, which is the expected loss in wealth to the agent relative to the insurance premium rate, can then be expressed as the ratio of expected losses over the premium rate:

$$lr = \frac{\pi(1-\rho) - H\rho}{\pi}. \tag{2.4}$$

For example, in the case where $W = \$250,000$ and $H = \$100,000$ where the cost of insurance was $\pi = \$2,000$ and $d = \$0$ with the likelihood of fire (ρ) equal to 1% the resulting loss ratio would be 50%.

If we compare only the expected wealth of each choice, given probabilities insurance will never be the optimal choice since by definition it will require a decline in expected wealth to be economically viable. To estimate the optimal choice of risk-averse investors dollar values must be transformed using a utility function that incorporates the degree of relative risk aversion.

In the case of insurance, two different wealth levels can be attained dependent upon two possible states (fire or no fire). In the case of fire, wealth (W_{IF}) is equal to initial wealth (W_0) minus the insurance premium (π) and deductible (d). In the case of no fire, no deductible is paid so that $W_{\mathrm{INF}} = W_0 - \pi$. In the expected utility framework, we can express expected utility as

$$\begin{aligned}\mathrm{EU}_I &= \rho * u(W_{\mathrm{IF}}) + [1-\rho] * u(W_{\mathrm{INF}}) \\ &= \rho * \frac{(W_0 - \pi - d)^{1-r}}{1-r} + [1-\rho] * \frac{(W_0 - \pi)^{1-r}}{1-r}.\end{aligned} \tag{2.5}$$

If we assume $r = 3$, then we obtain $\mathrm{EU}_I = -8.14e - 12$ units of utility. Alternatively, if the agent does not purchase insurance, $W_{\mathrm{NIF}} = W_0 - H$ and $W_{\mathrm{NINF}} = W_0$. In the case of no insurance, the expected utility ($\mathrm{EU}_{\mathrm{NI}}$) is equal to $-8.13e - 12$.

Because utility is an ordinal measure, we can conclude that even with a 50% loss ratio an individual with the specified preference function is better off purchasing insurance. The main reason is that the loss represents a very large share of wealth—in this case a loss of 40% of wealth. With this magnitude of loss an individual would only consider not buying insurance if the loss ratio fell beneath 45%. In the above example, if the initial wealth had instead been \$500,000, the maximum potential loss represents 20% of wealth. At 20% of wealth a household would be willing to buy insurance if the loss ratio is 60% or more.

The optimal retention limit is found by computing the point at which expected utility from retaining risk is exactly equal to the utility from transferring risk through insurance. At this point, π is set to the point where $\mathrm{EU}_I = \mathrm{EU}_{\mathrm{NI}}$. This can be rewritten as

$$\begin{aligned}\mathrm{EU}_I &= \rho(W - \pi - d)^{1-r} + [1-\rho](W - \pi)^{1-r} \\ &= \rho(W - H)^{1-r} + [1-\rho](W - d)^{1-r} \\ &= \mathrm{EU}_{\mathrm{NI}}.\end{aligned} \tag{2.6}$$

Even though an explicit analytical solution could not be determined for $\pi = f(W, d, r, \rho, H)$, the solution can be found numerically. With this example, the premium rate that solves the above equation is $\pi = 2,193$, which results in a loss ratio of 53.4% when the risk retention ratio is 20%. A greater loss ratio (or expected insurance claims from each premium dollar) will encourage greater transfer of risk as the cost of insurance is lower. More expensive insurance will encourage risk retention. A larger or catastrophic potential loss will lead to a greater willingness to insure. If the potential loss is small in terms of wealth, then the cost of insurance would need to be very low to induce a household to buy insurance.

Figure 1 illustrates the optimal tradeoff between risk transfer and risk retention given loss as a percentage of wealth and loss ratios. As the loss ratio declines, insurance becomes more expensive and individuals are only willing to insure if the loss represents a high percent

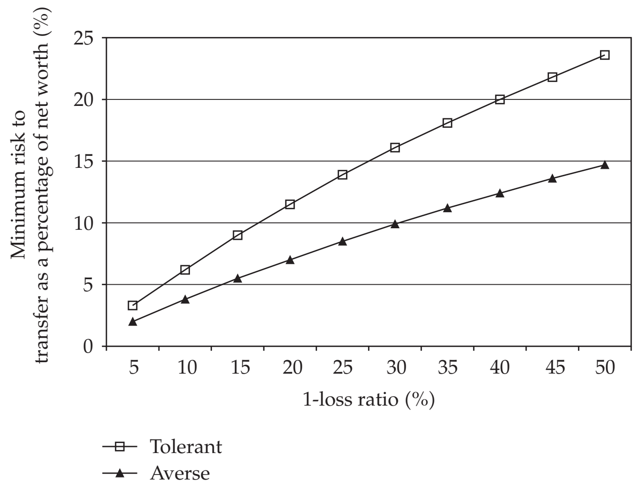


Figure 1: Optimal Risk Retention Limits.

of wealth. Risk tolerant individuals are assigned a coefficient of relative risk aversion of 3 and the risk-averse individuals are assigned a relative risk aversion of 5.

At a 40% loss ratio, characteristic of many property policies, it is only rational for individuals to insure losses that are about 1/7 of total wealth if they are risk averse. For the more risk tolerant an optimal risk retention limit is closer to 1/5 of total wealth. In practical terms these risk retention limits are enormous. A risk-averse household with a human wealth and wealth of $2 million would retain all risks below 10% of total wealth if the policy paid back 75% of premiums. With wealth this large products like comprehensive or collision insurance on a vehicle would have no appeal. It would also not make sense to carry anything other than the highest possible deductibles or to insure any assets whose value falls beneath $200,000.

While the conclusions of the model may appear extreme, they are valid in the sense that they are consistent with less intuitively extreme investment decision making. For example, retaining risk on property insurance either through very high deductibles or by not buying insurance provides an expected yield equal to the opportunity cost of insurance—which in the case of most property insurance is equivalent to a yield of at least 40%. The downside is that the household may subject themselves to a loss of, in this case, up to $200,000 to earn this return. However, expected returns on equities in the U.S. have been roughly 10%–12% and the tradeoff is exactly the same—the possibility of a large loss in wealth. In terms of wealth, there is no difference between a $100,000 loss due to property or casualty loss and a $100,000 loss in an investment portfolio. Both losses were the result of risk borne to maximize expected utility given uncertain outcomes.

Prudent risk management must also acknowledge the limitations of including human wealth when estimating risk retention limits. A wealth shock that falls beneath the estimated risk retention limit, for example, the loss of a $35,000 car to a 25-year-old, will be devastating if it wipes out liquid savings and if the ability to borrow against human capital is limited. Credit constraints in the face of a large loss can lead to a significant drop in consumption and a loss to illiquid projects that require a constant stream of cash to maintain (such as mortgages, business expenses, and student loan payments). Liquidity and access to credit are important considerations that impact optimal risk retention for those with high human wealth and few financial assets.

These results suggest that many are approaching the process of risk management by focusing on identifiable losses without recognizing the tradeoff of risk and return when choosing optimal risk management strategies. Most households are spending too much to prevent property and casualty risks while simultaneously retaining risk in their investment portfolio. This is neither wealth nor utility maximizing. However, the authors recognize that implementing this model will be difficult since many households are not prepared to retain risk they are accustomed to transferring.

3. Overweighting Losses

While the expected utility framework is rational in that it assumes disutility from wealth changes to be equal to the reduction in expected consumption, individuals appear to weigh gains and losses from risky choices differently. In fact, the persistent popularity of insurance products that protect against small losses suggests that individuals are willing to pay dearly to protect against minor losses while simultaneously paying insufficient attention to much larger risks (Kunreuther and Pauly [5]). Using results from experimental data, Kahneman and Tversky [1] point out three major limitations of expected utility theory which include the consistent overestimation of low probability events and underestimation of high probabilities. This would suggest that agents generally overinsure against rare events and underinsure against more common events, according to expected utility theory. The second finding is that utility functions (commonly referred to as value functions in the Prospect Theory literature) are generally concave concerning gains and convex for losses (Tversky and Kahneman [6]). Convexity for losses implies a large amount of disutility for relatively small losses and only a modest increase in disutility for larger losses. An individual with a prospect theory value function will place greater emphasis on avoiding small losses and a reduced emphasis on large losses than if they followed a conventional utility function.

The third problem with expected utility theory is that of absolute losses versus relative losses. As indicated by Kahneman and Tversky [1], individuals in experimental settings are shown to make decisions based on changes in their financial position rather than the impact on their final wealth. Kunreuther and Pauly [5] show that if an asset is only monetary, then it is rational to assess values based on absolute wealth changes. Instead, individuals appear to consider their current wealth a reference point and any loss from that reference point induces greater disutility than the dollar value of the loss would suggest.

Modeling a risky decision using prospect theory involves a reference point (R) from which gains and losses are assessed. The most relevant reference point is initial wealth ($R = W_0$). There is a small net loss associated with paying insurance premiums; however, there is an even larger net loss in the case of a fire under no insurance. If a fire occurs (F), there is a small net loss with and a large net loss without insurance; however, there is a no loss if insurance is not purchased and no fire occurs (NF). We define $X_i = W_i - R$, where $i = \{F, \mathrm{NF}\}$, as the net gains/losses. Because the utility function is assumed to be asymmetric around R, we then define the two parts of the value function which can be written as

$$v(\mathrm{X}_i) = \begin{cases} 0 & \text{if } \mathrm{X}_i \geq 0, \\ -\theta(\pi + d)^{\beta} & \text{if } \mathrm{X}_i < 0 \text{ with probability } \rho, \\ -\theta(\pi)^{\beta} & \text{if } \mathrm{X}_i < 0 \text{ with probability } (1 - \rho), \end{cases} \tag{3.1}$$

where θ is the loss aversion parameter and is greater than 1. The value function under the purchase of insurance (V_I) becomes

$$V_I = -g(\rho) * \theta(\pi + d)^{\beta} - g(1 - \rho) * \theta(\pi)^{\beta} \tag{3.2}$$

such that $g(\cdot)$ is a weighting function that accounts for overvaluing small probabilities and undervaluing large probabilities. The same reference point is used when we consider the case where no insurance is purchased. In this situation a fire results in a net loss and no fire results in a no loss. In this scenario we obtain

$$v(X_i) = \begin{cases} 0 & \text{if } X_i \geq 0, \\ -\theta(H)^{\beta} & \text{if } X_i < 0, \end{cases} \tag{3.3}$$

which then converts the value function (V_{NI}) to be

$$V_{\mathrm{NI}} = -g(\rho)\theta(H)^{\beta}. \tag{3.4}$$

To apply this function to the fire example above we assume $\beta = 0.88$ and $\theta = 2.25$ as suggested by Tversky and Kahneman [7]. To assess subjective probability biases we use the function derived by al-Nowaihi and Dhami [8] who derive their function from a more general form from Prelec [9], which is essentially

$$g(\rho) = e^{-\beta(-\ln(\rho))^{\alpha}}, \tag{3.5}$$

where we assume $\alpha = 0.80$. Using this function allows us to inflate our probability of 1% to 5% and deflate the probability of 99% to 98%. Given these adjusted probabilities and the downside risk parameter, we now look for the premium amount (π) that solves for $V_I = V_{\mathrm{NI}}$. At some point the premium will be high enough to outweigh the ability of the individual to manage the downside risk of a fire. For this scenario $\pi = 3,255$. Since both value functions are not functionally related to W, the optimal risk retention limit is unaffected by W_0. This differs from expected utility theory which assumes increasing risk tolerance with wealth from a potential loss of a given magnitude.

While prospect theory is useful as a means of understanding how individuals actually behave when faced with a decision to retain or transfer risk, it has little use as a normative tool to improve risk management practices. Estimates of optimal insurance using a prospect theory value function are high for small risks and low for more catastrophic risks. This is consistent with the current market for consumer insurance with its broad overuse of low deductibles, extended warranties, and protection of low-value tangible goods and simultaneous underuse of insurance products protecting more catastrophic risks of liability and loss of earnings. However, spending heavily to avoid small property and casualty risks while maintaining an optimal investment portfolio that requires acceptance of market risk (involving potential losses of a much greater magnitude) results in a wealth loss from framing these decisions separately. Prospect theory thus lends itself to modeling positive behavior but fails to guide practitioners or individuals in risk management.

Another possible shortcoming of Prospect Theory is that an insurance premium is itself a small-stakes loss. Payment of a premium assumes that an individual is willing to accept a small loss to avoid a possible larger loss; however, if the utility function is convex in losses (individuals are risk seeking when presented a choice that involves small and larger losses), then loss aversion becomes an even less plausible explanation unless the premium is not seen as a loss in itself. Such an individual may be induced into paying a relatively high premium to avoid a small-stakes loss since the loss from the random event creates greater disutility (per dollar) than the certain loss from the insurance premium. This appears to be the case in decisions involving the choice of homeowners' insurance deductibles. The choice between a $250 and a $500 deductible implies the payment of an added premium (say $25) to avoid an unlikely loss of the $250 difference between the two deductibles. For an individual with a total wealth of $100,000 (e.g., a destitute 80-year-old whose wealth consists of the present value of social security), choosing a $250 deductible implies a coefficient of relative risk aversion of over 500 (Sydnor [10]). Mehra and Prescott [11] describe the equity risk premium as a puzzle since it implies a coefficient of relative risk aversion of around 30. If the equity premium is a puzzle, then the risk retention premium is a mystery of inexplicable magnitude.

The only reasonable explanation for observed property insurance behavior is that the unexpected loss covered by insurance provides so much disutility that an individual is willing to give up large amounts of expected wealth to avoid it. This would be plausible if the losses caused painful regret when a peril caused a loss and if the premium itself was not viewed as a loss. Braun and Muermann [12] incorporate the unhappiness that is caused by regret into a utility function that can explain observed demand for insurance. Sydnor [10] modifies a utility function developed by Koszegi and Rabin [13] in which wealth losses are amplified by a factor that represents the relative pain felt by a random loss relative to a gain in wealth. The model suggests very strong amplification for low-stakes losses and more realistic amplification for large-scale loss, which appears consistent with observed behavior.

4. Conclusions

Proper household management of property and casualty risk requires an assessment of the dollar values of losses in various possible future states. When wealth in each state is transformed based on an individual's risk tolerance, it is possible to estimate the level of risk transfer through insurance. Using reasonable estimates of risk aversion, cost of property insurance, and initial wealth, optimal risk retention can be as high as 20% of initial household wealth. A risk retention limit of this magnitude would imply far higher deductibles in insurance policies and the abandonment of many popular policies that protect small losses. If human wealth is considered a component of total wealth, many young individuals would avoid insuring against all but catastrophic losses.

Advances in portfolio management and dissemination of normative investment science have led to broad acceptance of investment risk among households. For example, the percentage of U.S. households owning stock increased from 32% in 1989 to 52% in 2001 [14]. This increased acceptance of potential loss in investment portfolios has resulted in a significant improvement in household welfare. Holding a portfolio that is consistent with risk preferences implies an increase in expected utility relative to one that is excessively conservative. Similarly, dissemination of normative risk management science has the potential to improve welfare by illustrating the potential benefits of reducing costly protection of small losses and increasing protection against catastrophic risks. This study provides estimates of a large potential gain from increased acceptance of certain risks that are costly to insure.

The evidence from household risk retention preference through home insurance deductibles suggests that the market for property insurance reflects strong preferences for loss aversion. Benartzi and Thaler [15] provide evidence that Prospect Theory may also explain the observed high equity premium and an unwillingness to own risky assets. However, a clear welfare loss results from simultaneous ownership of risky securities and policy protection against small-scale risks. There is also evidence that individuals place insufficient weight on the utility loss from random large losses to total wealth by failing to insure adequately for potentially large losses such as large-scale liability losses and losses to human wealth (e.g., through disability income insurance). In the case of insurance, where there is little advice available (including from financial professionals) to maintain consistency among risky financial decisions, behavior may not accurately reveal preferences. If this is the case, then the application of a standard expected utility model may provide normative value that can help guide individuals and advisors toward making better decisions (Campbell [16]).

References

[1] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.

[2] R. G. Ibbotson, M. A. Milevsky, P. Chen, and K. X. Zhu, "Lifetime financial advice: human capital, asset allocation, and insurance," *The Research Foundation of CFA Institute*, 2007.

[3] S. Hanna, "Risk aversion and optimal insurance deductibles," in *Proceedings of the 31th Annual Conference on American Council on Consumer Interests (ACCI '89)*, pp. 141–147, Columbia, Mo, USA, 1989.

[4] L. Kaplow, "The value of a statistical life and the coefficient of relative risk aversion," *Journal of Risk and Uncertainty*, vol. 31, no. 1, pp. 23–34, 2005.

[5] H. Kunreuther and M. Pauly, "Insurance decision-making and market behavior," *Foundations and Trends in Microeconometrcs*, vol. 1, no. 2, pp. 63–127, 2005.

[6] A. Tversky and D. Kahneman, "Loss aversion in riskless choice: a reference dependent model," *Quarterly Journal of Economics*, vol. 106, pp. 1039–1062, 1991.

[7] A. Tversky and D. Kahneman, "Advances in prospect theory: cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[8] A. al-Nowaihi and S. Dhami, "A simple derivation of prelec's probability weighting function," Working Paper No. 05/20, University of Leicester, Department of Economics, 2005, http://www.le.ac.uk/economics/research/RePEc/lec/leecon/dp05-20.pdf.

[9] D. Prelec, "The probability weighting function," *Econometrica*, vol. 66, no. 3, pp. 497–527, 1998.

[10] J. Sydnor, "(Over)insuring modest risks," 2009, http://wsomfaculty.case.edu/sydnor/deductibles.pdf.

[11] R. Mehra and E. C. Prescott, "The equity premium: a puzzle," *Journal of Monetary Economics*, vol. 15, no. 2, pp. 145–161, 1985.

[12] M. Braun and A. Muermann, "The impact of regret on the demand for insurance," *The Journal of Risk and Insurance*, vol. 71, no. 4, pp. 737–767, 2004.

[13] B. Koszegi and M. Rabin, "Reference-dependent risk attitudes," *American Economic Review*, vol. 97, no. 4, pp. 1047–1073, 2007.

[14] E. Wolff, "Changes in Household Wealth in the 1980s and 1990s in the U.S.," in *International Perspectives on Household Wealth*, Elgar, 2004.

[15] S. Benartzi and R. Thaler, "Myopic loss aversion and the equity premium puzzle," *The Quarterly Journal of Economics*, vol. 110, no. 1, pp. 73–92, 1995.

[16] J. Campbell, "Household finance," *Journal of Finance*, vol. 61, no. 4, pp. 1553–1604, 2006.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 357321, 19 pages
doi:10.1155/2010/357321

Research Article

On Some Layer-Based Risk Measures with Applications to Exponential Dispersion Models

Olga Furman[1] and Edward Furman[2]

[1] *Actuarial Research Center, University of Haifa, Haifa 31905, Israel*
[2] *Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3*

Correspondence should be addressed to Edward Furman, efurman@mathstat.yorku.ca

Received 13 October 2009; Revised 21 March 2010; Accepted 10 April 2010

Academic Editor: Johanna Nešlehová

Copyright © 2010 O. Furman and E. Furman. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Layer-based counterparts of a number of well-known risk measures have been proposed and studied. Namely, some motivations and elementary properties have been discussed, and the analytic tractability has been demonstrated by developing closed-form expressions in the general framework of exponential dispersion models.

1. Introduction

Denote by $\mathcal{X}$ the set of (actuarial) risks, and let $0 \leq X \in \mathcal{X}$ be a random variable (rv) with cumulative distribution function (cdf) $F(x)$, decumulative distribution function (ddf) $\overline{F}(x) = 1 - F(x)$, and probability density function (pdf) $f(x)$. The functional $H : \mathcal{X} \to [0, \infty]$ is then referred to as a risk measure, and it is interpreted as the measure of risk inherent in X. Naturally, a quite significant number of risk measuring functionals have been proposed and studied, starting with the arguably oldest Value-at-Risk or VaR (cf. [1]), and up to the distorted (cf. [2–5]) and weighted (cf. [6, 7]) classes of risk measures.

More specifically, the Value-at-Risk risk measure is formulated, for every $0 < q < 1$, as

$$\mathrm{VaR}_q[X] = \inf\{x : F_X(x) \geq q\}, \tag{1.1}$$

which thus refers to the well-studied notion of the qth quantile. Then the family of distorted risk measures is defined with the help of an increasing and concave function $g : [0,1] \to [0,1]$, such that $g(0) = 0$ and $g(1) = 1$, as the following Choquet integral:

$$H_g[X] = \int_{\mathbf{R}_+} g\left(\overline{F}(x)\right) dx. \tag{1.2}$$

Last but not least, for an increasing nonnegative function $w : [0,\infty) \to [0,\infty)$ and the so-called weighted ddf $\overline{F}_w(x) = \mathbf{E}[\mathbf{1}\{X > x\}w(X)]/\mathbf{E}[w(X)]$ the class of weighted risk measures is given by

$$H_w[X] = \int_{\mathbf{R}_+} \overline{F}_w(x)dx. \tag{1.3}$$

Note that for at least once differentiable distortion function, we have that the weighted class contains the distorted one as a special case, that is, $H_g[X] = \mathbf{E}[Xg'(\overline{F}(X))]$ is a weighted risk measure with a dependent on $\overline{F}$ weight function.

Interestingly, probably in the view of the latter economic developments, the so-called "tail events" have been drawing increasing attention of insurance and general finance experts. Naturally therefore, tail-based risk measures have become quite popular, with the tail conditional expectation (TCE) risk measure being a quite remarkable example. For $0 < q < 1$ and thus $\overline{F}(\mathrm{VaR}_q[X]) \neq 0$, the TCE risk measure is formulated as

$$\mathrm{TCE}_q[X] = \frac{1}{\overline{F}(\mathrm{VaR}_q[X])} \int_{\mathrm{VaR}_q[X]}^{\infty} x\, dF(x). \tag{1.4}$$

Importantly, the TCE belongs in the class of distorted risk measures with the distortion function

$$g(x) = \frac{x}{1-q}\mathbf{1}(x < 1-q) + \mathbf{1}(x \geq 1-q), \tag{1.5}$$

where $\mathbf{1}$ denotes the indicator function (cf., e.g., [8]), as well as in the class of weighted risk measures with the weight function

$$w(x) = \mathbf{1}\{x \geq \mathrm{VaR}_q[X]\} \tag{1.6}$$

(cf., e.g., [6, 7]). The TCE risk measure is often referred to as the expected shortfall (ES) and the conditional Value-at-Risk (CVaR) when the pdf of X is continuous (cf., e.g., [9]).

Functional (1.4) can be considered a tail-based extension of the net premium $H[X] = \mathbf{E}[X]$. Furman and Landsman [10] introduced and studied a tail-based counterpart of the standard deviation premium calculation principle, which, for $0 < q < 1$, the tail variance

$$\mathrm{TV}_q[X] = \mathbf{Var}[X \mid X > \mathrm{VaR}_q[X]], \tag{1.7}$$

and a constant $\alpha \geq 0$, is defined as

$$\mathrm{TSD}_q[X] = \mathrm{TCE}_q[X] + \alpha \cdot \mathrm{TV}_q^{1/2}[X]. \tag{1.8}$$

For a discussion of various properties of the TSD risk measure, we refer to Furman and Landsman [10]. We note in passing that for $q \downarrow 0$, we have that $\mathrm{TSD}_q[X] \to \mathrm{SD}[X] = \mathbf{E}[X] + \alpha \cdot \mathbf{Var}^{1/2}[X]$.

The rest of the paper is organized as follows. In the next section we introduce and motivate layer-based extensions of functionals (1.4) and (1.8). Then in Sections 3 and 4 we analyze the aforementioned layer-based risk measures as well as their limiting cases in the general context of the exponential dispersion models (EDMs), that are to this end briefly reviewed in the appendix. Section 5 concludes the paper.

2. The Limited TCE and TSD Risk Measures

Let $0 < q < p < 1$ and let $X \in \mathcal{X}$ have a continuous and strictly increasing cdf. In many practical situations the degree of riskiness of the layer $(\mathrm{VaR}_q[X], \mathrm{VaR}_p[X])$ of an insurance contract is to be measured (certainly the layer width $\mathrm{VaR}_p[X] - \mathrm{VaR}_q[X] = \Delta_{q,p} > 0$). Indeed, the number of deductibles in a policy is often more than one, and/or there can be several reinsurance companies covering the same insured object. Also, there is the so-called "limited capacity" within the insurance industry to absorb losses resulting from, for example, terrorist attacks and catastrophes. In the context of the aforementioned events, the unpredictable nature of the threat and the size of the losses make it unlikely that the insurance industry can add enough capacity to cover them. In these and other cases neither (1.4) nor (1.8) can be applied since (1) both TCE and TSD are defined for one threshold, only, and (2) the aforementioned pair of risk measures is useless when, say, the expectations of the underlying risks are infinite, which can definitely be the case in the situations mentioned above.

Note 1. As noticed by a referee, the risk measure $H : \mathcal{X} \to [0, \infty]$ is often used to price (insurance) contracts. Naturally therefore, the limited TCE and TSD proposed and studied herein can serve as pricing functionals for policies with coverage modifications, such as, for example, policies with deductibles, retention levels, and so forth (cf., [11, Chapter 8]).

Next, we formally define the risk measures of interest.

Definition 2.1. Let $x_q = \mathrm{VaR}_q[X]$ and $x_p = \mathrm{VaR}_p[X]$, for $0 < q < p < 1$. Then the limited TCE and TSD risk measures are formulated as

$$\mathrm{LTCE}_{q,p}[X] = \mathbf{E}\left[X \mid x_q < X \le x_p\right], \tag{2.1}$$

and

$$\mathrm{LTSD}_{q,p}[X] = \mathbf{E}\left[X \mid x_q < X \le x_p\right] + \alpha \cdot \mathbf{Var}^{1/2}\left[X \mid x_q < X \le x_p\right], \tag{2.2}$$

respectively.

Clearly, the TCE and TSD are particular cases of their limited counterparts. We note in passing that the former pair of risk measures need not be finite for heavy tailed distributions, and they are thus not applicable. In this respect, limited variants (2.1) and (2.2) can provide a partial resolution. Indeed, for $k = 1, 2, \ldots$, we have that

$$\mathbf{E}\left[X^k \mid x_q < X \le x_p\right] = \frac{F(x_p)\mathbf{E}\left[X^k \mid X \le x_p\right] - F(x_q)\mathbf{E}\left[X^k \mid X \le x_q\right]}{F(x_p) - F(x_q)} < \infty, \tag{2.3}$$

regardless of the distribution of X.

We further enumerate some properties of the LTSD risk measure, which is our main object of study.

(1) *Translation Invariance.* For any constant $c \ge 0$, we have that

$$\mathrm{LTSD}_{q,p}[X + c] = \mathrm{LTSD}_{q,p}[X] + c. \tag{2.4}$$

(2) *Positive Homogeneity.* For any constant $d > 0$, we have that

$$\mathrm{LTSD}_{q,p}[d \cdot X] = d \cdot \mathrm{LTSD}_{q,p}[X]. \tag{2.5}$$

(3) *Layer Parity.* We call $X \in \mathcal{X}$ and $Y \in \mathcal{X}$ layer equivalent if for some $0 < q < p < 1$, such that $x_q = y_q$, $x_p = y_p$, and for every pair $\{(t_1, t_2) : q < t_1 < t_2 < p\}$, it holds that $\mathbf{P}[x_{t_1} < X \le x_{t_2}] = \mathbf{P}[y_{t_1} < Y \le y_{t_2}]$. In such a case, we have that

$$\mathrm{LTSD}_{t_1,t_2}[X] = \mathrm{LTSD}_{t_1,t_2}[Y]. \tag{2.6}$$

Literally, this property states that the LTSD risk measure for an arbitrary layer is only dependent on the cdf of that layer. Parity of the ddfs implies equality of LTSDs.

Although looking for original ways to assess the degree of (actuarial) riskiness is a very important task, subsequent applications of various theoretical approaches to a real-world data are not less essential. A significant number of papers have been devoted to deriving explicit formulas for some tail-based risk measures in the context of various loss distributions. The incomplete list of works discussing the TCE risk measure consists of, for example, Hürlimann [12] and Furman and Landsman [13], gamma distributions; Panjer [14], normal family; Landsman and Valdez [15], elliptical distributions; Landsman and Valdez [16], and Furman and Landsman [17], exponential dispersion models; and Chiragiev and Landsman [18], Vernic [19], Asimit et al. [20], Pareto distributions of the second kind.

As we have already noticed, the "unlimited" tail standard deviation risk measure has been studied in the framework of the elliptical distributions by Furman and Landsman [10]. Unfortunately, all members of the elliptical class are symmetric, while insurance risks are generally modeled by nonnegative and positively skewed random variables. These peculiarities can be fairly well addressed employing an alternative class of distribution laws. The exponential dispersion models include many well-known distributions such as normal, gamma, and inverse Gaussian, which, except for the normal, are nonsymmetric, have nonnegative supports, and can serve as adequate models for describing insurance risks' behavior. In this paper we therefore find it appropriate to apply both TSD and LTSD to EDM distributed risks.

3. The Limited Tail Standard Deviation Risk Measure for Exponential Dispersion Models

An early development of the exponential dispersion models is often attributed to Tweedie [21], however a more substantial and systematic investigation of this class of distributions was documented by Jørgensen [22, 23]. In his *Theory of dispersion models*, Jørgensen [24] writes that the main *raison d'étre* for the dispersion models is to serve as error distributions for generalized linear models, introduced by Nelder and Wedderburn [25]. Nowadays, EDMs play a prominent role in actuarial science and financial mathematics. This can be explained by the high level of generality that they enable in the context of statistical inference for widely popular distribution functions, such as normal, gamma, inverse Gaussian, stable, and many others. The specificity characterizing statistical modeling of actuarial subjects is that the underlying distributions mostly have nonnegative support, and many EDM members possess this important phenomenon, (for a formal definition of the EDMs, as well as for a brief review of some technical facts used in the sequel, cf., the appendix).

We are now in a position to evaluate the limited TSD risk measure in the framework of the EDMs. Recall that, for $0 < q < p < 1$, we denote by (x_q, x_p) an arbitrary layer having "attachment point" x_q and width $\Delta_{q,p}$. Also, let

$$h(x_q, x_p; \theta, \lambda) = \frac{\partial}{\partial \theta} \log\left(F(x_p; \theta, \lambda) - F(x_q; \theta, \lambda)\right) \tag{3.1}$$

denote the generalized layer-based hazard function, such that

$$\begin{aligned} h(x_q, x_1; \theta, \lambda) &= \frac{\partial}{\partial \theta} \log\left(\overline{F}(x_q; \theta, \lambda)\right) = h(x_q; \theta, \lambda), \\ h(x_0, x_p; \theta, \lambda) &= -\frac{\partial}{\partial \theta} \log\left(\overline{F}(x_p; \theta, \lambda)\right) = -h(x_p; \theta, \lambda), \end{aligned} \tag{3.2}$$

and thus

$$\begin{aligned} h(x_q, x_p; \theta, \lambda) = {} & \frac{\overline{F}(x_q; \theta, \lambda)}{\overline{F}(x_q; \theta, \lambda) - \overline{F}(x_p; \theta, \lambda)} h(x_q; \theta, \lambda) \\ & - \frac{\overline{F}(x_p; \theta, \lambda)}{\overline{F}(x_q; \theta, \lambda) - \overline{F}(x_p; \theta, \lambda)} h(x_p; \theta, \lambda). \end{aligned} \tag{3.3}$$

The next theorem derives expressions for the limited TCE risk measure, which is a natural precursor to deriving the limited TSD.

Theorem 3.1. *Assume that the natural exponential family (*NEF) which generates EDM is regular or at least steep (cf. [24, page 48]). Then the limited TCE risk measure

(i) for the reproductive EDM $Y \backsim$ ED(¯, œ^2) is given by

$$\mathrm{LTCE}_{q,p}[Y] = \mu + \sigma^2 \cdot h(x_q, x_p; \theta, \lambda) \tag{3.4}$$

and

(ii) for the additive EDM X $\backsim$ ED$^*(\theta, \lambda)$ is given by

$$\mathrm{LTCE}_{q,p}[X] = \lambda\kappa'(\theta) + h(x_q, x_p; \theta, \lambda). \tag{3.5}$$

Proof. We prove the reproductive case only, since the additive case follows in a similar fashion. By the definition of the limited TCE, we have that

$$\mathrm{LTCE}_{q,p}[Y] = \frac{\overline{F}(y_q)\mathbf{E}[Y \mid Y > y_q] - \overline{F}(y_p)\mathbf{E}[Y \mid Y > y_p]}{F(y_p) - F(y_q)}. \tag{3.6}$$

Further, following Landsman and Valdez [16], it can be shown that for every $0 < q < 1$, we have that

$$\mathbf{E}[Y \mid Y > y_q] = \mu + \sigma^2 \cdot h(y_q; \theta, \lambda), \tag{3.7}$$

which then, employing (3.1) and (3.3), yields

$$\begin{aligned} \mathrm{LTCE}_{q,p}[Y] &= \frac{\overline{F}(y_q; \theta, \lambda)(\mu + \sigma^2 \cdot h(y_q; \theta, \lambda)) - \overline{F}(y_p; \theta, \lambda)(\mu - \sigma^2 \cdot h(y_p; \theta, \lambda))}{\overline{F}(y_q; \theta, \lambda) - \overline{F}(y_p; \theta, \lambda)} \\ &= \mu + \sigma^2 \cdot h(y_q, y_p; \theta, \lambda) \end{aligned} \tag{3.8}$$

and hence completes the proof. □

In the sequel, we sometimes write $\mathrm{LTCE}_{q,p}[Y; \theta, \lambda]$ in order to emphasize the dependence on θ and λ.

Note 2. To obtain the results of Landsman and Valdez [16], we put p $\uparrow$ 1, and then, for instance, in the reproductive case, we end up with

$$\lim_{p \uparrow 1} \mathrm{LTCE}_{q,p}[Y] = \mu + \sigma^2 \cdot h(y_q; \theta, \lambda) = \mathrm{TCE}_q[Y], \tag{3.9}$$

subject to the existence of the limit.

Next theorem provides explicit expressions for the limited TSD risk measure for both reproductive and additive EDMs.

Theorem 3.2. *Assume that the* NEF which generates EDM is regular or at least steep. Then the limited TSD risk measure

(i) for the reproductive EDM $Y \backsim$ ED(¯, œ^2) is given by

$$\mathrm{LTSD}_{q,p}[Y] = \mathrm{LTCE}_{q,p}[Y] + \alpha \cdot \sqrt{\sigma^2 \frac{\partial}{\partial\theta} \mathrm{LTCE}_{q,p}[Y;\theta,\lambda]} \tag{3.10}$$

and

(ii) for the additive EDM $X \backsim \mathrm{ED}^*(\theta,\lambda)$ is given by

$$\mathrm{LTSD}_{q,p}[X] = \mathrm{LTCE}_{q,p}[X] + \alpha \cdot \sqrt{\frac{\partial}{\partial\theta} \mathrm{LTCE}_{q,p}[X;\theta,\lambda]}. \tag{3.11}$$

Proof. We again prove the reproductive case, only. Note that it has been assumed that $\kappa(\theta)$ is a differentiable function, and thus we can differentiate the following probability integral in θ under the integral sign (cf., the appendix):

$$\mathbf{P}(y_q < Y \le y_p) = \int_{y_q}^{y_p} e^{\lambda(\theta y - \kappa(\theta))} d\overline{\nu}_\lambda(y), \tag{3.12}$$

and hence, using Definition 2.1, we have that

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\left(\mathrm{LTCE}_{q,p}[Y;\theta,\lambda]\left(F(y_p;\theta,\lambda) - F(y_q;\theta,\lambda)\right)\right) \\
&\quad = \int_{y_q}^{y_p} \frac{\partial}{\partial\theta} y e^{\lambda(\theta y - \kappa(\theta))} d\overline{\nu}_\lambda(y) \\
&\quad = \lambda \int_{y_q}^{y_p} \left(y^2 e^{\lambda(\theta y - \kappa(\theta))} - y\kappa'(\theta) e^{\lambda(\theta y - \kappa(\theta))}\right) d\overline{\nu}_\lambda(y) \\
&\quad = \sigma^{-2}\left(\mathbf{E}\left[Y^2 \mid \mathbf{1}\{y_q < Y \le y_p\}\right] - \mu(\theta) \cdot \mathbf{E}\left[Y \mid \mathbf{1}\{y_q < Y \le y_p\}\right]\right),
\end{aligned} \tag{3.13}$$

with the last line following from the appendix. Further, by simple rearrangement and straightforward calculations, we obtain that

$$\begin{aligned}
\mathbf{E}\left[Y^2 \mid y_q < Y \le y_p\right] &= \frac{\int_{y_q}^{y_p} y^2 e^{\lambda(\theta y - \kappa(\theta))} d\overline{\nu}_\lambda(y)}{F(y_p;\theta,\lambda) - F(y_q;\theta,\lambda)} \\
&= \mu \cdot \mathrm{LTCE}_{q,p}[Y] + \sigma^2 \frac{(\partial/\partial\theta)\mathrm{LTCE}_{q,p}[Y;\theta,\lambda](F(y_p;\theta,\lambda) - F(y_q;\theta,\lambda))}{F(y_p;\theta,\lambda) - F(y_q;\theta,\lambda)} \\
&= \sigma^2 \frac{\partial}{\partial\theta}\mathrm{LTCE}_{q,p}[Y;\theta,\lambda] + \mathrm{LTCE}_{q,p}[Y]\left(\mu + \sigma^2 \frac{\partial}{\partial\theta} \log(F(y_p;\theta,\lambda) - F(y_q;\theta,\lambda))\right) \\
&= \sigma^2 \frac{\partial}{\partial\theta}\mathrm{LTCE}_{q,p}[Y;\theta,\lambda] + \left(\mathrm{LTCE}_{q,p}[Y;\theta,\lambda]\right)^2,
\end{aligned} \tag{3.14}$$

which along with the definition of the limited TSD risk measure completes the proof. □

We further consider two examples to elaborate on Theorem 3.2. We start with the normal distribution, which occupies a central role in statistical theory, and its position in statistical analysis of insurance problems is very difficult to underestimate, for example, due to the law of large numbers.

Example 3.3. Let $Y \backsim N(\mu, \sigma^2)$ be a normal random variable with mean μ and variance σ^2, then we can write the pdf of Y as

$$\begin{aligned}
f(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}y^2\right) \exp\left(\frac{1}{\sigma^2}\left(\mu y - \frac{1}{2}\mu^2\right)\right), \quad y \in \mathbf{R}.
\end{aligned} \tag{3.15}$$

If we take $\theta = \mu$ and $\lambda = 1/\sigma^2$, we see that the normal distribution is a reproductive EDM with cumulant function $\kappa(\theta) = \theta^2/2$. Denote by $\varphi(\cdot)$ and $\Phi(\cdot)$ the pdf and the cdf, respectively, of the standardized normal random variable. Then using Theorem 3.1, we obtain the following expression for the limited TCE risk measure for the risk Y:

$$\mathrm{LTCE}_{q,p}[Y] = \mu + \sigma \frac{\varphi(\sigma^{-1}(y_q - \mu)) - \varphi(\sigma^{-1}(y_p - \mu))}{\Phi(\sigma^{-1}(y_p - \mu)) - \Phi(\sigma^{-1}(y_q - \mu))}. \tag{3.16}$$

If we put $p \uparrow 1$, then the latter equation reduces to the result of Landsman and Valdez [16]. Namely, we have that

$$\lim_{p\uparrow 1} \mathrm{LTCE}_{q,p}[Y] = \mu + \sigma \frac{\varphi(\sigma^{-1}(y_q - \mu))}{1 - \Phi(\sigma^{-1}(y_q - \mu))} = \mathrm{TCE}_q[Y]. \tag{3.17}$$

Further, let $z_q = (y_q - \mu)/\sigma$ and $z_p = (y_p - \mu)/\sigma$. Then

$$\sigma^2 \frac{\partial}{\partial\theta} \mathrm{LTCE}_{q,p}[Y;\theta,\lambda] = \sigma^2 \left(1 + \frac{\varphi(z_q)z_q - \varphi(z_p)z_p}{\Phi(z_p) - \Phi(z_q)} - \left(\frac{\varphi(z_q) - \varphi(z_p)}{\Phi(z_p) - \Phi(z_q)} \right)^2 \right). \tag{3.18}$$

Consequently, the limited TSD risk measure is as follows:

$$\begin{aligned} &\mathrm{LTSD}_{q,p}[Y] \\ &\quad = \mu + \sigma \frac{\varphi(z_q) - \varphi(z_p)}{\Phi(z_p) - \Phi(z_q)} + \alpha \sqrt{\sigma^2 \left(1 + \frac{\varphi(z_q)z_q - \varphi(z_p)z_p}{\Phi(z_p) - \Phi(z_q)} - \left(\frac{\varphi(z_q) - \varphi(z_p)}{\Phi(z_p) - \Phi(z_q)} \right)^2 \right)}. \end{aligned} \tag{3.19}$$

We proceed with the gamma distributions, which have been widely applied in various fields of actuarial science. It should be noted that these distribution functions possess positive support and positive skewness, which is important for modeling insurance losses. In addition, gamma rvs have been well-studied, and they share many tractable mathematical properties which facilitate their use. There are numerous examples of applying gamma distributions for modeling insurance portfolios (cf., e.g., [12, 13, 26, 27]).

Example 3.4. Let $X \backsim \mathrm{Ga}(\gamma,\beta)$ be a gamma rv with shape and rate parameters equal γ and β, correspondingly. The pdf of X is

$$f(x) = \frac{1}{\Gamma(\gamma)} e^{-\beta x} x^{\gamma-1} \beta^{\gamma} = \frac{1}{\Gamma(\gamma)} x^{\gamma-1} \exp(-\beta x + \gamma \log(\beta)), \quad x > 0. \tag{3.20}$$

Hence the gamma rv can be represented as an additive EDM with the following pdf:

$$f(x) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} \exp(\theta x + \lambda \log(-\theta)), \tag{3.21}$$

where $x > 0$ and $\theta < 0$. The mean and variance of X are $\mathbf{E}[X] = -\lambda/\theta$ and $\mathbf{Var}[X] = \lambda/\theta^2$. Also, $\theta = -\beta$, $\lambda = \gamma$, and $\kappa(\theta) = -\log(-\theta)$. According to Theorem 3.1, the limited tail conditional expectation is

$$\mathrm{LTCE}_{q,p}[X] = -\frac{\lambda}{\theta} \frac{F(x_p;\theta,\lambda+1) - F(x_q;\theta,\lambda+1)}{F(x_p;\theta,\lambda) - F(x_q;\theta,\lambda)}. \tag{3.22}$$

Putting $p \uparrow 1$ we obtain that

$$\lim_{p\uparrow 1} \left(-\frac{\lambda}{\theta} \right) \frac{F(x_p;\theta,\lambda+1) - F(x_q;\theta,\lambda+1)}{F(x_p;\theta,\lambda) - F(x_q;\theta,\lambda)} = -\frac{\lambda}{\theta} \frac{\overline{F}(x_q;\theta,\lambda+1)}{\overline{F}(x_q;\theta,\lambda)} = \mathrm{TCE}_q[X], \tag{3.23}$$

which coincides with [13, page 643]. To derive an expression for the limited TSD risk measure, we use Theorem 3.2, that is,

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathrm{LTCE}_{q,p}[X;\theta,\lambda] &= \frac{\partial}{\partial\theta}\left(-\frac{\lambda}{\theta}\frac{F(x_p;\theta,\lambda+1)-F(x_q;\theta,\lambda+1)}{F(x_p;\theta,\lambda)-F(x_q;\theta,\lambda)}\right)\\ &= \frac{\lambda}{\theta^2}\frac{F(x_p;\theta,\lambda+1)-F(x_q;\theta,\lambda+1)}{F(x_p;\theta,\lambda)-F(x_q;\theta,\lambda)}\\ &\quad -\frac{\lambda}{\theta}\left(\frac{\partial}{\partial\theta}\frac{F(x_p;\theta,\lambda+1)-F(x_q;\theta,\lambda+1)}{F(x_p;\theta,\lambda)-F(x_q;\theta,\lambda)}\right).\end{aligned} \tag{3.24}$$

Further, since for $n = 1, 2, \ldots,$

$$\begin{aligned}&\frac{\partial}{\partial\theta}\left(F(x_p;\theta,\lambda+n)-F(x_q;\theta,\lambda+n)\right)\\ &\quad= \int_{x_q}^{x_p}\frac{\partial}{\partial\theta}\left(\frac{1}{\Gamma(\lambda+n)}x^{\lambda+n-1}\exp(\theta x+(\lambda+n)\log(-\theta))\right)dx\\ &\quad= \int_{x_q}^{x_p} f(x;\theta,\lambda+n)\left(x+\frac{\lambda+n}{\theta}\right)dx\\ &\quad= -\frac{\lambda+n}{\theta}\left(\int_{x_q}^{x_p} f(x;\theta,\lambda+n+1)dx-\int_{x_q}^{x_p} f(x;\theta,\lambda+n)dx\right),\end{aligned} \tag{3.25}$$

the limited TSD risk measure for gamma is given by

$$\begin{aligned}&\mathrm{LTSD}_{q,p}[X]\\ &= \left(-\frac{\lambda}{\theta}\right)\frac{F(x_p;\theta,\lambda+1)-F(x_q;\theta,\lambda+1)}{F(x_p;\theta,\lambda)-F(x_q;\theta,\lambda)}\\ &\quad+\alpha\sqrt{\frac{\lambda}{\theta^2}\left((\lambda+1)\frac{F(x_p;\theta,\lambda+2)-F(x_q;\theta,\lambda+2)}{F(x_p;\theta,\lambda)-F(x_q;\theta,\lambda)}-\lambda\left(\frac{F(x_p;\theta,\lambda+1)-F(x_q;\theta,\lambda+1)}{F(x_p;\theta,\lambda)-F(x_q;\theta,\lambda)}\right)^2\right)}.\end{aligned} \tag{3.26}$$

In the sequel, we consider gamma and normal risks with equal means and variances, and we explore them on the interval $(t, 350]$, with $50 < t < 350$. Figure 1 depicts the results. Note that both LTCE and LTSD imply that the normal distribution is riskier than gamma for lower attachment points and vice-versa, that is quite natural bearing in mind the tail behavior of the two.

Although the EDMs are of pivotal importance in actuarial mathematics, they fail to appropriately describe heavy-tailed (insurance) losses. To elucidate on the applicability of the layer-based risk measures in the context of the probability distributions possessing heavy tails, we conclude this section with a simple example.

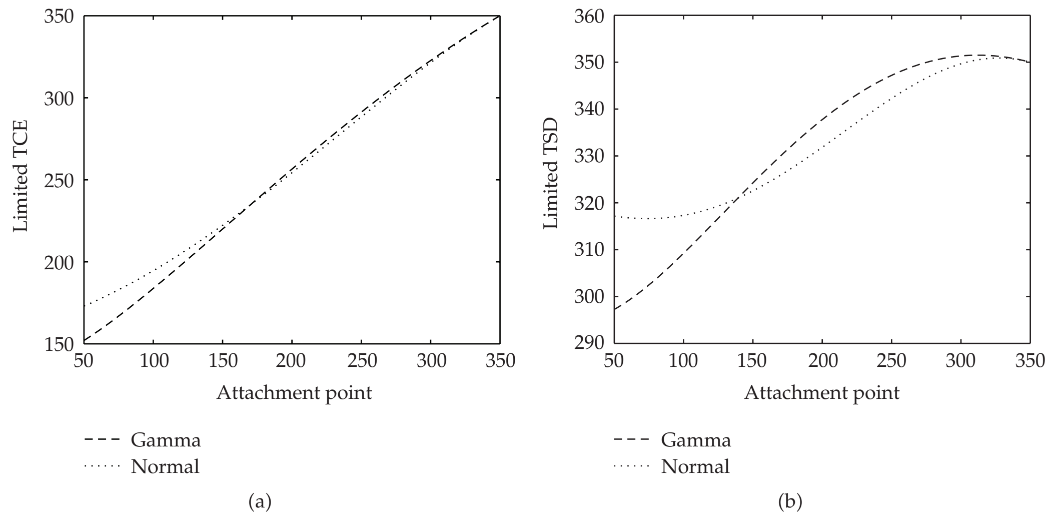


Figure 1: LTCE and LTSD for normal and gamma risks with means 150 and standard deviations 100, alpha = 2.

Example 3.5. Let X $\backsim$ Pa(γ, β) be a Pareto rv with the pdf

$$f(x) = \frac{\gamma\beta^{\gamma}}{x^{\gamma+1}}, \quad x > \beta > 0, \tag{3.27}$$

and $\gamma > 0$. Certainly, the Pareto rv is not a member of the EDMs, though it belongs to the log-exponential family (LEF) of distributions (cf. [7]). The LEF is defined by the differential equation

$$F(dx; \lambda, \nu) = \exp\{\lambda \log(x) - \kappa(\lambda)\}\nu(dx), \tag{3.28}$$

where λ is a parameter, ν is a measure, and $\kappa(\lambda) = \log \int_0^{\infty} x^{\lambda}\nu(dx)$ is a normalizing constant (the parameters should not be confused with the ones used in the context of the EDMs). Then X is easily seen to belong in LEF with the help of the reparameterization $\nu(dx) = x^{-1}dx$, and $\lambda = -\gamma$.

In this context, it is straightforward to see that $\mathbf{E}[X]$ is infinite for $\gamma \leq 1$, which thus implies infiniteness of the TCE risk measure. We can however readily obtain the limited variant as follows:

$$\mathrm{LTCE}_{q,p}[X] = \frac{1}{\mathbf{P}[x_q < X \leq x_p]} \int_{x_q}^{x_p} \frac{\gamma\beta^{\gamma}}{x^{\gamma}} dx = \frac{\gamma x_p x_q}{\gamma - 1} \left(\frac{x_p^{\gamma-1} - x_q^{\gamma-1}}{x_p^{\gamma} - x_q^{\gamma}} \right), \tag{3.29}$$

that is finite for any $\gamma > 0$. Also, since, for example, for $\gamma < 1$, we have that $x_p^{\gamma-1} - x_q^{\gamma-1} < 0$, the limited TCE risk measure is positive, as expected. The same is true for $\gamma \geq 1$.

We note in passing that, for $\gamma > 1$ and $p \uparrow 1$ and thus $x_p \rightarrow \infty$, we have that

$$\mathrm{TCE}_q[X] = \lim_{p \uparrow 1} \frac{\gamma x_p x_q}{\gamma - 1} \left(\frac{x_p^{\gamma-1} - x_q^{\gamma-1}}{x_p^{\gamma} - x_q^{\gamma}} \right) = \frac{\gamma}{\gamma - 1} x_q, \tag{3.30}$$

which confirms the corresponding expression in Furman and Landsman [8].

Except for the Pareto distribution, the LEF consists of, for example, the log-normal and inverse-gamma distributions, for which expressions similar to (3.29) can be developed in the context of the limited TCE and limited TSD risk measures, thus providing a partial solution to the heavy-tailness phenomenon.

4. The Tail Standard Deviation Risk Measure for Exponential Dispersion Models

The tail standard deviation risk measure was proposed in [10] as a possible quantifier of the so-called tail riskiness of the loss distribution. The above-mentioned authors applied this risk measure to elliptical class of distributions, which consists of such well-known pdfs as normal and student-t. Although the elliptical family is very useful in finance, insurance industry imposes its own restrictions. More specifically, insurance claims are always positive and mostly positively skewed. In this section we apply the TSD risk measure to EDMs.

The following corollary develops formulas for the TSD risk measure both in the reproductive and additive EDMs cases. Recall that we denote the ddf of say X by $\overline{F}(\cdot;\theta,\lambda)$ to emphasize the parameters θ and λ, and we assume that

$$\lim_{p \uparrow 1} \mathrm{LTSD}_{q,p}[X] < \infty. \tag{4.1}$$

The proof of the next corollary is left to the reader.

Corollary 4.1. *Under the conditions in Theorem 3.1, the tail standard deviation risk measure is*

$$\mathrm{TSD}_q[Y] = \mathrm{TCE}_q[Y] + \alpha \sqrt{\sigma^2 \frac{\partial}{\partial \theta} \mathrm{TCE}_q[Y;\theta,\lambda]} \tag{4.2}$$

in the context of the reproductive EDMs, and

$$\mathrm{TSD}_q[X] = \mathrm{TCE}_q[X] + \alpha \sqrt{\frac{\partial}{\partial \theta} \mathrm{TCE}_q[X;\theta,\lambda]} \tag{4.3}$$

in the context of the additive EDMs.

We further explore the TSD risk measure in some particular cases of EDMs, which seem to be of practical importance.

Example 4.2. Let $Y \backsim N(\mu,\sigma^2)$ be again some normal rv with mean μ and variance σ^2. Then we easily evaluate the TSD risk measure using Corollary 4.1 and Example 3.3 as follows:

$$\mathrm{TSD}_q[X] = \mu + \sigma\frac{\varphi(z_q)}{1-\Phi(z_q)} + \alpha\sqrt{\sigma^2\left(1+\frac{\varphi(z_q)}{1-\Phi(z_q)}z_q - \left(\frac{\varphi(z_q)}{1-\Phi(z_q)}\right)^2\right)}, \tag{4.4}$$

which coincides with [10].

Example 4.3. Let $X \backsim \mathrm{Ga}(\gamma,\beta)$ be a gamma rv with shape and scale parameters equal γ and β, correspondingly. Taking into account Example 3.4 and Corollary 4.1 leads to

$$\begin{aligned}
&\mathrm{TSD}_q[X]\\
&\quad= -\frac{\lambda}{\theta}\frac{\overline{F}(x_q;\theta,\lambda+1)}{\overline{F}(x_q;\theta,\lambda)} + \alpha\sqrt{\frac{\lambda}{\theta^2}\left((\lambda+1)\frac{\overline{F}(x_q;\theta,\lambda+2)}{\overline{F}(x_q;\theta,\lambda)} - \lambda\left(\frac{\overline{F}(x_q;\theta,\lambda+1)}{\overline{F}(x_q;\theta,\lambda)}\right)^2\right)}\\
&\quad= \frac{\gamma}{\beta}\frac{\overline{F}(x_q;\gamma+1,\beta)}{\overline{F}(x_q;\gamma,\beta)} + \alpha\sqrt{\frac{\gamma}{\beta^2}\left((\gamma+1)\frac{\overline{F}(x_q;\gamma+2,\beta)}{\overline{F}(x_q;\gamma,\beta)} - \gamma\left(\frac{\overline{F}(x_q;\gamma+1,\beta)}{\overline{F}(x_q;\gamma,\beta)}\right)^2\right)},
\end{aligned} \tag{4.5}$$

where the latter equation follows because of the reparameterization $\theta = -\beta$ and $\lambda = \gamma$.

We further discuss the inverse Gaussian distribution, which possesses heavier tails than, say, gamma distribution, and therefore it is somewhat more tolerant to large losses.

Example 4.4. Let $Y \backsim \mathrm{IG}(\mu,\lambda)$ be an inverse Gaussian rv. We then can write its pdf as

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}}\exp\left(\lambda\left(-\frac{y}{2\mu^2} - \frac{1}{2y} + \frac{1}{\mu}\right)\right), \quad y \in [0,\infty), \tag{4.6}$$

(cf. [24]), which means that Y belongs to the reproductive EDMs, with $\theta = -1/(2\mu^2)$ and $\kappa(\theta) = -1/\mu = -\sqrt{-2\theta}$. Further, due to Corollary 4.1 we need to calculate

$$\frac{\partial}{\partial\theta}\mathrm{TCE}_q[Y;\theta,\lambda] = \frac{\partial}{\partial\theta}\left(\mu(\theta) + \sigma^2\frac{\partial}{\partial\theta}\log\overline{F}_Y(y_q;\theta,\lambda)\right) = \mu'(\theta) + \sigma^2\frac{\partial}{\partial\theta}\frac{(\partial/\partial\theta)\overline{F}_Y(y_q;\theta,\lambda)}{\overline{F}_Y(y_q;\theta,\lambda)}. \tag{4.7}$$

To this end, note that the ddf of Y is

$$\overline{F}(y_q;\mu(\theta),\lambda) = \overline{\Phi}\left(\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)} - 1\right)\right) - e^{2\lambda/\mu(\theta)}\Phi\left(-\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)} + 1\right)\right) \tag{4.8}$$

(cf., e.g., [28]), where $\overline{\Phi}(\cdot)$ is the ddf of the standardized normal random variable. Hence, by simple differentiation and noticing that

$$\mu'(\theta) = (-2\theta)^{-3/2} = \mu(\theta)^3, \tag{4.9}$$

we obtain that

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\overline{F}(y_q;\mu(\theta),\lambda)\\
&\quad = \mu(\theta)\left(\sqrt{\lambda y_q}\varphi\left(\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}-1\right)\right) - e^{2\lambda/\mu(\theta)}\sqrt{\lambda y_q}\varphi\left(-\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}+1\right)\right)\right)\\
&\qquad + 2\lambda\mu(\theta)e^{2\lambda/\mu(\theta)}\Phi\left(-\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}+1\right)\right).
\end{aligned} \tag{4.10}$$

Notably,

$$\sqrt{\lambda y_q}\varphi\left(\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}-1\right)\right) = e^{2\lambda/\mu(\theta)}\sqrt{\lambda y_q}\varphi\left(-\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}+1\right)\right), \tag{4.11}$$

and therefore (4.10) results in

$$\frac{\partial}{\partial\theta}\overline{F}(y_q;\mu(\theta),\lambda) = 2\lambda\mu(\theta)e^{2\lambda/\mu(\theta)}\Phi\left(-\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}+1\right)\right). \tag{4.12}$$

Consequently, the expression for the TCE risk measure, obtained by Landsman and Valdez [16], is simplified to

$$\mathrm{TCE}_q[Y;\theta,\lambda] = \mu(\theta) + \frac{2\mu(\theta)}{\overline{F}(y_q;\mu(\theta),\lambda)}e^{2\lambda/\mu(\theta)}\Phi\left(-\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}+1\right)\right). \tag{4.13}$$

In order to derive the TSD risk measure we need to differentiate again, that is,

$$\begin{aligned}
\frac{\partial}{\partial\theta}\mathrm{TCE}_q[Y;\theta,\lambda] &= \frac{\partial}{\partial\theta}\left(\mu(\theta) + \frac{2\mu(\theta)}{\overline{F}(y_q;\mu(\theta),\lambda)}e^{2\lambda/\mu(\theta)}\Phi\left(-\sqrt{\frac{\lambda}{y_q}}\left(\frac{y_q}{\mu(\theta)}+1\right)\right)\right)\\
&= \mu(\theta)^3\left(1 + \frac{\partial}{\partial\theta}\frac{2\mu(\theta)e^{2\lambda/\mu(\theta)}\Phi\left(-\sqrt{\lambda/y_q}\left(y_q/\mu(\theta)+1\right)\right)}{\overline{F}(y_q;\mu(\theta),\lambda)}\right),
\end{aligned} \tag{4.14}$$

where we use $\mu'(\theta) = \mu(\theta)^3$. Further, we have that

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\frac{2\mu(\theta)e^{2\lambda/\mu(\theta)}\Phi\left(-\sqrt{\lambda/y_q}\left(y_q/\mu(\theta)+1\right)\right)}{\overline{F}\left(y_q;\mu(\theta),\lambda\right)} \\
&\quad = 2\frac{\mu(\theta)^3 e^{2\lambda/\mu(\theta)}\left(\Phi(\tilde{y}_q)\left(1-2\lambda/\mu(\theta)\right)+\left(\sqrt{\lambda y_q}/\mu(\theta)\right)\varphi(\tilde{y}_q)\right)}{\overline{F}\left(y_q;\mu(\theta),\lambda\right)} \\
&\quad\quad - \frac{\lambda\left(2\mu(\theta)e^{2\lambda/\mu(\theta)}\Phi(\tilde{y}_q)\right)^2}{\overline{F}\left(y_q;\mu(\theta),\lambda\right)^2},
\end{aligned} \tag{4.15}$$

where $\tilde{y}_q = -\sqrt{\lambda/y_q}\left(y_q/\mu(\theta)+1\right)$. Therefore

$$\begin{aligned}
\mathrm{TSD}_q[Y] = {}& \mu\left(1+\frac{\Phi(\tilde{y}_q)}{\overline{F}(y_q;\mu,\lambda)}2e^{2\lambda/\mu}\right) \\
&+\alpha\sqrt{\frac{\mu^3}{\lambda}\left(1+\frac{e^{2\lambda/\mu}\left(\Phi(\tilde{y}_q)\left(1-2\lambda/\mu\right)+\left(\sqrt{\lambda y_q}/\mu\right)\varphi(\tilde{y}_q)\right)}{\overline{F}(y_q;\mu,\lambda)}-\frac{\lambda\left(e^{2\lambda/\mu}\Phi(\tilde{y}_q)\right)^2}{\mu\overline{F}(y_q;\mu,\lambda)^2}\right)}
\end{aligned} \tag{4.16}$$

subject to $\mathbf{Var}[Y] = \mu^3/\lambda$.

5. Concluding Comments

In this work we have considered certain layer-based risk measuring functionals in the context of the exponential dispersion models. Although we have made an accent on the absolutely continuous EDMs, similar results can be developed for the discrete members of the class. Indeed, distributions with discrete supports often serve as frequency models in actuarial mathematics. Primarily in expository purposes, we further consider a very simple frequency distribution, and we evaluate the TSD risk measure for it. More encompassing formulas can however be developed with some effort for other EDM members of, say, the $(a, b, 0)$ class (cf., [11, Chapter 6]) as well as for limited TCE/TSD risk measures.

Example 5.1. Let $X \backsim \text{Poisson}(\mu)$ be a Poisson rv with the mean parameter μ. Then the probability mass function of X is written as

$$p(x) = \frac{1}{x!}\mu^x e^{-\mu} = \frac{1}{x!}\exp\left(x\log(\mu)-\mu\right), \quad x = 0, 1, \ldots, \tag{5.1}$$

which belongs to the additive EDMs in view of the reparametrization $\theta = \log(\mu)$, $\lambda = 1$, and $\kappa(\theta) = e^{\theta}$.

Motivated by Corollary 4.1, we differentiate (cf. [16], for the formula for the TCE risk measure)

$$\begin{aligned}
\frac{\partial}{\partial\theta}\mathrm{TCE}_q(X;\theta,\lambda) &= \frac{\partial}{\partial\theta}\left(e^{\theta}\left(1+\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}\right)\right)\\
&= e^{\theta}\left(1+\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}+\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}\left(x_q-e^{\theta}\right)-e^{\theta}\left(\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}\right)^2\right)\\
&= e^{\theta}\left(\frac{\overline{F}(x_q-1;\theta,1)}{\overline{F}(x_q;\theta,1)}+\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}\left(x_q-e^{\theta}\right)-e^{\theta}\left(\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}\right)^2\right),
\end{aligned} \tag{5.2}$$

where the latter equation follows because

$$\overline{F}(x_q;\theta,1)+p(x_q;\theta,1)=\overline{F}(x_q-1;\theta,1). \tag{5.3}$$

The formula for the TSD risk measure is then

$$\begin{aligned}
&\mathrm{TSD}_q(X)\\
&\quad= e^{\theta}\left(1+\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}\right)+\alpha\sqrt{e^{\theta}\left(\frac{\overline{F}(x_q-1;\theta,1)}{\overline{F}(x_q;\theta,1)}+\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}z_q-e^{\theta}\left(\frac{p(x_q;\theta,1)}{\overline{F}(x_q;\theta,1)}\right)^2\right)},
\end{aligned} \tag{5.4}$$

where $\mathbf{E}[X]=\mathbf{Var}[X]=e^{\theta}$ and $z_q=x_q-e^{\theta}$.

Appendix

A. Exponential Dispersion Models

Consider a σ-finite measure ν on $\mathbf{R}$ and assume that ν is nondegenerate. Next definition is based on [24].

Definition A.1. The family of distributions of $X \backsim \mathrm{ED}^*(\theta,\lambda)$ for $(\theta,\lambda)\in\Theta\times\Lambda$ is called the additive exponential dispersion model generated by ν. The corresponding family of distributions of $Y = X/\lambda \backsim \mathrm{ED}(\mu,\sigma^2)$, where $\mu=\tau(\theta)$ and $\sigma^2=1/\lambda$ are the mean value and the dispersion parameters, respectively, is called the reproductive exponential dispersion model generated by ν. Moreover, given some measure ν_λ the representation of $X \backsim \mathrm{ED}^*(\theta,\lambda)$ is as follows:

$$\exp(\theta x-\lambda\kappa(\theta))\nu_\lambda(dx). \tag{A.1}$$

If in addition the measure ν_λ has density $c^*(x;\lambda)$ with respect to some fixed measure (typically Lebesgue measure or counting measure), the density for the additive model is

$$f^*(x;\theta,\lambda) = c^*(x;\lambda)\exp(\theta x - \lambda\kappa(\theta)), \quad x \in \mathbf{R}. \tag{A.2}$$

Similarly, we obtain the following representation of $Y \backsim \mathrm{ED}(\mu,\sigma^2)$ as

$$\exp(\lambda(y\theta - \kappa(\theta)))\overline{\nu}_\lambda(dy), \tag{A.3}$$

where $\overline{\nu}_\lambda$ denotes ν_λ transformed by the duality transformation $X = Y/\sigma^2$. Again if the measure $\overline{\nu}_\lambda$ has density $c(y;\lambda)$ with respect to a fixed measure, the reproductive model has the following pdf:

$$f(y;\theta,\lambda) = c(y;\lambda)\exp(\lambda(\theta y - \kappa(\theta))), \quad y \in \mathbf{R}. \tag{A.4}$$

Note that θ and λ are called canonical and index parameters, $\Theta = \{\theta \in \mathbf{R} : \kappa(\theta) < \infty\}$ for some function $\kappa(\theta)$ called the cumulant, and Λ is the index set. Throughout the paper, we use $X \backsim \mathrm{ED}^*(\mu,\sigma^2)$ and $X \backsim \mathrm{ED}(\theta,\lambda)$ for the additive form with parameters μ and σ^2 and the reproductive form with parameters θ and λ, correspondingly, depending on which notation is more convenient.

We further briefly review some properties of the EDMs related to this work. Consider the reproductive form first, that is, $Y \backsim \mathrm{ED}(\mu,\sigma^2)$, then

(i) the cumulant generating function (cgf) of Y is, for $\theta' = \theta + t/\lambda$,

$$\begin{aligned} K(t) = \log \mathbf{E}\left[e^{tY}\right] &= \log\left(\int_{\mathbf{R}} \exp\left(\lambda\left(y\left(\theta + \frac{t}{\lambda}\right) - \kappa(\theta)\right)\right) d\overline{\nu}_\lambda(y)\right) \\ &= \log\left(\exp\left(\lambda\left(\kappa\left(\theta + \frac{t}{\lambda}\right) - \kappa(\theta)\right)\right)\int_{\mathbf{R}} \exp(\lambda[\theta' y - \kappa(\theta')]) d\overline{\nu}_\lambda(y)\right) \\ &= \lambda\left(\kappa\left(\theta + \frac{t}{\lambda}\right) - \kappa(\theta)\right), \end{aligned} \tag{A.5}$$

(ii) the moment generating function (mgf) of Y is given by

$$M(t) = \exp\left(\lambda\left(\kappa\left(\frac{\theta + t}{\lambda}\right) - \kappa(\theta)\right)\right), \tag{A.6}$$

(iii) the expectation of Y is

$$\mathbf{E}[Y] = \left.\frac{\partial K(t)}{\partial t}\right|_{t=0} = \kappa'(\theta) = \mu, \tag{A.7}$$

(iv) the variance of Y is

$$\mathbf{Var}[Y] = \left.\frac{\partial^2 K(t)}{\partial t^2}\right|_{t=0} = \sigma^2 \kappa^{(2)}(\theta). \tag{A.8}$$

Consider next an rv X following an additive EDM, that is, $X \backsim \mathrm{ED}^*(\theta, \lambda)$. Then, in a similar fashion,

(i) the cgf of X is

$$K(t) = \lambda(\kappa(\theta + t) - \kappa(\theta)), \tag{A.9}$$

(ii) the mgf of X is

$$M(t) = \exp(\lambda(\kappa(\theta + t) - \kappa(\theta))), \tag{A.10}$$

(iii) the expectation of X is

$$\mathbf{E}[X] = \lambda\kappa'(\theta), \tag{A.11}$$

(iv) the variance of X is

$$\mathbf{Var}[X] = \lambda\kappa^{(2)}(\theta). \tag{A.12}$$

For valuable examples of various distributions belonging in the EDMs we refer to Jørgensen [24].

Acknowledgments

This is a concluding part of the authors' research supported by the Zimmerman Foundation of Banking and Finance, Haifa, Israel. In addition, Edward Furman acknowledges the support of his research by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Also, the authors are grateful to two anonymous referees and the editor, Professor Johanna Nešlehová, for constructive criticism and suggestions that helped them to revise the paper.

References

[1] D. H. Leavens, "Diversification of investments," *Trusts and Estates*, vol. 80, no. 5, pp. 469–473, 1945.

[2] D. Denneberg, *Non-Additive Measure and Integral*, vol. 27 of *Theory and Decision Library. Series B: Mathematical and Statistical Methods*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[3] S. Wang, "Insurance pricing and increased limits ratemaking by proportional hazards transforms," *Insurance: Mathematics & Economics*, vol. 17, no. 1, pp. 43–54, 1995.

[4] S. S. Wang, "Premium calculation by transforming the layer premium density," *ASTIN Bulletin*, vol. 26, pp. 71–92, 1996.
[5] S. S. Wang, V. R. Young, and H. H. Panjer, "Axiomatic characterization of insurance prices," *Insurance: Mathematics & Economics*, vol. 21, no. 2, pp. 173–183, 1997.
[6] E. Furman and R. Zitikis, "Weighted premium calculation principles," *Insurance: Mathematics & Economics*, vol. 42, no. 1, pp. 459–465, 2008.
[7] E. Furman and R. Zitikis, "Weighted pricing functionals with applications to insurance: an overview," *North American Actuarial Journal*, vol. 13, no. 4, pp. 483–496, 2009.
[8] E. Furman and Z. Landsman, "On some risk-adjusted tail-based premium calculation principles," *Journal of Actuarial Practice*, vol. 13, pp. 175–191, 2006.
[9] W. Hürlimann, "Conditional value-at-risk bounds for compound Poisson risks and a normal approximation," *Journal of Applied Mathematics*, no. 3, pp. 141–153, 2003.
[10] E. Furman and Z. Landsman, "Tail variance premium with applications for elliptical portfolio of risks," *Astin Bulletin*, vol. 36, no. 2, pp. 433–462, 2006.
[11] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models from Data to Decisions*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, USA, 3rd edition, 2008.
[12] W. Hürlimann, "Analytical evaluation of economic risk capital for portfolios of gamma risks," *Astin Bulletin*, vol. 31, no. 1, pp. 107–122, 2001.
[13] E. Furman and Z. Landsman, "Risk capital decomposition for a multivariate dependent gamma portfolio," *Insurance: Mathematics & Economics*, vol. 37, no. 3, pp. 635–649, 2005.
[14] H. H. Panjer, "Measurement of risk, solvency requirements, and allocation of capital within financial conglomerates," Research Report 01-15, Institute of Insurance and Pension Research, University of Waterloo, 2002.
[15] Z. M. Landsman and E. A. Valdez, "Tail conditional expectations for elliptical distributions," *North American Actuarial Journal*, vol. 7, no. 4, pp. 55–71, 2003.
[16] Z. Landsman and E. A. Valdez, "Tail conditional expectations for exponential dispersion models," *Astin Bulletin*, vol. 35, no. 1, pp. 189–209, 2005.
[17] E. Furman and Z. Landsman, "Multivariate Tweedie distributions and some related capital-at-risk analyses," *Insurance: Mathematics and Economics*, vol. 46, no. 2, pp. 351–361, 2010.
[18] A. Chiragiev and Z. Landsman, "Multivariate Pareto portfolios: TCE-based capital allocation and divided differences," *Scandinavian Actuarial Journal*, vol. 2007, no. 4, pp. 261–280, 2007.
[19] R. Vernic, "Tail conditional expectation for the multivariate Pareto distribution of the second kind: another approach," *Methodology and Computing in Applied Probability*. In press.
[20] A. V. Asimit, E. Furman, and R. Vernic, "On a multivariate Pareto distribution," *Insurance: Mathematics and Economics*, vol. 46, no. 2, pp. 308–316, 2010.
[21] M. C. K. Tweedie, "Functions of a statistical variate with given means, with special reference to Laplacian distributions," *Proceedings of the Cambridge Philosophical Society*, vol. 49, pp. 41–49, 1947.
[22] B. Jørgensen, "Some properties of exponential dispersion models," *Scandinavian Journal of Statistics*, vol. 13, no. 3, pp. 187–197, 1986.
[23] B. Jørgensen, "Exponential dispersion models," *Journal of the Royal Statistical Society. Series B*, vol. 49, no. 2, pp. 127–162, 1987.
[24] B. Jørgensen, *The Theory of Dispersion Models*, vol. 76 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 1997.
[25] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society, Series A*, vol. 135, pp. 370–384, 1972.
[26] J. Rioux and S. Klugman, "Toward a unified approach to fitting loss models," 2004, http://www.iowaactuariesclub.org/library/lossmodels.ppt.
[27] E. Furman, "On a multivariate gamma distribution," *Statistics & Probability Letters*, vol. 78, no. 15, pp. 2353–2360, 2008.
[28] R. S. Chhikara and J. L. Folks, "Estimation of the inverse Gaussian distribution function," *Journal of the American Statistical Association*, vol. 69, pp. 250–254, 1974.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 726389, 29 pages
doi:10.1155/2010/726389

Research Article

Pricing Equity-Indexed Annuities under Stochastic Interest Rates Using Copulas

Patrice Gaillardetz

Department of Mathematics and Statistics, Concordia University, Montreal, QC, Canada H3G 1M8

Correspondence should be addressed to Patrice Gaillardetz, gaillardetz@hotmail.com

Received 1 October 2009; Accepted 18 February 2010

Academic Editor: Johanna Neslehova

Copyright © 2010 Patrice Gaillardetz. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We develop a consistent evaluation approach for equity-linked insurance products under stochastic interest rates. This pricing approach requires that the premium information of standard insurance products is given exogenously. In order to evaluate equity-linked products, we derive three martingale probability measures that reproduce the information from standard insurance products, interest rates, and equity index. These risk adjusted martingale probability measures are determined using copula theory and evolve with the stochastic interest rate process. A detailed numerical analysis is performed for existing equity-indexed annuities in the North American market.

1. Introduction

An equity-indexed annuity is an insurance product whose benefits are linked to the performance of an equity market. It provides limited participation in the performance of an equity index (e.g., S&P 500) while guaranteeing a minimum rate of return. Introduced by Keyport Life Insurance Co. in 1995, equity-indexed annuities have been the most innovative insurance product introduced in recent years. They have become increasingly popular since their debut and sales have broken the $20 billion barrier ($23.1 billion) in 2004, reaching $27.3 billion in 2005. Equity-indexed annuities have also reached a critical mass with a total asset of $93 billion in 2005 (2006 Annuity Fact Book (Tables 7-8) from the National Association for Variable Annuities (NAVA)). See the monograph by Hardy [1] for comprehensive discussions on these products.

The traditional actuarial pricing approach evaluates the premiums of standard life insurance products as the expected present value of its benefits with respect to a mortality law plus a security loading. Since equity-linked products are embedded with various types of financial guarantees, the actuarial approach is difficult to extend to these products and often produces premiums inconsistent with the insurance and financial markets. Many attempts have been made to provide consistent pricing approaches for equity-linked products using financial and economical approaches. For instance, Brennan and Schwartz [2] and Boyle and Schwartz [3] use option pricing techniques to evaluate life insurance products embedded with some financial guarantees. Bacinello and Ortu [4, 5] consider the case where the interest rate is stochastic. More recently, Møller [6] employs the risk-minimization method to evaluate equity-linked life insurances. Young and Zariphopoulou [7] evaluate these products using utility theory[1]. Particularly for equity-indexed annuities, Tiong [8] and Lee [9] obtain closed-form formulas for several annuities under the Black-Scholes-Merton framework. Moore [10] evaluates equity-indexed annuities based on utility theory. Lin and Tan [11] and Kijima and Wong [12] consider more general models for equity-indexed annuities, in which the external equity index and the interest rates are general stochastic differential equations.

The liabilities and premiums of standard insurance products are influenced by the insurer financial performance. Indeed, insurance companies adjust their premiums according to the realized return from their fixed income and other financial instruments as well as market pressure. Therefore, mortality security loadings underlying insurance pricing approach evolve with the financial market. With the current financial crisis, a flexible approach for equity-linked products that allows interdependency between risks should be used. Hence, we generalize the approach of Gaillardetz and Lin [13] to stochastic interest rates. Similarly to Wüthrich et al. [14], they introduce a market consistent valuation method for equity-linked products by combining probability measures using copulas. Indeed, the deterministic interest rate assumption may be adequate for short-term derivative products; however, it is undesirable to extrapolate for longer maturities as for the financial guarantees embedded in equity-linked products. Therefore, we use the approach of Gaillardetz [15] to model standard insurance products under stochastic interest rates. It supposes the conditional independence between the insurance and interest rate risks. Here, this approach is generalized to models that are based on copulas.

Similarly to Gaillardetz and Lin [13], we assume that the premium information of term life insurances, pure endowment insurances, and endowment insurances at all maturities is obtainable. We obtain martingale measures for each standard insurance product under stochastic interest rates. To this end, it is required to assume that the volatilities for standard insurance prices are given exogenously. Gaillardetz [15] provides additional structure to find an implicit volatilities for the standard insurance and annuity products. Then, the martingale probability measures for the insurance and interest rate risks are combined with the martingale measure from the equity index. These extend martingale measures are used to evaluate equity-linked insurance contracts and equity-indexed annuities in particular.

This paper is organized as follows. The next section presents financial models for the interest rates and equity index as well as insurance model. We then derive martingale measures for those standard insurance products under stochastic interest rates in Section 3. In Section 4, we derive the martingale measures for equity-linked products. Section 5 focuses on recursive pricing formulas for equity-linked contracts. Finally, we examine the implications of the proposed approaches on the EIAs by conducting a detailed numerical analysis in Section 6.

2. Underlying Financial Models

In this section, we present a multiperiod discrete model that describes the dynamic of a stock index and the interest rate. These lattice models have been intensively used to model stocks, stock indices, interest rates, and other financial securities due to their flexibility and tractability; see Panjer et al. [16] and Lin [17], for example. Moreover, as it often happens when working in a continuous framework, it becomes necessary to resort to simulation methods in order to obtain a solution to the problems we are considering. Moreover, the premiums obtained from discrete models converge rapidly to the premiums obtained with the corresponding continuous models when considering equity-indexed annuities.

2.1. Interest Rate Model

Similarly to Gaillardetz [15], it is assumed that the short-term ratefollows that of Black et al. [18] (BDT), which means that the short-term rate follows a lattice model that is recombining and Markovian. Particularly, the short-term rate can take exactly $t+1$ distinct values at year t denoted by $r(t,0), r(t,1), \ldots, r(t,t)$. Indeed, $r(t,l)$ represents the short-term rate between time t and $t+1$ that has made "l" up moves. The short-term rate today, $r(0)$, is equal to $r(0,0)$, and in the case where $r(t) = r(t,l)$, the short-term rate at time $t+1$, $r(t+1)$ can only take two values, either $r(t+1,l)$ (decrease) or $r(t+1,l+1)$ (increase). We consider the short-term rate process under the martingale measure Q and hence, the discounted value process $L(t,T)/B(t)$ is a martingale. $L(t,T)$ represents the price at time t of a default-free, zero-coupon bond paying one monetary unit at time T and $B(t)$, the money market account, represents one monetary unit ($B(0) = 1$) accumulated at the short-term rate

$$B(t) = \prod_{i=0}^{t-1}[1 + r(i)]. \tag{2.1}$$

Let $q(t,l)$ be the probability under Q that the short-term rate increases at time $t+1$ given $r(t) = r(t,l)$. That is

$$q(t,l) = Q[r(t+1) = r(t+1,l+1) \mid r(t) = r(t,l)], \tag{2.2}$$

for $0 \le l \le t$, which is set to be 0.5 under the BDT model. Figure 1 describes the dynamic of the short-term rate process.

The BDT model also assumes that short-term rate process matches an array of yields volatilities $(\sigma_r(1), \sigma_r(2), \ldots)$, which is assumed to be observable from the financial market. This vector is deterministic, specified at time 0, and each element is defined by

$$\begin{aligned}\sigma_r(t)^2 &= \operatorname{Var}[\ln r(t) \mid r(t-1) = r(t-1,l)] \\ &= \left[0.5\ln\left(\frac{r(t,l+1)}{r(t,l)}\right)\right]^2,\end{aligned} \tag{2.3}$$

for $l = 0,1,\ldots,t-1$ and $t = 1,\ldots$. Hence, $r(t,l+1)$ is larger than $r(t,l)$ thus, (2.3) may be rewritten as follows:

$$\sigma_r(t) = 0.5\ln\left(\frac{r(t,l+1)}{r(t,l)}\right). \tag{2.4}$$

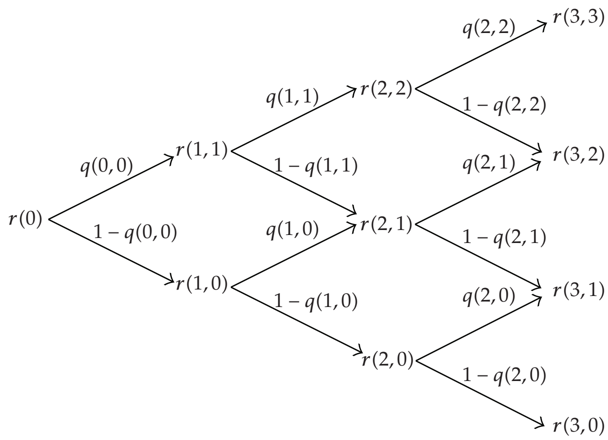


Figure 1: The probability tree of the BDT model process over 3 years.

Equation (2.4) holds for $l \in \{0, 1, \ldots, t-1\}$ and leads to

$$r(t, l) = r(t, 0)^{1-l/t} r(t, t)^{l/t}, \tag{2.5}$$

for $l = 0, 1, \ldots, t$. Equations (2.4) and (2.5) lead to

$$\sigma_r(t) = \frac{1}{2t} \ln\left(\frac{r(t, t)}{r(t, 0)}\right). \tag{2.6}$$

By matching the market prices and the model prices, we have

$$\begin{aligned} L(0, t) &= E\left[\frac{1}{B(t)}\right] \\ &= \frac{2^{-t+1}}{(1 + r(0))} \sum_{l_1=0}^{1} \sum_{l_2=l_1}^{l_1+1} \cdots \sum_{l_{t-1}=l_{t-2}}^{l_{t-2}+1} \prod_{m=1}^{t-1} (1 + r(m, l_m))^{-1}, \end{aligned} \tag{2.7}$$

where $E[\cdot]$ represents the expectation with respect to Q. Replacing $r(t, l)$, $l = 1, 2, \ldots, t-1$, in (2.7) using (2.5) leads to a system of two equations (2.6) and (2.7) with two unknowns $r(t, t)$ and $r(t, 0)$, which can be solved for all t.

2.2. Index Model

Similar to Gaillardetz and Lin [13], we suppose that each year is divided into N trading subperiods of equal length $\Delta = N^{-1}$, which means that the set of trading dates for the index is $\{0, \Delta, 2\Delta, \ldots\}$. We also assume a lattice index model such that the index process $S(k)$, $k = \Delta, 2\Delta, \ldots$, has two possible outcomes $S(k-\Delta)d(k)$ and $S(k-\Delta)u(k)$ given $S(k-\Delta)$ for the time period $[k-\Delta, k]$, where $S(0)$ is the initial level of the index. The index level $S(k-\Delta)d(k)$ at time k represents the index level when the index value goes down and $S(k-\Delta)u(k)$ represents the index level when the index values goes up. Since the short-term rate is a yearly process, we assume that the values $d(k)$ and $u(k)$ are constant for each year. Hence, we may write $d(k) = d(t)$ and $u(k) = u(t)$ for $k = t-1+\Delta, t-1+2\Delta, \dots, t$. Because of the number of trading dates per year, the time-k value of the money-market account is given by

$$B(k) = \prod_{i=\{0,\Delta,\dots,k-\Delta\}} [1 + r(\lfloor i \rfloor)]^{\Delta}, \tag{2.8}$$

for $k = 0, \Delta, 2\Delta, \dots$. Here, $\lfloor \cdot \rfloor$ is the floor function.

A martingale measure for the financial model needs to be determined for the valuation of equity-linked products. This martingale measures Q should be such that $L(t,T)/B(t)$, $t = 0, 1, \dots, T$, and $S(k)/B(k)$, $k = 0, \Delta, \dots, \tau$, remain martingales. Note that the goal of this section is to derive the conditional distribution of the index process. Hence, the constraints imposed by the martingale discounted value process $L(t,T)/B(t)$ are to be discussed later. Let $\pi(k,l)$ be the conditional probability that the index value goes up during the period $[k-\Delta, k]$ given $r(t-1) = r(t-1,l)$, that is,

$$\begin{aligned} Q[S(k) = S(k-\Delta)u(t) \mid S(k-\Delta), r(t-1) = r(t-1,l)] &= \pi(k,l), \\ Q[S(k) = S(k-\Delta)d(t)S(k-\Delta), r(t-1) = r(t-1,l)] &= 1 - \pi(k,l), \end{aligned} \tag{2.9}$$

for $k = t-1+\Delta,\ t-1+2\Delta, \dots, t$ and $t = 1, 2, \dots$. Supposing that the discounted value process $S(k)/B(k)$ is a martingale implies

$$\pi(k,l) = \frac{(1 + r(t-1,l))^{\Delta} - d(t)}{u(t) - d(t)}, \tag{2.10}$$

for $k = t-1+\Delta,\ t-1+2\Delta, \dots, t$ and $t = 1, 2 \dots$. From (2.10) it is obvious that $\pi(k,l)$ is constant over each year, that is, $\pi(k,l) = \pi(t,l)$ for $k = t-1+\Delta,\ t-1+2\Delta, \dots, t$. The no-arbitrage thus requires

$$d(t) < (1 + r(t-1,0))^{\Delta}, \qquad (1 + r(t-1,t-1))^{\Delta} < u(t), \tag{2.11}$$

for $t = 1, 2, \dots$. The previous conditions may not be respected for the BDT model when long maturity or high volatility are considered. In this case, the bounded trinomial model from Hull and White [19] would be more suitable.

Under this model, the ratio $(S(t))/(S(t-1))$ takes $N+1$ possible values denoted $\gamma(t,i)$, $i = 0, 1, \dots, N$, which are defined by

$$\gamma(t,i) = u(t)^{i} d(t)^{N-i}, \quad i = 0, 1, \dots, N. \tag{2.12}$$

Their corresponding conditional martingale probabilities are

$$Q\left[\frac{S(t)}{S(t-1)} = \gamma(t,i) \mid r(t-1) = r(t-1,l)\right] = \binom{N}{i} \pi(t,l)^{i} (1 - \pi(t,l))^{N-i}, \tag{2.13}$$

for $i = 0, \dots, N$.

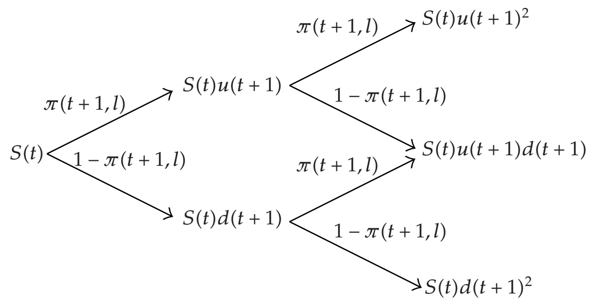


Figure 2: The probability tree of the index under stochastic interest rates over t and $t+1$ when $N = 2$ given $r(t) = r(t,l)$.

The model assumes the usual frictionless market: no tax, no transaction costs, and so forth. Furthermore, for practical implementation purposes, one may also use current forward rates for r.

Figure 2 presents the conditional index process tree under stochastic interest rates when $N = 2$ for the time period $[t, t+1]$.

For notational convenience, let

$$\mathbf{i}_t = \{i_0, \ldots, i_t\}, \tag{2.14}$$

which represents the index's realization up to time t with

$$S(t, \mathbf{i}_t) = S(0) \prod_{l=0}^{t} \gamma(l, i_l), \tag{2.15}$$

for $t = 0, 1, \ldots$, where $\gamma(0, i_0) = 1$.

2.3. Insurance Models

In this subsection, we introduce lattice models for the standard insurance products under stochastic interest rates. We will use the standard actuarial notation which can be found in Bowers et al. [20]. Let $T(x)$ be the future lifetime of insured (x) of age x and the curtate-future-lifetime

$$K(x) = \lfloor T(x) \rfloor \tag{2.16}$$

the number of future complete years lived by the insured (x) prior to death. For notational purposes, let

$$\mathbf{l}_t = \{l_0, l_1, \ldots, l_t\} \tag{2.17}$$

represent the realization of the short-term rate process up to time t with $r(i) = r(i, l_i)$, $i = 0, 1, \ldots, t$, where $l_0 = 0$.

For integers t and n $(t \leq n)$, let $V^{(j)}(x,t,n,\mathbf{l}_t\}$ denote, respectively, the time-t prices for the n-year term life insurance $(j = 1)$, n-year pure endowment insurance $(j = 2)$, and n-year endowment insurance $(j = 3)$ given that the short-term rate followed the path $\mathbf{l}_t$.

The value process $W^{(1)}(x,t,n,\mathbf{l}_{t-1})$ of n-year term life insurance is defined by

$$W^{(1)}(x,t,n,\mathbf{l}_{t-1}) = \begin{cases} \dfrac{B(t)}{B(K(x)+1)}, & K(x) < t, \\ V^{(1)}(x,t,n,\{\mathbf{l}_{t-1}, l_{t-1}+1\}), & K(x) \geq t,\ r(t) = r(t, l_{t-1}+1), \\ V^{(1)}(x,t,n,\{\mathbf{l}_{t-1}, l_{t-1}\}), & K(x) \geq t,\ r(t) = r(t, l_{t-1}), \end{cases} \tag{2.18}$$

with $V^{(1)}(x,n,n,\mathbf{l}_n) = 0$. Note that $\mathbf{l}_t$ represents the interest rate information known by the process, but does not stand as an indexing parameter.

Similarly, define $W^{(2)}(x,t,n,\mathbf{l}_{t-1})$ to be the value process of the n-year pure endowment insurance and it is given by

$$W^{(2)}(x,t,n,\mathbf{l}_{t-1}) = \begin{cases} 0, & K(x) < t, \\ V^{(2)}(x,t,n,\{\mathbf{l}_{t-1}, l_{t-1}+1\}), & K(x) \geq t,\ r(t) = r(t, l_{t-1}+1), \\ V^{(2)}(x,t,n,\{\mathbf{l}_{t-1}, l_{t-1}\}), & K(x) \geq t,\ r(t) = r(t, l_{t-1}), \end{cases} \tag{2.19}$$

with $V^{(2)}(x,n,n,\mathbf{l}_n) = 1$.

Finally, let $W^{(3)}(x,t,n,\mathbf{l}_{t-1})$ denote the value process generated by the n-year endowment insurance

$$W^{(3)}(x,t,n,\mathbf{l}_{t-1}) = \begin{cases} \dfrac{B(t)}{B(K(x)+1)}, & K(x) < t, \\ V^{(3)}(x,t,n,\{\mathbf{l}_{t-1}, l_{t-1}+1\}), & K(x) \geq t,\ r(t) = r(t, l_{t-1}+1), \\ V^{(3)}(x,t,n,\{\mathbf{l}_{t-1}, l_{t-1}\}), & K(x) \geq t,\ r(t) = r(t, l_{t-1}), \end{cases} \tag{2.20}$$

with $V^{(3)}(x,n,n,\mathbf{l}_n) = 1$

The processes $W^{(j)}(x,t,n)$, $j = 1,2,3$, represent the intrinsic values of the standard insurance products and are presented in Figure 3.

3. Martingale Measures for Insurance Models

In this section, we employ a method similar to the approach of Gaillardetz [15] to derive a martingale probability measure for each of the value processes introduced in the last section. Gaillardetz [15] derives these martingale measures under conditional independence assumptions. Here, we relax this assumption by using copulas to describe

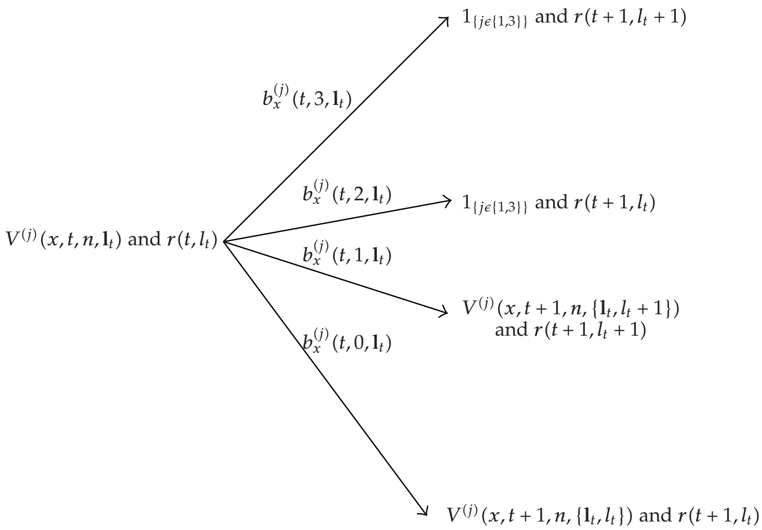


Figure 3: The probability tree of the combined insurance product ($j = 1,2,3$) and short-term rate processes between t and $t+1$ given $K(x) \ge t$ and $r(0), r(1), \ldots, r(t)$.

possible dependence structures between interest rates and insurance products. It is important to point out that these probabilities are age-dependent and include an adjustment for the mortality risk since we use the information from the insurance market.

The martingale measures $Q_x^{(j)}$, $j = 1,2,3$, are defined such that $W^{(j)}(x,t,n)/B(t)$ and $L(t,T)/B(t)$ are martingales. As mentioned in Section 2, we assume that the time-0 premiums $V^{(j)}(x,0,n)$, $j = 1,2,3$, of the term life insurance, pure endowment insurance, and endowment insurance are given exogenously. The annual short-term rate process $r(t)$ is governed by the BDT model with $q(t,l) = 0.5$ and volatilities $\sigma_r(t)$ are given exogenously for $l = 0,1,\ldots,t$ and $t = 1,2,\ldots$. The conditional martingale probability of each possible outcome is defined by

$$\begin{aligned}
b_x^{(j)}(t,0,\mathbf{l}_t) &= Q_x^{(j)}[K(x) > t,\ r(t+1) = r(t+1,l_t) \mid K(x) \ge t, \mathbf{l}_t], \\
b_x^{(j)}(t,1,\mathbf{l}_t) &= Q_x^{(j)}[K(x) > t,\ r(t+1) = r(t+1,l_t+1) \mid K(x) \ge t, \mathbf{l}_t], \\
b_x^{(j)}(t,2,\mathbf{l}_t) &= Q_x^{(j)}[K(x) = t,\ r(t+1) = r(t+1,l_t) \mid K(x) \ge t, \mathbf{l}_t], \\
b_x^{(j)}(t,3,\mathbf{l}_t) &= Q_x^{(j)}[K(x) = t,\ r(t+1) = r(t+1,l_t+1) \mid K(x) \ge t, \mathbf{l}_t],
\end{aligned} \tag{3.1}$$

for $j = 1,2,3$. These martingale probabilities are presented above each branch in Figure 3.

The main objective of this section is to determine $b_x^{(j)}$'s that will be used to evaluate equity-indexed annuities in later sections.

To ensure that the discounted value process $L(t,T)/B(t)$ is a martingale, we must have

$$b_x^{(j)}(t,0,\mathbf{l}_t) + b_x^{(j)}(t,2,\mathbf{l}_t) = \frac{1}{2}, \qquad b_x^{(j)}(t,1,\mathbf{l}_t) + b_x^{(j)}(t,3,\mathbf{l}_t) = \frac{1}{2}, \tag{3.2}$$

for $j = 1,2,3$, $t = 0,1,\ldots$, and all $\mathbf{l}_t$. Note that the martingale mortality and survival probabilities are given, respectively, by

$$\begin{aligned} q_x^{(j)}(t,\mathbf{l}_t) &= Q_x^{(j)}[K(x) = t \mid K(x) \geq t, \mathbf{l}_t] = b_x^{(j)}(t,2,\mathbf{l}_t) + b_x^{(j)}(t,3,\mathbf{l}_t), \\ p_x^{(j)}(t,\mathbf{l}_t) &= Q_x^{(j)}[K(x) > t \mid K(x) \geq t, \mathbf{l}_t] = b_x^{(j)}(t,0,\mathbf{l}_t) + b_x^{(j)}(t,1,\mathbf{l}_t). \end{aligned} \tag{3.3}$$

As in the short-term rate model, additional structure is needed to set the time-t premiums. Similar to Black et al. [18], we suppose that the volatilities of insurance liabilities $\sigma_x^{(j)}(t,l_{t-1})$ are defined at time t by

$$\sigma_x^{(1)}(t,\mathbf{l}_{t-1})^2 = \mathrm{Var}_x^{(1)}\left[\ln\left(W^{(1)}(x,t,n,\mathbf{l}_{t-1})\right) \mid K(x) \geq t, \mathbf{l}_{t-1}\right], \tag{3.4}$$

$$\sigma_x^{(j)}(t,\mathbf{l}_{t-1})^2 = \mathrm{Var}_x^{(j)}\left[\ln\left(1 - W^{(j)}(x,t,n,\mathbf{l}_{t-1})\right) \mid K(x) \geq t, \mathbf{l}_{t-1}\right], \tag{3.5}$$

for $j = 2,3$, $t = 1,2,\ldots$, and all $\mathbf{l}_{t-1}$. Here, $\mathrm{Var}_x^{(j)}[\cdot]$ represents the conditional variance with respect to $Q_x^{(j)}$. We assume that the volatilities are deterministic but vary over time and are given exogenously. Gaillardetz [15] uses the natural logarithm function to ensure that each process remains strictly positive. Since $W^{(1)}$ is close to 0, it directly uses $\ln(W^{(1)})$ to ensure that the process remains strictly greater than 0. On the other hand, it uses $\ln(1 - W^{(j)})$ for $j = 2,3$ to ensure that the processes are strictly smaller than 1 since $W^{(j)}$'s are closer to 1.

In order to identify the martingale probabilities $b_x^{(j)}$, Gaillardetz [15] assumes the independence or the conditional independence between the interest rate process and the insurer's life. Here, the additional structure is provided by the choice of copulas. Indeed, the dependence structure between the interest rates and the premiums of insurance products is modeled using a copula. The main advantage of using copulas is that they separate a joint distribution function in two parts: the dependence structure and the marginal distribution functions. We use them because of their mathematical tractability and, based on the Sklar's Theorem, they can express all multivariate distributions. A comprehensive introduction may be found in Joe [21] or Nelsen [22]. Frees and Valdez [23], Wang [24], and Venter [25] have given an overview of copulas and their applications to actuarial science. Cherubini et al. [26] present the applications of copulas in finance.

There exists a wide range of copulas that may define a joint cumulative distribution function. The simplest one is the independent copula

$$C^I(F_{Y_1}(y_1), F_{Y_2}(y_2)) = F_{Y_1}(y_1)F_{Y_2}(y_2), \tag{3.6}$$

where F_{Y_1} and F_{Y_2} are marginal cumulative distribution functions. Extreme copulas are defined using the upper and lower Frechet-Hoeffding bounds, which are given by

$$C^U(F_{Y_1}(y_1), F_{Y_2}(y_2)) = \min\left[F_{Y_1}(y_1), F_{Y_2}(y_2)\right], \tag{3.7}$$

$$C^D(F_{Y_1}(y_1), F_{Y_2}(y_2)) = \max\left[F_{Y_1}(y_1) + F_{Y_2}(y_2) - 1, 0\right]. \tag{3.8}$$

One of the most important families of copulas is the archimedean copulas. Among them, the Cook-Johnson (Clayton) copula is widely used in actuarial science because of its desirable properties and simplicity. The Cook-Johnson copula with parameter $\kappa > 0$ is given by

$$C_\kappa^{CJ}(F_{Y_1}(y_1), F_{Y_2}(y_2)) = \left[F_{Y_1}(y_1)^{-\kappa} + F_{Y_2}(y_2)^{-\kappa} - 1\right]^{-1/\kappa}. \tag{3.9}$$

The Gaussian $(-1 \le \kappa \le 1)$ copula, which is often used in finance, is defined as

$$C_\kappa^G(F_{Y_1}(y_1), F_{Y_2}(y_2)) = \Phi_\kappa\left(\Phi^{-1}(F_{Y_1}(y_1)), \Phi^{-1}(F_{Y_2}(y_2))\right), \tag{3.10}$$

where Φ_κ is the bivariate standard normal cumulative distribution function with correlation coefficient κ and Φ^{-1} is the inverse of the standard normal cumulative distribution function. Hence, the parameter κ in formulas (3.9) and (3.10) indicates the level of dependence between the insurance products and interest rates.

The joint cumulative distribution of $W^{(j)}$ and $r(t)$ is obtained using a copula $C_{\kappa(t)}$, that is,

$$\begin{aligned} &Q_x^{(j)}\left[W^{(j)}(x, t+1, n, \mathbf{l}_t) \le y_1, r(t+1) \le y_2 \mid K(x) \ge t, \mathbf{l}_t\right] \\ &\quad = C_{\kappa(t)}\left(Q_x^{(j)}\left[W^{(j)}(x, t+1, n, \mathbf{l}_t) \le y_1 \mid K(x) \ge t, \mathbf{l}_t\right], Q\left[r(t+1) \le y_2 \mid \mathbf{l}_t\right]\right), \end{aligned} \tag{3.11}$$

for $j = 1, 2, 3$, where the copula may be defined by either (3.6), (3.7), (3.8), (3.9), or (3.10).

The martingale probabilities have the following constraints:

$$\begin{aligned} &b_x^{(j)}(t, 2, \mathbf{l}_t) \\ &\quad = Q_x^{(j)}\left[W^{(j)}(x, t+1, n, \mathbf{l}_t) = 1, r(t+1) = r(t+1, l_t) \mid K(x) \ge t, \mathbf{l}_t\right] \\ &\quad = Q_x^{(j)}\left[W^{(j)}(x, t+1, n, \mathbf{l}_t) \le 1, r(t+1) \le r(t+1, l_t) \mid K(x) \ge t, \mathbf{l}_t\right] \\ &\qquad - Q_x^{(j)}\left[W^{(j)}(x, t+1, n, \mathbf{l}_t) \le V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t\}), r(t+1) \le r(t+1, l_t) \mid K(x) \ge t, \mathbf{l}_t\right]. \end{aligned} \tag{3.12}$$

It follows from (3.11) that

$$\begin{aligned}
&b_x^{(j)}(t,2,\mathbf{l}_t) \\
&\quad = Q_x^{(j)}[r(t+1) \le r(t+1,l_t) \mid K(x) \ge t,\mathbf{l}_t] \\
&\qquad - C_{\kappa(t)}\Big(Q_x^{(j)}\Big[W^{(j)}(x,t+1,n,\mathbf{l}_t) \le V^{(j)}(x,t+1,n,\{\mathbf{l}_t,l_t\}) \mid K(x) \ge t,\mathbf{l}_t\Big], \\
&\qquad\qquad Q[r(t+1) \le r(t+1,l_t) \mid \mathbf{l}_t]\Big).
\end{aligned} \tag{3.13}$$

Using the following inequality

$$V^{(j)}(x,t+1,n,\{\mathbf{l}_t,l_t\}) \ge V^{(j)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}) \tag{3.14}$$

in (3.13) leads to

$$\begin{aligned}
b_x^{(j)}(t,2,\mathbf{l}_t) &= 0.5 - C_{\kappa(t)}\Big(b_x^{(j)}(t,0,\mathbf{l}_t) + b_x^{(j)}(t,1,\mathbf{l}_t), 0.5\Big) \\
&= 0.5 - C_{\kappa(t)}\Big(p_x^{(j)}(t,\mathbf{l}_t), 0.5\Big),
\end{aligned} \tag{3.15}$$

for $j = 1, 3$, and we have for $j = 2$

$$\begin{aligned}
b_x^{(2)}(t,2,\mathbf{l}_t) &= Q_x^{(2)}\Big[W^{(2)}(x,t+1,n,\mathbf{l}_t) = 0,\ r(t+1) = r(t+1,l_t) \mid K(x) \ge t,\mathbf{l}_t\Big] \\
&= Q_x^{(2)}\Big[W^{(2)}(x,t+1,n,\mathbf{l}_t) \le 0,\ r(t+1) \le r(t+1,l_t) \mid K(x) \ge t,\mathbf{l}_t\Big].
\end{aligned} \tag{3.16}$$

It follows from (3.11) that

$$\begin{aligned}
b_x^{(2)}(t,2,\mathbf{l}_t) &= C_{\kappa(t)}\Big(Q_x^{(2)}\Big[W^{(2)}(x,t+1,n,\mathbf{l}_t) \le 0 \mid K(x) \ge t,\mathbf{l}_t\Big], Q[r(t+1) \le r(t+1,l_t) \mid \mathbf{l}_t]\Big) \\
&= C_{\kappa(t)}\Big(q_x^{(2)}(t,\mathbf{l}_t), 0.5\Big).
\end{aligned} \tag{3.17}$$

3.1. Term Life Insurance

Proposition 3.1. *For given $V^{(1)}(x,0,n,l_0)$ $(n = 1,2,\ldots,\tau)$, copulas $C_{\kappa(t)}$, and volatilities $\sigma_x^{(1)}(t,\mathbf{l}_{t-1})$ $(t = 1,2,\ldots,\tau$ and all $\mathbf{l}_{t-1})$, the age-dependent, mortality risk-adjusted martingale probabilities are given by*

$$b_x^{(1)}(t,2,\mathbf{l}_t) = 0.5 - C_{\kappa(t)}\Big(1 - V^{(1)}(x,t,t+1,\mathbf{l}_t)(1+r(t,l_t)),0.5\Big), \tag{3.18}$$

$$b_x^{(1)}(t,3,\mathbf{l}_t) = V^{(1)}(x,t,t+1,\mathbf{l}_t)(1+r(t,l_t)) - b_x^{(1)}(t,2,\mathbf{l}_t), \tag{3.19}$$

$$b_x^{(1)}(t,i,\mathbf{l}_t) = \frac{1}{2} - b_x^{(1)}(t,i+2,\mathbf{l}_t), \tag{3.20}$$

for $i = 0,1$, where the price at time t is defined recursively using

$$\begin{aligned} &V^{(1)}(x,t+1,n,\{\mathbf{l}_t,l_t\}) \\ &\quad= \frac{V^{(1)}(x,t,n,\mathbf{l}_t)(1+r(t,l_t)) - b_x^{(1)}(t,2,\mathbf{l}_t) - b_x^{(1)}(t,3,\mathbf{l}_t)}{b_x^{(1)}(t,0,\mathbf{l}_t) + b_x^{(1)}(t,1,\mathbf{l}_t)e^{-((b_x^{(1)}(t,0,\mathbf{l}_t)+b_x^{(1)}(t,1,\mathbf{l}_t))/\{b_x^{(1)}(t,0,\mathbf{l}_t)b_x^{(1)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(1)}(t+1,\mathbf{l}_t)}}, \end{aligned} \tag{3.21}$$

$$\begin{aligned} &V^{(1)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}) \\ &\quad= \frac{V^{(1)}(x,t,n,\mathbf{l}_t)(1+r(t,l_t)) - b_x^{(1)}(t,2,\mathbf{l}_t) - b_x^{(1)}(t,3,\mathbf{l}_t)}{b_x^{(1)}(t,1,\mathbf{l}_t) + b_x^{(1)}(t,0,\mathbf{l}_t)e^{((b_x^{(1)}(t,0,\mathbf{l}_t)+b_x^{(1)}(t,1,\mathbf{l}_t))/\{b_x^{(1)}(t,0,\mathbf{l}_t)b_x^{(1)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(1)}(t+1,\mathbf{l}_t)}}, \end{aligned} \tag{3.22}$$

for $t = 0,2,\ldots,n-2$ and all $\mathbf{l}_t$.

Proof. The proof is similar to the proof of Proposition 3.1 of Gaillardetz [15] and can be found in Gaillardetz [27]. □

With the martingale structure identified, the n-year term life insurance premiums may be reproduced as the expected discounted payoff of the insurance

$$V^{(1)}(x,0,n) = E_x^{(1)}\left[\frac{1_{\{K(x)<n\}}}{B(K(x)+1)}\right]. \tag{3.23}$$

3.2. Pure Endowment Insurance

Proposition 3.2. *For given $V^{(2)}(x,0,n,l_0)$ $(n = 1,2,\ldots,\tau)$, copulas $C_{\kappa(t)}$, and volatilities $\sigma_x^{(2)}(t,\mathbf{l}_{t-1})$ $(t = 1,2,\ldots,\tau$ and all $\mathbf{l}_{t-1})$, the age-dependent, mortality risk-adjusted martingale probabilities are given by*

$$b_x^{(2)}(t,2,\mathbf{l}_t) = C_{\kappa(t)}\Big(1 - V^{(2)}(x,t,t+1,\mathbf{l}_t)(1+r(t,l_t)),0.5\Big), \tag{3.24}$$

$$b_x^{(2)}(t,3,\mathbf{l}_t) = 1 - V^{(2)}(x,t,t+1,\mathbf{l}_t)(1+r(t,l_t)) - b_x^{(2)}(t,2,\mathbf{l}_t), \tag{3.25}$$

$$b_x^{(2)}(t,i,\mathbf{l}_t) = \frac{1}{2} - b_x^{(2)}(t,i+2,\mathbf{l}_t), \tag{3.26}$$

for $i = 0, 1$, *where the price at time* t *is defined recursively using*

$$\begin{aligned}
&V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t\}) \\
&= \frac{V^{(2)}(x,t,n,\mathbf{l}_t)(1+r(t,l_t)) - b_x^{(2)}(t,1,\mathbf{l}_t)\left(1 - e^{((b_x^{(2)}(t,0,\mathbf{l}_t)+b_x^{(2)}(t,1,\mathbf{l}_t))/\{b_x^{(2)}(t,0,\mathbf{l}_t)b_x^{(2)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(2)}(t+1,\mathbf{l}_t)}\right)}{b_x^{(2)}(t,0,\mathbf{l}_t)+b_x^{(2)}(t,1,\mathbf{l}_t)e^{((b_x^{(2)}(t,0,\mathbf{l}_t)+b_x^{(2)}(t,1,\mathbf{l}_t))/\{b_x^{(2)}(t,0,\mathbf{l}_t)b_x^{(2)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(2)}(t+1,\mathbf{l}_t)}}, \\
&V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}) \\
&= \frac{V^{(2)}(x,t,n,\mathbf{l}_t)(1+r(t,l_t)) - b_x^{(2)}(t,0,\mathbf{l}_t)\left(1 - e^{-((b_x^{(2)}(t,0,\mathbf{l}_t)+b_x^{(2)}(t,1,\mathbf{l}_t))/\{b_x^{(2)}(t,0,\mathbf{l}_t)b_x^{(2)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(2)}(t+1,\mathbf{l}_t)}\right)}{b_x^{(2)}(t,1,\mathbf{l}_t)+b_x^{(2)}(t,0,\mathbf{l}_t)e^{-((b_x^{(2)}(t,0,\mathbf{l}_t)+b_x^{(2)}(t,1,\mathbf{l}_t))/\{b_x^{(2)}(t,0,\mathbf{l}_t)b_x^{(2)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(2)}(t+1,\mathbf{l}_t)}},
\end{aligned} \tag{3.27}$$

for $t = 0, 2, \ldots, n-2$ *and all* $\mathbf{l}_t$.

Proof. The proof is similar to the proof of Proposition 3.2 of Gaillardetz [15] and can be found in Gaillardetz [27]. □

With the martingale structure identified, the n-year pure endowment insurance premiums may be reproduced as the expected discounted payoff of the insurance

$$V^{(2)}(x,0,n) = E_x^{(2)}\left[\frac{1_{\{K(x)\geq n\}}}{B(n)}\right]. \tag{3.28}$$

3.3. *Endowment Insurance*

There is no general solution for the endowment insurance products since the n-year endowment insurance price at time $n-2$ may not be expressed using only either mortality or survival probabilities. For the n-year term-life insurance, the time-$(n-1)$ price is determined based on the death martingale probabilities and the n-year pure endowment price may be obtained using the survival probabilities at time $n-1$. Therefore, once you combine both products to form an endowment insurance, there is no way to solve explicitly for the martingale probabilities. However, closed-from solutions may be derived for the independent, upper and lower copulas. Numerical methods need to be used for the Cook-Johnson and Gaussian copulas. Furthermore, the width of the participation rate bands for the unified approach is narrow under deterministic interest (see Gaillardetz and Lin [13]). For these reasons, we are focusing on the independent and Frechet-Hoeffding bounds.

Proposition 3.3. *For given $V^{(3)}(x,0,n,l_0)$ $(n=1,2,\ldots,\tau)$, copulas C, and volatilities $\sigma_x^{(3)}(t,\mathbf{l}_{t-1})$ $(t=1,2,\ldots,\tau$ and all $\mathbf{l}_{t-1})$, the age-dependent, mortality risk-adjusted martingale probabilities are given by*

$$b_x^{(3)}(t,2,\mathbf{l}_t)=\begin{cases}\left[V^{(3)}(x,t,t+2,\mathbf{l}_t)(1+r(t,l_t))\right.\\ \quad\left.-0.5\left(\dfrac{1}{1+r(t+1,l_t)}+\dfrac{1}{1+r(t+1,l_t+1)}\right)\right]\\ \div\left(2-\left(\dfrac{1}{1+r(t+1,l_t)}+\dfrac{1}{1+r(t+1,l_t+1)}\right)\right), & \text{Independent},\\ 0, & \text{Upper},\\ \left[V^{(3)}(x,t,t+2,\mathbf{l}_t)(1+r(t,l_t))\right.\\ \quad\left.-0.5\left(\dfrac{1}{1+r(t+1,l_t)}+\dfrac{1}{1+r(t+1,l_t+1)}\right)\right]\\ \div\left(1-\dfrac{1}{1+r(t+1,l_t)}\right), & \text{Lower},\end{cases}\tag{3.29}$$

$$b_x^{(3)}(t,3,\mathbf{l}_t)=\begin{cases}b_x^{(3)}(t,2,\mathbf{l}_t), & \text{Independent},\\ \left[V^{(3)}(x,t,t+2,\mathbf{l}_t)(1+r(t,l_t))\right.\\ \quad\left.-0.5\left(\dfrac{1}{1+r(t+1,l_t)}+\dfrac{1}{1+r(t+1,l_t+1)}\right)\right]\\ \div\left(1-\dfrac{1}{1+r(t+1,l_t+1)}\right), & \text{Upper},\\ 0, & \text{Lower},\end{cases}\tag{3.30}$$

$$b_x^{(3)}(t,i,\mathbf{l}_t)=0.5-b_x^{(3)}(t,i+2,\mathbf{l}_t),\tag{3.31}$$

for $i=0,1$, where the price at time t is defined recursively using

$$\begin{aligned}V^{(3)}&(x,t+1,n,\{\mathbf{l}_t,l_t\})\\ &=\Big[V^{(3)}(x,t,n,\mathbf{l}_t)(1+r(t,l_t))-b_x^{(3)}(t,2,\mathbf{l}_t)-b_x^{(3)}(t,3,\mathbf{l}_t)\\ &\quad-b_x^{(3)}(t,1,\mathbf{l}_t)\left(1-e^{((b_x^{(3)}(t,0,\mathbf{l}_t)+b_x^{(3)}(t,1,\mathbf{l}_t))/\{b_x^{(3)}(t,0,\mathbf{l}_t)b_x^{(3)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(3)}(t+1,\mathbf{l}_t)}\right)\Big]\end{aligned}$$

$$\div\left(b_x^{(3)}(t,0,\mathbf{l}_t)+b_x^{(3)}(t,1,\mathbf{l}_t)e^{((b_x^{(3)}(t,0,\mathbf{l}_t)+b_x^{(3)}(t,1,\mathbf{l}_t))/\{b_x^{(3)}(t,0,\mathbf{l}_t)b_x^{(3)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(3)}(t+1,\mathbf{l}_t)}\right),$$

$$\begin{aligned}
&V^{(3)}(x,t+1,n,\{\mathbf{l}_t,l_t\})\\
&\quad=\Big[V^{(3)}(x,t,n,\mathbf{l}_t)(1+r(t,l_t))-b_x^{(3)}(t,2,\mathbf{l}_t)-b_x^{(3)}(t,3,\mathbf{l}_t)\\
&\qquad-b_x^{(3)}(t,0,\mathbf{l}_t)\left(1-e^{-((b_x^{(3)}(t,0,\mathbf{l}_t)+b_x^{(3)}(t,1,\mathbf{l}_t))/\{b_x^{(3)}(t,0,\mathbf{l}_t)b_x^{(3)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(3)}(t+1,\mathbf{l}_t)}\right)\Big]\\
&\quad\div\left(b_x^{(3)}(t,1,\mathbf{l}_t)+b_x^{(3)}(t,0,\mathbf{l}_t)e^{-((b_x^{(3)}(t,0,\mathbf{l}_t)+b_x^{(3)}(t,1,\mathbf{l}_t))/\{b_x^{(3)}(t,0,\mathbf{l}_t)b_x^{(3)}(t,1,\mathbf{l}_t)\}^{0.5})\sigma_x^{(3)}(t+1,\mathbf{l}_t)}\right),
\end{aligned}\tag{3.32}$$

for $t=0,2,\ldots,n-3$ *and all* $\mathbf{l}_t$.

Proof. The proof can be found in Gaillardetz [27]. □

Since we suppose that the time-0 insurance prices, the insurance volatilities, the zero-coupon bond prices, and the interest rate volatility are given exogenously, it is possible to extract the stochastic structure of each insurance products using Propositions 3.1, 3.2, and 3.3. There are constraints on the parameters because the martingale probabilities should be strictly positive. However, there is no closed-form solution for the stochastic interest models.

Theoretically, there exists a natural hedging between the insurance and annuity products. However, Gaillardetz and Lin [13] argue that it is reasonable to evaluate insurances and annuities separatelysince in practice due to certain regulatory and accounting constraints and issues such as moral hazard and anti-selection.

3.4. Determination of Insurance Volatility Structure

For implementation purposes, we now relax the assumption of exogenous insurance volatilities. In Subsections 3.1, 3.2, and 3.3, the volatilities of insurance liabilities $\sigma_x^{(j)}(t,l_{t-1})$ defined by either (3.4) or (3.5) were supposed to be known. However, identifying these volatilities is extremely challenging due to the lack of empirical data and studies. Similar to Gaillardetz [15], we extract an implied volatility from the insurance market under certain assumptions.

There are three different sources that define the insurance volatilities: the interest rates, the insurance prices, and the martingale probabilities. The implied insurance volatilities is obtained assuming that the short-term rate has no impact on the martingale probabilities. Thus, we extract the insurance volatility such that the martingale probabilities in the case of an up move from the interest rate process are equal to the martingale probabilities in the case of a down move. Let $\sigma_x^{(j)'}(t,\mathbf{l}_{t-1})$ ($j=1,2,3$, $t=1,2,\ldots$, and all $\mathbf{l}_{t-1}$) denote the implied volatilities defined by (3.4) for $j=1$ and (3.5) for $j=2,3$, under the following constraint:

$$q_x^{(j)}(t+1,\{\mathbf{l}_t,l_t\})=q_x^{(j)}(t+1,\{\mathbf{l}_t,l_t+1\}),\tag{3.33}$$

for $j = 1, 2, 3$. In other words, insurance companies that do not react to the interest rate change should have an insurance volatility close to $\sigma_x^{(j)'}$. Gaillardetz [13, 27] explain that behavior of insurance companies facing the interest rate shifts could be understood through these volatilities. They also describe recursive formulas to obtain numerically the implied volatilities. In the following examples, equity-indexed annuity contracts are evaluated using the implied volatilities, which are obtained from (3.33).

4. Martingale Measures for Equity-Linked Products

Due to their unique designs, equity-linked products involve mortality and financial risks since these type of contracts provide both death and accumulation/survival benefits. Moreover, the level of these benefits are linked to the financial market performance and an equity index in particular. Hence, it is natural to assume that equity-linked products belong to a combined insurance and financial markets since they are simultaneously subject to the interest rate, equity, and mortality risks. Similar to Section 3, we evaluate these types of products by evaluating the death benefits and survival benefits separately. Under this approach, two martingale measures again need to be generated: one for death benefits and another for survival benefits. Furthermore, these martingale measures should be such that they reproduce the index values in Section 2 and the premiums of insurance products under stochastic interest rates in Section 3. In other words, the marginal probabilities derived in the previous sections should be preserved, and the martingale measures $Q_x^{(j)+}$, $j = 1, 2, 3$ are such that $\{W^{(j)}(x, t, n)/B(t),\ t = 0, 1, \ldots\}$, $\{L(t, T)/B(t),\ t = 0, 1, \ldots, T$ and $T = 1, \ldots\}$, and $\{S(k)/B(k),\ k = 0, \Delta, \ldots\}$ will remain martingales. Let $e_x^{(j)}(t, i, \mathbf{i}_t, \mathbf{l}_t)$ denote the martingale probability under $Q_x^{(j)+}$ such that (x) survives and

$$\begin{cases} r(t+1) = r(t+1, l_t), & \dfrac{S(t+1)}{S(t)} = \gamma(t+1, i), \quad i = 0, \ldots, N, \\ r(t+1) = r(t+1, l_t+1), & \dfrac{S(t+1)}{S(t)} = \gamma(t+1, i-(N+1)), \quad i = N+1, \ldots, 2N+1 \end{cases} \tag{4.1}$$

or the martingale probability such that (x) dies and

$$\begin{cases} r(t+1) = r(t+1, l_t), & \dfrac{S(t+1)}{S(t)} = \gamma(t, i-(2N+2)), \quad i = 2N+2, \ldots, 3N+2, \\ r(t+1) = r(t+1, l_t+1), & \dfrac{S(t+1)}{S(t)} = \gamma(t, i-(3N+3)), \quad i = 3N+3, \ldots, 4N+3, \end{cases} \tag{4.2}$$

between t and $t+1$ given $S(t)$, $K(x) \geq t$, and $\mathbf{l}_t$ as illustrated in Figure 4. The function γ is given explicitly by (2.12).

What remains is to determine the probabilities $e_x^{(j)}$'s for all $\mathbf{i}_t$ and $\mathbf{l}_t$. We introduce the dependency between the index process, the short-term rate, and the premiums of insurance products using copulas. Let $G_x^{(j)}$, $j = 1,2,3$, denote this joint conditional cumulative distribution function over time t and $t+1$. That is

$$G_x^{(j)}(y_1, y_2, y_3; \mathbf{i}_t, \mathbf{l}_t) = Q_x^{(j)+}\Big[S(t+1) \le y_1, W^{(j)}(x, t+1, n, \mathbf{l}_t) \le y_2, \\ r(t+1) \le y_3 \mid K(x) \ge t, \mathbf{i}_t, \mathbf{l}_t\Big]. \tag{4.3}$$

As explained, the marginal cumulative distribution functions of the insurance products and the index are preserved under the extended measures, that is,

$$\begin{aligned} &G_x^{(j)}(\infty, y_2, y_3; \mathbf{i}_t, \mathbf{l}_t) \\ &\quad = Q_x^{(j)}\Big[W^{(j)}(x, t+1, n, \mathbf{l}_t) \le y_2, r(t+1) \le y_3 \mid \mathbf{i}_t, \mathbf{l}_t\Big], \\ &G_x^{(j)}(y_1, \infty, \infty; \mathbf{i}_t, \mathbf{l}_t) = Q\big[S(t+1) \le y_1 \mid \mathbf{i}_t, \mathbf{l}_t\big], \end{aligned} \tag{4.4}$$

which are determined using (3.18), (3.19), and (3.20) for $j = 1$, (3.24), (45), and (3.26) for $j = 2$, (3.29), (3.30), as well as (3.31) for $j = 3$, and (2.13) for the index. Let $C_{\kappa(t)}$ be the choice of copula, then the cumulative distribution function $G_x^{(j)}$ is defined by

$$G_x^{(j)}(y_1, y_2, y_3; \mathbf{i}_t, \mathbf{l}_t) = C_{\kappa(t)}\Big(G_x^{(j)}(y_1, \infty, \infty; \mathbf{i}_t, \mathbf{l}_t), G_x^{(j)}(\infty, y_2, y_3; \mathbf{i}_t, \mathbf{l}_t)\Big), \tag{4.5}$$

where $\kappa(t)$ represents the free parameter between t and $t+1$ that indicate the level of dependence between the insurance product, interest rate, and the index processes. Here, the copula $C_{\kappa(t)}$ could be defined using either (3.6), (3.7), (3.8), (3.9), or (3.10). Note that in some cases, for example, the lower copula (3.8), the function $G_x^{(j)}$ would not be a cumulative distribution function. We also remark that $G_x^{(j)}$'s are functions of $K(x) \ge t$, but for notational simplicity we suppress $K(x)$.

The martingale probabilities can be obtained from the cumulative distribution function and are given by

$$\begin{aligned} e_x^{(j)}(t, i, \mathbf{i}_t, \mathbf{l}_t) = {} & G_x^{(j)}\Big(S(t)\gamma(t+1, i), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t\}), r(t+1, l_t); \mathbf{i}_t, \mathbf{l}_t\Big) \\ & - G_x^{(j)}\Big(S(t)\gamma(t+1, i-1), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t\}), r(t+1, l_t); \mathbf{i}_t, \mathbf{l}_t\Big), \end{aligned} \tag{4.6}$$

for $i = 0, \ldots, N$,

$$\begin{aligned} e_x^{(j)}(t, i, \mathbf{i}_t, \mathbf{l}_t) = {} & G_x^{(j)}\Big(S(t)\gamma(t+1, i-N-1), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t+1\}), r(t+1, l_t+1); \mathbf{i}_t, \mathbf{l}_t\Big) \\ & - G_x^{(j)}\Big(S(t)\gamma(t+1, i-N-2), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t+1\}), r(t+1, l_t+1); \mathbf{i}_t, \mathbf{l}_t\Big), \end{aligned} \tag{4.7}$$

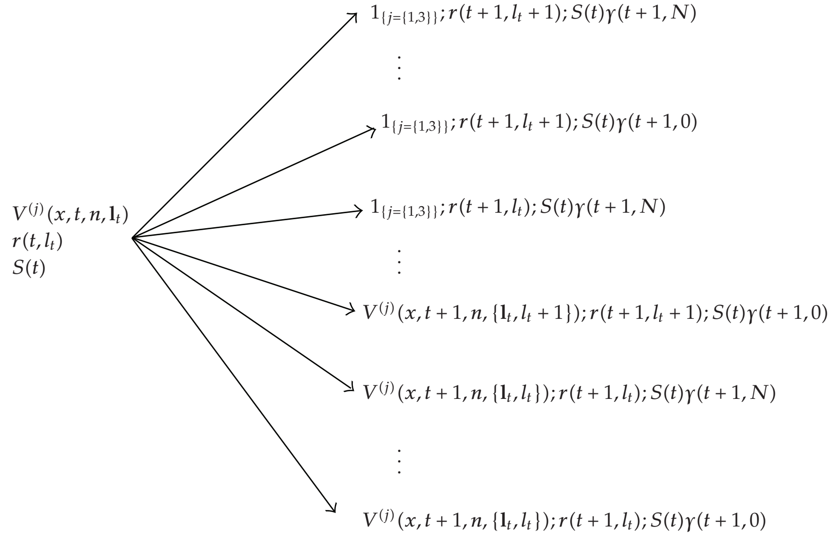


Figure 4: The probability tree of the combined insurance product ($j = 1, 2, 3$), short-term rate, and index processes between t and $t+1$ given that $K(x) \geq t$, $\mathbf{l}_t$, and $\mathbf{i}_t$.

for $i = N+1, \ldots, 2N+1$,

$$\begin{aligned} e_x^{(j)}(t, i, \mathbf{i}_t, \mathbf{l}_t) = {} & G_x^{(j)}(S(t)\gamma(t+1, i-2N-2), 1, r(t+1, l_t); \mathbf{i}_t, \mathbf{l}_t) \\ & - G_x^{(j)}(S(t)\gamma(t+1, i-2N-3), 1, r(t+1, l_t); \mathbf{i}_t, \mathbf{l}_t) \\ & - G_x^{(j)}\Big(S(t)\gamma(t+1, i-2N-2), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t\}), r(t+1, l_t); \mathbf{i}_t, \mathbf{l}_t\Big) \\ & + G_x^{(j)}\Big(S(t)\gamma(t+1, i-2N-3), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t\}), r(t+1, l_t); \mathbf{i}_t, \mathbf{l}_t\Big), \end{aligned} \tag{4.8}$$

for $i = 2N+2, \ldots, 3N+2$, and

$$\begin{aligned} e_x^{(j)}(t, i, \mathbf{i}_t, \mathbf{l}_t) = {} & G_x^{(j)}(S(t)\gamma(t+1, i-3N-3), 1, r(t+1, l_t+1); \mathbf{i}_t, \mathbf{l}_t) \\ & - G_x^{(j)}(S(t)\gamma(t+1, i-3N-4), 1, r(t+1, l_t+1); \mathbf{i}_t, \mathbf{l}_t) \\ & - G_x^{(j)}\Big(S(t)\gamma(t+1, i-3N-3), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t\}), r(t+1, l_t+1); \mathbf{i}_t, \mathbf{l}_t\Big) \\ & - G_x^{(j)}(S(t)\gamma(t+1, i-3N-3), 1, r(t+1, l_t); \mathbf{i}_t, \mathbf{l}_t) \\ & + G_x^{(j)}\Big(S(t)\gamma(t+1, i-3N-4), V^{(j)}(x, t+1, n, \{\mathbf{l}_t, l_t\}), r(t+1, l_t+1); \mathbf{i}_t, \mathbf{l}_t\Big) \end{aligned}$$

$$+ G_x^{(j)}(S(t)\gamma(t+1,i-3N-4),1,r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t)$$
$$+ G_x^{(j)}\Big(S(t)\gamma(t+1,i-3N-3),V^{(j)}(x,t+1,n,\{\mathbf{l}_t,l_t\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big)$$
$$- G_x^{(j)}\Big(S(t)\gamma(t+1,i-3N-4),V^{(j)}(x,t+1,n,\{\mathbf{l}_t,l_t\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big), \quad (4.9)$$

for $i = 3N+3,\ldots,4N+3$ and $j = 1,3$, where $G_x^{(j)}(S(t)\gamma(t+1,-1),\ldots;\mathbf{i}_t,\mathbf{l}_t) = 0$ and $G_x^{(j)}(\ldots;\mathbf{i}_t,\mathbf{l}_t)$ is obtained using (4.5). Similarly, for $j = 2$,

$$\begin{aligned} e_x^{(2)}(t,i,\mathbf{i}_t,\mathbf{l}_t) = {} & G_x^{(2)}\Big(S(t)\gamma(t+1,i),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big) \\ & - G_x^{(2)}\Big(S(t)\gamma(t+1,i-1),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big) \\ & - G_x^{(2)}\Big(S(t)\gamma(t+1,i),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big) \\ & + G_x^{(2)}\Big(S(t)\gamma(t+1,i-1),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big), \end{aligned} \quad (4.10)$$

for $i = 0,\ldots,N$,

$$\begin{aligned} e_x^{(2)}(t,i,\mathbf{i}_t,\mathbf{l}_t) = {} & G_x^{(2)}\Big(S(t)\gamma(t+1,i-N-1),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}),r(t+1,l_t+1);\mathbf{i}_t,\mathbf{l}_t\Big) \\ & - G_x^{(2)}\Big(S(t)\gamma(t+1,i-N-2),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}),r(t+1,l_t+1);\mathbf{i}_t,\mathbf{l}_t\Big) \\ & - G_x^{(2)}(S(t)\gamma(t+1,i-N-1),0,r(t+1,l_t+1);\mathbf{i}_t,\mathbf{l}_t) \\ & - G_x^{(2)}\Big(S(t)\gamma(t+1,i-N-1),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big) \\ & + G_x^{(2)}(S(t)\gamma(t+1,i-N-2),0,r(t+1,l_t+1);\mathbf{i}_t,\mathbf{l}_t) \\ & + G_x^{(2)}\Big(S(t)\gamma(t+1,i-N-2),V^{(2)}(x,t+1,n,\{\mathbf{l}_t,l_t+1\}),r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\Big) \\ & + G_x^{(2)}(S(t)\gamma(t+1,i-N-1),0,r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t) \\ & - G_x^{(2)}(S(t)\gamma(t+1,i-N-2),0,r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t), \end{aligned} \quad (4.11)$$

for $i = N+1,\ldots,2N+1$,

$$\begin{aligned} e_x^{(2)}(t,i,\mathbf{i}_t,\mathbf{l}_t) = {} & G_x^{(2)}(S(t)\gamma(t+1,i-2N-2),0,r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t) \\ & - G_x^{(2)}(S(t)\gamma(t+1,i-2N-3),0,r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t), \end{aligned} \quad (4.12)$$

for $i = 2N+2,\dots,3N+2$, and

$$\begin{aligned} e_x^{(2)}(t,i,\mathbf{i}_t,\mathbf{l}_t) &= G_x^{(2)}\left(S(t)\gamma(t+1,i-3N-3),0,r(t+1,l_t+1);\mathbf{i}_t,\mathbf{l}_t\right) \\ &\quad - G_x^{(2)}\left(S(t)\gamma(t+1,i-3N-4),0,r(t+1,l_t+1);\mathbf{i}_t,\mathbf{l}_t\right) \\ &\quad - G_x^{(2)}\left(S(t)\gamma(t+1,i-3N-3),0,r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\right) \\ &\quad + G_x^{(2)}\left(S(t)\gamma(t+1,i-3N-4),0,r(t+1,l_t);\mathbf{i}_t,\mathbf{l}_t\right), \end{aligned} \tag{4.13}$$

for $i = 3N+3,\dots,4N+3$.

Consider now an equity-linked product that pays

$$\begin{cases} D(K(x)+1) & if\ K(x) = 0,1,\dots,n-1, \\ D(n) & if\ K(x) \geq n. \end{cases} \tag{4.14}$$

For notational convenience, we sometimes use $D(t,\mathbf{i}_t)$ to specify the index's realization.

Let $P^{(1)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ denote the premium at time t of the equity-linked contract death benefit given that (x) is still alive and the index and short-term rate processes have taken the path $\mathbf{i}_t$ and $\mathbf{l}_t$, respectively. With the martingale structure identified by (4.6), (4.7), (4.8), and (4.9), $P^{(1)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ may be obtained as the expected discounted payoffs

$$P^{(1)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t) = E_x^{(1)+}\left[\frac{D(K(x)+1)1_{\{K(x)<n\}}}{B(K(x)+1)}B(t)\,\middle|\,\mathbf{i}_t,\mathbf{l}_t,K(x)\geq t\right], \tag{4.15}$$

where $E_x^{(1)+}[\cdot]$ represents the expectation with respect to $Q_x^{(1)+}$.

On the other hand, let $P^{(2)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ denote the premium at time t of the equity-linked product accumulation benefit given that (x) is still alive and the index process has taken the path $\mathbf{i}_t$. With the martingale structure identified by (4.10), (4.11), (4.12), and (4.13), $P^{(2)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ may be obtained as the expected discounted payoffs

$$P^{(2)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t) = E_x^{(2)+}\left[\frac{D(n)1_{\{K(x)\geq n\}}}{B(n)}B(t)\,\middle|\,\mathbf{i}_t,\mathbf{l}_t,K(x)\geq t\right]. \tag{4.16}$$

Let $P(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ denote the premium at time t of an n-year equity-linked product issue to (x) with its payoff defined by (4.14). In particular, $P(x,0,n,\mathbf{i}_0,\mathbf{l}_0)$ is the amount invested by an insured or, from insurers' point of view, the premium paid by the policyholder. We assume that $P(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ may be decomposed in two different premiums; $P^{(1)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ is the premium to cover the death benefit and $P^{(2)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ is the premium to cover the accumulation benefit. That is, we assume that

$$P(x,t,n,\mathbf{i}_t,\mathbf{l}_t) = P^{(1)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t) + P^{(2)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t), \tag{4.17}$$

where $P^{(1)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ and $P^{(2)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ are obtained using (4.15) and (4.16), respectively.

An alternative approach is to evaluate equity-linked products containing death and accumulation benefits in a unified manner, using the pricing information from the endowment insurance products. In this case, the n-year equity-linked product premium $P(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ may be obtained as expected discounted payoffs

$$P(x,t,n,\mathbf{i}_t,\mathbf{l}_t) = E_x^{(3)+}\left[\left(\frac{D(K(x)+1)1_{\{K(x)<n-1\}}}{B(K(x)+1)} + \frac{D(n)1_{\{K(x)\geq n-1\}}}{B(n)}\right)B(t)\middle|\mathbf{i}_t,\mathbf{l}_t,K(x)\geq t\right]. \tag{4.18}$$

Figure 4 presents the dynamic of the equity-linked premiums for time period $[t,t+1]$.

Bear in mind that the first approach presented in this section evaluates equity-linked products by loading the death and survival probabilities separately. The second approach evaluates the equity-linked product using unified loaded probabilities.

5. Evaluation of Equity-Linked Products

In this section, we evaluate equity-linked contracts using recursive algorithms. It follows from (4.15) and (4.16) that

$$\begin{aligned}
&P^{(1)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t)\\
&= \frac{1}{1+r(t,l_t)}\Bigg[\sum_{v=0}^{N}\Big(e_x^{(1)}(t,v+2N+2,\mathbf{i}_t,\mathbf{l}_t)+e_x^{(1)}(t,v+3N+3,\mathbf{i}_t,\mathbf{l}_t)\Big)D(t+1,\{\mathbf{i}_t,v\})\\
&\qquad + \sum_{v=0}^{N}\Big(e_x^{(1)}(t,v,\mathbf{i}_t,\mathbf{l}_t)P^{(1)}(x,t+1,n,\{\mathbf{i}_t,v\},\{\mathbf{l}_t,l_t\})\\
&\qquad\qquad + e_x^{(1)}(t,v+N+1,\mathbf{i}_t,\mathbf{l}_t)P^{(1)}(x,t+1,n,\{\mathbf{i}_t,v\},\{\mathbf{l}_t,l_t+1\})\Big)\Bigg],
\end{aligned} \tag{5.1}$$

with $P^{(1)}(x,n,n,\mathbf{i}_n,\mathbf{l}_n)=0$ and

$$\begin{aligned}
P^{(2)}(x,t,n,\mathbf{i}_t,\mathbf{l}_t) = \frac{1}{1+r(t,l_t)}\Bigg[&\sum_{v=0}^{N}\Big(e_x^{(2)}(t,v,\mathbf{i}_t,\mathbf{l}_t)P^{(2)}(x,t+1,n,\{\mathbf{i}_t,v\},\{\mathbf{l}_t,l_t\})\\
&+e_x^{(2)}(t,v+N+1,\mathbf{i}_t,\mathbf{l}_t)P^{(2)}(x,t+1,n,\{\mathbf{i}_t,v\},\{\mathbf{l}_t,l_t+1\})\Big)\Bigg],
\end{aligned} \tag{5.2}$$

with $P^{(2)}(x,n,n,\mathbf{i}_n,\mathbf{l}_n)=D(n,\mathbf{i}_n)$.

Table 1: Point-to-point with term-end design for various interest rate volatilities.

		Decomposed approach							Unified approach		
σ	$\sigma_r(t)$	C^I	C^U	C^D	$C_{0.5}^{CJ}$	C_2^{CJ}	$C_{-0.1}^G$	$C_{0.3}^G$	C^I	C^U	C^D
3% Minimum guarantee on 90% premium											
	0%	62.17	56.27	68.58	59.94	58.38	63.08	59.59	70.28	69.98	70.05
20%	4%	62.17	55.85	69.17	59.79	58.07	63.12	59.46	70.27	69.40	70.64
	8%	62.15	55.44	69.75	59.63	57.75	63.15	59.31	70.25	68.83	71.23
30%	0%	48.95	43.20	55.15	47.02	45.63	49.79	46.54	55.46	55.14	55.25
	4%	48.95	42.93	55.54	46.93	45.43	49.82	46.45	55.45	54.74	55.68
	8%	48.94	42.66	55.93	46.83	45.23	49.85	46.37	55.45	54.35	56.11
3% Minimum guarantee on 100% premium											
	0%	44.58	41.32	47.87	43.45	42.65	45.06	43.20	53.25	53.12	52.88
20%	4%	44.57	40.93	48.38	43.32	42.36	45.09	43.07	53.25	52.56	53.46
	8%	44.55	40.53	48.89	43.18	42.07	45.12	42.94	53.24	52.00	54.03
	0%	32.20	29.12	35.38	31.25	30.56	32.63	30.96	39.27	39.13	38.98
30%	4%	32.20	28.89	35.69	31.18	30.40	32.65	30.89	39.27	38.77	39.36
	8%	32.19	28.65	36.00	31.10	30.23	32.67	30.81	39.27	38.41	39.75

A recursive formula to evaluate $P(x,t,n,\mathbf{i}_t,\mathbf{l}_t)$ under $Q_x^{(3)+}$ is determined using (4.18), that is,

$$\begin{aligned} &P(x,t,n,\mathbf{i}_t,\mathbf{l}_t) \\ &= \frac{1}{1+r(t,l_t)}\Bigg[\sum_{v=0}^{N}\Big(e_x^{(3)}(t,v+2N+2,\mathbf{i}_t,\mathbf{l}_t)+e_x^{(3)}(t,v+3N+3,\mathbf{i}_t,\mathbf{l}_t)\Big)D(t+1,\{\mathbf{i}_t,v\}) \\ &\qquad + \sum_{v=0}^{N}\Big(e_x^{(3)}(t,v,\mathbf{i}_t,\mathbf{l}_t)P^{(3)}(x,t+1,n,\{\mathbf{i}_t,v\},\{\mathbf{l}_t,l_t\}) \\ &\qquad\qquad + e_x^{(3)}(t,v+N+1,\mathbf{i}_t,\mathbf{l}_t)P(x,t+1,n,\{\mathbf{i}_t,v\},\{\mathbf{l}_t,l_t+1\})\Big)\Bigg], \end{aligned} \tag{5.3}$$

for $t=0,\ldots,n-2$, where

$$P(x,n-1,n,\mathbf{i}_{n-1},\mathbf{l}_{n-1}) = \sum_{v=0}^{N}\frac{\pi(n,l_{n-1})^{v}(1-\pi(n,l_{n-1}))^{N-v}}{1+r(n-1,l_{n-1})}D(n,\{\mathbf{i}_{n-1},v\}). \tag{5.4}$$

Note that the surrender options for equity-linked products under stochastic interest rates are evaluated in Gaillardetz [27].

6. Valuation of Equity-Indexed Annuities: Numerical Examples

This section implements numerically the methods we developed previously by considering two types of equity-indexed annuities. They appeal to investors because they offer the same protection as conventional annuities by limiting the financial risk, but also provide participation in the equity market. From Lin and Tan [11] and Tiong [8], EIA designs may be generally grouped in two broad classes: Annual Reset and Point-to-Point. The index growth on an EIA with the former is measured and locked in each year. Particularly, the index growth with a term-end point design is calculated using the index value at the beginning and the end of each year. In the latter, the index growth is based on the growth between two time points over the entire term of the annuity. In the case of the term-end point design, the growth is evaluated using the beginning and ending values of the index. The cost of the EIA contract is reflected through the participation rate. Hence, the participation rate is expected to be lower for expensive designs.

Our examples involve five-year EIAs issued to a male-aged 55 with minimum interest rate guarantee of either 3% on 100% of the premium or 3% on 90% of the premium. For illustration purposes, we assume that the insurance product values, $V^{(j)}(x,0,n)$ ($j = 1,2,3$ and $n = 1,2,3,4,5$), are determined using the standard deviation premium principle (see Bowers et al. [20]) with a loading factor of 5.00% based on the 1980 US CSO table (see http://www.soa.org/). We also assume that the short-term rate $r(t)$ follows the BDT where the volatility is either 0%, 4%, or 8%. The observed price of the zero-coupon bond $L(0,T)$ is assumed to be equal to $(1.05)^{-T}$ for $T = 1,2,\ldots,5$. Hence, the interest rate model may be calibrated using (2.6) and (2.7). For simplification purposes, the index will be governed by the Cox et al. [28] model where $S(0) = 1$ and the number of trading dates N is 3. In this recombining model, the index at time k $S(k)$ has two possible outcomes at time $k+\Delta$: it is either increasing to $S(k+\Delta) = US(k)$ or decreasing to $S(k+\Delta) = dS(k)$. The increasing and decreasing factors u and d are supposed to be constant and are obtained from the volatility of the index σ. This volatility is assumed to be constant and is either 20% or 30%. In other words, $u = e^{\sigma/\sqrt{N}}(\sigma = 0.2, 0.3)$ and $d = u^{-1}$. The index conditional martingale probability structure is obtained using (2.10). The conditional joint distribution of the interest rates and the insurance products are obtained using Propositions 3.1, 3.2, and 3.2. Here, these martingale probabilities are determined based on the implied insurance volatilities, which are derived numerically under the constraint given in (3.33).

The analysis is performed using the point-to-point and reset EIA classes with term-end point design.

6.1. Point-to-Point

We first consider one of the simplest classes of EIAs, known as the point-to-point. Their payoffs in year t can be represented by

$$D(t) = \max\left[\min\left[1+\alpha R(t), (1+\zeta)^t\right], \beta(1+g)^t\right], \tag{6.1}$$

where α represents the participation rate and the "gain" $R(t)$ need to be defined depending on the design. It also provides a protection against the loss from a down market $\beta(1+g)^t$. The cap rate $(1+\zeta)^t$ reduces the cost of such contract since it imposes an upper bound on the maximum return.

As explained in Lin and Tan [11], an EIA is evaluated through its participation rate α. Without loss of generality, we suppose that the initial value of EIA contracts is one monetary unit. The present value of the EIA is a function of the participation rate through the payoff function D. We then solve numerically for α, the critical participation rate, such

Table 2: Point-to-point with term-end design and 15% cap rate for various interest rate volatilities.

		Decomposed approach							Unified approach		
σ	$\sigma_r(t)$	C^I	C^U	C^D	$C_{0.5}^{CJ}$	C_2^{CJ}	$C_{-0.1}^G$	$C_{0.3}^G$	C^I	C^U	C^D
3% Minimum guarantee on 90% premium											
20%	0%	65.65	59.00	72.98	62.95	61.16	66.69	62.73	75.95	76.61	75.43
	4%	65.65	58.61	73.40	62.82	60.88	66.72	62.62	75.95	76.10	75.85
	8%	65.64	58.23	73.85	62.68	60.61	66.75	62.50	75.94	75.64	76.29
30%	0%	58.94	51.26	67.42	55.97	54.00	60.11	55.62	69.87	70.94	69.33
	4%	58.94	51.10	67.54	55.91	53.89	60.12	55.59	69.88	70.66	69.52
	8%	58.94	50.96	67.70	55.86	53.79	60.13	55.56	69.90	70.45	69.76
3% Minimum guarantee on 100% premium											
20%	0%	45.20	41.85	48.58	44.00	43.18	45.69	43.78	54.60	54.93	54.10
	4%	45.19	41.44	49.07	43.87	42.89	45.72	43.66	54.60	54.35	54.63
	8%	45.17	41.03	49.56	43.73	42.59	45.74	43.52	54.59	53.80	55.16
30%	0%	34.25	30.86	37.78	33.12	32.34	34.73	32.87	43.20	43.99	42.60
	4%	34.25	30.63	38.04	33.06	32.19	34.75	32.81	43.20	43.46	42.96
	8%	34.25	30.40	38.31	32.99	32.04	34.77	32.74	43.21	42.97	43.45

that $P(x,0,n,\mathbf{i}_0,\mathbf{l}_0) = 1$, where $P(x,0,n,\mathbf{i}_0,\mathbf{l}_0)$ is obtained using (4.17) for the first approach or using (4.18) for the second approach by holding all other parameter values constant.

6.1.1. Term-End Point

In practice, various designs for $R(t)$ have been proposed. The term-end point design is the simplest crediting method. It measures the index growth from the start to the end of a term. The index on the day the contract is issued is taken as the starting index, and the index on the day the policy matures or the time of death is taken as the ending index. Hence, the "gain" provided by the point-to-point EIA with term-end point may expressed as

$$R(t) = \frac{S(t)}{S(0)} - 1. \tag{6.2}$$

The EIA payoff given in (6.1) is defined by

$$D(t,\mathbf{i}_t) = \max\left[\min\left[1 + \alpha\left(\prod_{l=0}^{t}\gamma(l,i_l) - 1\right), (1+\zeta)^t\right], \beta(1+g)^t\right]. \tag{6.3}$$

Tables 1 and 2 give the critical participation rates based on (5.1) and (5.2) for the decomposed approach as well as (5.3) for the unified approach over different short-term rate volatilities (0%, 4%, and 8%). The index volatility is set to either 20% or 30%. We present the participation rates of 5-year EIA contracts with the term-end design without cap rate ($\zeta = \infty$) in Table 1 and 15% cap rate in Table 2. We consider two types of minimum guarantees: $\beta = 90\%$ and $\beta = 100\%$ and both with $g = 3\%$.

The participation rates obtained for $\sigma_r(t) = 0\%$ are consistent with the corresponding participation rates under deterministic interest rates presented in Gaillardetz and Lin [13].

Table 3: Annual reset with term-end point design for various interest rate volatilities.

		Decomposed approach							Unified approach		
σ	$\sigma_r(t)$	C^I	C^U	C^D	$C^{CJ}_{0.5}$	C^{CJ}_{2}	$C^{G}_{-0.1}$	$C^{G}_{0.3}$	C^I	C^U	C^D
		3% Minimum guarantee on 90% premium									
20%	0%	38.91	37.19	40.51	38.42	38.02	39.16	38.19	43.26	43.25	43.27
	4%	38.91	37.27	40.41	38.44	38.07	39.15	38.21	43.26	43.31	43.21
	8%	38.91	37.35	40.32	38.46	38.12	39.14	38.23	43.27	43.37	43.17
30%	0%	28.23	26.72	29.59	27.86	27.54	28.42	27.65	31.39	31.38	31.39
	4%	28.23	26.79	29.50	27.88	27.59	28.41	27.67	31.39	31.43	31.34
	8%	28.23	26.87	29.41	27.89	27.63	28.41	27.69	31.39	31.50	31.30
		3% Minimum guarantee on 100% premium									
20%	0%	35.42	33.93	36.74	35.04	34.71	35.62	34.82	40.59	40.57	40.55
	4%	35.42	33.92	36.76	35.03	34.70	35.62	34.82	40.59	40.52	40.60
	8%	35.41	33.90	36.77	35.02	34.69	35.61	34.81	40.59	40.49	40.66
30%	0%	25.19	23.87	26.36	24.91	24.65	25.35	24.71	28.89	28.86	28.87
	4%	25.19	23.88	26.34	24.91	24.66	25.35	24.72	28.89	28.85	28.89
	8%	25.19	23.89	26.32	24.91	24.67	25.34	24.72	28.90	28.85	28.91

As expected, the participation rates for the independent copulas decrease as the interest rate volatility increases; however, this effect is negligible for 5-year contracts.

The independent copula may be obtained by letting $\kappa \rightarrow 0$ in the Cook-Johnson copula. Similarly, the Frechet-Hoeffding upper bound is obtained by letting $\kappa \rightarrow \infty$. This explains that the participation rates with $\kappa(t) = 0.5$ are closer to the independent one than the participation rates obtained using $\kappa(t) = 2$, which are closer to the upper copula. Setting $\kappa(t) = 0$ in the Gaussian copula also leads to the independent copula. The participation rates are between the independent copula and the lower copula when $\kappa(t) = -0.1$. On the other hand, when $\kappa(t) = 0.3$ the participation rates are between the independent copula and the upper copula.

The width of participation rate bands for the decomposed approach increases as the short-term rate volatility increases. Here, the participation rate band represents the difference between the participation rates obtained from the lower copula and the upper copula. Indeed, the participation rate for the upper copula ($\beta = 90\%$ and $\sigma = 20\%$) decreases from 56.27% ($\sigma_r(t) = 0\%$) to 55.44% ($\sigma_r(t) = 8\%$), meanwhile under the lower copula the participation rate passes from 68.58% ($\sigma_r(t) = 0\%$) to 69.75% ($\sigma_r(t) = 0\%$). This is due to the fact that increasing σ_r introduces more uncertainty in the model.

As we increase the volatility of the index, the participation rate decreases since a higher volatility leads to more valuable embedded financial options. As expected, the participation rates for $\beta = 100\%$ are lower than the corresponding values with $\beta = 90\%$.

The dependence effects for the unified approach are negligible since there is a natural "hedging" between the death and accumulation benefits. The introduction of stochastic interest rates has more impact when $\beta = 90\%$ than when $\beta = 100\%$ because the participation rates are higher. Although the participation rates are higher when $\sigma = 20\%$, the dependence has relatively more impact if $\sigma = 30\%$ because the model is more risky.

As expected, imposing a ceiling on the equity return that can be credited increases the participation rates. Furthermore, the magnitude of the increments is more significant

Table 4: Annual reset with term-end point design and 15% cap rate for various interest rate volatilities.

		Decomposed Approach							Unified Approach		
σ	$\sigma_r(t)$	C^I	C^U	C^D	$C^{CJ}_{0.5}$	C^{CJ}_{2}	$C^G_{-0.1}$	$C^G_{0.3}$	C^I	C^U	C^D
3% Minimum Guarantee on 90% Premium											
20%	0%	42.49	38.58	45.86	41.35	40.44	43.04	40.86	52.63	52.62	52.65
	4%	42.49	38.78	45.65	41.39	40.56	43.02	40.91	52.64	52.76	52.53
	8%	42.49	38.98	45.44	41.44	40.69	43.00	40.97	52.66	52.91	52.44
30%	0%	35.80	33.10	38.12	35.02	34.40	36.18	34.69	42.82	42.80	42.83
	4%	35.81	33.28	37.93	35.06	34.50	36.17	34.73	42.83	42.94	42.72
	8%	35.81	33.47	37.75	35.09	34.60	36.16	34.78	42.84	43.09	42.62
3% Minimum Guarantee on 100% Premium											
20%	0%	35.42	33.93	37.69	35.04	34.71	35.62	34.82	47.46	47.45	47.35
	4%	35.42	33.92	37.73	35.03	34.70	35.62	34.82	47.46	47.36	47.46
	8%	35.41	33.90	37.77	35.02	34.69	35.61	34.81	47.47	47.28	47.60
30%	0%	30.40	27.60	32.72	29.73	29.17	30.75	29.36	39.41	39.45	39.34
	4%	30.40	27.63	32.66	29.74	29.19	30.75	29.37	39.42	39.45	39.36
	8%	30.40	27.67	32.62	29.75	29.22	30.74	29.38	39.43	39.48	39.40

in a high volatility market. This is because the effect of the volatility diminishes as the cap rate decreases and hence the behavior of the EIA payoff is similar for different ranges of volatilities. This is particularly observable when $\beta = 90\%$.

6.2. *Annual Reset*

We now consider the most popular class of EIAs, known as the annual reset. They appeal to investors because they offer similar features as the point-to-point class; however, the interest credited to a annual reset EIA contract cannot be lost. This "lock-in" feature protects the investor against a poor performance of the index over a particular year. The payoff of this type of EIA contracts is defined by

$$D(t) = \max\left[\prod_{l=1}^{t} \max[\min[1 + \alpha R(l) - \nu, (1 + \zeta)], 1], \beta(1 + g)^t\right], \tag{6.4}$$

where $R(l)$ represents the realized "gain" in year l, which varies from product to product.

The cases where ν is set to 0 are known as annual reset EIAs and the cases where $\nu > 0$ are known as annual yield spreads. Furthermore, the participation levels in those cases $\nu > 0$ are typically 100%. As mentioned previously, in the case of annual reset, we fix $\nu = 0$ and determine the critical participation rate α while fixing g, β, and ζ. In the traditional yield spread ν needs to be determined such that the cost of EIA embedded options is covered by the initial premium while fixing $\alpha = 100\%$, g, β, and ζ.

6.2.1. Term-End Point

In the case of annual reset EIA with term-end point, the index return is calculated each year by comparing the indices at the beginning and ending policy anniversaries. Hence, the participation rate may be expressed as

$$R(t) = \frac{S(t)}{S(t-1)} - 1. \tag{6.5}$$

In this case, the EIA payoff given in (6.4) is defined by

$$D(t, \mathbf{i}_t) = \max\left[\prod_{l=1}^{t} \max\left[\min\left[1 + \alpha\left(\gamma(l, i_l) - 1\right) - v, (1+\zeta)\right], 1\right], \beta(1+g)^t\right]. \tag{6.6}$$

Tables 3 and 4 consider an annual reset EIA with term-end point design for various cap rates ($\zeta = \infty$ and $\zeta = 15\%$). In this numerical illustration, we consider the same set of parameters; particularly, the short-term rate volatility is either equal to 0%, 4%, or 8% and the index volatility is set to either 20% or 30% with $N = 3$. We find α such that (4.17) for the decomposed approach and (4.18) for the unified approach are equal to 1 by setting $\nu = 0\%$.

The annual reset with term-end point design is more expensive than the point-to-point with the term-end point design. The participation rates from the upper copula increase while the ones from the lower copula decrease as $\sigma_r(t)$ increases for both approaches when $\beta = 90\%$. This leads to narrower participation rate bands for the decomposed approach. These behaviors are inverted for $\beta = 100\%$. In that case, it leads to wider participation rate bands for the decomposed approach. The imposition of a 15% cap rate may increase the participation rate as much as 10%. However, there is no impact on the participation rate under the decomposed approach when $\sigma = 20\%$ and $\beta = 100\%$.

7. Conclusions

The purpose of this paper is to generalize the approach presented by Gaillardetz and Lin [13] under stochastic interest rates. To this end, martingale probability measures for each of the term life, pure endowment, and endowment insurances are introduced under stochastic interest rates. Using the insurance market information, we obtain equity-linked martingale measures that combined the insurance, interest rates, and index information. Although the choice of copulas is somewhat arbitrary, with additional premium information from certain equity-linked products, we would be able to narrow down the choices. We present two different pricing approaches for equity-linked products. The first approach evaluates death benefits and accumulation/survival benefits separately. In the second approach, we evaluate the death benefits and the survival benefits in a unified manner by using the endowment insurance products to define the martingale measure. A detailed numerical analysis is then performed for existing EIAs in the North American market.

Our methodology may be used to evaluate variable annuities (segregated fund contracts in Canada) because of the similarity in payoff structure between EIAs and VAs. Furthermore, our approach may also be used to evaluate Universal Life insurances, variable Universal Life insurances, and others equity-linked products.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada and a Ph. D. grant from the Casualty Actuarial Society and the Society of Actuaries. The author is very grateful to Professor X. Sheldon Lin for his valuable comments and suggestions.

Endnotes

1. Utility theory is also used to price standard life insurance products.

References

[1] M. R. Hardy, *Investment Guarantees: Modeling and Risk Management for Equity- Linked Life Insurance*, John Wiley & Sons, Hoboken, NJ, USA, 2003.
[2] M. J. Brennan and E. S. Schwartz, "The pricing of equity-linked life insurance policies with an asset value guarantee," *Journal of Financial Economics*, vol. 3, pp. 195–213, 1976.
[3] P. Boyle and E. Schwartz, "Equilibrium prices of equity linked insurance policies with an asset value guarantee," *Journal of Risk and Insurance*, vol. 44, pp. 639–660, 1977.
[4] A. R. Bacinello and F. Ortu, "Pricing guaranteed securities-linked life insurance under interest-rate risk," in *Proceedings of the 3rd AFIR International Colloquium on Actuarial Approach For Financial Risks*, pp. 35–55, Rome, Italy, April 1993.
[5] A. R. Bacinello and F. Ortu, "Single and periodic premiums for guaranteed equitylinked life insurance under interest-rate risk: the "lognormal + Vasicek" case," in *Financial Modelling*, pp. 1–25, Physica, Berlin, Germany, 1994.
[6] T. Møller, "Risk-minimizing hedging strategies for unit-linked life insurance contracts," *Astin Bulletin*, vol. 28, no. 1, pp. 17–47, 1998.
[7] V. R. Young and T. Zariphopoulou, "Pricing dynamic insurance risks using the principle of equivalent utility," *Scandinavian Actuarial Journal*, no. 4, pp. 246–279, 2002.
[8] S. Tiong, "Valuing equity-indexed annuities," *North American Actuarial Journal*, vol. 4, no. 4, pp. 149–163, 2000.
[9] H. Lee, "Pricing equity-indexed annuities with path-dependent options," *Insurance: Mathematics & Economics*, vol. 33, no. 3, pp. 677–690, 2003.
[10] K. S. Moore, "Optimal surrender strategies for equity-indexed annuity investors," *Insurance: Mathematics & Economics*, vol. 44, no. 1, pp. 1–18, 2009.
[11] X. S. Lin and K. S. Tan, "Valuation of equity-Indexed annuities under stochastic interest rate," *North American Actuarial Journal*, vol. 7, no. 3, pp. 72–91, 2003.
[12] M. Kijima and T. Wong, "Pricing of ratchet equity-indexed annuities under stochastic interest rates," *Insurance: Mathematics & Economics*, vol. 41, no. 3, pp. 317–338, 2007.
[13] P. Gaillardetz and X. S. Lin, "Valuation of equity-linked insurance and annuity products with binomial models," *North American Actuarial Journal*, vol. 10, no. 4, pp. 117–144, 2006.
[14] M.V. Wüthrich, H. Bühlmann, and H. Furrer, *Market-Consistent Actuarial Valuation*, EAA Lecture Notes, Springer, Berlin, Germany, 2008.
[15] P. Gaillardetz, "Valuation of life insurance products under stochastic interest rates," *Insurance: Mathematics & Economics*, vol. 42, no. 1, pp. 212–226, 2008.
[16] H. H. Panjer, et al., *Financial Economics with Applications to Investment, Insurance and Pensions*, The Actuarial Foundation, Schaumburg, Ill, USA, 1998.
[17] X. S. Lin, *Introductory Stochastic Analysis for Finance and Insurance*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, USA, 2006.
[18] F. Black, E. Derman, and W. Toy, "A one-factor model of interest rates and its application to treasury bond options," *Financial Analysts Journal*, vol. 46, pp. 33–39, 1990.
[19] J. Hull and A. White, "Numerical procedures for implementing term structure models I: single-factor models," *Journal of Derivatives*, vol. 2, no. 1, p. 716.

[20] N. L. Bowers Jr., U. G. Hans, J. C. Hickman, D. A. Jones, and Nesbitt, *Actuarial Mathematics*, Society of Actuaries, Schaumburg, Ill, USA, 2nd edition, 1997.
[21] H. Joe, *Multivariate Models and Dependence Concepts*, vol. 73 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London, UK, 1997.
[22] R. B. Nelsen, *An Introduction to Copulas*, vol. 139 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1999.
[23] E. W. Frees and E. A. Valdez, "Understanding relationships using copulas," *North American Actuarial Journal*, vol. 2, no. 1, pp. 1–25, 1998.
[24] S. Wang, "Aggregation of correlated risk portfolios: models and algorithms," in *Proceedings of Casualty Actuarial Society*, pp. 848–939, Arlington, Va, USA, 1998.
[25] G. G. Venter, "Tails of copulas," in *Proceedings of Casualty Actuarial Society*, pp. 68–113, 2000.
[26] G. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*, Wiley Finance Series, John Wiley & Sons, Chichester, UK, 2004.
[27] P. Gaillardetz, *Equity-linked annuities and insurances*, Ph.D. thesis, University of Toronto, Toronto, Ontario, Canada, 2006.
[28] J. C. Cox, S. A. Ross, and M. Rubinstein, "Option pricing: a simplified approach," *Journal of Financial Economics*, vol. 7, no. 3, pp. 229–263, 1979.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 754851, 26 pages
doi:10.1155/2010/754851

Research Article

Local Likelihood Density Estimation and Value-at-Risk

Christian Gourieroux[1] and Joann Jasiak[2]

[1] *CREST and University of Toronto, Canada*
[2] *York University, Canada*

Correspondence should be addressed to Joann Jasiak, jasiakj@econ.yorku.ca

Received 5 October 2009; Accepted 9 March 2010

Academic Editor: Ričardas Zitikis

Copyright © 2010 C. Gourieroux and J. Jasiak. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a new nonparametric method for computing the conditional Value-at-Risk, based on a local approximation of the conditional density function in a neighborhood of a predetermined extreme value for univariate and multivariate series of portfolio returns. For illustration, the method is applied to intraday VaR estimation on portfolios of two stocks traded on the Toronto Stock Exchange. The performance of the new VaR computation method is compared to the historical simulation, variance-covariance, and J. P. Morgan methods.

1. Introduction

The Value-at-Risk (VaR) is a measure of market risk exposure for portfolios of assets. It has been introduced by the Basle Committee on Banking Supervision (BCBS) and implemented in the financial sector worldwide in the late nineties. By definition, the VaR equals the Dollar loss on a portfolio that will not be exceeded by the end of a holding time with a given probability. Initially, the BCBS has recommended a 10-day holding time (and allowed for computing the VaR at horizon 10 days by rescaling the VaR at a shorter horizon) and loss probability 1%; (see, [1], page 3), Banks use the VaR to determine the required capital to be put aside for coverage of potential losses. (The required capital reserve is defined as $RC_t = \mathrm{Max}[\mathrm{VaR}_t, (M+m)1/60 \sum_{h=1}^{60} \mathrm{VaR}_{t-h}]$, (see, [1], page 14 and [2], page 2), where M is a multiplier set equal to 3, and m takes a value between 0 and 1 depending on the predictive quality of the internal model used by the bank.) The VaR is also used in portfolio management and internal risk control. Therefore, some banks compute intradaily VaRs, at horizons of one or two hours, and risk levels of 0.5%, or less.

Formally, the conditional Value-at-Risk is the lower-tail conditional quantile and satisfies the following expression:

$$P_t[x_{t+1} < -\mathrm{VaR}_t(\alpha)] = \alpha, \tag{1.1}$$

where x_t is the portfolio return between $t-1$ and t, α denotes the loss probability, and P_t represents the conditional distribution of x_{t+1} given the information available at time t. Usually, the information set contains the lagged values $x_t, x_{t-1}, \ldots$ of portfolio returns. It can also contain lagged returns on individual assets, or on the market portfolio.

While the definition of the VaR as a market risk measure is common to all banks, the VaR computation method is not. In practice, there exist a variety of parametric, semiparametric, and nonparametric methods, which differ with respect to the assumptions on the dynamics of portfolio returns. They can be summarized as follows (see, e.g., [3]).

(a) Marginal VaR Estimation

The approach relies on the assumption of i.i.d. returns and comprises the following methods.

(1) Gaussian Approach

The VaR is the α-quantile, obtained by inverting the Gaussian cumulative distribution function

$$\mathrm{VaR}(\alpha) = -Ex_{t+1} - \Phi^{-1}(\alpha)(Vx_{t+1})^{1/2}, \tag{1.2}$$

where Ex_{t+1} is the expected return on a portfolio, Vx_{t+1} is the variance of portfolio returns, and $\Phi^{-1}(\alpha)$ is the α-quantile of the standard normal distribution. This method assumes the normality of returns and generally underestimates the VaR. The reason is that the tails of the normal distribution are much thinner than the tails of an empirical marginal distribution of portfolio returns.

(2) Historical Simulation (see [1])

$\mathrm{VaR}(\alpha)$ is approximated from a sample quantile at probability α, obtained from historical data collected over an observation period not shorter than one year. The advantage of this method is that it relaxes the normality assumption. Its major drawback is that it provides poor approximation of small quantiles at α's such as 1%, for example, as extreme values are very infrequent. Therefore, a very large sample is required to collect enough information about the true shape of the tail. (According to the asymptotic properties of the empirical quantile, at least 200–300 observations, that is, one year, approximately, are needed for $\alpha = 5\%$ and at least 1000, that is, 4 years are needed for $\alpha = 1\%$, both for a Gaussian tail. For fatter tails, even more observations can be required (see, e.g., the discussion in [3]).

(3) Tail Model Building

The marginal quantile at a small risk level α is computed from a parametric model of the tail and from the sample quantile(s) at a larger α. For example, McKinsey Inc. suggests to infer the 99th quantile from the 95th quantile by multiplying the latter one by 1.5, which is the weight based on a zero-mean Gaussian model of the tail. This method is improved by considering two tail quantiles. If a Gaussian model with mean μ and variance σ is assumed to fit the tail for $\alpha < 10\%$, then the $\widehat{\mathrm{VaR}}(\alpha)$, for any $\alpha < 10\%$, can be calculated as follows, Let $\widehat{\mathrm{VaR}}(10\%)$ and $\widehat{\mathrm{VaR}}(5\%)$ denote the sample quantiles at risk levels 5% and 10%. From (1.2), the estimated mean and variance in the tail arise as the solutions of the system

$$\begin{aligned} \widehat{\mathrm{VaR}}(10\%) &= -\widehat{m} - \Phi^{-1}(10\%)\widehat{\sigma}, \\ \widehat{\mathrm{VaR}}(5\%) &= -\widehat{m} - \Phi^{-1}(5\%)\widehat{\sigma}. \end{aligned} \tag{1.3}$$

The marginal VaR at any loss probability α less than 10% is calculated as

$$\widehat{\mathrm{VaR}}(\alpha) = -\widehat{m} - \Phi^{-1}(\alpha)\widehat{\sigma}, \tag{1.4}$$

where $\widehat{m}, \widehat{\sigma}$ are solutions of the above system. Equivalently, we get

$$\frac{\widehat{\mathrm{VaR}}(\alpha) - \widehat{\mathrm{VaR}}(10\%)}{\widehat{\mathrm{VaR}}(5\%) - \widehat{\mathrm{VaR}}(10\%)} = \frac{\Phi^{-1}(\alpha) - \Phi^{-1}(10\%)}{\Phi^{-1}(5\%) - \Phi^{-1}(10\%)}. \tag{1.5}$$

Thus, $\widehat{\mathrm{VaR}}(\alpha)$ is a linear combination of sample quantiles $\widehat{\mathrm{VaR}}(10\%)$ and $\widehat{\mathrm{VaR}}(5\%)$ with the weights determined by the Gaussian model of the tail.

This method is parametric as far as the tail is concerned and nonparametric for the central part of the distribution, which is left unspecified.

The marginal VaR estimation methods discussed so far do not account for serial dependence in financial returns, evidenced in the literature. (These methods are often applied by rolling, that is, by averaging observations over a window of fixed length, which implicitly assumes independent returns, with time dependent distributions.)

(b) Conditional VaR Estimation

These methods accommodate serial dependence in financial returns.

(1) J. P. Morgan

The VaR at 5% is computed by inverting a Gaussian distribution with conditional mean zero and variance equal to an estimated conditional variance of returns. The conditional variance is estimated from a conditionally Gaussian IGARCH-type model of volatility σ_t^2, called the Exponentially Weighted Moving Average, where $\sigma_t^2 = \theta\sigma_{t-1}^2 + (1-\theta)x_{t-1}^2$, and parameter θ is arbitrarily fixed at 0.94 for any portfolio [4].

(2) CaViar [5]

The CaViar model is an autoregressive specification of the conditional quantile. The model is estimated independently for each value of α, and is nonparametric in that respect.

Table 1: Computation of the VaR.

	Parametric	Semi-parametric	Nonparametric
i.i.d.	Gaussian approach	tail model building (advisory firms)	historical simulation (regulators)
Serial Dependence	IGARCH by J. P. Morgan DAQ		CaViar

(3) Dynamic Additive Quantile (DAQ) [6]

This is a parametric, dynamic factor model of the conditional quantile function.

Table 1 summarizes all the aforementioned methods.

This paper is intended to fill in the empty cell in Table 1 by extending the tail model building method to the conditional Value-at-Risk. To do that, we introduce a parametric pseudomodel of the conditional portfolio return distribution that is assumed valid in a neighbourhood of the VaR of interest. Next, we estimate locally the pseudodensity, and use this result for calculating the conditional VaRs in the tail.

The local nonparametric approach appears preferable to the fully parametric approaches for two reasons. First, the nonparametric methods are too sensitive to specification errors. Second, even if the theoretical rate of convergence appears to be smaller than that of a fully parametric method (under the assumption of no specification error in the latter one), the estimator proposed in this paper is based on a local approximation of the density in a neighborhood where more observations are available than at the quantile of interest.

The paper is organized as follows. Section 2 presents the local estimation of a probability density function from a misspecified parametric model. By applying this technique to a Gaussian pseudomodel, we derive the local drift and local volatility, which can be used as inputs in expression (1.2). In Section 3, the new method is used to compute the intraday conditional Value-at-Risk for portfolios of two stocks traded on the Toronto Stock Exchange. Next, the performance of the new method of VaR computation is compared to other methods, such as the historical simulation, Gaussian variance-covariance method, J. P. Morgan IGARCH, and ARCH-based VaR estimation in Monte Carlo experiments. Section 4 discusses the asymptotic properties of the new nonparametric estimator of the log-density derivatives. Section 5 concludes the paper. The proofs are gathered in Appendices.

2. Local Analysis of the Marginal Density Function

The local analysis of a marginal density function is based on a family of pseudodensities. Among these, we define the pseudodensity, which is locally the closest to the true density. Next, we define the estimators of the local pseudodensity, and show the specific results obtained for a Gaussian family of pseudodensities.

2.1. Local Pseudodensity

Let us consider a univariate or multivariate random variable Y, with unknown density f_0, and a parametric multivariate family of densities $\mathcal{F} = \{f(y,\theta),\ \theta \in \Theta\}$, called the family of pseudodensities where the parameter set $\Theta \subset \mathbb{R}^p$. This family is generally misspecified. Our method consists in finding the pseudodensity $f(y;\theta_0^*)$, which is locally the closest to the true density. To do that we look for the local pseudo-true value of parameter θ.

In the first step, let us assume that variable Y is univariate and consider an approximation on an interval $A = [c-h, c+h]$, centered at some value c of variable Y. The pseudodensity is derived by optimizing the Kullback-Leibler criterion evaluated from the pseudo and true densities truncated over A. The pseudo-true value of θ is

$$\begin{aligned}
\tilde{\theta}_{c,h} &= \operatorname*{Argmax}_{\theta}\Big[E_0\left[1_{c-h<Y<c+h}\log f(Y;\theta)\right] \\
&\qquad -E_0\left[1_{c-h<Y<c+h}\right]\log\int 1_{c-h<y<c+h}\, f(y;\theta)dy\Big] \\
&= \operatorname*{Argmax}_{\theta}\ E_0\left[\frac{1}{2h}1_{c-h<Y<c+h}\ \log f(Y;\theta)\right] \\
&\qquad -E_0\left[\frac{1}{2h}1_{c-h<Y<c+h}\right]\log\int\frac{1}{2h}1_{c-h<y<c+h}\, f(y;\theta)dy,
\end{aligned} \tag{2.1}$$

where E_0 denotes the expectation taken with respect to the true probability density function (pdf, henceforth) f_0. The pseudo-true value depends on the pseudofamily, the true pdf, the bandwidth, and the location c. The above formula can be equivalently rewritten in terms of the uniform kernel $K(u) = (1/2)1_{[-1,1]}(u)$. This leads to the following extended definition of the pseudo-true value of the parameter which is valid for vector Y of any dimension d, kernel K, bandwidth h, and location c:

$$\begin{aligned}
\tilde{\theta}_{c,h} &= \operatorname*{Argmax}_{\theta} E_0\left[\frac{1}{h^d}K\left(\frac{Y-c}{h}\right)\log f(Y;\theta)\right] \\
&\qquad -E_0\left[\frac{1}{h^d}K\left(\frac{Y-c}{h}\right)\right]\log\int\frac{1}{h^d}K\left(\frac{y-c}{h}\right)f(y;\theta)dy.
\end{aligned} \tag{2.2}$$

Let us examine the behavior of the pseudo-true value when the bandwidth tends to zero.

Definition 2.1. (i) The *local parameter function (l.p.f.)* is the limit of $\tilde{\theta}_{c,h}$ when h tends to zero, given by

$$\tilde{\theta}(c, f_0) = \lim_{h\to 0}\tilde{\theta}_{c,h}, \tag{2.3}$$

when this limit exists.

(ii) The local pseudodensity is $f[y;\tilde{\theta}(c, f_0)]$.

The local parameter function provides the set of local pseudo-true values indexed by c, while the local pseudodensity approximates the true pdf in a neighborhood of c. Let us now discuss some properties of the l.p.f.

Proposition 2.2. *Let one assume the following:*

(A.1) *There exists a unique solution to the objective function maximized in* (2.2) *for any h, and the limit $\widetilde{\theta}(c, f_0)$ exists.*

(A.2) *The kernel K is continuous on $\mathbb{R}^d$, of order 2, such that $\int K(u)du = 1$, $\int uK(u)du = 0$, $\int uu'K(u)du = \eta^2$, positive definite.*

(A.3) *The density functions $f(y, \theta)$ and $f_0(y)$ are positive and third-order differentiable with respect to y.*

(A.4) $\dim\theta = p \geq d$*, and, for any c in the support of f_0,*

$$\left\{ \frac{\partial \log f(c;\theta)}{\partial y}, \theta \in \Theta \right\} = \mathbb{R}^d. \tag{2.4}$$

(A.5) *For h small and any c, the following integrals exist: $\int K(u) \log f(c+uh;\theta)\, f_0(c+uh)du$, $\int K(u) f_0(c+uh)du$, $\int K(u) f(c+uh;\theta)du$, and are twice differentiable under the integral sign with respect to h.*

Then, the local parameter function is a solution of the following system of equations:

$$\frac{\partial \log f\left[c; \widetilde{\theta}(c; f_0)\right]}{\partial y} = \frac{\partial \log f_0(c)}{\partial y}, \quad \forall c. \tag{2.5}$$

Proof. See Appendix A. □

The first-order conditions in Proposition 2.2 show that functions $f[y, \widetilde{\theta}(c, f_0)]$ and $f_0(c)$ have the same derivatives at c. When p is strictly larger than d, the first-order conditions are not sufficient to characterize the l.p.f.

Assumption (A.1) is a local identification condition of parameter θ. As shown in the application given later in the text, it is verified to hold for standard pseudofamilies of densities such as the Gaussian, where $\widetilde{\theta}(c, f_0)$ has a closed form. (The existence of a limit $\widetilde{\theta}(c, f_0)$ is assumed for expository purpose. However, the main result concerning the first-order conditions is easily extended to the case when $\widetilde{\theta}_{c,h}$ exists, with a compact parameter set Θ. The proof in Appendix A shows that, even if the $\lim_{h\to 0}\widetilde{\theta}(c, f_0)$ does not exist, we get $\lim_{h\to 0}(\partial \log f[c, \widetilde{\theta}_{c,h}]/\partial y) = \partial \log f_0(c)/\partial y$, $\forall c$. This condition would be sufficient to define a local approximation to the log-derivative of the density.)

It is known that a distribution is characterized by the log-derivative of its density due to the unit mass restriction. This implies the following corollary.

Corollary 2.3. *The local parameter function characterizes the true distribution.*

2.2. Estimation of the Local Parameter Function and of the Log-Density Derivative

Suppose that $y_1, \ldots, y_T$ are observations on a strictly stationary process (Y_t) of dimension d. Let us denote by f_0 the true marginal density of Y_t and by $\{f(y : \theta), \theta \in \Theta\}$ a (misspecified) pseudoparametric family used to approximate f_0. We now consider the l.p.f. characterization of f_0, and introduce nonparametric estimators of the l.p.f. and of the marginal density.

The estimator of the l.p.f. is obtained from formula (2.2), where the theoretical expectations are replaced by their empirical counterparts:

$$\tilde{\theta}_T(c) = \underset{\theta}{\operatorname{Argmax}}\left[\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right)\log f(y_t;\theta) - \sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right)\log\int\frac{1}{h^d}K\left(\frac{y-c}{h}\right)f(y;\theta)dy\right]. \tag{2.6}$$

The above estimator depends on the selected kernel and bandwidth. This estimator allows us to derive from Proposition 2.2 a new nonparametric consistent estimator of the log-density derivative defined as

$$\frac{\partial\log\hat{f}_T(c)}{\partial y} = \frac{\partial\log f\left[c;\tilde{\theta}_T(c)\right]}{\partial y}. \tag{2.7}$$

The asymptotic properties of the estimators of the l.p.f. and log-density derivatives are discussed in Section 4, for the exactly identified case $p = d$. In that case, $\tilde{\theta}_T(c)$ is characterized by the system of first-order conditions (2.7).

The quantity $f[c,\tilde{\theta}_T(c)]$ is generally a nonconsistent estimator of the density $f_0(c)$ at c (see, e.g., [7] for a discussion of such a bias in an analogous framework). However, a consistent estimator of the log-density (and thus of the density itself, obtained as the exponential function of the log-density) is derived directly by integrating the estimated log-density derivatives under the unit mass restriction. This offers a correction for the bias, and is an alternative to including additional terms in the objective function (see, e.g., [7, 8]).

2.3. Gaussian Pseudomodel

A Gaussian family is a natural choice of pseudomodel for local analysis, as the true density is locally characterized by a local mean and a local variance-covariance matrix. Below, we provide an interpretation of the Gaussian local density approximation. Next, we consider a Gaussian pseudomodel parametrized by the mean only, and show the relationship between the l.p.f. estimator and two well-known nonparametric estimators of regression and density, respectively.

(i) Interpretation

For a Gaussian pseudomodel indexed by mean m and variance Σ, we have

$$\frac{\partial\log f(y;m,\Sigma)}{\partial y} = \Sigma^{-1}(y-m). \tag{2.8}$$

Thus, the approximation associated with a Gaussian pseudofamily is the standard one, where the partial derivatives of the log-density are replaced by a family of hyperplanes parallel to the tangent hyperplanes. These tangent hyperplanes are not independently defined, due to the Schwartz equality

$$\frac{\partial^2 \log f(y)}{\partial y_i \partial y_j} = \frac{\partial^2 \log f(y)}{\partial y_j \partial y_i}; \quad \forall i \neq j. \tag{2.9}$$

The Schwartz equalities are automatically satisfied by the approximated densities because of the symmetry of matrix Σ^{-1}.

(ii) Gaussian Pseudomodel Parametrized by the Mean and Gaussian Kernel

Let us consider a Gaussian kernel: $K(\cdot) = \phi(\cdot)$ of dimension d, where ϕ denotes the pdf of the standard Normal $N(0, Id)$.

Proposition 2.4. *The l.p.f. estimator for a Gaussian pseudomodel parametrized by the mean and with a Gaussian kernel can be written as*

$$\tilde{\theta}_T(c) = c + \frac{1+h^2}{h^2}[\tilde{m}_T(c) - c] = c + \left(1+h^2\right)\frac{\partial \log}{\partial c}\tilde{f}_T(c), \tag{2.10}$$

where

$$\tilde{m}_T(c) = \frac{\left(\sum_{t=1}^{T}(1/h^d)\phi((y_t - c)/h)y_t\right)}{\left(\sum_{t=1}^{T}(1/h^d)\phi((y_t - c)/h)\right)} \tag{2.11}$$

is the Nadaraya-Watson estimator of the conditional mean $m(c) = E[Y \mid Y = c] = c$, and

$$\tilde{f}_T(c) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}\phi\left(\frac{y_t - c}{h}\right) \tag{2.12}$$

is the Gaussian kernel estimator of the unknown value of the true marginal pdf at c.

Proof. See Appendix B. □

In this special case, the asymptotic properties of $\tilde{\theta}_T(c)$ follow directly from the asymptotic properties of $\tilde{f}_T(c)$ and $\partial\tilde{f}_T(c)/\partial c$ [9]. In particular, $\tilde{\theta}_T(c)$ converges to $c + \partial \log f_0(c)/\partial y$, when T and h tend to infinity and zero, respectively, with $Th^{d+2} \to 0$.

Alternatively, the asymptotic behavior can be inferred from the Nadaraya-Watson estimator [10, 11] in the degenerate case when the regressor and the regressand are identical. Section 5 will show that similar relationships are asymptotically valid for non-Gaussian pseudofamilies.

2.4. Pseudodensity over a Tail Interval

Instead of using the local parameter function and calibrating the pseudodensity locally about a value, one could calibrate the pseudodensity over an interval in the tail. (We thank an anonymous referee for this suggestion.) More precisely, we could define a pseudo-true parameter value

$$\theta^*(c, f_0) = \underset{\theta}{\operatorname{Argmax}} E_0 \left\{ 1_{Y>c} \log \left[\frac{f(y;\theta)}{S(c,\theta)} \right] \right\}, \tag{2.13}$$

where S denotes the survival function, and consider an approximation of the true distribution over a tail interval $f[y;\theta^*(c, f_0)]$, for $y > c$. From a theoretical point of view, this approach can be criticized as it provides different approximations of $f_0(y)$ depending on the selected value of c, $c < y$.

3. From Marginal to Conditional Analysis

Section 2 described the local approach to marginal density estimation. Let us now show the passage from the marginal to conditional density analysis and the application to the conditional VaR.

3.1. *General Approach to VaR Computation*

The VaR analysis concerns the future return on a given portfolio. Let x_t denote the return on that portfolio at date t. In practice, the prediction of x_t is based on a few summary statistics computed from past observations, such as a lagged portfolio return, realized market volatility, or realized idiosyncratic volatility in a previous period. The application of our method consists in approximating locally the joint density of series $y_t = (y'_{1t}, y'_{2t})'$, whose component y_{1t} is x_t, and component y_{2t} contains the summary statistics, denoted by z_{t-1}. Next, from the marginal density of y_t, that is, the joint density of y_{1t} and y_{2t}, we derive the conditional density of y_{1t} given y_{2t}, and the conditional VaR.

The joint density is approximated locally about c which is a vector of two components, $c = (c'_1, c'_2)'$. The first component c_1 is a tail value of portfolio returns, such as the 5% quantile of the historical distribution of portfolio returns, for example, if the conditional VaR at $\alpha < 5\%$ needs to be found. The second component c_2 is the value of the conditioning set, which is fixed, for example, at the last observed value of the summary statistics in $y_{2t} = z_{t-1}$. Due to the difference in interpretation, the bandwidths for c_1 and c_2 need to be different.

The approach above does not suffer from the curse of dimensionality. Indeed, in practice, y_1 is univariate, and the number of summary statistics is small (often less than 3), while the number of observations is sufficiently large (250 per year) for a daily VaR.

3.2. *Gaussian Pseudofamily*

When the pseudofamily is Gaussian, the local approximation of the density of y_t is characterized by the local mean and variance-covariance matrix. For $y_t = (y'_{1t}, y'_{2t})'$, these moments are decomposed by blocks as follows:

$$\mu(c) = \begin{pmatrix} \mu_1(c) \\ \mu_2(c) \end{pmatrix}, \qquad \Sigma(c) = \begin{pmatrix} \Sigma_{11}(c) & \Sigma_{12}(c) \\ \Sigma_{21}(c) & \Sigma_{22}(c) \end{pmatrix}. \tag{3.1}$$

The local conditional first and second-order moments are functions of these joint moments:

$$\mu_{1|2}(c) = \mu_1(c) - \Sigma_{12}(c)\Sigma_{22}^{-1}(c)\mu_2(c), \tag{3.2}$$

$$\Sigma_{1|2}(c) = \Sigma_{11}(c) - \Sigma_{12}(c)\Sigma_{22}^{-1}(c)\Sigma_{21}(c). \tag{3.3}$$

When $y_{1t} = x_t$ is univariate, these local conditional moments can be used as inputs in the basic Gaussian VaR formula (1.2).

The method is convenient for practitioners, as it suggests them to keep using the misspecified Gaussian VaR formula. The only modifications are the inputs, which become the local conditional mean and variance in the tail that are easy to calculate given the closed-form expressions given above.

Even though the theoretical approach is nonparametric, its practical implementation is semi-parametric. This is because, once an appropriate location c has been selected, the local pseudodensity estimated at c is used to calculate any VaR in the tail. Therefore, the procedure can be viewed as a model building method, in which the two benchmark loss probabilities are arbitrarily close. As compared with other model building approaches, it allows for choosing a location c with more data-points in its neighborhood than the quantile of interest.

4. Application to Value-at-Risk

The nonparametric feature of our localized approach requires the availability of a sufficient number of observations in a neighborhood of the selected c. This requirement is easily satisfied when high-frequency data are used and an intraday VaR is computed. We first consider an application of this type. It is followed by a Monte-Carlo study, which provides information on the properties of the estimator when the number of observations is about 200, which is the sample size used in practice for computing the daily VaR.

4.1. Comparative Study of Portfolios

We apply the local conditional mean and variance approach to intraday data on financial returns and calculate the intraday Value-at-Risk. The financial motivation for intraday risk analysis is that internal control of the trading desks and portfolio management is carried out continuously by banks, due to the use of algorithmic trading that implements automatic portfolio management, based on high-frequency data. Also, the BCBS in [2, page 3], suggests that a weakness of the current (daily) risk measure is that it is based on the end-of-day positions, and disregards the intraday trading risk. It is known that intraday stock price variation can be often as high as the variation of the market closure prices over 5 to 6 consecutive days.

Our analysis concerns two stocks traded on the Toronto Stock Exchange: the Bank of Montreal (BMO) and the Royal Bank (ROY) from October 1st to October 31, 1998, and all portfolios with nonnegative allocations in these two stocks. This approach under the no-short-sell constraint will suffice to show that allocations of the least risky portfolios differ, depending on the method of VaR computation.

From the tick-by-tick data, we select stock prices at a sampling interval of two minutes, and compute the two minute returns $x_t = (x_{1t}, x_{2t})'$. The data contain a large proportion of zero price movements, which are not deleted from the sample, because the current portfolio values have to be computed from the most recent trading prices.

The BMO and ROY sample consists of 5220 observations on both returns from October 1 to October 31, 1998. The series have equal means of zero. The standard deviations are 0.0015 and 0.0012 for BMO and ROY, respectively. To detect the presence of fat tails, we calculate the kurtosis, which is 5.98 for BMO and 3.91 for ROY, and total range, which is 0.0207 for BMO and 0.0162 for ROY. The total range is approximately 50 (for BMO) and 20 (for ROY) times greater than the interquartile range, equal to 0.0007 in both samples.

The objective is to compute the VaR for any portfolio that contains these two assets. Therefore, $y_t = (y_{1t},\ y_{2t})$ has two components; each of which is a bivariate vector. We are interested in finding a local Gaussian approximation of the conditional distribution of $y_{1t} = x_t$ given $y_{2t} = x_{t-1}$ in a neighborhood of values $c_1 = (c_{11}, c_{12})$ of x_t and $c_2 = (c_{21},\ c_{22})$ of x_{t-1} (which does not mean that the conditional distribution itself is Gaussian) . We fix $c_{21} = c_{22} = 0$. Because a zero return is generally due to nontrading, by conditioning on zero past returns, we investigate the occurrence of extreme price variations after a non-trading period. As a significant proportion of returns is equal to zero, we eliminate smoothing with respect to these conditioning values in our application.

The local conditional mean and variance estimators were computed from formulae (3.2)-(3.3) for $c_{11} = 0.00188$ and $c_{12} = 0.00154$, which are the 90% upper percentiles of the sample distribution of each return on the dates preceded by zero returns. The bandwidth for x_t was fixed at $h = 0.001$, proportionally to the difference between the 10% and 1% quantiles. The estimates are

$$\begin{gathered}\mu_1 = -6.54\,10^{-3}, \qquad \mu_2 = -0.48\,10^{-3}, \\ \sigma_{11} = 10.2\,10^{-6}, \qquad \sigma_{22} = 1.33\,10^{-6}, \qquad \rho = \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} = -0.0034.\end{gathered} \tag{4.1}$$

They can be compared to the global conditional moments of the returns, which are the moments computed from the whole sample, $\overline{\mu} = E(x_t \mid x_{t-1} = 0)$, $\overline{\Sigma} = V(x_t \mid x_{t-1} = 0)$. Their estimates are

$$\begin{gathered}\overline{\mu}_1 = -2.057\,10^{-5}, \qquad \overline{\mu}_2 = -1.359\,10^{-4}, \\ \overline{\sigma}_{11} = 2.347\,10^{-6}, \qquad \overline{\sigma}_{22} = 1.846\,10^{-6}, \qquad \overline{\rho} = \frac{\overline{\sigma}_{12}}{\overline{\sigma}_{11}\overline{\sigma}_{22}} = 0.0976.\end{gathered} \tag{4.2}$$

As the conditional distribution of x_t given $x_{t-1} = 0$ has a sharp peak at zero, it comes as no surprise that the global conditional moments estimators based on the whole sample lead to smaller Values-at-Risk than the localized ones. More precisely, for loss probability 5% and a portfolio with allocations $a, 1-a$, $0 \leq a \leq 1$, in the two assets, the Gaussian VaR is given by

$$\operatorname{VaR}(5\%, a) = -(a, 1-a)\mu + 1.64\left[(a, 1-a)\Sigma(a, 1-a)'\right]^{1/2}, \tag{4.3}$$

and determines the required capital reserve for loss probability 5%. Figure 1 presents the Values-at-Risk computed from the localized and unlocalized conditional moments, for any admissible portfolios of nonnegative allocations. The proportion a invested in the BMO is measured on the horizontal axis.

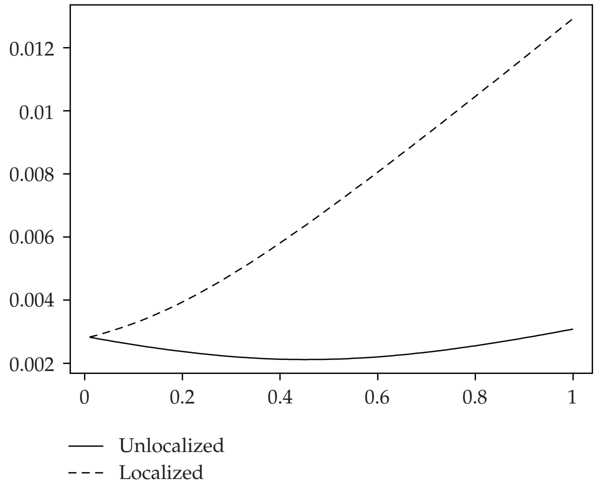


Figure 1: Localized and Unlocalized VaRs.

As expected, the localized VaR lies far above the unlocalized one. This means that the localized VaR implies a larger required capital reserve. We also note that, under the unlocalized VaR, the least risky portfolio contains equal allocations in both assets. In contrast, the localized measure suggests to invest the whole portfolio in a single asset to avoid extreme risks (under the no-short-sell constraint).

4.2. Monte-Carlo Study

The previous application was based on a quite large number of data (more than 5000) on trades in October 1998 and risk level of 5%. It is natural to assess the performance of the new method in comparison to other methods of VaR computation, for smaller samples, such as 200 (resp. 400) observations that correspond to one year (resp., two years) of daily returns and for a smaller risk level of 1%.

A univariate series of 1000 simulated portfolio returns is generated from an ARCH(1) model, with a double exponential (Laplace) error distribution. More precisely, the model is

$$x_t = \left(0.4 + 0.95x_{t-1}^2\right)^{1/2} u_t, \tag{4.4}$$

where the errors u_t are i.i.d. with pdf

$$g(u) = \frac{1}{2}\exp(-|u|). \tag{4.5}$$

The error distribution has exponential tails that are slightly heavier than the tails of a Gaussian distribution. The data generating process are assumed to be unknown to the person who estimates the VaR. In practice, that person will apply a method based on a misspecified model (such as the i.i.d. Gaussian model of returns in the Gaussian variance-covariance method or the IGARCH model of squared returns by J. P. Morgan with an ad-hoc fixed parameter 0.94). Such a procedure leads to either biased, or inefficient estimators of the VaR level.

The following methods of VaR computation at risk level of 1% are compared. Methods 1 to 4 are based on standard routines used in banks, while method 5 is the one proposed in this paper.

(1) The historical simulation based on a rolling window of 200 observations. We will see later (Figure 2) that this approach results in heavy smoothing with respect to time. A larger bandwidth would entail even more smoothing.

(2) The Gaussian variance-covariance approach based on the same window.

(3) The IGARCH-based method by J. P. Morgan:

$$\widehat{\mathrm{VaR}}_t = -\Phi^{-1}(1\%)0.06 \sum_{h=0}^{\infty} (0.94)^h x_{t-h}^2. \tag{4.6}$$

(4) Two conditional ARCH-based procedures that consist of the following steps. First, we consider a subset of observations to estimate an ARCH(1) model:

$$x_t = \left(a_0 + a_1 x_{t-1}^2\right)^{1/2} v_t, \tag{4.7}$$

where v_t are i.i.d. with an unknown distribution. First, the parameters a_0 and a_1 are estimated by the quasi-maximum likelihood, and the residuals are computed. From the residuals we infer the empirical 1% quantile $\widehat{q}$, say. The VaR is computed as $\widehat{\mathrm{VaR}}_t = -\left(\widehat{a}_0 + \widehat{a}_1 x_t^2\right)^{1/2} \widehat{q}$. We observe that the ARCH parameter estimators are very inaccurate, which is due to the exponential tails of the error distribution. Two subsets of data were used to estimate the ARCH parameters and the 1%-quantile. The estimator values based on a sample of 200 observations are $\widehat{a}_0 = 8.01$, $\widehat{a}_1 = 0.17$, and $\widehat{q} = -3.85$. The estimator values based on a sample of 800 observations are $\widehat{a}_0 = 4.12$, $\widehat{a}_1 = 0.56$, and $\widehat{q} = -2.78$. We find that the ratios $\widehat{a}_1/\widehat{a}_0$ are quite far from the true value 0.95/0.4 used to generate the data, which is likely due to fat tails.

(5) Localized VaR.

We use a Gaussian pseudofamily, a Gaussian kernel, and two different bandwidths for the current and lagged value of returns, respectively. The bandwidths were set proportional to the difference between the 10% and 1% quantiles, and the bandwidth for the lagged return is 4 times the bandwidth for the current return. Their values are 1.16 and 4.64, respectively. We use a Gaussian kernel (resp., a simple bandwidth) instead of an optimal kernel (resp., an optimal bandwidth) for the sake of robustness. Indeed, an optimal approach may not be sufficiently robust for fixing the required capital. Threshold c is set equal to the 3%-quantile of the marginal empirical distribution. The localized VaR's are computed by rolling with a window of 400 observations.

For each method, Figures 2, 3, 4, 5, 6 and 7 report the evolution of the true VaR corresponding to the data generating model along with the evolution of the estimated VaR. For clarity, only 200 data points are plotted.

The true VaR features persistence and admits extreme values. The rolling methods such as the historical simulation and variance-covariance method produce stepwise patterns of VaR, as already noted, for example, by Hull and White [12]. These patterns result from the i.i.d. assumption that underlies the computations. The J. P. Morgan IGARCH approach creates spurious long memory in the estimated VaR and is not capable to recover the dynamics of the true VaR series. The comparison of the two ARCH-based VaR's shows that the estimated

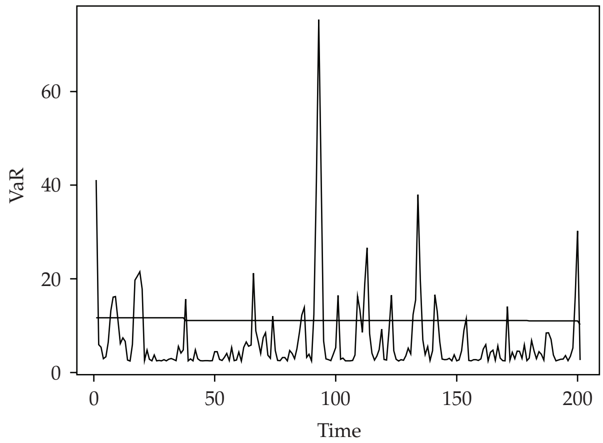


Figure 2: True and Historical Simulation-Based VaRs.

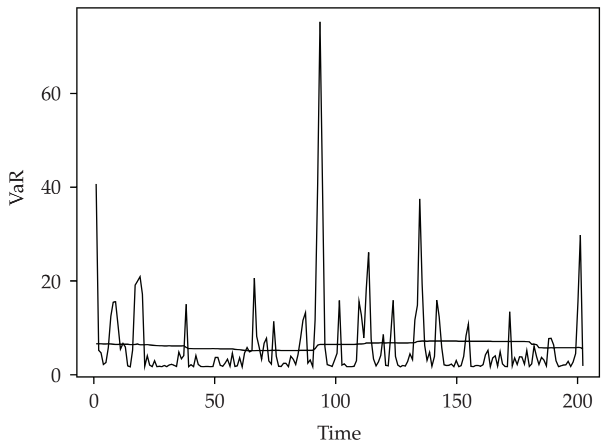


Figure 3: True and Variance-Covariance Based VaRs.

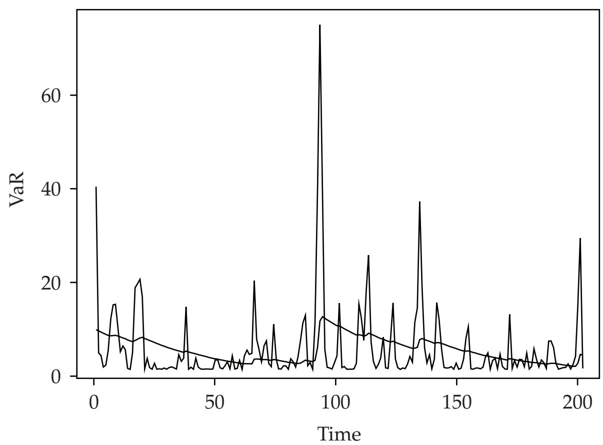


Figure 4: True and IGARCH-Based VaR.

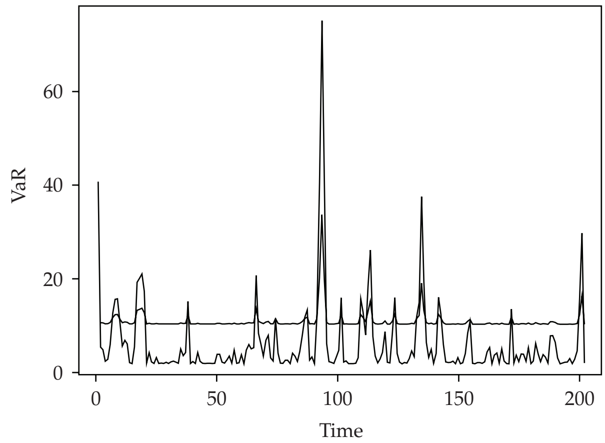


Figure 5: True and ARCH-Based VaR (200 obs).

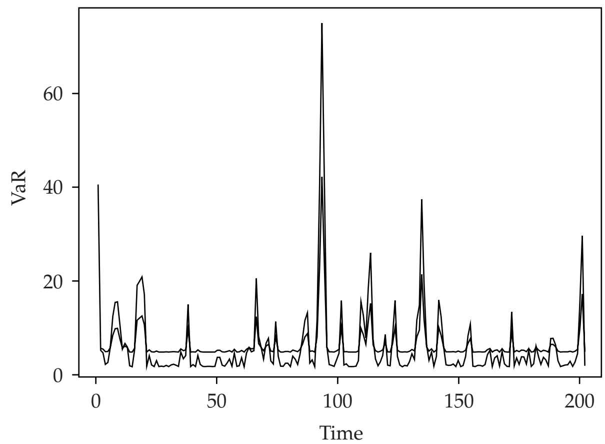


Figure 6: True and ARCH-Based VaR (800 obs).

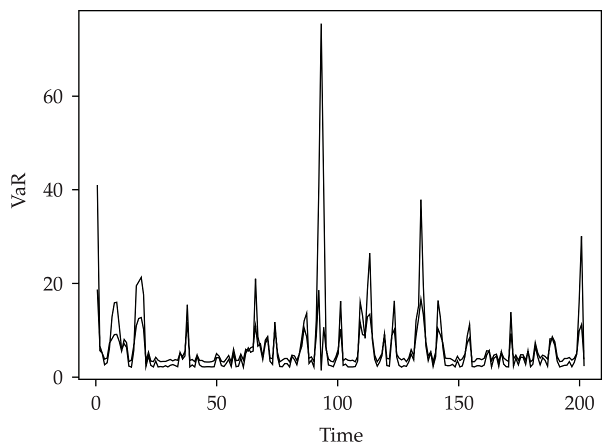


Figure 7: True and Localized VaR.

paths strongly depend on the estimated ARCH coefficients. When the estimators are based on 200 observations, we observe excess smoothing. When the estimators are based on 800 observations, the model is able to recover the general pattern, but overestimates the VaR when it is small and underestimates the VaR when it is large. The outcomes of the localized VaR method are similar to the second ARCH model, with a weaker tendency to overestimate the VaR when it is small.

The comparison of the different approaches shows the good mimicking properties of the ARCH-based methods and of the localized VaR. However, these methods need also to be compared with respect to their tractability. It is important to note that the ARCH parameters were estimated only once and were kept fixed for future VaR computations. The approach would become very time consuming if the ARCH model was reestimated at each point in time. In contrast, it is very easy to regularly update the localized VaR.

5. Properties of the Estimator of the Local Parameter Function

5.1. Asymptotic Properties

In this section, we discuss the asymptotic properties of the local pseudomaximum likelihood estimator under the following strict stationarity assumption.

Assumption 5.1. The process $Y = (Y_t)$ is strictly stationary, with marginal pdf f_0.

Let us note that the strict stationarity assumption is compatible with nonlinear dynamics, such as in the ARCH-GARCH models, stochastic volatility models, and so forth, All proofs are gathered in Appendices.

The asymptotic properties of the local P. M. L. estimator of θ are derived along the following lines. First, we find the asymptotic equivalents of the objective function and estimator, that depend only on a limited number of kernel estimators. Next, we derive the properties of the local P. M. L. estimator from the properties of these basic kernel estimators. As the set of assumptions for the existence and asymptotic normality of the basic kernel estimators for multivariates dependent observations can be found in the literature (see the study by Bosq in [13]), we only list in detail the additional assumptions that are necessary to satisfy the asymptotic equivalence. The results are derived under the assumption that θ is exactly identified (see Assumptions 5.2 and 5.3). (In the overidentified case $p > d$, the asymptotic analysis can be performed by considering the terms of order h^3, h^4 in the expansion of the objective function (see Appendix A), which is out of the scope of this paper.)

Let us introduce the additional assumptions.

Assumption 5.2. The parameter set Θ is a compact set and $p = d$.

Assumption 5.3. (i) There exists a unique solution $\widetilde{\theta}(c; f_0)$ of the system of equations:

$$\frac{\partial \log f(c;\theta)}{\partial y} = \frac{\partial \log f_0(c)}{\partial y}, \tag{5.1}$$

and this solution belongs to the interior of Θ.

(ii) The matrix $\partial^2 \log f[c, \widetilde{\theta}(c, f_0)]/\partial\theta\partial y'$ is nonsingular.

Assumption 5.4. The following kernel estimators are strongly consistent:

(i) $(1/T)\sum_{t=1}^{T}(1/h^d)K((y_t-c)/h) \overset{a.s.}{\to} f_0(c)$,

(ii) $(1/T)\sum_{t=1}^{T}(1/h^d)K((y_t-c)/h)((y_t-c)/h)((y_t-c)/h)' \overset{a.s.}{\to} \eta^2 f_0(c)$,

(iii) $(1/Th)\sum_{t=1}^{T}(1/h^d)K((y_t-c)/h)((y_t-c)/h) \overset{a.s.}{\to} f_0(c)\eta^2(\partial \log f_0(c)/\partial y)$.

Assumption 5.5. In any neighbourhood of θ, the third-order derivatives $\partial^3 \log f(y;\theta)/\partial y_i \partial y_j \partial y_k, i, j, k$ varying, are dominated by a function $a(y)$ such that $\|y\|^3 a(y)$ is integrable.

Proposition 5.6. *The local pseudomaximum likelihood estimator* $\widetilde{\theta}_T(c)$ *exists and is strongly consistent for the local parameter function* $\widetilde{\theta}(c; f_0)$ *under Assumptions 5.1–5.5.*

Proof. See Appendix C. □

It is possible to replace the set of Assumptions 5.4 by sufficient assumptions concerning directly the kernel, the true density function f_0, the bandwidth h, and the Y process. In particular it is common to assume that the process Y is geometrically strong mixing, and that $h \to 0, Th^d/(\log T)^2 \to +\infty$, when T tends to infinity (see [13–15]).

Proposition 5.7. *Under Assumptions 5.1–5.5 the local pseudomaximum likelihood estimator is asymptotically equivalent to the solution* $\widetilde{\widetilde{\theta}}_T(c)$ *of the equation:*

$$\frac{\partial \log f\left[c; \widetilde{\widetilde{\theta}}_T(c)\right]}{\partial y} = \left(\eta^2\right)^{-1}\frac{1}{h^2}\left(\widetilde{m}_T(c)-c\right), \tag{5.2}$$

where:

$$\widetilde{m}_T(c) = \frac{\sum_{t=1}^{T} K((y_t-c)/h)y_t}{\sum_{t=1}^{T} K((y_t-c)/h)} \tag{5.3}$$

is the Nadaraya-Watson estimator of $m(c) = E(Y \mid Y = c) = c$ *based on the kernel* K.

Proof. See Appendix D. □

Therefore the asymptotic distribution of $\widetilde{\theta}_T(c)$ may be derived from the properties of $\widetilde{m}_T(c) - c$, which are the properties of the Nadaraya-Watson estimator in the degenerate case when the regressand and the regressor are identical. Under standard regularity conditions [13], the numerator and denominator of $1/h^2(\widetilde{m}_T(c) - c)$ have the following asymptotic properties.

Assumption 5.8. If $T \to \infty$, $h \to 0$, $Th^{d+2} \to \infty$, and $Th^{d+4} \to 0$, we have the limiting distribution

$$\begin{pmatrix} \sqrt{Th^{d+2}}\left[\dfrac{1}{Th^{d+2}}\displaystyle\sum_{t=1}^{T} K\left(\dfrac{y_t - c}{h}\right)(y_t - c) - \eta^2 \dfrac{\partial f_0(c)}{\partial y}\right] \\ \sqrt{Th^d}\left[\dfrac{1}{Th^d}\displaystyle\sum_{t=1}^{T} K\left(\dfrac{y_t - c}{h}\right) - f_0(c)\right] \end{pmatrix} \xrightarrow{d} N\left(0, f_0(c) \begin{bmatrix} \int uu'K^2(u)du & \int uK^2(u)du \\ \int u'K^2(u)du & \int K^2(u)du \end{bmatrix}\right). \tag{5.4}$$

The formulas of the first- and second-order asymptotic moments are easy to verify (see Appendix E). (Assumption 5.8 is implied by sufficient conditions concerning the kernel, the process... (see, [13]). In particular it requires some conditions on the multivariate distribution of the process such as $\sup_{t_1<t_2} \|f_{t_1,t_2} - f \otimes f\|_\infty < \infty$, where f_{t_1,t_2} denotes the joint p.d.f. of (Y_{t_1}, Y_{t_2}) and $f \otimes f$ the associated product of marginal distributions, and $\sup_{t_1<t_2<t_3<t_4} \|f_{t_1,t_2,t_3,t_4}\|_\infty < \infty$, where f_{t_1,t_2,t_3,t_4} denotes the joint p.d.f of $(Y_{t_1}, Y_{t_2}, Y_{t_3}, Y_{t_4})$.) Note that the rate of convergence of the numerator is slower than the rate of convergence of the denominator since we study a degenerate case, when the Nadaraya-Watson estimator is applied to a regression with the regressor equal to the regressand.

We deduce that the asymptotic distribution of

$$\sqrt{Th^{d+2}}\left(\frac{1}{h^2}[\tilde{m}_T(c) - c] - \eta^2 \frac{\partial \log f_0(c)}{\partial y}\right) \tag{5.5}$$

is equal to the asymptotic distribution of

$$\sqrt{Th^{d+2}} \frac{1}{f_0(c)}\left(\frac{1}{Th^{d+2}}\sum_{t=1}^{T} K\left(\frac{y_t - c}{h}\right)(y_t - c) - \eta^2 \frac{\partial f_0(c)}{\partial y}\right), \tag{5.6}$$

which is $N[0, (1/f_0(c)) \int uu'K^2(u)du]$.

By the δ-method we find the asymptotic distribution of the local pseudomaximum likelihood estimator and the asymptotic distribution of the log-derivative of the true p.d.f..

Proposition 5.9. *Under Assumptions 5.1–5.8 one has the following.*

(i)

$$\sqrt{Th^{d+2}}\left(\frac{\partial \log f\left(c; \tilde{\tilde{\theta}}_T(c)\right)}{\partial y} - \frac{\partial \log f_0(c)}{\partial y}\right) \xrightarrow{d} N\left[0, \frac{(\eta^2)^{-1}}{f_0(c)} \int uu'K^2(u)du\left[\eta^2\right]^{-1}\right]. \tag{5.7}$$

(ii)

$$\sqrt{Th^{d+2}}\left(\tilde{\tilde{\theta}}_T(c)-\tilde{\theta}(c;f_0)\right)\xrightarrow{d}N\left[0,\left[\frac{\partial^2\log f\left[c;\tilde{\theta}(c;f_o)\right]}{\partial\theta\partial y'}\right]^{-1}\frac{\left[\eta^2\right]^{-1}}{f_0(c)}\right.$$
$$\left.\times\int uu'K^2(u)du\left[\eta^2\right]^{-1}\left[\frac{\partial^2\log f\left[c;\tilde{\theta}(c;f_0)\right]}{\partial y\partial\theta'}\right]^{-1}\right] \tag{5.8}$$

The first-order asymptotic properties of the estimator of the log-derivative of the density function do not depend on the pseudofamily, whereas the value of the estimator does. (It is beyond the scope of this paper to discuss the effect of the pseudofamily when dimension p is strictly larger than d. Nevertheless, by analogy to the literature on local estimation of nonparametric regression and density functions (see, e.g., the discussion in [7]), we expect that the finite sample bias in the associated estimator of the density will diminish when the pseudofamily is enlarged, that is, when the dimension of the pseudoparameter vector increases.) For a univariate proces (y_t), the functional estimator of the log-derivative $\partial\log f_0(c)/\partial y$ may be compared to the standard estimator

$$\frac{\partial\log\hat{f}_0(c)}{\partial y}=\frac{\sum_{t=1}^{T}(1/h)\dot{K}\left((y_t-c)/h\right)}{\sum_{t=1}^{T}(1/h)K\left((y_t-c)/h\right)}, \tag{5.9}$$

where $\dot{K}$ is the derivative of the kernel of the standard estimator. The standard estimator has a rate of convergence equal to that of the estimator introduced in this paper and the following asymptotic distribution:

$$\sqrt{Th^3}\left(\frac{\partial\log\hat{f}_0(c)}{\partial y}-\frac{\partial\log f_0(c)}{\partial y}\right)\xrightarrow{d}N\left[0,\frac{1}{\eta^4 f_0(c)}\int\dot{K}(u)^2du\right]. \tag{5.10}$$

The asymptotic distributions of the two estimators of the log-derivative of the density function are in general different, except when $|dK(u)/du|=|uK(u)|$, which, in particular, arises when the kernel is Gaussian. In such a case the asymptotic distributions of the estimators are identical.

5.2. Asymptotic versus Finite Sample Properties

In kernel-based estimation methods, the asymptotic distributions of estimators do not depend on serial dependence and are computed as if the data were i.i.d. However, serial dependence affects the finite sample properties of estimators and the accuracy of the theoretical approximation. Pritsker [16] (see also work by Conley et al. in [17]) illustrates this point by considering the finite sample properties of Ait-Sahalia's test of continuous time model of the short-term interest rate [18] in an application to data generated by the Vasicek model.

The impact of serial correlation depends on the parameter of interest, in particular on whether this parameter characterizes the marginal or the conditional density. This problem is not specific to the kernel-based approaches, but arises also in other methods such as the OLS. To see that, consider a simple autoregressive model $y_t = \rho y_{t-1} + \epsilon_t$, where ϵ_t is $\mathrm{IIN}(0, \sigma^2)$. The expected value of y_t is commonly estimated from the empirical mean $\widehat{m} = \overline{y}_T$ that has asymptotic variance $V\widehat{m} \approx (\eta^2/T)(1/(1-\rho)-1)$, where $\eta^2 = Vy_t = \sigma^2/(1-\rho^2)$. In contrast, the autoregressive coefficient is estimated by $\widehat{\rho} = \sum_t y_t y_{t-1} / \sum_t y_{t-1}^2$ and has asymptotic variance $V\widehat{\rho} \approx (1-\rho^2)/T$.

If serial dependence is disregarded, both estimators $\widehat{m}$ and $\widehat{\rho}$ have similar asymptotic efficiencies that are η^2/T and $1/T$, respectively. However, when ρ tends to one while η^2 remains fixed, the variance of $\widehat{m}$ tends to infinity whereas the variance of $\widehat{\rho}$ tends to zero. This simple example shows that omission of serial dependence does not have the same effect on the marginal parameters as opposed to the conditional ones. Problems considered by Conley et al. [17] or Pritsker [16] concern the marginal (long run) distribution of y_t, while our application is focused on a conditional parameter, which is the conditional VaR. This parameter is derived from the analysis of the joint pdf $f(y_t, y_{t-1})$ as in the previous example $\widehat{\rho}$ was derived from the bivariate vector $((1/T)\sum_t y_t y_{t-1}, (1/T)\sum_t y_{t-1}^2)$. Due to cointegration between y_t and y_{t-1} in the case of extreme persistence, we can reasonably expect that the estimator of the conditional VaR has good finite sample properties, even when the point estimators $\widehat{f}(y_t, y_{t-1})$ do not. The example shows that in finite sample the properties of the estimator of a conditional parameter can be even better than those derived under the i.i.d. assumption.

6. Conclusions

This paper introduces a local likelihood method of VaR computation for univariate or multivariate data on portfolio returns. Our approach relies on a local approximation of the unknown density of returns by means of a misspecified model. The method allows us to estimate locally the conditional density of returns, and to find the local conditional moments, such as a tail mean and tail variance. For a Gaussian pseudofamily, these tail moments can replace the global moments in the standard Gaussian formula used for computing the VaR's. Therefore, our method based on the Gaussian pseudofamily is convenient for practitioners, as it justifies computing the VaR from the standard Gaussian formula, although with a different input, which accommodates both the thick tails and path dependence of financial returns. The Monte-Carlo experiments indicate that tail-adjusted VaRs are more accurate than other VaR approximations used in the industry.

Appendices

A. Proof of Proposition 2.2

Let us derive the expansion of the objective function

$$\begin{aligned} A_h(\theta) = {} & E_0\left[\frac{1}{h^d} K\left(\frac{Y-c}{h}\right) \log f(Y;\theta)\right] \\ & - E_0\left[\frac{1}{h^d} K\left(\frac{Y-c}{h}\right)\right] \log \int \frac{1}{h^d} K\left(\frac{y-c}{h}\right) f(y;\theta)\, dy, \end{aligned} \tag{A.1}$$

when h approaches zero. By using the equivalence (see Assumption (A.1))

$$\int \frac{1}{h^d} K\left(\frac{y-c}{h}\right) g(y) dy = g(c) + \frac{h^2}{2} \operatorname{Tr}\left[\frac{\partial^2 g(c)}{\partial y \partial y'} \eta^2\right] + o\left(h^2\right), \tag{A.2}$$

where Tr is the trace operator, we find that

$$\begin{aligned}
A_h(\theta) &= f_0(c) \log f(c;\theta) + \frac{h^2}{2} \operatorname{Tr}\left\{\frac{\partial^2}{\partial y \partial y'} \left[f_0(c) \log f(c;\theta)\right] \eta^2\right\} \\
&\quad - \left[f_0(c) + \frac{h^2}{2} \operatorname{Tr}\left(\eta^2 \frac{\partial^2 f_0(c)}{\partial y \partial y'}\right)\right] \log\left\{f(c;\theta) + \frac{h^2}{2} \operatorname{Tr}\left[\eta^2 \frac{\partial^2 f(c;\theta)}{\partial y \partial y'}\right]\right\} + o\left(h^2\right) \\
&= -\frac{h^2}{2} \frac{\partial \log f_0(c)}{\partial y'} \eta^2 \frac{\partial \log f_0(c)}{\partial y} \\
&\quad + \frac{h^2}{2} \left[\frac{\partial \log f(c;\theta)}{\partial y'} - \frac{\partial \log f_0(c)}{\partial y'}\right] \eta^2 \left[\frac{\partial \log f(c;\theta)}{\partial y} - \frac{\partial \log f_0(c)}{\partial y}\right] + o\left(h^2\right).
\end{aligned} \tag{A.3}$$

The result follows.

The expansion above provides a local interpretation of the asymptotic objective function at order h^2 as a distance between the first-order derivatives of the logarithms of the pseudo and true pdf's. In this respect the asymptotic objective function clearly differs from the objective function proposed by Hjort and Jones [7], whose expansion defines an l^2-distance between the true and pseudo pdfs.

B. Proof of Proposition 2.4

For a Gaussian kernel $K(\cdot) = \phi(\cdot)$ of dimension d, we get

$$\begin{aligned}
\int \frac{1}{h^d} \phi\left(\frac{y-c}{h}\right) f(y;\theta) dy &= \int \frac{1}{h^d} \phi\left(\frac{c-y}{h}\right) \phi(y-\theta) dy \\
&= \frac{1}{(1+h^2)^{d/2}} \phi\left(\frac{c-\theta}{\sqrt{1+h^2}}\right).
\end{aligned} \tag{B.1}$$

We have

$$\begin{aligned}\tilde{\theta}_T(c) = \operatorname*{Argmax}_{\theta} \Bigg\{ &\sum_{t=1}^{T} \frac{1}{h^d} \phi\left(\frac{y_t - c}{h}\right) \left[-\frac{d}{2} \log 2\pi - \frac{\|y_t - \theta\|^2}{2} \right] \\ &- \sum_{t=1}^{T} \frac{1}{h^d} \phi\left(\frac{y_t - c}{h}\right) \left[-\frac{d}{2} \log 2\pi - \frac{d}{2} \log \left(1 + h^2\right) - \frac{\|c - \theta\|^2}{2(1 + h^2)} \right] \Bigg\} \end{aligned} \tag{B.2}$$

$$\begin{aligned} &= \frac{1 + h^2}{h^2} \frac{\sum_{t=1}^{T} (1/h^d) \phi((y_t - c)/h) y_t}{\sum_{t=1}^{T} (1/h^d) \phi((y_t - c)/h)} - \frac{1}{h^2} c \\ &= \frac{1 + h^2}{h^2} \tilde{m}_T(c) - \frac{1}{h^2} c. \end{aligned}$$

Moreover we have

$$\begin{aligned} \tilde{\theta}_T(c) - c &= \frac{1 + h^2}{h^2} (\tilde{m}_T(c) - c) \\ &= \frac{1 + h^2}{h^2} \frac{\sum_{t=1}^{T} (1/h^d) (y_t - c)\ \phi((y_t - c)/h)}{\sum_{t=1}^{T} (1/h^d)\ \phi((y_t - c)/h)} \\ &= \left(1 + h^2\right) \frac{(\partial/\partial c) \left[\sum_{t=1}^{T} (1/h^d)\ \phi((y_t - c)/h)\right]}{\left[\sum_{t=1}^{T} (1/h^d)\ \phi((y_t - c)/h)\right]} \\ &= \left(1 + h^2\right) \frac{\partial \log}{\partial c} \tilde{f}_T(c). \end{aligned} \tag{B.3}$$

C. Consistency

Let us consider the normalized objective function

$$\begin{aligned} \tilde{A}_{T,h}(\theta) = \frac{1}{Th^2} \Bigg[&\sum_{t=1}^{T} \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \log f(y_t; \theta) \\ &- \sum_{t=1}^{T} \frac{1}{h^d} K\left(\frac{y_t - c}{h}\right) \log \int \frac{1}{h^d} K\left(\frac{y - c}{h}\right) f(y; \theta) dy \Bigg]. \end{aligned} \tag{C.1}$$

It can be written as

$$\begin{aligned}
\widetilde{A}_{T,h}(\theta) = {} & \frac{1}{Th^2}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right) \\
& \times\left[\log f(c;\theta)+\frac{\partial \log f(c;\ \theta)}{\partial y'}(y_t-c)+\frac{1}{2}\left(\frac{y_t'-c}{h}\right)\frac{\partial^2 \log f(c;\ \theta)}{\partial y \partial y'}\left(\frac{y_t-c}{h}\right)\right. \\
& \left. +\ R_1(y_t;\theta)\left\|y_t-c\right\|^3\right]-\frac{1}{Th^2}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right) \\
& \times\left[\log f(c;\theta)+\frac{h^2}{2}\frac{1}{f(c;\theta)}\operatorname{Tr}\left[\eta^2\frac{\partial^2 f(c;\theta)}{\partial y \partial y'}\right]+h^3R_2(\theta,h)\right]
\end{aligned} \tag{C.2}$$

where $R_1(y_t;\ \theta), R_2(\theta;h)$ are the residual terms in the expansion. We deduce:

$$\begin{aligned}
\widetilde{A}_{T,h}(\theta) = {} & \frac{1}{Th^2}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right)\frac{\partial \log f(c;\theta)}{\partial y'}(y_t-c) \\
& +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right)\frac{1}{2}\left(\frac{y_t-c}{h}\right)'\frac{\partial^2 \log f(c;\theta)}{\partial y \partial y'}\left(\frac{y_t-c}{h}\right) \\
& -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right)\frac{1}{f(c;\theta)}\operatorname{Tr}\left[\eta^2\frac{\partial^2 f(c;\theta)}{\partial y \partial y'}\right]+\text{residual terms.}
\end{aligned} \tag{C.3}$$

Under the assumptions of Proposition 5.7, the residual terms tend almost surely to zero, uniformly on Θ, while the main terms tend almost surely uniformly on Θ to

$$\begin{aligned}
A_\infty = {} & -\frac{\partial \log f_0(c)}{\partial y'}\eta^2\frac{\partial \log f_0(c)}{\partial y} \\
& +\frac{1}{2}\left[\frac{\partial \log f(c;\theta)}{\partial y'}-\frac{\partial \log f_0(c)}{\partial y'}\right]\eta^2\left[\frac{\partial \log f(c;\theta)}{\partial y}-\frac{\partial \log f_0(c)}{\partial y}\right],
\end{aligned} \tag{C.4}$$

which is identical to $\lim_{h\to o}(A_h(\theta)/h^2)$ (see Appendix A).

Then, by Jennrich theorem [19] and the identifiability condition, we conclude that the estimator $\widetilde{\theta}_T(c)$ exists and is strongly consistent of $\theta(c; f_0)$.

D. Asymptotic Equivalence

The main part of the objective function may also be written as

$$\begin{aligned}\tilde{A}_{T,h}(\theta) \approx & \frac{1}{Th^2}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right)\frac{\partial \log f(c;\theta)}{\partial y'}(y_t-c) \\ & -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{y_t-c}{h}\right)\frac{\partial \log f(c;\theta)}{\partial y'}\eta^2\frac{\partial \log f(c;\theta)}{\partial y}.\end{aligned} \tag{D.1}$$

We deduce that the local parameter function can be asymptotically replaced by the solution $\tilde{\tilde{\theta}}_T(c)$ of

$$\frac{\partial \log f\left(c;\tilde{\tilde{\theta}}_T(c)\right)}{\partial y} = \left(\eta^2\right)^{-1}\frac{\left(1/Th^2\right)\sum_{t=1}^{T}(1/h^d)K((y_t-c)/h)(y_t-c)}{(1/T)\sum_{t=1}^{T}(1/h^d)K((y_t-c)/h)}. \tag{D.2}$$

E. The First- and Second-Order Asymptotic Moments

Let us restrict the analysis to the numerator term $(1/Th^{d+2})\sum_{t=1}^{T}K((Y_t-c)/h)(Y_t-c)$, which implies the nonstandard rate of convergence.

(1) First-Order Moment

We get

$$\begin{aligned}E\left[\frac{1}{Th^{d+2}}\sum_{t=1}^{T}K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right] &= \frac{1}{h^{d+2}}E\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right] \\ &= \frac{1}{h^{d+2}}\int K\left(\frac{y-c}{h}\right)(y-c)f_0(y)dy \\ &= \frac{1}{h}\int K(u)\ u\ f_0(c+uh)du \\ &= \frac{1}{h}\int K(u)u\left[f_0(c)+h\frac{\partial f_0(c)}{\partial y'}u+\frac{h^2}{2}u'\frac{\partial^2 f_0(c)}{\partial y\partial y'}u+o\left(h^2\right)\right]du \\ &= \eta^2\frac{\partial f_0(c)}{\partial y}+\frac{h}{2}\int K(u)uu'\frac{\partial^2 f_0(c)}{\partial y\partial y'}udu+o(h).\end{aligned} \tag{E.1}$$

(2) Asymptotic Variance

We have

$$
\begin{aligned}
V\left[\frac{1}{Th^{d+2}}\sum_{t=1}^{T}K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right] &= \frac{1}{Th^{2d+4}}V\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right] \\
&= \frac{1}{Th^{2d+4}}\ \left\{E\left[K^2\left(\frac{Y_t-c}{h}\right)(Y_t-c)(Y_t-c)'\right]\right. \\
&\quad \left. -E\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]E\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]'\right\} \\
&= \frac{1}{Th^{d+2}}\left[\int K^2(u)uu'f_0(c+uh)du - h^{d+2}\eta^2\frac{\partial f_0(c)}{\partial y}\frac{\partial f_0(c)}{\partial y'}\eta^2\right] \\
&= \frac{1}{Th^{d+2}}f_0(c)\int uu'K^2(u)du + o\left(\frac{1}{Th^{d+2}}\right),
\end{aligned}
\tag{E.2}
$$

which provides the rate of convergence $\left(Th^{d+2}\right)^{-1/2}$ of the standard error. Moreover the second term of the bias will be negligible if $h\left(Th^{d+2}\right)^{1/2} \to 0$ or $Th^{d+4} \to 0$.

(3) Asymptotic Covariance

Finally we have also to consider:

$$
\begin{aligned}
&\operatorname{Cov}\ \left[\frac{1}{Th^{d+2}}\sum_{t=1}^{T}K\left(\frac{Y_t-c}{h}\right)(Y_t-c),\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{Y_t-c}{h}\right)\right] \\
&\quad= \frac{1}{Th^{d+3}}\operatorname{Cov}\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c),K\left(\frac{Y_t-c}{h}\right)\right] \\
&\quad= \frac{1}{Th^{d+3}}\left\{E\left[K^2\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]-E\left[K\left(\frac{Y_t-c}{h}\right)(Y_t-c)\right]EK\left(\frac{Y_t-c}{h}\right)\right\} \\
&\quad= \frac{1}{Th^{d+3}}\left\{h^2\int K^2(u)uf_0(c+uh)du+O\left(h^4\right)\right\} \\
&\quad= \frac{1}{Th^{d+1}}f_0(c)\int uK^2(u)du+o\left(\frac{1}{Th^{d+1}}\right).
\end{aligned}
\tag{E.3}
$$

Acknowledgment

The authors gratefully acknowledge financial support of NSERC Canada and of the Chair AXA/Risk Foundation: "Large Risks in Insurance".

References

[1] Basle Committee on Banking Supervision, *An Internal Model-Based Approach to Market Risk Capital Requirements*, Basle, Switzerland, 1995.

[2] Basle Committee on Banking Supervision, *Overview of the Amendment to the Capital Accord to Incorporate Market Risk*, Basle, Switzerland, 1996.

[3] C. Gourieroux and J. Jasiak, "Value-at-risk," in *Handbook of Financial Econometrics*, Y. Ait-Sahalia and L. P. Hansen, Eds., pp. 553–609, Elsevier, Amsterdam, The Netherlands, 2009.

[4] J. P. Morgan, *RiskMetrics Technical Manual*, J.P. Morgan Bank, New York, NY, USA, 1995.

[5] R. F. Engle and S. Manganelli, "CAViaR: conditional autoregressive value at risk by regression quantiles," *Journal of Business & Economic Statistics*, vol. 22, no. 4, pp. 367–381, 2004.

[6] C. Gourieroux and J. Jasiak, "Dynamic quantile models," *Journal of Econometrics*, vol. 147, no. 1, pp. 198–205, 2008.

[7] N. L. Hjort and M. C. Jones, "Locally parametric nonparametric density estimation," *The Annals of Statistics*, vol. 24, no. 4, pp. 1619–1647, 1996.

[8] C. R. Loader, "Local likelihood density estimation," *The Annals of Statistics*, vol. 24, no. 4, pp. 1602–1618, 1996.

[9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, UK, 1986.

[10] E. Nadaraya, "On estimating regression," *Theory of Probability and Its Applications*, vol. 10, pp. 186–190, 1964.

[11] G. S. Watson, "Smooth regression analysis," *Sankhyā. Series A*, vol. 26, pp. 359–372, 1964.

[12] J. Hull and A. White, "Incorporating volatility updating into the historical simulation for VaR," *The Journal of Risk*, vol. 1, pp. 5–19, 1998.

[13] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, vol. 110 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1996.

[14] G. G. Roussas, "Nonparametric estimation in mixing sequences of random variables," *Journal of Statistical Planning and Inference*, vol. 18, no. 2, pp. 135–149, 1988.

[15] G. G. Roussas, "Asymptotic normality of the kernel estimate under dependence conditions: application to hazard rate," *Journal of Statistical Planning and Inference*, vol. 25, no. 1, pp. 81–104, 1990.

[16] M. Pritsker, "Nonparametric density estimation and tests of continuous time interest rate models," *Review of Financial Studies*, vol. 11, no. 3, pp. 449–487, 1998.

[17] T. G. Conley, L. P. Hansen, and W.-F. Liu, "Bootstrapping the long run," *Macroeconomic Dynamics*, vol. 1, no. 2, pp. 279–311, 1997.

[18] Y. Aït-Sahalia, "Testing continuous-time models of the spot interest rate," *Review of Financial Studies*, vol. 9, no. 2, pp. 385–426, 1996.

[19] R. I. Jennrich, "Asymptotic properties of non-linear least squares estimators," *The Annals of Mathematical Statistics*, vol. 40, pp. 633–643, 1969.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 718905, 26 pages
doi:10.1155/2010/718905

Research Article

Zenga's New Index of Economic Inequality, Its Estimation, and an Analysis of Incomes in Italy

Francesca Greselin,[1] Leo Pasquazzi,[1] and Ričardas Zitikis[2]

[1] *Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali, Università di Milano, Bicocca 20126, Milan, Italy*
[2] *Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON, Canada N6A 5B7*

Correspondence should be addressed to Ričardas Zitikis, zitikis@stats.uwo.ca

Received 2 October 2009; Accepted 28 February 2010

Academic Editor: Madan L. Puri

Copyright © 2010 Francesca Greselin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For at least a century academics and governmental researchers have been developing measures that would aid them in understanding income distributions, their differences with respect to geographic regions, and changes over time periods. It is a fascinating area due to a number of reasons, one of them being the fact that different measures, or indices, are needed to reveal different features of income distributions. Keeping also in mind that the notions of poor and rich are relative to each other, Zenga (2007) proposed a new index of economic inequality. The index is remarkably insightful and useful, but deriving statistical inferential results has been a challenge. For example, unlike many other indices, Zenga's new index does not fall into the classes of L-, U-, and V-statistics. In this paper we derive desired statistical inferential results, explore their performance in a simulation study, and then use the results to analyze data from the Bank of Italy Survey on Household Income and Wealth (SHIW).

1. Introduction

Measuring and analyzing incomes, losses, risks, and other random outcomes, which we denote by X, has been an active and fruitful research area, particularly in the fields of econometrics and actuarial science. The Gini index is arguably the most popular measure of inequality, with a number of extensions and generalizations available in the literature. Keeping in mind that the notions of poor and rich are relative to each other, Zenga [1] constructed an index that reflects this relativity. We will next recall the definitions of the Gini and Zenga indices.

Let $F(x) = \mathbf{P}[X \leq x]$ denote the cumulative distribution function (cdf) of the random variable X, which we assume to be nonnegative throughout the paper. Let $F^{-1}(p) = \inf\{x : F(x) \geq p\}$ denote the corresponding quantile function. The Lorenz curve $L_F(p)$ is given by the formula (see [2])

$$L_F(p) = \frac{1}{\mu_F}\int_0^p F^{-1}(s)ds, \tag{1.1}$$

where $\mu_F = \mathbf{E}[X]$ is the unknown true mean of X. Certainly, from the rigorous mathematical point of view we should call $L_F(p)$ the Lorenz *function*, but this would deviate from the widely accepted usage of the term "Lorenz curve". Hence, curves and functions are viewed as synonyms throughout this paper.

The classical Gini index G_F can now be written as follows:

$$G_F = \int_0^1 \left(1 - \frac{L_F(p)}{p}\right)\psi(p)dp, \tag{1.2}$$

where $\psi(p) = 2p$. Note that $\psi(p)$ is a density function on $[0, 1]$. Given the usual econometric interpretation of the Lorenz curve [3], the function

$$G_F(p) = 1 - \frac{L_F(p)}{p}, \tag{1.3}$$

which we call the Gini curve, is a relative measure of inequality (see [4]). Indeed, $L_F(p)/p$ is the ratio between (i) the mean income of the poorest $p \times 100\%$ of the population and (ii) the mean income of the entire population: the closer to each other these two means are, the lower is the inequality.

Zenga's [1] index Z_F of inequality is defined by the formula

$$Z_F = \int_0^1 Z_F(p)dp, \tag{1.4}$$

where the Zenga curve $Z_F(p)$ is given by

$$Z_F(p) = 1 - \frac{L_F(p)}{p} \cdot \frac{1-p}{1-L_F(p)}. \tag{1.5}$$

The Zenga curve measures the inequality between (i) the poorest $p \times 100\%$ of the population and (ii) the richer remaining $(1-p) \times 100\%$ part of the population by comparing the mean incomes of these two disjoint and exhaustive subpopulations. We will elaborate on this interpretation later, in Section 5.

The Gini and Zenga indices G_F and Z_F are (weighted) averages of the Gini and Zenga curves $G_F(p)$ and $Z_F(p)$, respectively. However, while in the case of the Gini index the weight function (i.e., the density) $\psi(p) = 2p$ is employed, in the case of the Zenga index the uniform weight function $\psi(p) = 1$ is used. As a consequence, the Gini index underestimates comparisons between the very poor and the whole population, and emphasizes comparisons which involve almost identical population subgroups. From this point of view, the Zenga index is more impartial: it is based on all comparisons between complementary disjoint population subgroups and gives the same weight to each comparison. Hence, the Zenga index Z_F detects, with the same sensibility, all deviations from equality in any part of the distribution.

To illustrate the Gini curve $G_F(p)$ and its weighted version $g_F(p) = G_F(p)\psi(p)$, and to also facilitate their comparisons with the Zenga curve $Z_F(p)$, we choose the Pareto distribution

$$F(x) = 1 - \left(\frac{x_0}{x}\right)^{\theta}, \quad x \geq x_0, \tag{1.6}$$

where $x_0 > 0$ and $\theta > 0$ are parameters. Later in this paper, we will use this distribution in a simulation study, setting $x_0 = 1$ and $\theta = 2.06$. Note that when $\theta > 2$, then the second moment of the distribution is finite. The "heavy-tailed" case $1 < \theta < 2$ is also of interest, especially when modeling incomes of countries with very high economic inequality. We will provide additional details on the case in Section 5.

Note 1. Pareto distribution (1.6) is perhaps the oldest model for income distributions. It dates back to Pareto [5], and Pareto [6]. Pareto's original empirical research suggested him that the number of tax payers with income x is roughly proportional to $x^{-(\theta+1)}$, where θ is a parameter that measures inequality. For historical details on the interpretation of this parameter in the context of measuring economic inequality, we refer to Zenga [7]. We can view the parameter $x_0 > 0$ as the lowest taxable income. In addition, besides being the greatest lower bound of the distribution support, x_0 is also the scale parameter of the distribution and thus does not affect our inequality indices and curves, as we will see in formulas below.

Note 2. The Pareto distribution is positively supported, $x \geq x_0 > 0$. In real surveys, however, in addition to many positive incomes we may also observe some zero and negative incomes. This happens when evaluating net household incomes, which are the sums of payroll incomes (net wages, salaries, fringe benefits), pensions and net transfers (pensions, arrears, financial assistance, scholarships, alimony, gifts). Paid alimony and gifts are subtracted in forming the incomes. However, negative incomes usually happen in the case of very few statistical units. For example, in the 2006 Bank of Italy survey we observe only four households with nonpositive incomes, out of the total of 7,766 households. Hence, it is natural to fit the Pareto model to the positive incomes and keep in mind that we are actually dealing with a conditional distribution. If, however, it is desired to deal with negative, null, and positive incomes, then instead of the Pareto distribution we may switch to different ones, such as Dagum distributions with three or four parameters [8–10].

Corresponding to Pareto distribution (1.6), the Lorenz curve is given by the formula $L_F(p) = 1-(1-p)^{1-1/\theta}$ (see [11]), and thus the Gini curve becomes $G_F(p) = ((1-p)^{1-1/\theta} - (1-p))/p$. In Figure 1(a) we have depicted the Gini and weighted Gini curves. The corresponding Zenga curve is equal to $Z_F(p) = (1-(1-p)^{1/\theta})/p$ and is depicted in Figure 1(b), alongside the Gini curve $G_F(p)$ for an easy comparison. Figure 1(a) allows us to appreciate how the Gini weight function $\psi(p) = 2p$ disguises the high inequality between the mean income of the very poor and that of the whole population, and overemphasizes comparisons between almost

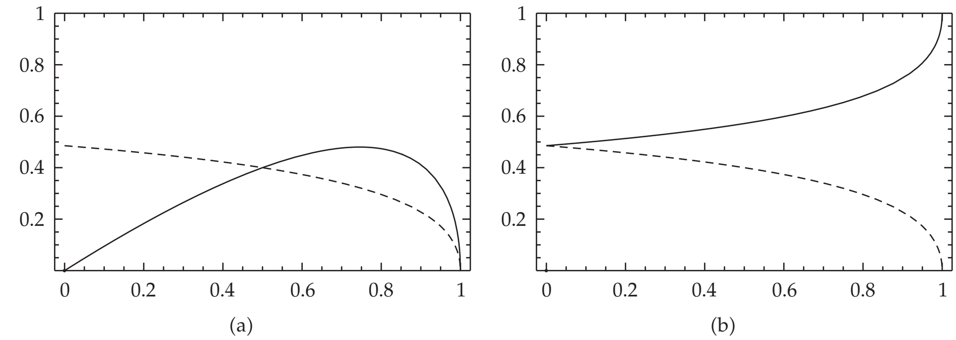


Figure 1: The Gini curve $G_F(p)$(dashed; (a) and (b)), the weighted Gini curve $g_F(p)$(solid; (a)), and the Zenga curve $Z_F(p)$(solid; (b)) in the Pareto case with $x_0 = 1$ and $\theta = 2.06$.

identical subgroups. The outcome is that the Gini index G_F underestimates inequality. In Figure 1(b) we see the difference between the Gini and Zenga inequality curves. For example, $G_F(p)$ for $p = 0.8$ yields 0.296, which tells us that the mean income of the poorest 80% of the population is 29.6% lower than the mean income of the whole population, while the corresponding ordinate of the Zenga curve is $Z_F(0.8) = 0.678$, which tells us that the mean income of the poorest 80% of the population is 67.8% lower than the mean income of the remaining (richer) part of the population.

The rest of this paper is organized as follows. In Section 2 we define two estimators of the Zenga index Z_F and develop statistical inferential results. In Section 3 we present results of a simulation study, which explores the empirical performance of two Zenga estimators, $\widehat{Z}_n$ and $\widetilde{Z}_n$, including coverage accuracy and length of several types of confidence intervals. In Section 4 we present an analysis of the the Bank of Italy Survey on Household Income and Wealth (SHIW) data. In Section 5 we further contribute to the understanding of the Zenga index Z_F by relating it to lower and upper conditional expectations, as well as to the conditional tail expectation (CTE), which has been widely used in insurance. In Section 6 we provide a theoretical background of the aforementioned two empirical Zenga estimators. In Section 7 we justify the definitions of several variance estimators as well as their uses in constructing confidence intervals. In Section 8 we prove Theorem 2.1 of Section 2, which is the main technical result of the present paper. Technical lemmas and their proofs are relegated to Section 9.

2. Estimators and Statistical Inference

Unless explicitly stated otherwise, our statistical inferential results are derived under the assumption that data are outcomes of independent and identically distributed (i.i.d.) random variables.

Hence, let $X_1, \ldots, X_n$ be independent copies of X. We use two nonparametric estimators for the Zenga index Z_F. The first one [12] is given by the formula

$$\widehat{Z}_n = 1 - \frac{1}{n}\sum_{i=1}^{n-1} \frac{i^{-1}\sum_{k=1}^{i} X_{k:n}}{(n-i)^{-1}\sum_{k=i+1}^{n} X_{k:n}}, \tag{2.1}$$

where $X_{1:n} \leq \cdots \leq X_{n:n}$ are the order statistics of $X_1, \ldots, X_n$. With $\overline{X}$ denoting the sample mean of $X_1, \ldots, X_n$, the second estimator of the Zenga index Z_F is given by the formula

$$\begin{aligned}
\widetilde{Z}_n = &-\sum_{i=2}^{n} \frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}} \log\left(\frac{i}{i-1}\right) \\
&+ \sum_{i=1}^{n-1} \left(\frac{\overline{X}}{X_{i:n}} - 1 - \frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}}\right) \log\left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^{n} X_{k:n}}\right).
\end{aligned} \tag{2.2}$$

The two estimators $\widehat{Z}_n$ and $\widetilde{Z}_n$ are asymptotically equivalent. However, despite the fact that the estimator $\widetilde{Z}_n$ is more complex, it will nevertheless be more convenient to work with when establishing asymptotic results later in this paper.

Unless explicitly stated otherwise, we assume throughout that the cdf $F(x)$ of X is a continuous function. We note that continuous cdf's are natural choices when modeling income distributions, insurance risks, and losses (see, e.g., [13]).

Theorem 2.1. *If the moment* $\mathbf{E}[X^{2+\alpha}]$ *is finite for some* $\alpha > 0$, *then one has the asymptotic representation*

$$\sqrt{n}\left(\widetilde{Z}_n - Z_F\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i) + o_{\mathbf{P}}(1), \tag{2.3}$$

where $o_{\mathbf{P}}(1)$ *denotes a random variable that converges to* 0 *in probability when* $n \to \infty$, *and*

$$h(X_i) = \int_0^\infty (\mathbf{1}\{X_i \leq x\} - F(x)) w_F(F(x))\, dx \tag{2.4}$$

with the weight function

$$w_F(t) = -\frac{1}{\mu_F} \int_0^t \left(\frac{1}{p} - 1\right) \frac{L_F(p)}{(1 - L_F(p))^2} dp + \frac{1}{\mu_F} \int_t^1 \left(\frac{1}{p} - 1\right) \frac{1}{1 - L_F(p)} dp. \tag{2.5}$$

In view of Theorem 2.1, the asymptotic distribution of $\sqrt{n}\,(\widetilde{Z}_n - Z_F)$ is centered normal with the variance $\sigma_F^2 = \mathbf{E}[h^2(X)]$, which is finite (see Theorem 7.1) and can be written as follows:

$$\sigma_F^2 = \int_0^\infty \int_0^\infty (\min\{F(x), F(y)\} - F(x)F(y)) w_F(F(x)) w_F(F(y))\, dx\, dy. \tag{2.6}$$

Alternatively,

$$\sigma_F^2 = \int_0^1 \left(\int_{[0,u)} t w_F(t)\, dF^{-1}(t) - \int_{[u,1)} (1-t) w_F(t)\, dF^{-1}(t) \right)^2 du. \tag{2.7}$$

The latter expression of σ_F^2 is particularly convenient when working with distributions for which the first derivative (when it exists) of the quantile $F^{-1}(t)$ is a relatively simple function, as is the case for a large class of distributions (see, e.g., [14]). However, irrespectively of what expression for the variance σ_F^2 we use, the variance is unknown since the cdf $F(x)$ is unknown, and thus σ_F^2 needs to be estimated empirically.

2.1. One Sample Case

Replacing the population cdf everywhere on the right-hand side of (2.6) by the empirical cdf $F_n(x) = n^{-1}\sum_{i=1}^{n}\mathbf{1}\{X_i \le x\}$, where $\mathbf{1}$ denotes the indicator function, we obtain (Theorem 7.2) the following estimator of the variance σ_F^2:

$$\begin{aligned} S_{X,n}^2 = &\sum_{k=1}^{n-1}\sum_{l=1}^{n-1}\left(\frac{\min\{k,l\}}{n} - \frac{k}{n}\frac{l}{n}\right) \\ &\times w_{X,n}\left(\frac{k}{n}\right) w_{X,n}\left(\frac{l}{n}\right)(X_{k+1:n}-X_{k:n})(X_{l+1:n}-X_{l:n}), \end{aligned} \tag{2.8}$$

where

$$w_{X,n}\left(\frac{k}{n}\right) = -\sum_{i=1}^{k} I_{X,n}(i) + \sum_{i=k+1}^{n} J_{X,n}(i) \tag{2.9}$$

with the following expressions for the summands $I_{X,n}(i)$ and $J_{X,n}(i)$: first,

$$I_{X,n}(1) = -\frac{\sum_{k=2}^{n} X_{k:n} - (n-1)X_{1:n}}{\left(\sum_{k=1}^{n} X_{k:n}\right)\left(\sum_{k=2}^{n} X_{k:n}\right)} + \frac{1}{X_{1,n}}\log\left(1 + \frac{X_{1:n}}{\sum_{k=2}^{n} X_{k:n}}\right). \tag{2.10}$$

Furthermore, for every $i = 2, \ldots, n-1$,

$$\begin{aligned} I_{X,n}(i) = {}& n\frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{\left(\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}\right)^2}\log\left(\frac{i}{i-1}\right) \\ &- \frac{\left(\sum_{k=i+1}^{n} X_{k:n} - (n-i)X_{i:n}\right)\left(\sum_{k=1}^{n} X_{k:n}\right)}{\left(\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}\right)\left(\sum_{k=i+1}^{n} X_{k:n}\right)\left(\sum_{k=i}^{n} X_{k:n}\right)} \\ &+ \left(\frac{1}{X_{i:n}} + n\frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{\left(\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}\right)^2}\right)\log\left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^{n} X_{k:n}}\right), \end{aligned} \tag{2.11}$$

$$\begin{aligned} J_{X,n}(i) = {}& \frac{n}{\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}}\log\left(\frac{i}{i-1}\right) \\ &- \frac{\sum_{k=i+1}^{n} X_{k:n} - (n-i)X_{i:n}}{X_{i:n}\left(\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}\right)}\log\left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^{n} X_{k:n}}\right). \end{aligned} \tag{2.12}$$

Finally,

$$J_{X,n}(n) = \frac{1}{X_{n,n}} \log\left(\frac{n}{n-1}\right). \tag{2.13}$$

With the just defined estimator $S^2_{X,n}$ of the variance σ^2_F, we have the asymptotic result:

$$\frac{\sqrt{n}\left(\widetilde{Z}_n - Z_F\right)}{S_{X,n}} \longrightarrow_d \mathcal{N}(0,1), \tag{2.14}$$

where $\rightarrow_d$ denotes convergence in distribution.

2.2. Two Independent Samples

We now discuss a variant of statement (2.14) in the case of two populations when samples are independent. Namely, let the random variables $X_1, \ldots, X_n \sim F$ and $Y_1, \ldots, Y_m \sim H$ be independent within and between the two samples. Just like in the case of the cdf $F(x)$, here we also assume that the cdf $H(x)$ is continuous and $\mathbf{E}[Y^{2+\alpha}] < \infty$ for some $\alpha > 0$. Furthermore, we assume that the sample sizes n and m are comparable, which means that there exists $\eta \in (0,1)$ such that

$$\frac{m}{n+m} \longrightarrow \eta \in (0,1) \tag{2.15}$$

when both n and m tend to infinity. From statement (2.3) and its counterpart for $Y_i \sim H$ we then have that the quantity $\sqrt{nm/(n+m)}\ ((\widetilde{Z}_{X,n} - \widetilde{Z}_{Y,m}) - (Z_F - Z_H))$ is asymptotically normal with mean zero and the variance $\eta\sigma^2_F + (1-\eta)\sigma^2_H$. To estimate the variances σ^2_F and σ^2_H, we use $S^2_{X,n}$ and $S^2_{Y,n}$, respectively, and obtain the following result:

$$\frac{\left(\widetilde{Z}_{X,n} - \widetilde{Z}_{Y,m}\right) - (Z_F - Z_H)}{\sqrt{(1/n)S^2_{X,n} + (1/m)S^2_{Y,m}}} \longrightarrow_d \mathcal{N}(0,1). \tag{2.16}$$

2.3. Paired Samples

Consider now the case when the two samples $X_1, \ldots, X_n \sim F$ and $Y_1, \ldots, Y_m \sim H$ are paired. Thus, we have that $m = n$, and we also have that the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed. Nothing is assumed about the joint distribution of (X, Y). As before, the cdf's $F(x)$ and $H(y)$ are continuous and both have finite moments of order $2 + \alpha$, for some $\alpha > 0$. From statement (2.3) and its analog for Y we have that $\sqrt{n}\ ((\widetilde{Z}_{X,n} - \widetilde{Z}_{Y,n}) - (Z_F - Z_H))$ is asymptotically normal with mean zero and the variance $\sigma^2_{F,H} = \mathbf{E}[(h(X) - h(Y))^2]$. The latter variance can of course be written as $\sigma^2_F - 2\mathbf{E}[h(X)h(Y)] + \sigma^2_H$. Having already constructed estimators $S^2_{X,n}$ and $S^2_{Y,n}$, we are only left to construct an estimator for $\mathbf{E}[h(X)h(Y)]$. (Note that when X and Y are independent, then $\mathbf{P}[X \leq x, Y \leq y] = F(x)H(y)$ and thus the expectation $\mathbf{E}[h(X)h(Y)]$ vanishes.) To this end, we write the equation

$$\mathbf{E}[h(X)h(Y)] = \int_0^\infty \int_0^\infty (\mathbf{P}[X \leq x, Y \leq y] - F(x)H(y)) w_F(F(x)) w_H(H(y)) dx\, dy. \tag{2.17}$$

Replacing the cdf's $F(x)$ and $H(y)$ everywhere on the right-hand side of the above equation by their respective empirical estimators $F_n(x)$ and $H_n(y)$, we have (Theorem 7.3)

$$\begin{aligned} S_{X,Y,n} = {} & \sum_{k=1}^{n-1}\sum_{l=1}^{n-1}\left(\frac{1}{n}\sum_{i=1}^{k}\mathbf{1}\{Y_{(i,n)} \leq Y_{l:n}\} - \frac{k}{n}\frac{l}{n}\right) \\ & \times w_{X,n}\left(\frac{k}{n}\right) w_{Y,n}\left(\frac{l}{n}\right)(X_{k+1:n} - X_{k:n})(Y_{l+1:n} - Y_{l:n}), \end{aligned} \tag{2.18}$$

where $Y_{(1,n)}, \ldots, Y_{(n,n)}$ are the induced (by $X_1, \ldots, X_n$) order statistics of $Y_1, \ldots, Y_n$. (Note that when $Y \equiv X$, then $Y_{(i,n)} = Y_{i:n}$ and so the sum $\sum_{i=1}^{k} \mathbf{1}\{Y_{(i,n)} \leq Y_{l:n}\}$ is equal to $\min\{k, l\}$; hence, estimator (2.18) coincides with estimator (2.8), as expected.) Consequently, $S^2_{X,n} - 2S_{X,Y,n} + S^2_{Y,n}$ is an empirical estimator of $\sigma^2_{F,H}$, and so we have that

$$\frac{\sqrt{n}\left(\widetilde{Z}_{X,n} - \widetilde{Z}_{Y,n}\right) - (Z_F - Z_H)}{\sqrt{S^2_{X,n} - 2S_{X,Y,n} + S^2_{Y,n}}} \longrightarrow_d \mathcal{N}(0,1). \tag{2.19}$$

We conclude this section with a note that the above established asymptotic results (2.14), (2.16), and (2.19) are what we typically need when dealing with two populations, or two time periods, but extensions to more populations and/or time periods would be a worthwhile contribution. For hints and references on the topic, we refer to Jones et al. [15] and Brazauskas et al. [16].

3. A Simulation Study

Here we investigate the numerical performance of the estimators $\widehat{Z}_n$ and $\widetilde{Z}_n$ by simulating data from Pareto distribution (1.6) with $x_0 = 1$ and $\theta = 2.06$. These choices give the value $Z_F = 0.6$, which is approximately seen in real income distributions. As to the (artificial) choice $x_0 = 1$, we note that since x_0 is the scale parameter in the Pareto model, the inequality indices and curves are invariant to it. Hence, all results to be reported in this section concerning the coverage accuracy and size of confidence intervals will not be affected by the choice $x_0 = 1$.

Following Davison and Hinkley [17, Chapter 5], we compute four types of confidence intervals: normal, percentile, BCa, and t-bootstrap. For normal and studentized bootstrap confidence intervals we estimate the variance using empirical influence values. For the estimator $\widetilde{Z}_n$, the influence values $h(X_i)$ are obtained from Theorem 2.1, and those for the estimator $\widehat{Z}_n$ using numerical differentiation as in Greselin and Pasquazzi [12].

In Table 1 we report coverage percentages of 10,000 confidence intervals, for each of the four types: normal, percentile, BCa, and t-bootstrap. Bootstrap-based approximations

Table 1: Coverage proportions of confidence intervals from the Pareto parent distribution with $x_0 = 1$ and $\theta = 2.06$ ($Z_F = 0.6$).

	$\widehat{Z}_n$				$\widetilde{Z}_n$			
	0.9000	0.9500	0.9750	0.9900	0.9000	0.9500	0.9750	0.9900
n	Normal confidence intervals							
200	0.7915	0.8560	0.8954	0.9281	0.7881	0.8527	0.8926	0.9266
400	0.8059	0.8705	0.9083	0.9409	0.8047	0.8693	0.9078	0.9396
800	0.8256	0.8889	0.9245	0.9514	0.8246	0.8882	0.9237	0.9503
n	Percentile confidence intervals							
200	0.7763	0.8326	0.8684	0.9002	0.7629	0.8190	0.8567	0.8892
400	0.8004	0.8543	0.8919	0.9218	0.7934	0.8487	0.8864	0.9179
800	0.8210	0.8777	0.9138	0.9415	0.8168	0.8751	0.9119	0.9393
n	BCa confidence intervals							
200	0.8082	0.8684	0.9077	0.9383	0.8054	0.867	0.9047	0.9374
400	0.8205	0.8863	0.9226	0.9531	0.8204	0.886	0.9212	0.9523
800	0.8343	0.8987	0.9331	0.9634	0.8338	0.8983	0.9323	0.9634
n	t-bootstrap confidence intervals							
200	0.8475	0.9041	0.9385	0.9658	0.8485	0.9049	0.9400	0.9675
400	0.8535	0.9124	0.9462	0.9708	0.8534	0.9120	0.9463	0.9709
800	0.8580	0.9168	0.9507	0.9758	0.8572	0.9169	0.9504	0.9754

have been obtained from 9,999 resamples of the original samples. As suggested by Efron [18], we have approximated the acceleration constant for the BCa confidence intervals by one-sixth times the standardized third moment of the influence values. In Table 2 we report summary statistics concerning the size of the 10,000 confidence intervals. As expected, the confidence intervals based on $\widehat{Z}_n$ and $\widetilde{Z}_n$ exhibit similar characteristics. We observe from Table 1 that all confidence intervals suffer from some undercoverage. For example, with sample size 800, about 97.5% of the studentized bootstrap confidence intervals with 0.99 nominal confidence level contain the true value of the Zenga index. It should be noted that the higher coverage accuracy of the studentized bootstrap confidence intervals (when compared to the other ones) comes at the cost of their larger sizes, as seen in Table 2. Some of the studentized bootstrap confidence intervals extend beyond the range $[0,1]$ of the Zenga index Z_F, but this can easily be fixed by taking the minimum between the currently recorded upper bounds and 1, which is the upper bound of the Zenga index Z_F for every cdf F. We note that for the BCa confidence intervals, the number of bootstrap replications of the original sample has to be increased beyond 9,999 if the nominal confidence level is high. Indeed, for samples of size 800, it turns out that the upper bound of 1,598 (out of 10,000) of the BCa confidence intervals based on $\widehat{Z}_n$ and with 0.99 nominal confidence level is given by the largest order statistics of the bootstrap distribution. For the confidence intervals based on $\widetilde{Z}_n$, the corresponding figure is 1,641.

4. An Analysis of Italian Income Data

In this section we use the Zenga index Z_F to analyze data from the Bank of Italy Survey on Household Income and Wealth (SHIW). The sample of the 2006 wave of this survey contains 7,768 households, with 3,957 of them being panel households. For detailed information on

Table 2: Size of the 95% asymptotic confidence intervals from the Pareto parent distribution with $x_0 = 1$ and $\theta = 2.06$ ($Z_F = 0.6$).

	$\widehat{Z}_n$			$\widetilde{Z}_n$		
	min	mean	max	min	mean	max
n	Normal confidence intervals					
200	0.0680	0.1493	0.7263	0.0674	0.1500	0.7300
400	0.0564	0.1164	0.7446	0.0563	0.1167	0.7465
800	0.0462	0.0899	0.6528	0.0462	0.0900	0.6535
n	Percentile confidence intervals					
200	0.0673	0.1456	0.4751	0.0667	0.1462	0.4782
400	0.0561	0.1140	0.4712	0.0561	0.1143	0.4721
800	0.0467	0.0883	0.4110	0.0468	0.0884	0.4117
n	BCa confidence intervals					
200	0.0668	0.1491	0.4632	0.0661	0.1497	0.4652
400	0.0561	0.1183	0.4625	0.0558	0.1186	0.4629
800	0.0465	0.0925	0.4083	0.0467	0.0927	0.4085
n	t-bootstrap confidence intervals					
200	0.0677	0.2068	2.4307	0.0680	0.2099	2.5148
400	0.0572	0.1550	2.0851	0.0573	0.1559	2.1009
800	0.0473	0.1159	2.2015	0.0474	0.1162	2.2051

the survey, we refer to the Bank of Italy [19] publication. In order to treat data correctly in the case of different household sizes, we work with equivalent incomes, which we have obtained by dividing the total household income by an equivalence coefficient, which is the sum of weights assigned to each household member. Following the modified Organization for Economic Cooperation and Development (OECD) equivalence scale, we give weight 1 to the household head, 0.5 to the other adult members of the household, and 0.3 to the members under 14 years of age. It should be noted, however, that—as is the case in many surveys concerning income analysis—households are selected using complex sampling designs. In such cases, statistical inferential results are quite complex. To alleviate the difficulties, in the present paper we follow the commonly accepted practice and treat income data as if they were i.i.d.

In Table 3 we report the values of $\widehat{Z}_n$ and $\widetilde{Z}_n$ according to the geographic area of the households, and we also report confidence intervals for Z_F based on the two estimators. We note that two households in the sample had negative incomes in 2006, and so we have not included them in our computations.

Note 3. Removing the negative incomes from our current analysis is important as otherwise we would need to develop a much more complex methodology than the one offered in this paper. To give a flavour of technical challenges, we note that the Gini index may overestimate the economic inequality when negative, zero, and positive incomes are considered. In this case the Gini index needs to be renormalized as demonstrated by, for example, Chen et al. [20]. Another way to deal with the issue would be to analyze the negative incomes and their concentration separately from the zero and positive incomes and their concentration.

Consequently, the point estimates of Z_F are based on 7,766 equivalent incomes with $\widehat{Z}_n = 0.6470$ and $\widetilde{Z}_n = 0.6464$. As pointed out by Maasoumi [21], however, good care is

Table 3: Confidence intervals for Z_F in the 2006 Italian income distribution.

	$\widehat{Z}_n$ estimator				$\widetilde{Z}_n$ estimator			
	95%		99%		95%		99%	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Northwest: $n = 1988$, $\widehat{Z}_n = 0.5953$, $\widetilde{Z}_n = 0.5948$								
Normal	0.5775	0.6144	0.5717	0.6202	0.5771	0.6138	0.5713	0.6196
Student	0.5786	0.6168	0.5737	0.6240	0.5791	0.6172	0.5748	0.6243
Percent	0.5763	0.6132	0.5710	0.6193	0.5758	0.6124	0.5706	0.6185
BCa	0.5789	0.6160	0.5741	0.6234	0.5785	0.6156	0.5738	0.6226
Northeast: $n = 1723$, $\widehat{Z}_n = 0.6108$, $\widetilde{Z}_n = 0.6108$								
Normal	0.5849	0.6393	0.5764	0.6478	0.5849	0.6393	0.5764	0.6479
Student	0.5874	0.6526	0.5796	0.6669	0.5897	0.6538	0.5836	0.6685
Percent	0.5840	0.6379	0.5773	0.6476	0.5839	0.6379	0.5772	0.6475
BCa	0.5894	0.6478	0.5841	0.6616	0.5894	0.6479	0.5842	0.6615
Center: $n = 1574$, $\widehat{Z}_n = 0.6316$, $\widetilde{Z}_n = 0.6316$								
Normal	0.5957	0.6708	0.5839	0.6826	0.5956	0.6708	0.5838	0.6827
Student	0.5991	0.6991	0.5897	0.7284	0.6036	0.7016	0.5977	0.7311
Percent	0.5948	0.6689	0.5864	0.6818	0.5948	0.6688	0.5863	0.6818
BCa	0.6024	0.6850	0.5963	0.7021	0.6024	0.6850	0.5963	0.7020
South: $n = 1620$, $\widehat{Z}_n = 0.6557$, $\widetilde{Z}_n = 0.6543$								
Normal	0.6358	0.6770	0.6293	0.6834	0.6346	0.6756	0.6282	0.6820
Student	0.6371	0.6805	0.6313	0.6902	0.6371	0.6796	0.6320	0.6900
Percent	0.6351	0.6757	0.6286	0.6828	0.6337	0.6742	0.6276	0.6812
BCa	0.6375	0.6793	0.6325	0.6888	0.6363	0.6778	0.6315	0.6873
Islands: $n = 861$, $\widehat{Z}_n = 0.6109$, $\widetilde{Z}_n = 0.6095$								
Normal	0.5918	0.6317	0.5856	0.6380	0.5910	0.6302	0.5848	0.6364
Student	0.5927	0.6339	0.5864	0.6405	0.5928	0.6330	0.5874	0.6401
Percent	0.5897	0.6297	0.5839	0.6360	0.5885	0.6275	0.5831	0.6340
BCa	0.5923	0.6324	0.5868	0.6414	0.5914	0.6307	0.5860	0.6394
Italy (entire population): $n = 7766$, $\widehat{Z}_n = 0.6470$, $\widetilde{Z}_n = 0.6464$								
Normal	0.6346	0.6596	0.6307	0.6636	0.6341	0.6591	0.6302	0.6630
Student	0.6359	0.6629	0.6327	0.6686	0.6358	0.6627	0.6331	0.6683
Percent	0.6348	0.6597	0.6314	0.6640	0.6343	0.6592	0.6309	0.6635
BCa	0.6363	0.6619	0.6334	0.6676	0.6358	0.6613	0.6330	0.6669

needed when comparing point estimates of inequality measures. Indeed, direct comparison of the point estimates corresponding to the five geographic areas of Italy would lead us to the conclusion that the inequality is higher in the central and southern areas when compared to the northern area and the islands. But as we glean from pairwise comparisons of the confidence intervals, only the differences between the estimates corresponding to the northwestern and southern areas and perhaps to the islands and the southern area may be deemed statistically significant.

Moreover, we have used the paired samples of the 2004 and 2006 incomes of the 3,957 panel households in order to check whether during this time period there was a change in inequality among households. In Table 4 we report the values of $\widetilde{Z}_n$ based on the panel households for these two years, and the 95% confidence intervals for the difference between the values of the Zenga index for the years 2006 and 2004. These computations have been

Table 4: 95% confidence intervals for the difference of the Zenga indices between 2006 and 2004 in the Italian income distribution.

	Northwest (926 pairs)		Northeast (841 pairs)		Center (831 pairs)	
	$\widetilde{Z}_n^{(2006)}$	0.5797	$\widetilde{Z}_n^{(2006)}$	0.6199	$\widetilde{Z}_n^{(2006)}$	0.5921
	$\widetilde{Z}_n^{(2004)}$	0.5955	$\widetilde{Z}_n^{(2004)}$	0.6474	$\widetilde{Z}_n^{(2004)}$	0.5766
	Difference	−0.0158	Difference	−0.0275	Difference	0.0155
	Lower	Upper	Lower	Upper	Lower	Upper
Normal	−0.0426	0.0102	−0.0573	0.0003	−0.0183	0.0514
Student	−0.0463	0.0103	−0.0591	0.0017	−0.0156	0.0644
Percent	−0.0421	0.0108	−0.0537	0.0040	−0.0183	0.0505
BCa	−0.0440	0.0087	−0.0551	0.0022	−0.0130	0.0593
	South (843 pairs)		Islands (512 pairs)		Italy (3953 pairs)	
	$\widetilde{Z}_n^{(2006)}$	0.6200	$\widetilde{Z}_n^{(2006)}$	0.6179	$\widetilde{Z}_n^{(2006)}$	0.6362
	$\widetilde{Z}_n^{(2004)}$	0.6325	$\widetilde{Z}_n^{(2004)}$	0.6239	$\widetilde{Z}_n^{(2004)}$	0.6485
	Difference	−0.0125	Difference	−0.0060	Difference	−0.0123
	Lower	Upper	Lower	Upper	Lower	Upper
Normal	−0.0372	0.0129	−0.0333	0.0213	−0.0259	0.0007
Student	−0.0365	0.0166	−0.0351	0.0222	−0.0264	0.0013
Percent	−0.0372	0.0131	−0.0333	0.0214	−0.0253	0.0016
BCa	−0.0351	0.0162	−0.0331	0.0216	−0.0255	0.0013

based on formula (2.19). Having removed the four households with at least one negative income in the paired sample, we were left with a total of 3,953 observations. We see that even though we deal with large sample sizes, the point estimates alone are not reliable. Indeed, for Italy as the whole and for all geographic areas except the center, the point estimates suggest that the Zenga index decreased from the year 2004 to 2006. However, the 95% confidence intervals in Table 4 suggest that this change is not significant.

5. An Alternative Look at the Zenga Index

In various contexts we have notions of rich and poor, large and small, risky and secure. They divide the underlying populations into two parts, which we view as subpopulations. The quantile $F^{-1}(p)$, for some $p \in (0,1)$, usually serves as a boundary separating the two subpopulations. For example, we may define rich if $X > F^{-1}(p)$ and poor if $X \le F^{-1}(p)$. Calculating the mean value of the former subpopulation gives rise to the upper conditional expectation $\mathbf{E}[X \mid X > F^{-1}(p)]$, which is known in the actuarial risk theory as the conditional tail expectation (CTE). Calculating the mean value of the latter subpopulation gives rise to the lower conditional expectation $\mathbf{E}[X \mid X \le F^{-1}(p)]$, which is known in the econometric literature as the absolute Bonferroni curve, as a function of p.

Clearly, the ratio

$$R_F(p) = \frac{\mathbf{E}[X \mid X \le F^{-1}(p)]}{\mathbf{E}[X \mid X > F^{-1}(p)]} \tag{5.1}$$

of the lower and upper conditional expectations takes on values in the interval $[0,1]$. When X is equal to any constant, which can be interpreted as the egalitarian case, then $R_F(p)$ is equal to 1. The ratio $R_F(p)$ is equal to 0 for all $p \in (0,1)$ when the lower conditional expectation is equal to 0 for all $p \in (0,1)$. This means extreme inequality in the sense that, loosely speaking, there is only one individual who possesses the entire wealth. Our wish to associate the egalitarian case with 0 and the extreme inequality with 1 leads to function $1-R_F(p)$, which coincides with the Zenga curve (1.5) when the cdf $F(x)$ is continuous. The area

$$1-\int_0^1 \frac{\mathbf{E}[X \mid X \le F^{-1}(p)]}{\mathbf{E}[X \mid X > F^{-1}(p)]}dp \tag{5.2}$$

beneath the function $1-R_F(p)$ is always in the interval $[0,1]$. Quantity (5.2) is a measure of inequality and coincides with the earlier defined Zenga index Z_F when the cdf $F(x)$ is continuous, which we assume throughout the paper.

Note that under the continuity of $F(x)$, the lower and upper conditional expectations are equal to the absolute Bonferroni curve $p^{-1}AL_F(p)$ and the dual absolute Bonferroni curve $(1-p)^{-1}(\mu_F - AL_F(p))$, respectively, where

$$AL_F(p) = \int_0^p F^{-1}(t)dt \tag{5.3}$$

is the absolute Lorenz curve. This leads us to the expression of the Zenga index Z_F given by (1.4), which we now rewrite in terms of the absolute Lorenz curve as follows:

$$Z_F = 1-\int_0^1 \left(\frac{1}{p}-1\right)\frac{AL_F(p)}{\mu_F - AL_F(p)}dp. \tag{5.4}$$

We will extensively use expression (5.4) in the proofs below. In particular, we will see in the next section that the empirical Zenga index $\widetilde{Z}_n$ is equal to Z_F with the population cdf $F(x)$ replaced by the empirical cdf $F_n(x)$.

We are now in the position to provide additional details on the earlier noted Pareto case $1 < \theta < 2$, when the Pareto distribution has finite $\mathbf{E}[X]$ but infinite $\mathbf{E}[X^2]$. The above derived asymptotic results and thus the statistical inferential theory fail in this case. The required adjustments are serious and rely on the use of the extreme value theory, instead of the classical central limit theorem (CLT). Specifically, the task can be achieved by first expressing the absolute Lorenz curve $AL_F(p)$ in terms of the conditional tail expectation (CTE):

$$\mathrm{CTE}_F(p) = \frac{1}{1-p}\int_p^1 F^{-1}(t)dt \tag{5.5}$$

using the equation $AL_F(p) = \mu_F - (1-p)\mathrm{CTE}_F(p)$. Hence, (5.4) becomes

$$Z_F = 1-\int_0^1 \frac{1}{p}\left(\frac{\mathrm{CTE}_F(0)}{\mathrm{CTE}_F(p)} - (1-p)\right)dp, \tag{5.6}$$

where $\mathrm{CTE}_F(0)$ is of course the mean μ_F. Note that replacing the population cdf $F(x)$ by its empirical counterpart $F_n(x)$ on the right-hand side of (5.6) would not lead to an estimator that would work when $\mathbf{E}[X^2] = \infty$, and thus when the Pareto parameter $1 < \theta < 2$. A solution to this problem is provided by Necir et al. [22], who have suggested a new estimator of the conditional tail expectation $\mathrm{CTE}_F(p)$ for heavy-tailed distributions. Plugging in that estimator instead of the CTE on the right-hand side of (5.6) produces an estimator of the Zenga index when $\mathbf{E}[X^2] = \infty$. Establishing asymptotic results for the new "heavy-tailed" Zenga estimator would, however, be a complex technical task, well beyond the scope of the present paper, as can be seen from the proofs of Necir et al. [22].

6. A Closer Look at the Two Zenga Estimators

Since samples are "discrete populations", (5.2) and (5.4) lead to slightly different empirical estimators of Z_F. If we choose (5.2) and replace all population-related quantities by their empirical counterparts, then we will arrive at the estimator $\widehat{Z}_n$, as seen from the proof of the following theorem.

Theorem 6.1. *The empirical Zenga index $\widehat{Z}_n$ is an empirical estimator of Z_F.*

Proof. Let U be a uniform on $[0,1]$ random variable independent of X. The cdf of $F^{-1}(U)$ is F. Hence, we have the following equations:

$$Z_F = 1 - \mathbf{E}_U\left(\frac{\mathbf{E}_X\left[X \mid X \le F^{-1}(U)\right]}{\mathbf{E}_X\left[X \mid X > F^{-1}(U)\right]}\right)$$
$$= 1 - \int_{(0,\infty)} \frac{1-F(x)}{F(x)} \frac{\mathbf{E}[X\,\mathbf{1}\{X \le x\}]}{\mathbf{E}[X\,\mathbf{1}\{X > x\}]} dF(x) \tag{6.1}$$
$$= 1 - \int_{(0,\infty)} \frac{1-F(x)}{F(x)} \frac{\int_{(0,x]} y\, dF(y)}{\int_{(x,\infty)} y\, dF(y)} dF(x).$$

Replacing every F on the right-hand side of (6.1) by F_n, we obtain

$$1 - \frac{1}{n}\sum_{i=1}^{n-1} \frac{1-F_n(X_{i:n})}{F_n(X_{i:n})} \frac{\sum_{k=1}^{n} X_{k:n}\mathbf{1}\{X_{k:n} \le X_{i:n}\}}{\sum_{k=1}^{n} X_{k:n}\mathbf{1}\{X_{k:n} > X_{i:n}\}}, \tag{6.2}$$

which simplifies to

$$1 - \frac{1}{n}\sum_{i=1}^{n-1} \frac{1-i/n}{i/n} \frac{\sum_{k=1}^{i} X_{k:n}}{\sum_{k=i+1}^{n} X_{k:n}}. \tag{6.3}$$

This is the estimator $\widehat{Z}_n$ [12]. □

If, on the other hand, we choose (5.4) as the starting point for constructing an empirical estimator for Z_F, then we first replace the quantile $F^{-1}(p)$ by its empirical counterpart

$$\begin{aligned} F_n^{-1}(p) &= \inf\{x : F_n(x) \geq p\} \\ &= X_{i:n} \quad \text{when } p \in \left(\frac{(i-1)}{n}, \frac{i}{n}\right] \end{aligned} \tag{6.4}$$

in the definition of $AL_F(p)$, which leads to the empirical absolute Lorenz curve $AL_n(p)$, and then we replace each $AL_F(p)$ on the right-hand side of (5.4) by the just constructed $AL_n(p)$. (Note that $\mu_F = AL_F(1) \approx AL_n(1) = \overline{X}$.) These considerations produce the empirical Zenga index $\widetilde{Z}_n$, as seen from the proof of the following theorem.

Theorem 6.2. *The empirical Zenga index $\widetilde{Z}_n$ is an estimator of Z_F.*

Proof. By construction, the estimator $\widetilde{Z}_n$ is given by the equation:

$$\widetilde{Z}_n = 1 - \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_n(p)}{\overline{X} - AL_n(p)} dp. \tag{6.5}$$

Hence, the proof of the lemma reduces to verifying that the right-hand sides of (2.2) and (6.5) coincide. For this, we split the integral in (6.5) into the sum of integrals over the intervals $((i-1),\ i/n)$ for $i = 1, \ldots, n$. For every $p \in ((i-1)/n, i/n)$, we have $AL_n(p) = C_{i,n} + pX_{i:n}$, where

$$C_{i,n} = \frac{1}{n}\sum_{k=1}^{i-1} X_{k:n} - \frac{i-1}{n} X_{i:n}. \tag{6.6}$$

Hence, (6.5) can be rewritten as $\widetilde{Z}_n = \sum_{i=1}^n \zeta_{i,n}$, where

$$\zeta_{i,n} = \frac{1}{n} - \int_{(i-1)/n}^{i/n} \left(\frac{1}{p} - 1\right) \frac{\Lambda_{i,n} + p}{\Psi_{i,n} - p} dp \tag{6.7}$$

with

$$\Lambda_{i,n} = \frac{C_{i,n}}{X_{i:n}}, \qquad \Psi_{i,n} = \frac{\overline{X} - C_{i,n}}{X_{i:n}}. \tag{6.8}$$

Consider first the case $i = 1$. We have $C_{1,n} = 0$ and thus $\Lambda_{1,n} = 0$, which implies

$$\zeta_{1,n} = \left(\frac{\overline{X}}{X_{1:n}} - 1\right) \log\left(1 + \frac{X_{1:n}}{\sum_{k=2}^n X_{k:n}}\right). \tag{6.9}$$

Next, we consider the case $i = n$. We have $C_{n,n} = \overline{X} - X_{n:n}$ and thus $\Psi_{n,n} = 1$, which implies

$$\zeta_{n,n} = \left(1 - \frac{\overline{X}}{X_{n:n}}\right) \log\left(\frac{n}{n-1}\right). \tag{6.10}$$

When $2 \le i \le n-1$, then the integrand in the definition of $\zeta_{i,n}$ does not have any singularity, since $\Psi_{i,n} > i/n$ due to $\sum_{k=i+1}^{n} X_{k:n} > 0$ almost surely. Hence, after simple integration we have that, for $i = 2, \ldots, n-1$,

$$\begin{aligned}\zeta_{i,n} &= \frac{(i-1)X_{i:n} - \sum_{k=1}^{i-1} X_{k:n}}{\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}} \log\left(\frac{i}{i-1}\right)\\ &\quad+ \left(\frac{\overline{X}}{X_{i:n}} - 1 + \frac{(i-1)X_{i:n} - \sum_{k=1}^{i-1} X_{k:n}}{\sum_{k=i+1}^{n} X_{k:n} + iX_{i:n}}\right) \log\left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^{n} X_{k:n}}\right).\end{aligned} \tag{6.11}$$

With the above formulas for $\zeta_{i,n}$ we easily check that the sum $\sum_{i=1}^{n} \zeta_{i,n}$ is equal to the right-hand side of (2.2). This completes the proof of Theorem 6.2. □

7. A Closer Look at the Variances

Following the formulation of Theorem 2.1 we claimed that the asymptotic distribution of $\sqrt{n}\,(\widetilde{Z}_n - Z_F)$ is centered normal with the *finite* variance $\sigma_F^2 = \mathbf{E}[h^2(X)]$. The following theorem proves this claim.

Theorem 7.1. *When $\mathbf{E}[X^{2+\alpha}] < \infty$ for some $\alpha > 0$, then $n^{-1/2}\sum_{i=1}^{n} h(X_i)$ converges in distribution to the centered normal random variable*

$$\Gamma = \int_0^\infty \mathcal{B}(F(x)) w_F(F(x))dx, \tag{7.1}$$

where $\mathcal{B}(p)$ is the Brownian bridge on the interval $[0,1]$. The variance of Γ is finite and equal to σ_F^2.

Proof. Note that $n^{-1/2}\sum_{i=1}^{n} h(X_i)$ can be written as $\int_0^\infty e_n(F(x)) w_F(F(x))dx$, where $e_n(p) = \sqrt{n}(E_n(p) - p)$ is the empirical process based on the uniform on $[0,1]$ random variables $U_i = F(X_i)$, $i = 1, \ldots, n$. We will next show that

$$\int_0^\infty e_n(F(x)) w_F(F(x))dx \longrightarrow_d \int_0^\infty \mathcal{B}(F(x)) w_F(F(x))dx. \tag{7.2}$$

The proof is based on the well-known fact that, for every $\varepsilon > 0$, the following weak convergence of stochastic processes takes place:

$$\left\{\frac{e_n(p)}{p^{1/2-\varepsilon}(1-p)^{1/2-\varepsilon}},\ 0 \le p \le 1\right\} \Longrightarrow \left\{\frac{\mathcal{B}(p)}{p^{1/2-\varepsilon}(1-p)^{1/2-\varepsilon}},\ 0 \le p \le 1\right\}. \tag{7.3}$$

Hence, in order to prove statement (7.2), we only need to check that the integral

$$\int_0^\infty F(x)^{1/2-\varepsilon}(1-F(x))^{1/2-\varepsilon} w_F(F(x))dx \tag{7.4}$$

is finite. For this, by considering, for example, the two cases $p \le 1/2$ and $p > 1/2$ separately, we first easily verify the bound $|w_F(p)| \le c + c\log(1/p) + c\log(1/(1-p))$. Hence, for every $\varepsilon > 0$, there exists a constant $c < \infty$ such that, for all $p \in (0,1)$,

$$|w_F(p)| \le \frac{c}{p^\varepsilon (1-p)^\varepsilon}. \tag{7.5}$$

Bound (7.5) implies that integral (7.4) is finite when $\int_0^\infty (1-F(x))^{1/2-2\varepsilon}dx < \infty$, which is true since the moment $\mathbf{E}[X^{2+\alpha}]$ is finite for some $\alpha > 0$ and the parameter $\varepsilon > 0$ can be chosen as small as desired. Hence, $n^{-1/2}\sum_{i=1}^n h(X_i) \to_d \Gamma$ with Γ denoting the integral on the right-hand side of statement (7.2). The random variable Γ is normal because the Brownian bridge $\mathcal{B}(p)$ is a Gaussian process. Furthermore, Γ has mean zero because $\mathcal{B}(p)$ has mean zero for every $p \in [0,1]$. The variance of Γ is equal to σ_F^2 because $\mathbf{E}[\mathcal{B}(p)\mathcal{B}(q)] = \min\{p,q\} - pq$ for all $p, q \in [0,1]$. We are left to show that $\mathbf{E}[\Gamma^2] < \infty$. For this, we write the bound:

$$\begin{aligned} \mathbf{E}\left[\Gamma^2\right] &= \int_0^\infty \int_0^\infty \mathbf{E}[\mathcal{B}(F(x))\mathcal{B}(F(y))] w_F(F(x)) w_F(F(y)) dx\, dy \\ &\le \left(\int_0^\infty \sqrt{\mathbf{E}[\mathcal{B}^2(F(x))]}\; w_F(F(x))dx\right)^2. \end{aligned} \tag{7.6}$$

Since $\mathbf{E}[\mathcal{B}^2(F(x))] = F(x)(1-F(x))$, the finiteness of the integral on the right-hand side of bound (7.6) follows from the earlier proved statement that integral (7.4) is finite. Hence, $\mathbf{E}[\Gamma^2] < \infty$ as claimed, which concludes the proof of Theorem 7.1. □

Theorem 7.2. *The empirical variance $S_{X,n}^2$ is an estimator of σ_F^2.*

Proof. We construct an empirical estimator for σ_F^2 by replacing every F on the right-hand side of (2.6) by the empirical F_n. Consequently, we replace the function $w_F(t)$ by its empirical version

$$w_{X,n}(t) = -\int_0^t \left(\frac{1}{p} - 1\right) \frac{AL_n(p)}{\left(\overline{X} - AL_n(p)\right)^2} dp + \int_t^1 \left(\frac{1}{p} - 1\right) \frac{1}{\overline{X} - AL_n(p)} dp. \tag{7.7}$$

We denote the resulting estimator of σ_F^2 by $S_{X,n}^2$. The rest of the proof consists of verifying that this estimator coincides with the one defined by (2.8). Note that $\min\{F_n(x), F_n(y)\} - F_n(x)F_n(y) = 0$ when $x \in [0, X_{1:n}) \cup [X_{n:n}, \infty)$ and/or $y \in [0, X_{1:n}) \cup [X_{n:n}, \infty)$. Hence, the just defined $S_{X,n}^2$ is equal to

$$\int_{X_{1:n}}^{X_{n:n}} \int_{X_{1:n}}^{X_{n:n}} (\min\{F_n(x), F_n(y)\} - F_n(x)F_n(y)) w_{X,n}(F_n(x)) w_{X,n}(F_n(y)) dx\, dy. \tag{7.8}$$

Since $F_n(x) = k/n$ when $x \in [X_{k:n}, X_{k+1:n})$, we therefore have that

$$\begin{aligned} S_{X,n}^2 = &\sum_{k=1}^{n-1}\sum_{l=1}^{n-1}\left(\frac{\min\{k,l\}}{n} - \frac{k}{n}\frac{l}{n}\right) \\ &\times w_{X,n}\left(\frac{k}{n}\right) w_{X,n}\left(\frac{l}{n}\right)(X_{k+1:n} - X_{k:n})(X_{l+1:n} - X_{l:n}). \end{aligned} \tag{7.9}$$

Furthermore,

$$\begin{aligned} w_{X,n}\left(\frac{k}{n}\right) &= -\int_0^{k/n}\left(\frac{1}{p} - 1\right)\frac{AL_n(p)}{\left(\overline{X} - AL_n(p)\right)^2}dp + \int_{k/n}^1\left(\frac{1}{p} - 1\right)\frac{1}{\overline{X} - AL_n(p)}dp \\ &= -\sum_{i=1}^{k} I_{X,n}(i) + \sum_{i=k+1}^{n} J_{X,n}(i), \end{aligned} \tag{7.10}$$

where, using notations (6.6) and (6.8), the summands on the right-hand side of (7.10) are

$$I_{X,n}(i) = \frac{1}{X_{i:n}}\int_{(i-1)/n}^{i/n}\left(\frac{1}{p} - 1\right)\frac{\Lambda_{i,n} + p}{\left(\Psi_{i,n} - p\right)^2}dp \tag{7.11}$$

for all $i = 1, \ldots, n-1$, and

$$J_{X,n}(i) = \frac{1}{X_{i:n}}\int_{(i-1)/n}^{i/n}\left(\frac{1}{p} - 1\right)\frac{1}{\Psi_{i,n} - p}dp \tag{7.12}$$

for all $i = 2, \ldots, n$. When $i = 1$, then $\Lambda_{i,n} = 0$. Hence, we immediately arrive at the expression for $I_{X,n}(1)$ given by (2.10). When $2 \le i \le n-1$, then

$$\begin{aligned} I_{X,n}(i) = &\frac{\Lambda_{i,n}}{X_{i:n}\Psi_{i,n}^2}\log\left(\frac{i}{i-1}\right) - \frac{(\Lambda_{i,n} + \Psi_{i,n})(\Psi_{i,n} - 1)}{nX_{i:n}\Psi_{i,n}(\Psi_{i,n} - (i-1)/n)(\Psi_{i,n} - i/n)} \\ &+ \frac{1}{X_{i:n}}\left(1 + \frac{\Lambda_{i,n}}{\Psi_{i,n}^2}\right)\log\left(\frac{\Psi_{i,n} - (i-1)/n}{\Psi_{i,n} - i/n}\right), \end{aligned} \tag{7.13}$$

and, after some algebra, we arrive at the right-hand side of (2.11). When $2 \le i \le n-1$, then we have the expression

$$J_{X,n}(i) = \frac{1}{X_{i:n}\Psi_{i,n}}\log\left(\frac{i}{i-1}\right) - \frac{1}{X_{i:n}}\left(1 - \frac{1}{\Psi_{i,n}}\right)\log\left(\frac{\Psi_{i,n} - (i-1)/n}{\Psi_{i,n} - i/n}\right), \tag{7.14}$$

which, after some algebra, becomes the expression recorded in (2.12). When $i = n$, then $\Psi_{i,n} = 1$, and so we see that $J_{X,n}(n)$ is given by (2.13). This completes the proof of Theorem 7.2. □

Theorem 7.3. *The empirical mixed moment $S_{X,Y,n}$ is an estimator of $\mathbf{E}[h(X)h(Y)]$.*

Proof. We proceed similarly to the proof of Theorem 7.2. We estimate the integrand $\mathbf{P}[X \le x, Y \le y] - F(x)H(y)$ using

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \le x, Y_i \le y\} - \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \le x\}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{Y_i \le y\}. \tag{7.15}$$

After some rearrangement of terms, estimator (7.15) becomes

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_{i:n} \le x, Y_{(i,n)} \le y\} - \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_{i:n} \le x\}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{Y_{i:n} \le y\}. \tag{7.16}$$

When $x \in [X_{k:n}, X_{k+1:n})$ and $y \in [Y_{l:n}, Y_{l+1:n})$, then estimator (7.16) is equal to $n^{-1}\sum_{i=1}^{k}\mathbf{1}\{Y_{(i,n)} \le Y_{l:n}\} - (k/n)(l/n)$, which leads us to the estimator $S_{X,Y,n}$. This completes the proof of Theorem 7.3. $\square$

8. Proof of Theorem 2.1

Throughout the proof we use the notation $AL_F^*(p)$ for the dual absolute Lorenz curve $\int_p^1 F^{-1}(t)dt$, which is equal to $\mu_F - AL_F(p)$. Likewise, we use the notation $AL_n^*(p)$ for the *empirical* dual absolute Lorenz curve.

Proof. Simple algebra gives the equations

$$\begin{aligned}\sqrt{n}\left(\tilde{Z}_n - Z_F\right) &= -\sqrt{n}\int_0^1\left(\frac{1}{p}-1\right)\left(\frac{AL_n(p)}{AL_n^*(p)} - \frac{AL_F(p)}{AL_F^*(p)}\right)dp \\ &= -\sqrt{n}\int_0^1\left(\frac{1}{p}-1\right)\frac{AL_n(p) - AL_F(p)}{AL_F^*(p)}dp \\ &\quad + \sqrt{n}\int_0^1\left(\frac{1}{p}-1\right)\frac{AL_F(p)}{AL_F^{*2}(p)}\left(AL_n^*(p) - AL_F^*(p)\right)dp \\ &\quad + O_{\mathbf{P}}(r_{n,1}) + O_{\mathbf{P}}(r_{n,2})\end{aligned} \tag{8.1}$$

with the remainder terms

$$\begin{aligned}r_{n,1} &= \sqrt{n}\int_0^1\left(\frac{1}{p}-1\right)\left(AL_n(p) - AL_F(p)\right)\left(\frac{1}{AL_n^*(p)} - \frac{1}{AL_F^*(p)}\right)dp, \\ r_{n,2} &= \sqrt{n}\int_0^1\left(\frac{1}{p}-1\right)\frac{AL_F(p)}{AL_F^*(p)}\left(AL_n^*(p) - AL_F^*(p)\right)\left(\frac{1}{AL_n^*(p)} - \frac{1}{AL_F^*(p)}\right)dp.\end{aligned} \tag{8.2}$$

We will later show (Lemmas 9.1 and 9.2) that the remainder terms $r_{n,1}$ and $r_{n,2}$ are of the order $o_{\mathbf{P}}(1)$. Hence, we now proceed with our analysis of the first two terms on the right-hand side of (8.1), for which we use the (general) Vervaat process

$$V_n(p) = \int_0^p \left(F_n^{-1}(t) - F^{-1}(t)\right)dt + \int_0^{F^{-1}(p)} (F_n(x) - F(x))dx \tag{8.3}$$

and its dual version

$$V_n^*(p) = \int_p^1 \left(F_n^{-1}(t) - F^{-1}(t)\right)dt + \int_{F^{-1}(p)}^{\infty} (F_n(x) - F(x))dx. \tag{8.4}$$

For mathematical and historical details on the Vervaat process, see Zitikis [23], Davydov and Zitikis [24], Greselin et al. [25], and references therein. Since $\int_0^1 (F_n^{-1}(t) - F^{-1}(t))dt = \overline{X} - \mu_F$ and $\int_0^\infty (F_n(x) - F(x))dx = -(\overline{X} - \mu_F)$, adding the right-hand sides of (8.3) and (8.4) gives the equation $V_n^*(p) = -V_n(p)$. Hence, whatever upper bound we have for $|V_n(p)|$, the same bound holds for $|V_n^*(p)|$. In fact, the absolute value can be dropped from $|V_n(p)|$ since $V_n(p)$ is always nonnegative. Furthermore, we know that $V_n(p)$ does not exceed $(p - F_n(F^{-1}(p)))(F_n^{-1}(p) - F^{-1}(p))$. Hence, with the notation $e_n(p) = \sqrt{n}(F_n(F^{-1}(p)) - p)$, which is the uniform on $[0,1]$ empirical process, we have that

$$\sqrt{n}\; V_n(p) \le \left|e_n(p)\right| \left|F_n^{-1}(p) - F^{-1}(p)\right|. \tag{8.5}$$

Bound (8.5) implies the following asymptotic representation for the first term on the right-hand side of (8.1):

$$\begin{aligned} &-\sqrt{n}\int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_n(p) - AL_F(p)}{AL_F^*(p)} dp \\ &\quad = \sqrt{n}\int_0^1 \left(\frac{1}{p} - 1\right) \frac{1}{AL_F^*(p)} \left(\int_0^{F^{-1}(p)} (F_n(x) - F(x))dx\right) dp + O_{\mathbf{P}}(r_{n,3}), \end{aligned} \tag{8.6}$$

where

$$r_{n,3} = \int_0^1 \left(\frac{1}{p} - 1\right) \frac{1}{AL_F^*(p)} \left|e_n(p)\right| \left|F_n^{-1}(p) - F^{-1}(p)\right| dp. \tag{8.7}$$

We will later show (Lemma 9.3) that $r_{n,3} = o_{\mathbf{P}}(1)$. Furthermore, we have the following asymptotic representation for the second term on the right-hand side of (8.1):

$$\begin{aligned} &\sqrt{n}\int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} (AL_n^*(p) - AL_F^*(p)) dp \\ &\quad = -\sqrt{n}\int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} \left(\int_{F^{-1}(p)}^{\infty} (F_n(x) - F(x))dx\right) dp + O_{\mathbf{P}}(r_{n,4}), \end{aligned} \tag{8.8}$$

where

$$r_{n,4} = \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} \left|e_n(p)\right| \left|F_n^{-1}(p) - F^{-1}(p)\right| dp. \tag{8.9}$$

We will later show (Lemma 9.4) that $r_{n,4} = o_{\mathbf{P}}(1)$. Hence, (8.1), (8.6) and (8.8) together with the aforementioned statements that $r_{n,1}, \ldots, r_{n,4}$ are of the order $o_{\mathbf{P}}(1)$ imply that

$$\begin{aligned}
\sqrt{n}\left(\widetilde{Z}_n - Z_F\right) &= \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{1}{AL_F^*(p)} \left(\int_0^{F^{-1}(p)} (F_n(x) - F(x))dx\right) dp \\
&\quad - \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} \left(\int_{F^{-1}(p)}^{\infty} (F_n(x) - F(x))dx\right) dp + o_{\mathbf{P}}(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i) + o_{\mathbf{P}}(1).
\end{aligned} \tag{8.10}$$

This completes the proof of Theorem 2.1. □

9. Negligibility of Remainder Terms

The following four lemmas establish the above noted statements that the remainder terms $r_{n,1}, \ldots, r_{n,4}$ are of the order $o_{\mathbf{P}}(1)$. In the proofs of the lemmas we will use a parameter $\delta \in (0, 1/2]$, possibly different from line to line but never depending on n. Furthermore, we will frequently use the fact that

$$\mathbf{E}[X^q] < \infty \quad \text{implies} \quad \int_0^1 \left|F_n^{-1}(t) - F^{-1}(t)\right|^q dt = o_{\mathbf{P}}(1). \tag{9.1}$$

Another technical result that we will frequently use is the fact that, for any $\varepsilon > 0$ as small as desired,

$$\sup_{x \in \mathbf{R}} \frac{\sqrt{n}|F_n(x) - F(x)|}{F(x)^{1/2-\varepsilon}(1 - F(x))^{1/2-\varepsilon}} = O_{\mathbf{P}}(1) \tag{9.2}$$

when $n \to \infty$.

Lemma 9.1. *Under the conditions of Theorem 2.1, $r_{n,1} = o_{\mathbf{P}}(1)$.*

Proof. We split the remainder term $r_{n,1} = \sqrt{n}\int_0^1 \cdots dp$ into the sum of $r_{n,1}^*(\delta) = \sqrt{n}\int_0^{1-\delta} \cdots dp$ and $r_{n,1}^{**}(\delta) = \sqrt{n}\int_{1-\delta}^1 \cdots dp$. The lemma follows if

(1) for every $\delta > 0$, the statement $r_{n,1}^*(\delta) = o_{\mathbf{P}}(1)$ holds when $n \to \infty$,

(2) $r_{n,1}^{**}(\delta) = h(\delta)O_{\mathbf{P}}(1)$ for a deterministic $h(\delta) \downarrow 0$ when $\delta \downarrow 0$, where $O_{\mathbf{P}}(1)$ does not depend on δ.

To prove part (1), we first note that when $0 < p < 1-\delta$, then $AL_F^*(p) \geq \int_{1-\delta}^1 F^{-1}(t)dt$, which is positive, and $AL_n^*(p) \geq \int_{1-\delta}^1 F^{-1}(t)dt + o_{\mathbf{P}}(1)$ due to statement (9.1) with $q = 1$. Hence, we are left to show that, when $n \to \infty$,

$$\sqrt{n}\int_0^{1-\delta} \frac{1}{p} \left|AL_n(p) - AL_F(p)\right|\left|AL_n^*(p) - AL_F^*(p)\right|dp = o_{\mathbf{P}}(1). \tag{9.3}$$

Since $AL_n^*(p) - AL_F^*(p) = (\overline{X} - \mu_F) - (AL_n(p) - AL_F(p))$, statement (9.3) follows if

$$\sqrt{n}\left|\overline{X} - \mu_F\right| \int_0^{1-\delta} \frac{1}{p}\left|AL_n(p) - AL_F(p)\right|dp = o_{\mathbf{P}}(1), \tag{9.4}$$

$$\sqrt{n}\int_0^{1-\delta} \frac{1}{p}\left|AL_n(p) - AL_F(p)\right|^2 dp = o_{\mathbf{P}}(1). \tag{9.5}$$

We have $\sqrt{n}\,|\overline{X} - \mu_F| = O_{\mathbf{P}}(1)$ and $|AL_n(p) - AL_F(p)| \leq \sqrt{p}\,(\int_0^1 |F_n^{-1}(p) - F^{-1}(p)|^2 dp)^{1/2}$. Since $\int_0^1 |F_n^{-1}(p) - F^{-1}(p)|^2 dp = o_{\mathbf{P}}(1)$ and $\int_0^{1-\delta} p^{-1}\sqrt{p}\, dp < \infty$, we have statement (9.4). To prove statement (9.5), we use bound (8.5) and reduce the proof to showing that

$$\frac{1}{\sqrt{n}}\int_0^{1-\delta} \frac{1}{p}\left|\int_0^{F^{-1}(p)} \sqrt{n}\,(F_n(x) - F(x))dx\right|^2 dp = o_{\mathbf{P}}(1), \tag{9.6}$$

$$\frac{1}{\sqrt{n}}\int_0^{1-\delta} \frac{1}{p}\,\left|e_n(p)\right|^2\left|F_n^{-1}(p) - F^{-1}(p)\right|^2 dp = o_{\mathbf{P}}(1). \tag{9.7}$$

To prove statement (9.6), we use statement (9.2) and observe that

$$\int_0^{1-\delta} \frac{1}{p}\left(\int_0^{F^{-1}(p)} F(x)^{1/2-\varepsilon}dx\right)^2 dp \leq c(F,\delta)\int_0^{1-\delta} \frac{1}{p}\, p^{1-2\varepsilon}dp < \infty. \tag{9.8}$$

To prove statement (9.7), we use the uniform on $[0,1]$ version of statement (9.2) and Hölder's inequality, and in this way reduce the proof to showing that

$$\frac{1}{\sqrt{n}}\left(\int_0^{1-\delta} \frac{1}{p^{2\varepsilon a}}dp\right)^{1/a}\left(\int_0^{1-\delta} \left|F_n^{-1}(p) - F^{-1}(p)\right|^{2b}dp\right)^{1/b} = o_{\mathbf{P}}(1) \tag{9.9}$$

for some $a, b > 1$ such that $a^{-1} + b^{-1} = 1$. We choose the parameters a and b as follows. First, since $\mathbf{E}[X^{2+\alpha}] < \infty$, we set $b = (2+\alpha)/2$. Next, we choose $\varepsilon > 0$ on the left-hand side of statement (9.9) so that $2\varepsilon a < 1$, which holds when $\varepsilon < \alpha/(4+2\alpha)$ in view of the equation $a^{-1} + b^{-1} = 1$. Hence, statement (9.9) holds and thus statement (9.7) follows. This completes the proof of part (1).

To establish part (2), we first estimate $|r_{n,1}^{**}(\delta)|$ from above using the bounds $AL_F^*(p) \geq (1-p)F^{-1}(1/2)$ and $AL_n^*(p) \geq (1-p)F_n^{-1}(1/2)$, which hold since $\delta \leq 1/2$. Hence, we have reduced our task to verifying the statement $\sqrt{n}\int_{1-\delta}^{1}|AL_n(p)-AL_F(p)|dp = h(\delta)O_{\mathbf{P}}(1)$. Using the Vervaat process $V_n(p)$ and bound (8.5), we reduce the proof of the statement to showing that the integrals

$$\int_{1-\delta}^{1}\left(\int_{0}^{F^{-1}(p)}\sqrt{n}\,|F_n(x)-F(x)|dx\right)dp, \tag{9.10}$$

$$\int_{1-\delta}^{1}|e_n(p)|\left|F_n^{-1}(p)-F^{-1}(p)\right|dp \tag{9.11}$$

are of the order $h(\delta)O_{\mathbf{P}}(1)$ with possibly different $h(\delta)\downarrow 0$ in each case. In view of statement (9.2), we have the desired statement for integral (9.10) if the quantity

$$\int_{1-\delta}^{1}\left(\int_{0}^{F^{-1}(p)}(1-F(x))^{1/2-\varepsilon}dx\right)dp \tag{9.12}$$

converges to 0 when $\delta\downarrow 0$, in which case we use it as $h(\delta)$. The inner integral of (9.12) does not exceed $\int_0^\infty(1-F(x))^{1/2-\varepsilon}dx$, which is finite for all sufficiently small $\varepsilon>0$ since $\mathbf{E}[X^{2+\alpha}]<\infty$ for some $\alpha>0$. This completes the proof that quantity (9.10) is of the order $h(\delta)O_{\mathbf{P}}(1)$. To show that quantity (9.11) is of a similar order, we use the uniform on $[0,1]$ version of statement (9.2) and reduce the task to showing that $\int_{1-\delta}^{1}|F_n^{-1}(p)-F^{-1}(p)|dp$ is of the order $h(\delta)O_{\mathbf{P}}(1)$. By the Cauchy-Bunyakowski-Schwarz inequality, we have that

$$\int_{1-\delta}^{1}\left|F_n^{-1}(p)-F^{-1}(p)\right|dp \le \sqrt{\delta}\left(\int_{0}^{1}\left|F_n^{-1}(p)-F^{-1}(p)\right|^2dp\right)^{1/2}. \tag{9.13}$$

Since $\mathbf{E}[X^2]<\infty$, we have $\int_0^1|F_n^{-1}(p)-F^{-1}(p)|^2dp = o_{\mathbf{P}}(1)$, and so setting $h(\delta)=\sqrt{\delta}$ establishes the desired asymptotic result for integral (9.11). This also completes the proof of part (2), and also of Lemma 9.1. □

Lemma 9.2. *Under the conditions of Theorem 2.1, $r_{n,2}=o_{\mathbf{P}}(1)$.*

Proof. Like in the proof of Lemma 9.1, we split the remainder term $r_{n,2}=\sqrt{n}\int_0^1\cdots dp$ into the sum of $r_{n,2}^{*}(\delta)=\sqrt{n}\int_0^{1-\delta}\cdots dp$ and $r_{n,2}^{**}(\delta)=\sqrt{n}\int_{1-\delta}^{1}\cdots dp$. To prove the lemma, we need to show the following.

(1) For every $\delta>0$, the statement $r_{n,2}^{*}(\delta)=o_{\mathbf{P}}(1)$ holds when $n\to\infty$.

(2) $r_{n,2}^{**}(\delta)=h(\delta)O_{\mathbf{P}}(1)$ for a deterministic $h(\delta)\downarrow 0$ when $\delta\downarrow 0$, where $O_{\mathbf{P}}(1)$ does not depend on δ.

To prove part (1), we first estimate $|r_{n,2}^{*}(\delta)|$ from above using the bounds $p^{-1}AL_F(p)\le F^{-1}(1-\delta)<\infty$, $AL_F^{*}(p)\ge\int_{1-\delta}^{1}F^{-1}(t)dt>0$, and $AL_n^{*}(p)\ge\int_{1-\delta}^{1}F^{-1}(t)dt+o_{\mathbf{P}}(1)$. This reduces our task to showing that, for every $\delta>0$,

$$\sqrt{n}\int_0^{1-\delta}\left|AL_n^{*}(p)-AL_F^{*}(p)\right|^2dp = o_{\mathbf{P}}(1). \tag{9.14}$$

Since $AL_n^*(p) - AL_F^*(p) = (\overline{X} - \mu_F) - (AL_n(p) - AL_F(p))$ and $\sqrt{n}\ (\overline{X} - \mu_F)^2 = o_{\mathbf{P}}(1)$, statement (9.14) follows from

$$\sqrt{n}\int_0^{1-\delta} \left|AL_n(p) - AL_F(p)\right|^2 dp = o_{\mathbf{P}}(1), \tag{9.15}$$

which is an elementary consequence of statement (9.5). This establishes part (1).

To prove part (2), we first estimate $|r_{n,2}^{**}(\delta)|$ from above using the bounds $AL_F^*(p) \geq (1-p)F^{-1}(1/2)$ and $AL_n^*(p) \geq (1-p)F_n^{-1}(1/2)$, and in this way reduce the task to showing that

$$\sqrt{n}\int_{1-\delta}^{1} \frac{1}{1-p}\left|AL_n^*(p) - AL_F^*(p)\right| dp = h(\delta)O_{\mathbf{P}}(1). \tag{9.16}$$

Using the Vervaat process, statement (9.16) follows if

$$\int_{1-\delta}^{1} \frac{1}{1-p}\left(\int_{F^{-1}(p)}^{\infty} \sqrt{n}\ |F_n(x) - F(x)| dx\right) dp = h(\delta)O_{\mathbf{P}}(1), \tag{9.17}$$

$$\int_{1-\delta}^{1} \frac{1}{1-p}\ \left|e_n(p)\right|\left|F_n^{-1}(p) - F^{-1}(p)\right| dp = h(\delta)O_{\mathbf{P}}(1) \tag{9.18}$$

with possibly different $h(\delta) \downarrow 0$ in each case. Using statement (9.2), we have that statement (9.17) holds with $h(\delta)$ defined as the integral

$$\int_{1-\delta}^{1} \frac{1}{1-p}\left(\int_{F^{-1}(p)}^{\infty} (1-F(x))^{1/2-\varepsilon} dx\right) dp, \tag{9.19}$$

which converges to 0 when $\delta \downarrow 0$ as the following argument shows. First, we write the integrand as the product of $(1-F(x))^{\varepsilon}$ and $(1-F(x))^{1/2-2\varepsilon}$. Then we estimate the first factor by $(1-p)^{\varepsilon}$. The integral $\int_0^{\infty}(1-F(x))^{1/2-2\varepsilon}dx$ is finite for all sufficiently small $\varepsilon > 0$ since $\mathbf{E}[X^{2+\alpha}] < \infty$ for some $\alpha > 0$. Since $\int_{1-\delta}^{1}(1-p)^{-1+\varepsilon}dp \downarrow 0$ when $\delta \downarrow 0$, integral (9.19) converges to 0 when $\delta \downarrow 0$. The proof of statement (9.17) is finished.

We are left to prove statement (9.18). Using the uniform on $[0,1]$ version of statement (9.2), we reduce the task to showing that

$$\int_{1-\delta}^{1} \frac{1}{(1-p)^{1/2+\varepsilon}}\left|F_n^{-1}(p) - F^{-1}(p)\right| dp = h(\delta)O_{\mathbf{P}}(1). \tag{9.20}$$

In fact, we will see below that $O_{\mathbf{P}}(1)$ can be replaced by $o_{\mathbf{P}}(1)$. Using Hölder's inequality, we have that the right-hand side of (9.20) does not exceed

$$\left(\int_{1-\delta}^{1} \frac{1}{(1-p)^{(1/2+\varepsilon)a}} dp\right)^{1/a}\left(\int_{1-\delta}^{1} \left|F_n^{-1}(p) - F^{-1}(p)\right|^b dp\right)^{1/b} \tag{9.21}$$

for some $a, b > 1$ such that $a^{-1} + b^{-1} = 1$. We choose the parameters a and b as follows. Since $\mathbf{E}[X^{2+\alpha}] < \infty$, we set $b = 2 + \alpha$, and so the right-most integral of (9.21) is of the order $o_{\mathbf{P}}(1)$. Furthermore, $a = (2 + \alpha)/(1 + \alpha) < 2$, which can be made arbitrarily close to 2 by choosing sufficiently small $\alpha > 0$. Choosing $\varepsilon > 0$ so small that $(1/2+\varepsilon)a < 1$, we have that the left-most integral in (9.21) converges to 0 when $\delta \downarrow 0$. This establishes statement (9.18) and completes the proof of Lemma 9.2. □

Lemma 9.3. *Under the conditions of Theorem 2.1,* $r_{n,3} = o_{\mathbf{P}}(1)$.

Proof. We split the remainder term $r_{n,3} = \int_0^1 \cdots dp$ into the sum of $r^*_{n,3} = \int_0^{1/2} \cdots dp$ and $r^{**}_{n,3} = \int_{1/2}^1 \cdots dp$. The lemma follows if the two summands are of the order $o_{\mathbf{P}}(1)$.

To prove $r^*_{n,3} = o_{\mathbf{P}}(1)$, we use the bound $AL^*_F(p) \ge \int_{1/2}^1 F^{-1}(p)dp$ and the uniform on $[0, 1]$ version of statement (9.2), and in this way reduce our task to showing that

$$\int_0^{1/2} \frac{1}{p^{1/2+\varepsilon}} \left| F_n^{-1}(p) - F^{-1}(p) \right| dp = o_{\mathbf{P}}(1). \tag{9.22}$$

This statement can be established following the proof of statement (9.20), with minor modifications.

To prove $r^{**}_{n,3} = o_{\mathbf{P}}(1)$, we use the bound $AL^*_F(p) \ge (1 - p)F^{-1}(1/2)$, the fact that $\sup_t|e_n(t)| = O_{\mathbf{P}}(1)$, and statement (9.1) with $q = 1$. The desired result for $r^{**}_{n,3}$ follows, which finishes the proof of Lemma 9.3. □

Lemma 9.4. *Under the conditions of Theorem 2.1,* $r_{n,4} = o_{\mathbf{P}}(1)$.

Proof. We split $r_{n,4} = \int_0^1 \cdots dp$ into the sum of $r^*_{n,4} = \int_0^{1/2} \cdots dp$ and $r^{**}_{n,4} = \int_{1/2}^1 \cdots dp$, and then show that the two summands are of the order $o_{\mathbf{P}}(1)$.

To prove $r^*_{n,4} = o_{\mathbf{P}}(1)$, we use the bounds $p^{-1}AL_F(p) \le F^{-1}(1/2) < \infty$ and $AL^*_F(p) \ge \int_{1/2}^1 F^{-1}(p)dp > 0$ together with the uniform on $[0, 1]$ version of statement (9.2). This reduces our task to showing that $\int_0^{1/2} |F_n^{-1}(p) - F^{-1}(p)|dp = o_{\mathbf{P}}(1)$, which holds due to statement (9.1) with $q = 1$.

To prove $r^{**}_{n,4} = o_{\mathbf{P}}(1)$, we use the bound $AL^*_F(p) \ge (1 - p)F^{-1}(1/2)$ and the uniform on $[0, 1]$ version of statement (9.2), and in this way reduce the proof to showing that

$$\int_{1/2}^1 \frac{1}{(1-p)^{1/2+\varepsilon}} \left| F_n^{-1}(p) - F^{-1}(p) \right| dp = o_{\mathbf{P}}(1). \tag{9.23}$$

This statement can be established following the proof of statement (9.20). The proof of Lemma 9.4 is finished. □

Acknowledgments

The authors are indebted to two anonymous referees and the editor in charge of the manuscript, Madan L. Puri, for their constructive criticism and suggestions that helped them to improve the paper. The research has been partially supported by the 2009 F.A.R. (Fondo di Ateneo per la Ricerca) at the University of Milan Bicocca, and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] M. Zenga, “Inequality curve and inequality index based on the ratios between lower and upper arithmetic means,” *Statistica & Applicazioni*, vol. 5, pp. 3–27, 2007.
[2] G. Pietra, “Delle relazioni fra indici di variabilitá, note I e II,” *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, vol. 74, pp. 775–804, 1915.
[3] M. O. Lorenz, “Methods of measuring the concentration of wealth,” *Journal of the American Statistical Association*, vol. 9, pp. 209–219, 1905.
[4] C. Gini, “Sulla misura della concentrazione e della variabilità dei caratteri,” in *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti. Anno Accademico*, vol. 48, part 2, pp. 1201–1248, Premiate Officine Grafiche Carlo Ferrari, Venezia, Italy, 1914.
[5] V. Pareto, “La legge della domanda,” *Giornale degli Economisti*, vol. 10, pp. 59–68, 1895.
[6] V. Pareto, “Ecrits sur la courve de la répartition de la richesse,” in *Complete Works of V. Pareto*, G. Busino, Ed., Librairie Droz, Genève, Switzerland, 1965.
[7] M. Zenga, “Il contributo degli italiani allo studio della concentrazione,” in *La Distribuzione Personale del Reddito: Problemi di Formazione, di Ripartizione e di Misurazione*, M. Zenga, Ed., Vita e Pensiero, Milano, Italy, 1987.
[8] C. Dagum, “A new model of personal distribution: specification and estimation,” *Economie Appliquée*, vol. 30, pp. 413–437, 1977.
[9] C. Dagum, “The generation and distribution of income. The Lorenz curve and the Gini ratio,” *Economie Appliquée*, vol. 33, pp. 327–367, 1980.
[10] C. Dagum, “A model of net wealth distribution specified for negative, null and positive wealth. A case of study: Italy,” in *Income and Wealth Distribution, Inequality and Poverty*, C. Dagum and M. Zenga, Eds., pp. 42–56, Springer, Berlin, Germany, 1990.
[11] J. L. Gastwirth, “A general definition of the Lorenz curve,” *Econometrica*, vol. 39, pp. 1037–1039, 1971.
[12] F. Greselin and L. Pasquazzi, “Asymptotic confidence intervals for a new inequality measure,” *Communications in Statistics: Simulation and Computation*, vol. 38, no. 8, pp. 1742–1756, 2009.
[13] C. Kleiber and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ, USA, 2003.
[14] Z. A. Karian and E. J. Dudewicz, *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Method*, CRC Press, Boca Raton, Fla, USA, 2000.
[15] B. L. Jones, M. L. Puri, and R. Zitikis, “Testing hypotheses about the equality of several risk measure values with applications in insurance,” *Insurance: Mathematics & Economics*, vol. 38, no. 2, pp. 253–270, 2006.
[16] V. Brazauskas, B. L. Jones, M. L. Puri, and R. Zitikis, “Nested L-statistics and their use in comparing the riskiness of portfolios,” *Scandinavian Actuarial Journal*, no. 3, pp. 162–179, 2007.
[17] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Application*, vol. 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, UK, 1997.
[18] B. Efron, “Better bootstrap confidence intervals,” *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 171–200, 1987.
[19] Bank of Italy, “Household income and wealth in 2004,” *Supplements to the Statistical Bulletin, Sample Surveys*, vol. 16, no. 7, 2006.
[20] C. N. Chen, T. W. Tsaur, and T. S. Rhai, “The Gini coefficient and negative income,” *Oxford Economic Papers*, vol. 34, pp. 473–478, 1982.
[21] E. Maasoumi, “Empirical analysis of welfare and inequality,” in *Handbook of Applied Econometrics, Volume II: Microeconomics*, M. H. Pesaran and P. Schmidt, Eds., Blackwell, Oxford, UK, 1994.
[22] A. Necir, A. Rassoul, and R. Zitikis, “Estimating the conditional tail expectation in the case of heavy-tailed losses,” *Journal of Probability and Statistics*. In press.
[23] R. Zitikis, “The Vervaat process,” in *Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997)*, B. Szyszkowicz, Ed., pp. 667–694, North-Holland, Amsterdam, The Netherlands, 1998.
[24] Y. Davydov and R. Zitikis, “Convex rearrangements of random elements,” in *Asymptotic Methods in Stochastics*, vol. 44 of *Fields Institute Communications*, pp. 141–171, American Mathematical Society, Providence, RI, USA, 2004.
[25] F. Greselin, M. L. Puri, and R. Zitikis, “L-functions, processes, and statistics in measuring economic inequality and actuarial risks,” *Statistics and Its Interface*, vol. 2, no. 2, pp. 227–245, 2009.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 214358, 17 pages
doi:10.1155/2010/214358

Research Article

Risk Navigator SRM: An Applied Risk Management Tool

Dana L. K. Hoag and Jay Parsons

Department of Agricultural and Resource Economics, Colorado State University, Campus 1172, Fort Collins, CO 80523-1172, USA

Correspondence should be addressed to Dana L. K. Hoag, dhoag@colostate.edu

Received 3 November 2009; Revised 25 January 2010; Accepted 16 March 2010

Academic Editor: Ricardas Zitikis

Copyright © 2010 D. L. K. Hoag and J. Parsons. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Risk Navigator SRM is a ten-step risk management program for agricultural producers, which is based on the strategic planning process. The ten steps are designed to integrate disparate and difficult risk management concepts into a single system that is easy to use, yet still effective. With the aid of computers and the internet, producers can work through each step toward a final comprehensive plan. The website includes 25 decision tools that help producers accomplish each step and provides links to complementary educational programs, like a national agricultural risk education library, the award-winning risk management simulation program called Ag Survivor, and a recently published book that describes the program and provides additional depth and explanations. The ten-step program has been presented in over 200 workshops with over 90 percent approval by participants. The website has averaged over 1,000 unique visitors per month from 120 countries.

1. Risk Navigator SRM: An Applied Risk Management Tool for Agriculture

Risk management technology has drastically outpaced the ability of most practitioners to adopt innovations. Studies of farmers and ranchers, for example, consistently show that they do not utilize what is available; some even seem to virtually disregard risk altogether [1, 2]. One survey of nearly 1,000 farmers for Farm Futures [3] found that only 5 percent use available tools. That survey also showed that those that did manage risk tended to be high-end managers with skills not typical of the industry. Agricultural producers were said to resist change because there are too many decisions and too little time.

Congress considered agricultural risks so important that in 1996 it created the Risk Management Agency (RMA) in the US Department of Agriculture. Its purpose is to promote, support, and regulate sound risk management solutions for agricultural

Table 1: Risk management sources and controls; see source in [6].

Risk	Defined	Sources	Management Controls
Production	Uncontrollable events such as weather, pests, or disease make yields unpredictable. Changing technology makes a manager or capital obsolescent. Inputs are unavailable or low quality	Weather, extreme temperatures, pests, disease, technology, genetics, inputs (availability, quality, price), equipment failure, labor, . . .	Diversification, insurance (crop, revenue), buildings, storage, vaccines, extra labor, production contracts (e.g., ensure input supply and quality), new technologies (e.g., automate watering)
Marketing and Price	Prices of inputs or outputs change after a producer commits to a plan of action. Price fluctuations stem from domestic and international supplies or substantial changes in demand	Product quality (genetics, disease, handling, input/feed) Product price (quality, timing, global market, weather, government policy, contracts, . . .)	Futures and options, forward contracting, retained ownership, quality controls, storage (timing), cooperatives, niche/value-added marketing, . . .
Financial	Stems form the way a business is financed Borrowed funds leverage business equity but increase business risks	Market, production, legal and human risk, interest rate changes, natural disasters (drought), land market changes, foreign exchange, loan calls, . . .	Cash reserves, equity, borrowing capacity, reducing other types of risk (production, marketing, etc.), insurance
Institutional	Government or other institutional rules, regulations, and policies effect profitability through costs or returns	Taxes, contract disputes, regulations, government policies, law suits, ambiguous and/or unwritten agreements, neighbors, environmental programs, . . .	Estate planning, tax planning, contracts, bonds (e.g., environmental liability), research and education about local laws, . . .
Human Resources	The character, health, or behavior of people introduces risk. This could include theft, illness, death in the family, loss of an employee, or a divorce for example	Ambiguous and/or unwritten agreements, poor planning, miscommunication, health or other family disasters, . . .	Family planning, including labor planning, clear contracts, training and goal setting, communication, estate planning, . . .

producers [4]. According to RMA, there are at least five major forms of agricultural risks [5]: production, market/price, financial, institutional, and human. As shown in Table 1, agricultural producers face many sources of risk and a multitude of ways to manage them across and within these five categories. Price risk, for example, can be affected by product quality, exogenous supply, and government policies; it has at least seven management options, including futures, forward pricing, and storage. RMA efforts have substantially boosted the output of risk information and education to address these risks.

While there is little doubt about its importance for decision making, the “challenge is to know how to describe, measure, and communicate risk” [7, page 4]. Consider the parallels to understanding and using probability, which is itself an important component for risk management. Myerson [8] concludes that there is a disconnection between theory and practice because formulas traditionally taught in probability courses are hard to apply to real problems. He suggests that recent advances in computer technology can help overcome these disconnects. Aven [7] goes further by suggesting that a common unifying framework is also needed. For example, websites now combine a framework and computer technology to make managing a stock portfolio relatively simple, even for those people with only a minimal understanding of price analysis. These frameworks require integrating multiple fields, like economics, finance, and statistics, and finding an acceptable balance between the precision and usability.

Agricultural economics and related fields have contributed greatly to developing innovative and effective tools for managing risk in agriculture. But, as is the case with probability, these sometimes disparate theories and concepts can be difficult to understand, which may make integration difficult. Numerous books, articles, and materials are available but are generally inaccessible except by specialists [2]. A 2007 study in Nevada, for example, showed that after six months only 50% of program participants planned to incorporate what they had learned in a risk management workshop [9]. A 2007 study [10] showed that older producers had less knowledge about risk management tools, compared to younger producers, and had less interest in learning more about them. This supports the Farm Futures survey [3] in that many producers view the human capital investment required to learn how to properly incorporate the use of risk management tools into their operation as significant.

The purpose of this manuscript is to describe a new framework for risk management called Risk Navigator SRM. The program is too involved to fully describe here, and it was prepared for education and extension programs, rather than basic research. Nevertheless, researchers might be interested in how disparate parts were integrated to strike a balance between precision and usability, with the purpose of making the components of risk management more usable through a synergistic and reinforcing framework. Precision can interfere with usability and vice versa. That balance in Risk Navigator SRM is based on interaction with producers at over 200 meetings in over a dozen states. This includes a description of supporting software tools made available by website. Our focus here is on showing how probability is integrated with other risk concepts to make risk management more accessible to producers. All ten steps of the process were fully applied to a case study, EWS Farms, which also may be of interest. EWS Farms produce primarily corn and wheat in Northeastern Colorado [6].

We proceed with a description of Risk Navigator SRM and explain the SRM process, which has ten steps. A brief summary is provided for all ten steps, but a more complete description is provided about steps viewed to be of more interest to this readership. We also provide examples of the computer tools available to help producers with each step where appropriate.

2. Risk Navigator

Risk Navigator SRM is a program developed to make risk management accessible to typical agricultural producers. The process is general enough that it can be applied to other applications, but all of the examples and tools are customized for agriculture. SRM stands for Strategic Risk Management. Strategic planning [11, 12] is an umbrella framework used to organize and integrate risk management concepts and tools for farmers and ranchers. The SRM process has been taught to hundreds of farmers and ranchers under the brand name Risk Navigator SRM or under previous incarnations branded “RightRisk”. Risk Navigator SRM is housed on a public website at http://www.RiskNavigatorSRM.com/. The website includes detailed descriptions about a ten-step strategic risk management program and 25 customized computer tools to help decision makers with each step. The tools are matched to the steps and standardized through Xcelsius flash files from Excel where possible. Some files are pdf or Excel spreadsheets. The site also includes links to complementary educational materials, such as a risk practice simulator called Ag Survivor, http://www.AgSurvivor.com/, a comprehensive book titled *Applied Risk Management in Agriculture* [6], and a general education website called RightRisk, at http://www.RightRisk.org/. The book includes a comprehensive discussion about risk management, matched to the ten steps, and includes a case study to demonstrate each step. The RightRisk.org website has publications, workshops, and access to resources such as the National Ag Risk Education Library.

3. The Strategic Risk Management Process

The framework chosen to make risk management more accessible is strategic planning [6]. The foundation of strategic planning is based on three major phases: strategic, tactical, and operational [12, 13]. The strategic phase of strategic planning is designed to set boundaries based on resources, opportunities, and threats, and to set goals. The tactical phase is designed to evaluate various alternatives for reaching the strategic vision and goals and to develop a solid plan to achieve them. Implementation, evaluation, and replanning occur in the operational stage.

The strategic, tactical, and operational phases have also been linked to agriculture in previous studies. For example, Aull-Hyde and Tadesse [14] modeled strategic differences when long- and short-term risks are considered in agricultural production. Fountas et al. [15] modeled the effects of information flows for strategic, tactical, and operational differences in the context of precision agriculture. Other models review specific details of tactical or operational implementation, while assuming that the strategic phase of the model has already been conducted (Ahumada and Villalobos [16]).

Not surprisingly, risk management researchers have proposed frameworks that intuitively capture the strategic planning process. For example, Clemens and Reilly [17] propose the following steps that are typical of many risk researchers (e.g., [2, 18, 19]):

(1) identify decision,

(2) identify alternatives,

(3) decompose the problem,

(4) choose best alternative,

(5) conduct sensitivity analysis,

(6) repeat if necessary.

These are the same principles found in the tactical and operational phases of strategic planning. The strategic phase is more commonly addressed separately through risk preference and tolerance (e.g., [17, 19]).

Risk Navigator SRM provides formality to agricultural risk management and condenses this information in a way that is understandable to agricultural operators by fitting risk management into the strategic management framework. Hoag [20] developed ten steps that map existing risk management concepts into the three phases (Figure 1). There are three steps in the first phase: (1) determine financial health, (2) determine risk preference, and (3)

Figure 1: Strategic risk management process; see source in the study of Hoag in [6].

establish risk goals. These steps are not typically covered in risk management frameworks, but were inspired by the strategic planning process. The first two of these steps were chosen to set boundaries on tactical choices, which are then used to set goals in Step 3, as set forth by the strategic planning process.

There are four steps in the next phase, the tactical stage, which constitute Steps 4–7 in the SRM process: (4) determine risk sources, (5) identify management alternatives, (6) estimate risk probabilities, and (7) rank management alternatives. This tactical phase is based on another framework, the payoff matrix (Table 2), which is commonly used in risk analysis to capture frameworks like that shown above [17, 18]. Each of these steps is designed to elicit a component of a payoff matrix. The payoff matrix is a construct that displays payoffs, usually profits, by management actions (e.g., cash sale, contract sale, hedging on the futures market, etc.) and states of nature (e.g., normal weather resulting in a normal crop and typical crop prices, or bad weather, resulting in a short US crop and high crop prices). Probability is displayed next to each state of nature. The matrix is designed to show risk dimensions in a way that helps decision makers rank risks based on their risk personality, which is further described in Step 7. Summary statistics can be displayed at the bottom of the table to provide more information, such as expected value and standard deviation. For the purposes of illustration, the EWS Farms case study manages corn price risk. There are three marketing management alternatives: selling on the cash market, forward pricing, or hedging. The source of risk is the likelihood of a short U.S. crop.

One limitation of the payoff matrix is that it only addresses one risk at a time. It can accommodate complex problems with multiple management alternatives, but only one source. Our attempts to discuss joint distributions in risk training workshops, such as price with yield, reduced comprehension and acceptance by producers when presented; therefore we chose to focus on addressing multiple management options for a single risk. Joint distributions and extensions of the model are described by Hoag in [6]. In addition, Ag Survivor risk simulations are based on joint distributions, where appropriate.

Table 2: Payoff matrix for EWS farms' corn pricing decision; see source in [6].

			Whole Farm Returns for Management Alternatives		
			no. 1	no. 2	no. 3
	Outcome			Forward	
Risk states	*$/bushel corn*	Probability	Cash market	Contract	Hedge
Normal crop	$2.10/bu	.35	$287,700	$305,900	$298,200
Short crop	$2.50/bu	.65	$342,500	$332,700	$339,000
Expected value			$323,320	$323,320	$324,720
Standard deviation			$38,749	$18,900	$28,850

The last phase, the operational stage, utilizes three steps intended to carry out the plans made in the tactical stage: (8) implement plans, (9) monitor and adjust, and (10) replan. The first step focuses on the day-to-day activities to assure that good planning efforts are carried out. Monitoring can provide the information needed to determine whether plans should be adjusted. Re-planning takes the decision maker around the circle to start over.

4. The Strategic Phase

We proceed with a brief description of each step. An application to the case study and examples of tools available at the RiskNavigatorSRM.com website are included, where space is appropriate for this audience.

Step 1 (Determine Financial Health). The first step is to determine financial health in order to determine a person's financial capability to take on risks. Financial health refers to assessing the well-being of a business's financial resources, with respect to their ability to take on risk. Educational programs about financial management are widely available in agricultural extension programs. There are six tools available at http://www.RiskNavigatorSRM.com/ that were based on an extension program in Montana developed by Duane Griffith (http://www.montana.edu/softwaredownloads/financialmgtdownloads.html). This includes typical tools to develop not only commonly used financial statements, like a balance sheet and cash flow statement, but also tools specifically designed to help people understand how health affects risk resilience. Specifically, the "RDfinancial" tool and "Sweet Sixteen Ratio Analyzer" tool identify strengths and weaknesses of a decision maker's financial position. RDfinancial and the Ratio Analyzer provide a plethora of financial information, including the sixteen financial ratios commonly used to describe financial health. RDfinancial also contains a credit scoring model.

Step 2 (Determine Risk Preference). The second step involves assessing a person's risk preferences, which affects a person's attitude about taking on risks. There are many limitations regarding the elicitation of risk preferences [17, 18] but, sometimes, it is worth living with these problems if producers need the extra information at the decision margin. The Risk Navigator SRM's "Risk Preference Calculator" tool offers three different methods to help people gauge their preferences. One method, shown in Figure 2, computes a relative Pratt-Arrow risk preference score [21, 22]. The coefficient of absolute risk aversion, $r_a(W)$, is the negative of the second derivative of utility, U, for wealth, W, divided by the first

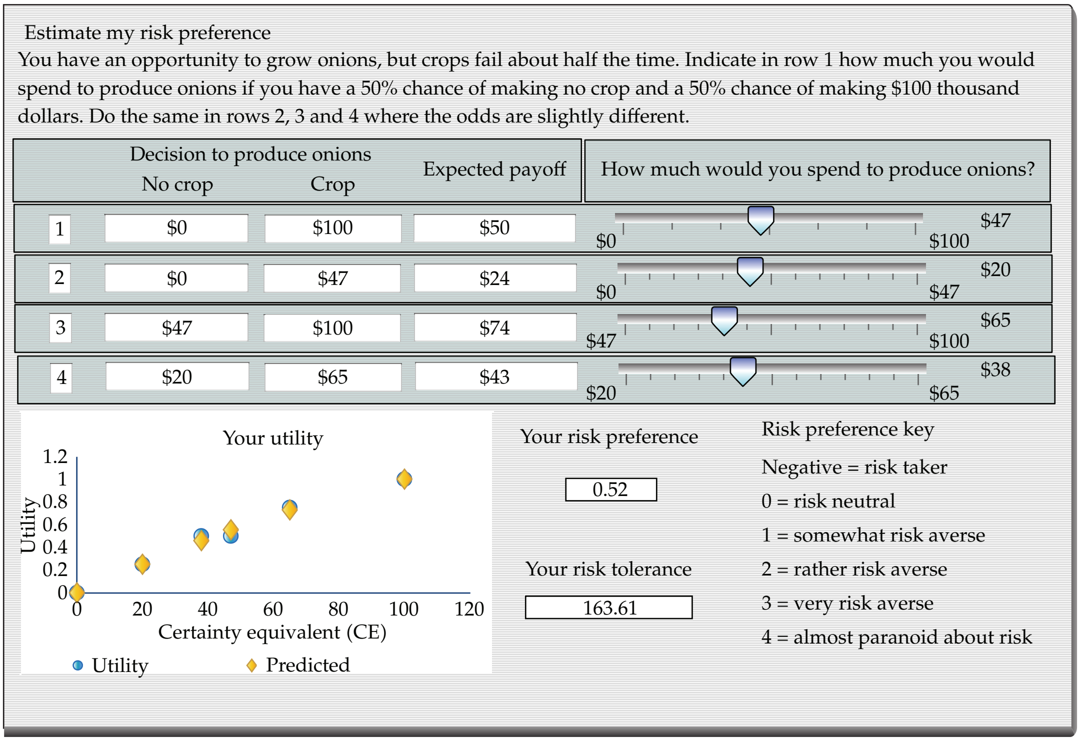


Figure 2: Elicitation of risk preference/tolerance in Risk Preference Calculator. See source in [6].

derivative of utility for wealth:

$$r_a(W) = -\frac{U''(W)}{U'(W)}. \tag{4.1}$$

This coefficient is positive for risk-averse individuals, zero for risk neutral individuals, and negative for risk-loving individuals. People that are risk averse would pay a premium to avoid risk. Risk-neutral individuals maximize expected values and ignore risk. Since "r" changes with the size of the gamble, the concept of a coefficient of relative risk aversion (RAC) was created simply by multiplying the coefficient of absolute risk aversion by wealth:

$$\text{RAC} = W r_a(W). \tag{4.2}$$

The RAC equals 0 for someone that is risk neutral. It varies from about 0.5 to 4.0 for risk-averse people, as suggested by Hardaker et al. [18]:

(i) $\text{RAC}(W) = 0.5$: hardly risk averse at all,

(ii) $\text{RAC}(W) = 1.0$: somewhat risk averse (normal),

(iii) $\text{RAC}(W) = 2.0$: rather risk averse,

(iv) $\text{RAC}(W) = 3.0$: very risk averse,

(v) $\text{RAC}(W) = 4.0$: almost paranoid about risk.

A common, yet limited, functional form used for utility is the negative exponential utility function:

$$U(W) = 1 - \exp^{(-rW)}. \tag{4.3}$$

This is a convenient functional form for illustration since the coefficient of absolute risk aversion is r.

The RAC can be found by eliciting a person's utility function, which can be estimated by the ELCE (Equally Likely Certainty Equivalents) method as described by Hardaker et al. in [18] (Figure 2). ELCE elicits equally likely certainty equivalents by asking a series of questions that present 50-50 bets, which can be used to sketch a utility function like that shown in the lower left quadrant of Figure 2. The certainty equivalent is the certain amount that a person would be indifferent to receiving compared to an expected value with risk. To provide a more realistic scenario, onion production was chosen for the ELCE questions in the Risk Preference Calculator tool, since the crop tends to produce boom or bust returns (so we could use the 50-50 technique). These "bets" are presented by slider bars as shown in Figure 2. This allowed the program to remain simple and realistic, and fit well with a 50-50 bet for agricultural producers. Using the Hardaker et al. scale, the EWS case study farmer turned out to be quite risk tolerant, with a risk preference score of 0.52 (somewhat risk averse). This score is based on derivatives from the utility curve fitted in the lower left quadrant of Figure 2.

The second risk preference assessment method (not shown) in the Risk Preference Calculator provides users the option of taking a short risk quiz designed by Grable and Lytton [23]. The quiz asks 13 questions, which are tabulated to provide a score from "low risk tolerance" to "high risk tolerance." More than one method is offered to counteract inherent difficulties in measuring risk preferences. In this case, psychological research offers a completely unique approach in the form of this quiz.

The third method links risk preference to risk tolerance, which is the amount $\$X$, where a person would be indifferent between an equal chance of receiving $\$X$ and losing $\$X/2$ [8]. A risk-averse person that would pay a certainty equivalent of \$7,000 for a risky, 50-50 bet of receiving either 0 or \$20,000, for example, has a risk tolerance of \$15,641; risk tolerance increases to \$99,833 for someone willing to pay \$9,900 for that same bet. A risk-neutral producer would of course be willing to pay \$10,000. Risk tolerance uses a different means than ELCE to elicit tolerance, but is closely linked to preference. Relative risk tolerance (RRT) can be derived from the RAC, since absolute risk tolerance, R, is the inverse of absolute risk preference, r, in the negative exponential utility function. The scale of RRT spans from 0.25 (1/4) for a person who is almost paranoid about risk to 2 (1/0.5) for someone who is rarely risk averse to 10 (1/.1) for someone that is almost risk neutral. This allows the Risk Preference Calculator to provide a comparable estimate of the relative risk preference based on risk tolerance, which opens the door to an entirely separate and more prevalent literature. For example, Howard [24] defined R for firms that he looked at in terms of annual sales, equity, and income. He found that $R = 1.24$ multiplied by net income, or 6.4% of sales, or 15.7% of equity for the businesses that he examined. The Risk Preference Calculator can therefore provide a parallel estimate of risk preference by simply asking the producer for net income, sales, or equity.

Utilizing three methods to elicit risk preferences makes the Risk Preference Calculator tool more accessible to users with differing tastes. It also provides some continuity and comparability for those users willing to apply all three methods, which helps combat the inherent problems with estimating risk preferences [17].

Step 3 (Establish Risk Goals). There are seven goal planning tools on the Risk Navigator SRM website, including action planning, team roster, mission statement, time management, and transition planning. These classic tools are updated to assist decision makers with developing their goals in the "Strategic Goal Worksheet." EWS Farms' strategic goals were [6] the following.

(i) Strategic Goal no. 1 (Financial): Ensure short- and long-term financial success by maintaining business profitability, while expanding the overall business financial resource base.

(ii) Strategic Goal no. 2 (Family): Continue to live, work, and grow our families in a rural, agricultural environment. Encourage individual development and exploration in a manner that is consistent and flexible in order to allow all individuals to reach their full potential.

(iii) Strategic Goal no. 3 (Organizational): Continue to pursue organizational structures that fit the family dynamics of the operation, as well as allow for strategic goal attainment. Also, increase the business activities efficiency of the operation.

(iv) Strategic Goal no. 4 (Integrated Farm Management): Manage our farm as a cointegrated unit, while providing a step-by-step process for developing a strategic risk management plan.

A comprehensive description of all of their goals and mission statement can be found in Applied Risk Management in Agriculture [6].

5. The Tactical Phase

Step 4 (Determine Risk Sources). The four steps in the tactical phase are those required to build a payoff matrix. The first step is to determine risk sources. In addition to determining risk sources, it is also important to prioritize risks so that management efforts can be focused. EWS Farms chose to focus on price risk for corn. This was interesting since it was not even one of their risk management goals. However, it is not inconsistent with observations at risk education extension meetings. Men that have attended the Risk Navigator SRM workshops have overwhelmingly chosen price and yield risk (revenue) over all other types of risk when doing this exercise. Decision makers are encouraged in this step to review their goals and information with those available in the study by Hoag in [6], like those shown in Table 1.

Several methods are demonstrated in the book titled *Applied Risk Management in Agriculture* [6], such as influence diagrams and SWOT ((S)trengths, (W)eaknesses, (O)pportunities, and (T)hreats) analysis. However, the method chosen for the Risk Navigator tool, "Risk Influence Calculator," uses a risk-influence diagram to help decision makers sort out and prioritize risk. The Risk Influence Calculator is simple and only asks producers to list each risk, then to rate it on a scale of 1 (low) to 10 (high) for (A) probability of occurring, (B) impact if it occurs, and (C) influence to stop it from occurring.

After reviewing his goals and reading what others said about risks (e.g., Table 1), EWS Farms identified the following strategic risks:

(i) *Market/Price*

 (a) Corn Price. Will my price cover my costs?

(ii) *Production*

 (a) Weather. Will rainfall support my crop stand?
 (b) Hail. Will hail destroy half my crop?
 (c) Input (seed). Will good corn seed be available at a reasonable price?

(iii) *Financial*

 (a) Expansion. Can the operation generate enough profit to cover new land payments?

(iv) *Human*

 (a) Family. Will my dad retire?

(v) *Institutional*

 (a) Water. Will irrigation water be restricted?

All risks identified are placed in a graph that plots influence against risk, so the decision maker can prioritize risks where the biggest impact can be made. The Risk Influence tool has been one of the most popular in Navigator risk management workshops, as it creates a lot of discussion and rethinking about priorities.

Step 5 (Identify Management Alternatives). The book, *Applied Risk Management in Agriculture* [6], describes four main techniques to manage risks: (1) avoid it, (2) transfer it, (3) assume or retain it, or (4) reduce it. To keep things understandable, we use a risk profile. A risk profile is a multidimensional representation of the consequences that stem from each management action. For simplicity, we use a probability density function (PDF) as the risk profile, since it is familiar to most people and it contains information that is relevant to managing risk, such as mean, mode, maximum, and minimum. We can show the consequences of the four basic management actions mentioned above through simple manipulations of the PDF such as skewing, truncating, changing variance (squishing), or changing the mean (moving). This simple representation of a risk profile (and terms like squishing) is meant to build basic skills and understanding in the participants; however, perhaps more importantly, the graphical depictions engage the audience in what may often be perceived as a dry subject. It also ties the concept of management alternatives to PDFs for the next step concerning likelihood. Like the previous step, the book also provides an extensive list and brief discussion about specific techniques commonly used in agriculture, like crop insurance and the futures market (e.g., Table 1).

Decision makers at EWS Farms identified the following possibilities to manage price risk in their corn:

(i) Cash market sales (selling everything in the market at harvest),

(ii) Forward contracting to a local grain elevator (contracting for a fixed price before harvest),

(iii) Hedging on the futures market (selling a contract on the futures market for a fixed price),

(iv) Spreading out crop sales across the year (multiple marketing points),

(v) Maintaining flexibility on timing of sales (storing, then selling opportunistically).

After more consideration, EWS Farms chose three management alternatives: cash market, forward contracting, and hedging. They developed a comprehensive marketing plan using the "Marketing Plan" tool in Risk Navigator SRM. Six components exist in this marketing plan, and each component has its own worksheet: (1) The Relationship between the Strategic Risk Management Plan and the Marketing Plan; (2) Production History and Expectations; (3) Expected Prices; (4) Production Costs; (5) Price, Date, and Quantity Targets; (6) Review and Evaluation. Each of their marketing alternatives is carefully planned, including the distribution of the sale over time. The marketing alternatives are also based on ten years of local data on EWS Farms and for their local elevator (including basis adjustments for prices).

To help EWS Farms prepare for the remaining steps, the book recommends putting their information into a decision tree and then a payoff matrix. The main value of starting with the decision tree, shown in Figure 3, is in its visual construction [25], which requires the decision maker to identify all relevant courses of action, events, and payoffs in a clear and understandable manner. It also makes it easier to process information to put into a payoff matrix, as shown in Table 2.

Step 6 (Estimate Probabilities). The concept of probability has been cultivated throughout the ten-step process by encouraging decision makers to represent risk with a "risk profile" and to think of a risk profile in terms that can be easily understood from the basic shape of a PDF. They are encouraged in Step 2 to determine their "risk personality" (preference) so they can find the risk profile that best suits them. In Step 5 they are shown how a risk profile is affected by risk management alternatives, where the random variable is usually income or cost. In this section we show how to tie the concept of the risk profile to a probability density function (PDF) or a cumulative density function (CDF) more directly. Much of the book chapter [6] is dedicated toward a basic lesson about the PDF and CDF, including concepts like mean, mode, median, variance, standard deviation, and coefficient of variation. Also included are descriptions about how to interpret basic shapes of the PDF. The discussion is very basic, aimed at making the PDF and CDF concepts that people could use to quickly and intuitively interpret the basic statistical components that are important for risk management (e.g., measures of central tendency, spread, and range). For example, it is easy to convey that management tools like insurance or the futures market "squish," "move," or "truncate," a PDF.

The "Risk Profiler" tool makes it relatively easy for producers to build probability density functions (PDF's), which then can be used to provide information for probability in the payoff matrix. The "art" of eliciting probability is fraught with limitations (e.g., [17–19]),

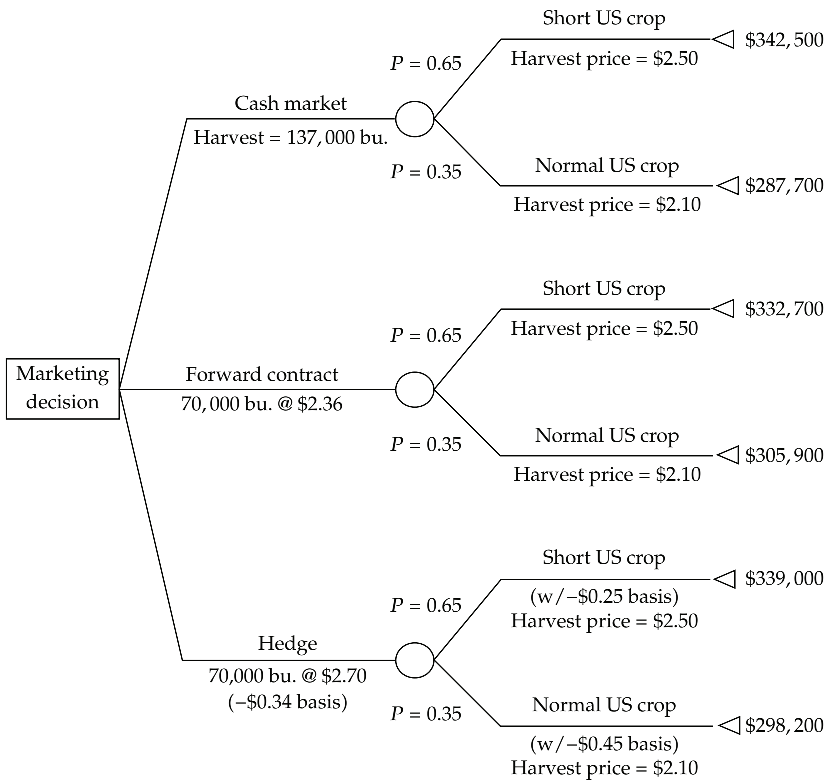


Figure 3: Decision tree for EWS Farms corn pricing decision; see source in [6].

but the PDF's are used here to simplify the process of estimating probability in the payoff matrix and to provide a stable mechanism to tie the steps together. Furthermore, Hoag [6] discusses some of the common problems, like anchoring, and how to avoid them.

There are at least two major ways to elicit probability information (6,10): asking the expert about his or her degree of belief about an outcome or present the expert with a lottery that reveals their probability values. Risk Profiler provides three options to elicit a PDF by asking the expert about what they believe will happen. The most straightforward method assumes that the future will look like the past ten periods. Figure 4 shows ten annual corn price entries for EWS Farms and resulting PDF, CDF, histogram, and summary statistics. The PDF is assumed to be Normal for simplicity. However, for increased education, a histogram is also drawn to help decision makers understand what might be hidden behind the normalized function. For example, a bimodal distribution is hidden by the normal distribution in the case of corn price on EWS farms. A few summary statistics are also provided for each estimation method.

The second method offered by Risk Profiler to elicit a PDF involves having the user "describe profile features." In this case, a PDF can be drawn based only on the minimum, most likely, and maximum values elicited from the decision maker for the random variable. The PERT distribution is applied as follows:

$$\mathrm{PERT}(a, b, c) = \mathrm{Beta}(\alpha_1, \alpha_2) * (c - a) + a, \tag{5.1}$$

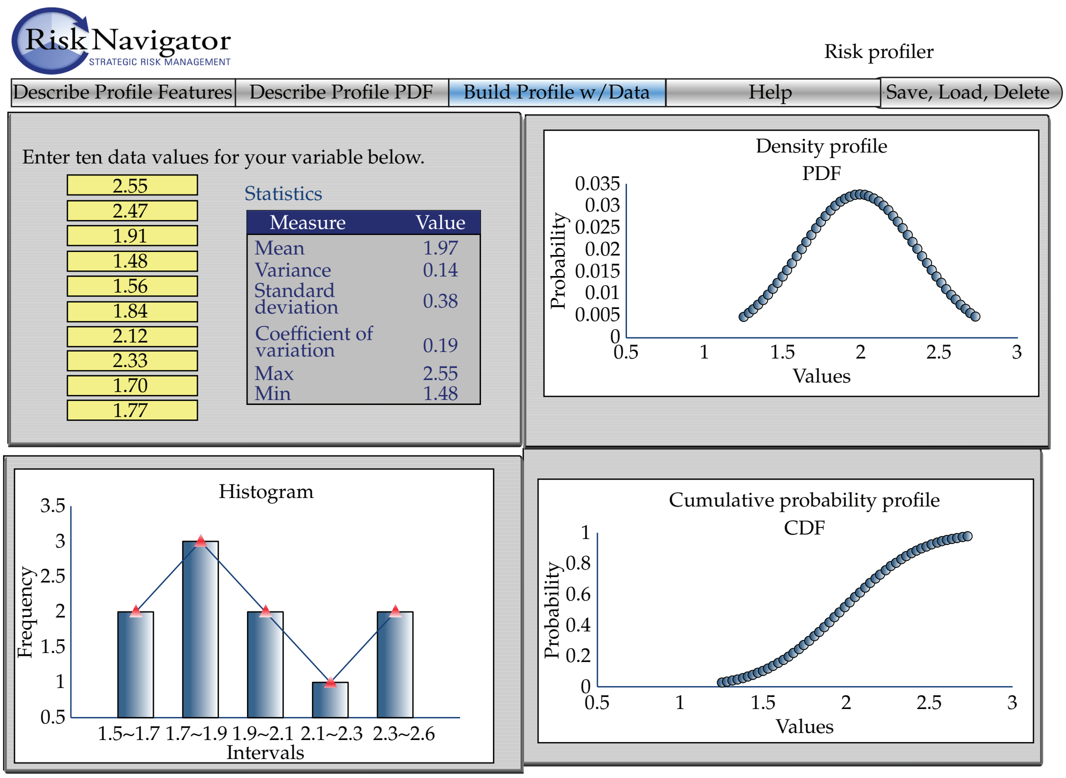


Figure 4: Building a PDF with Risk Profiler; see source in [6].

where "a" is the minimum value, "b" is the most likely value, "c" is the maximum value, and "Beta" is the beta distribution [19].

Furthermore,

$$\alpha_1 = \frac{(\mu - a) * (2b - a - c)}{(b - \mu) * (c - a)},$$
$$\alpha_2 = \frac{\alpha_1 * (c - \mu)}{(\mu - a)}. \tag{5.2}$$

In this case, the resulting PDF and related information are pictured on a screen exactly like that shown in Figure 4, but the upper left quadrant is replaced with a section to collect the PERT input data.

The last method is "describe profile PDF." In this method, people provide five sets of probabilities and values for the random variable, see Table 3.

This is a variation on the fractal method [17, 18]; in eliciting the PDF for prices for a given crop, for example, a decision maker might be asked to pick several price values and give an associated probability for each one. Asking for the producer to supply the probability and values simultaneously is a combination of what Frey [26] called the fixed value method and fixed probability method. The fixed value method asks an expert the probability that

Table 3

Probability	Price
.30	$2.00
.20	$2.25
.40	$2.50
.20	$2.75
.10	$3.00

the actual value is higher (or lower) than some arbitrary number for the random variable. The fixed probability method has the decision maker specify the value range for a random variable that corresponds to a specific probability. The Risk Profiler method lets the user simultaneously mix and match the fixed value and fixed probability methods by allowing them to enter either probabilities or values.

Step 7 (Rank Management Alternatives). The final step of the tactical stage is to rank the various alternatives considered to this point and select those with the most desirable outcomes. Risk Navigator SRM offers three tools to help decision makers rank risks. “Value at Risk” (VaR) is a popular method for capturing the downside risk in financial decision making. It is an evaluation of what you stand to lose with an investment. VaR answers the questions “What is my worst-case scenario?” or “How much could I potentially lose in a really bad month?” [27]. This strategic tool considers only the undesirable parts of dispersion, those on the negative side of the mean, as opposed to the standard deviation, for example. The VaR tool is simple to use and involves only one screen (not pictured here). A second tool, the “Risk Efficiency Tool,” uses Stochastic Efficiency with Respect to a Function to rank outcomes for all levels of risk preference, except risk preferring.

The easiest and most effective tool, “Risk Ranker,” uses the payoff matrix to link risk personalities with the risk profiles. This tool allows a user to directly compare risk profiles, which can provide a lot of information by itself, and offers an instant comparison of the management alternatives under consideration with seven ranking rules that cater to different risk personalities (e.g., someone that is avoiding risk as compared to someone that wishes to maximize expected value).

After filling out a payoff matrix for up to five management alternatives, the program can be used to compare risk profiles. For example, on the second tab, “Compare Profiles,” shown in Figure 5, the payoff matrix entered by the decision maker is reprinted in the upper left corner. The first column paired with any management alternative (e.g., cash) replicates the information entered in “Risk Profiler,” which provides continuity to the program and reinforces how the risk profile integrates with the payoff matrix. All five PDFs and CDFs for EWS Farms are plotted in one graph and summary statistics are provided in tabular form. Decision makers are provided with information about how to rank alternatives with methods that use only distributions, such as stochastic dominance [6], and may therefore use this method alone to rank risks.

In many, if not most, cases risks cannot be ranked by visual inspection of the PDF or CDF. Therefore, the next tab over, “Risk Ranker” displays seven different risk ranking measures for the payoff matrix, all on one screen: Maximize EV, Maximax, Most likely, Minimax regret, Hurwicz, Maximin, and the Laplace Insufficient Reason Index (Table 4). Each of these techniques ranks risks based on different aspects of the payoff that might, or might not match a decision maker’s risk personality. For example, Maximize EV (expected

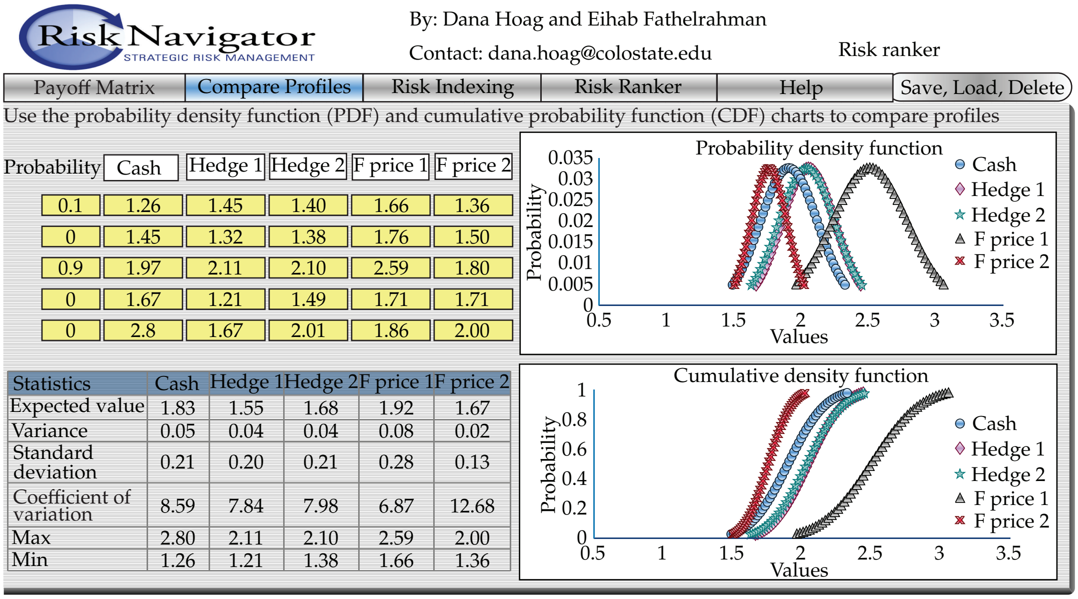


Figure 5: Risk Ranker—compare profiles—see source in [6].

Table 4: Attributes of basic decision rules[a].

Decision Rule	Mean	Variation	Low	High	Other
Maximized Expected Value—Choose alternative with highest expected value	x				
Maximax—Choose alternative with best outcome				x	
Most Likely—Choose alternative that is most likely to occur (subjective)					x
Minimax Regret—Choose alternative with least hindsight regret		x	x		
Hurwicz—Weighted average of maximax and minimax			x	x	
Maximin—Choose alternative with best of the worst outcomes			x		
Laplace—Simple average	x				

[a]See a more detailed description in [6]. An x is placed in the column where the rule is primarily focused.

value) is for risk neutral producers. Maximin chooses the alternative with the maximum, minimum outcome and is therefore designed for very risk-averse individuals.

Table 4 shows that each of the rules focuses on different areas. None is comprehensive, and many are oversimplified. For example, Maximin would choose A in the two five-year income streams A and B shown below, since it focuses only on the minimum; however, most people, if not all, would choose B if they could:

$$\begin{aligned} A &= (10, 10, 10, 10, 10), \\ B &= (100, 100, 100, 9, 100). \end{aligned} \tag{5.3}$$

The table presented in the tool makes it handy to compare rankings quickly and easily so that many dimensions can be considered, and is very effective at getting across the message that people need to match their risk management personality to their risk ranking techniques.

6. The Operational Phase

Steps 8–10.

The operational stage involves putting plans into action. The first step, implementation, involves acquiring the necessary resources, scheduling the tasks to be completed, and overseeing all aspects of the plan [6]. As plans are implemented, the second step in this phase kicks in as resources need to be monitored so adjustments can be made as needed. The last step in the phase, and step ten overall, is re-planning. If not actively brought to the attention of those involved with the program, this step may be easily ignored by many managers since it could highlight what was not achieved. Re-planning is also the first step in preparation for starting the cycle over. There are five SRM tools available on the website for the operational phase: (1) Action Planning Worksheet, (2) Critical Success Indicator Worksheet, (3) Resource Flow Plan, (4) Risk Management Worksheet, and (5) Time Management Worksheet. A detailed description of how EWS Farms applied this step is provided by Hoag in [6].

7. Conclusion

Risk Navigator SRM integrates risk management techniques into one place and provides farmers and ranchers the resources to use the products from their home office. It is both a learning tool as well as a means to help producers actually manage risks, financial and other, in their operations. It is also very practical for students in economics and business. The manageable steps allow for farmers and ranchers to learn the tools at their own speed, while providing the opportunities to customize the data for their own farm/ranch. We find that the difference between this and other programs designed to manage risk is the integration of basic risk management principles into a structured and easy to learn format. This allows people to use concepts that have been individually available for decades, but inaccessible, because they are most valuable in a framework, which typically requires expertise to build. The book, *Applied Risk Management in Agriculture* [6], can supplement the website by providing detailed descriptions of each step and by providing additional educational opportunities. Navigator workshops have been presented in formats from 45 minutes to two days. The program has grown and the tools can be continuously upgraded since they are offered at the Navigator website. A blog has been added and a free, ten-step online education program is nearly complete.

Acknowledgments

The authors thank to the USDA Risk Management Agency for partial funding and the anonymous reviewers for their helpful comments.

References

[1] J. Harwood, R. Heifner, K. Coble, J. Perry, and A. Somwaru, "Managing risk in farming: concepts, research and analysis," Report 774, U.S. Department of Agriculture-Economic Research Service, Market and Trade Economics Division and Resource Economics Division, Washington, DC, USA, 1999.

[2] B. Fleisher, *Agricultural Risk Management*, Lynne Rienner Publishers, Boulder, Colo, USA, 1990.
[3] Risk Management Agency—U.S. Department of Agriculture, *Building a Risk Management Plan: Risk-Reducing Ideas that Work*, U.S. Department of Agriculture, Washington, DC, USA, 1998.
[4] Risk Management Agency—U.S. Department of Agriculture, *A Risk Management Agency Fact Sheet About the Risk Management Agency, Program Aid 1667-02*, U.S. Department of Agriculture, Washington, DC, USA, 2009.
[5] Risk Management Agency—U.S. Department of Agriculture, *Introduction to Risk Management: Understanding Agricultural Risks: Production, Marketing, Financial, Legal and Human Resources*, U.S. Department of Agriculture, Washington, DC, USA, 1997.
[6] D. L. Hoag, *Applied Risk Management in Agriculture*, CRC Press-Taylor Francis, Boca Rotan, Fla, USA, 2009.
[7] T. Aven, *Foundations of Risk Analysis: A Knowledge and Decision-Oriented*, John Wiley & Sons, Chichester, UK, 2003.
[8] R. Myerson, *Probability Models for Economic Decisions*, Thomson, Brooks/Cole, Belmont, Calif, USA, 2005.
[9] K. Curtis, "Risk management programming: is it effective at creating change," in *Proceedings of the National Extension Risk Management Education Conference*, Phoenix, Ariz, USA, April 2007, http://www.agrisk.umn.edu/conference/agenda.aspx?ConfID=4&Output=Sum.
[10] G. F. Patrick, A. J. Peiter, T. O. Knight, K. H. Coble, and A. E. Baquet, "Hog producers' risk management attitudes and desire for additional risk management education," *Journal of Agricultural and Applied Economics*, vol. 39, no. 3, pp. 671–687, 2007.
[11] G. Steiner, *A Step by Step Guide to Strategic Planning: What Every Manager Must Know*, The Free Press, New York, NY, USA, 1979.
[12] G. Toft, "Synoptic (one best way) approaches of strategic management," in *Handbook of Strategic Management*, J. Rabin, G. Miller, and W. B. Hildreth, Eds., pp. 545–560, Marcel Dekker, New York, NY, USA, 2nd edition, 2000.
[13] B. Bilgen and I. Ozkarahan, "Strategic tactical and operational production-distribution models: a review," *International Journal of Technology Management*, vol. 28, no. 2, pp. 151–171, 2004.
[14] R. Aull-Hyde and S. Tadesse, "A strategic agricultural production model with risk and return considerations," *Agricultural and Resource Economics Review*, vol. 23, no. 1, pp. 37–46, 1994.
[15] S. Fountas, D. Wulfsohn, B. S. Blackmore, H. L. Jacobsen, and S. M. Pedersen, "A model of decision-making and information flows for information-intensive agriculture," *Agricultural Systems*, vol. 87, no. 2, pp. 192–210, 2006.
[16] O. Ahumada and J. R. Villalobos, "Planning the Production and Distribution of Fresh Produce," *Annals of Operations Research*. In Press.
[17] R. T. Clemens and T. Reilly, *Making Hard Decisions with Decision Tools®*, Duxbury, Pacific Grove, Calif, USA, 2001.
[18] J. B. Hardaker, R. M. Huirne, J. R. Anderson, and G. Lien, *Coping with Risk in Agriculture*, CABI Publishing, Wallingford, UK, 3rd edition, 2004.
[19] D. Vose, *Risk Analysis: A Quantitative Guide*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.
[20] D. L. Hoag, "The strategic risk management process," Unpublished proposal, Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, Colo, USA, 2005.
[21] K. J. Arrow, *Aspects of the Theory of Risk-Bearing*, Academic Bookstore, Helsinki, Finland, 1965.
[22] J. W. Pratt, "Risk aversion in the small and the large," *Econometrica*, vol. 32, pp. 122–136, 1964.
[23] J. E. Grable and R. H. Lytton, "Financial risk tolerance revisited: the development of a risk assessment instrument," *Financial Services Review*, vol. 8, no. 3, pp. 163–181, 1999.
[24] R. A. Howard, "Decision analysis: practice and promise," *Management Science*, vol. 34, no. 6, pp. 679–695, 1988.
[25] A. N. Rae, *Agricultural Management Economics: Activity Analysis and Decision Making*, CABI Publishing, Wallingford, UK, 1994.
[26] C. Frey, "Briefing paper part 1: introduction to uncertainty analysis," Department of Civil Engineering, North Carolina State University, 1998, http://legacy.ncsu.edu/classes/ce456001/www/Background1.html.
[27] D. Harper, "Introduction to value at risk (VaR)," 2008, http://www.investopedia.com/articles/04/092904.asp.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 707146, 34 pages
doi:10.1155/2010/707146

Research Article

Estimating L-Functionals for Heavy-Tailed Distributions and Application

Abdelhakim Necir and Djamel Meraghni

Laboratory of Applied Mathematics, Mohamed Khider University of Biskra, 07000 Biskra, Algeria

Correspondence should be addressed to Abdelhakim Necir, necirabdelhakim@yahoo.fr

Received 5 October 2009; Accepted 21 January 2010

Academic Editor: Ričardas Zitikis

Copyright © 2010 A. Necir and D. Meraghni. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

L-functionals summarize numerous statistical parameters and actuarial risk measures. Their sample estimators are linear combinations of order statistics (L-statistics). There exists a class of heavy-tailed distributions for which the asymptotic normality of these estimators cannot be obtained by classical results. In this paper we propose, by means of extreme value theory, alternative estimators for L-functionals and establish their asymptotic normality. Our results may be applied to estimate the trimmed L-moments and financial risk measures for heavy-tailed distributions.

1. Introduction

1.1. L-Functionals

Let X be a real random variable (rv) with continuous distribution function (df) F. The corresponding L-functionals are defined by

$$L(J) := \int_0^1 J(s)Q(s)ds, \tag{1.1}$$

where $Q(s) := \inf\{x \in \mathbb{R} : F(x) \geq s\}, 0 < s \leq 1$, is the quantile function pertaining to df F and J is a measurable function defined on $[0, 1]$ (see, e.g. Serfling, [1]). Several authors have used the quantity $L(J)$ to solve some statistical problems. For example, in a work by Chernoff et al. [2] the L-functionals have a connection with optimal estimators of location and scale parameters in parametric families of distributions. Hosking [3] introduced the L-moments as a new approach of statistical inference of location, dispersion, skewness, kurtosis, and other aspects of shape of probability distributions or data samples having finite means. Elamir and Seheult [4] have defined the trimmed L-moments to answer some questions related to *heavy-tailed* distributions for which means do not exist, and therefore the L-moment method cannot be applied. In the case where the trimming parameter equals one, the first four theoretical trimmed L-moments are

$$m_i := \int_0^1 J_i(s)Q(s)ds, \quad i = 1,2,3,4, \tag{1.2}$$

where

$$J_i(s) := s(1-s)\phi_i(s), \quad 0 < s < 1, \tag{1.3}$$

with ϕ_i polynomials of order $i-1$ (see Section 4). A partial study of statistical estimation of trimmed L-moments was given recently by Hosking [5].

Deriving asymptotics of complex statistics is a challenging problem, and this was indeed the case for a decade since the introduction of the distortion risk measure by Denneberg [6] and Wang [7]; see also Wang [8]. The breakthrough in the area was offered by Jones and Zitikis [9], who revealed a fundamental relationship between the distortion risk measure and the classical L-statistic, thus opening a broad gateway for developing statistical inferential results in the area (see, e.g., Jones and Zitikis [10, 11]; Brazauskas et al. [12, 13] and Greselin et al. [14]). These works mainly discuss CLT-type results. We have been utilizing the aforementioned relationship between distortion risk measures and L-statistics to develop a statistical inferential theory for distortion risk measures in the case of heavy-tailed distributions.

Indeed L-functionals have many applications in actuarial risk measures (see, e.g., Wang [8, 15, 16]). For example, if $X \geq 0$ represents an insurance loss, the distortion risk premium is defined by

$$\Pi(X) := \int_0^\infty g(1-F(x))dx, \tag{1.4}$$

where g is a non decreasing concave function with $g(0) = 0$ and $g(1) = 1$. By a change of variables and integration by parts, $\Pi(X)$ may be rewritten into

$$\Pi(X) = \int_0^1 g'(1-s)Q(s)ds, \tag{1.5}$$

where g' denotes the Lebesgue derivative of g. For heavy-tailed claim amounts, the empirical estimation with confidence bounds for $\Pi(X)$ has been discussed by Necir et al. [17] and Necir and Meraghni [18]. If $X \in \mathbb{R}$ represents financial data such as asset log-returns, the distortion risk measures are defined by

$$H(X) := \int_{-\infty}^0 (g(1-F(x))-1)dx + \int_0^\infty g(1-F(x))dx. \tag{1.6}$$

Likewise, by integration by parts it is shown that

$$H(X) = \int_0^1 g'(1-s)Q(s)ds. \tag{1.7}$$

Wang [8] and Jones and Zitikis [9] have defined the *risk two-sided deviation* by

$$\Delta_r(X) := \int_0^1 J_r(s)Q(s)ds, \quad 0 < r < 1, \tag{1.8}$$

with

$$J_r(s) := \frac{r}{2}\frac{s^{1-r} - (1-s)^{1-r}}{s^{1-r}(1-s)^{1-r}}, \quad 0 < s < 1. \tag{1.9}$$

As we see, $\Pi(X), H(X)$, and $\Delta_r(X)$ are L-functionals for specific weight functions. For more details about the distortion risk measures one refers to Wang [8, 16]. A discussion on their empirical estimation is given by Jones and Zitikis [9].

1.2. Estimation of L-Functionals and Motivations

In the sequel let $\xrightarrow{p}$ and $\xrightarrow{\mathcal{D}}$, respectively, stand for convergence in probability and convergence in distribution and let $\mathcal{N}(0,\eta^2)$ denote the normal distribution with mean 0 and variance η^2.

The natural estimators of quantity $L(J)$ are linear combinations of order statistics called L-statistics. For more details on this kind of statistics one refers to Shorack and Wellner [19, page 260]. Indeed, let $(X_1,\ldots,X_n)$ be a sample of size $n \geq 1$ from an rv X with df F, then the sample estimator of $L(J)$ is

$$\widehat{L}_n(J) := \int_0^1 J(s)Q_n(s)ds, \tag{1.10}$$

where $Q_n(s) := \inf\{x \in \mathbb{R} : F_n(x) \geq s\}$, $0 < s \leq 1$, is the empirical quantile function that corresponds to the empirical df $F_n(x) := n^{-1}\sum_{i=1}^n \mathbb{I}\{X_i \leq x\}$ for $x \in \mathbb{R}$, pertaining to the sample $(X_1,\ldots,X_n)$ with $\mathbb{I}(\cdot)$ denoting the indicator function. It is clear that $\widehat{L}_n(J)$ may be rewritten into

$$\widehat{L}_n(J) = \sum_{i=1}^n a_{i,n}X_{i,n}, \tag{1.11}$$

where $a_{i,n} := \int_{(i-1)/n}^{i/n} J(s)ds$, $i = 1,\ldots,n$, and $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the order statistics based upon the sample $(X_1,\ldots,X_n)$. The first general theorem on the asymptotic normality of $\widehat{L}_n(J)$ is established by Chernoff et al. [2]. Since then, a large number of authors have studied the asymptotic behavior of L-statistics. A partial list consists of Bickel [20], Shorack [21, 22], Stigler [23, 24], Ruymgaart and Van Zuijlen [25], Sen [26], Boos [27], Mason [28], and Singh [29]. Indeed, we have

$$\sqrt{n}\left(\widehat{L}_n(J) - L(J)\right) \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, \sigma^2(J)\right), \quad \text{as } n \longrightarrow \infty, \tag{1.12}$$

provided that

$$\sigma^2(J) := \int_0^1 \int_0^1 (\min(s,t) - st) J(s) J(t) dQ(s) Q(t) < \infty. \tag{1.13}$$

In other words, for a given function J, condition (1.13) excludes the class of distributions F for which $\sigma^2(J)$ is infinite. For example, if we take $J = 1$, $L(J)$ is equal to the expected value EX and hence the natural estimator of $\widehat{L}_n(J)$ is the sample mean $\overline{X}_n$. In this case, result (1.12) corresponds to the classical central limit theorem which is valid only when the variance of F is finite. How then can be construct confidence bounds for the mean of a df when its variance is infinite? This situation arises when df F belongs to the domain of attraction of α-stable laws (heavy-tailed) with characteristic exponent $\alpha \in (1, 2)$; see Section 2. This question was answered by Peng [30, 31] who proposed an alternative asymptotically normal estimator for the mean. Remark 3.3 below shows that this situation also arises for the sample trimmed L-moments m_i when $1/2 < \alpha < 2/3$ and for the sample risk two-sided deviation $\Delta_r(X)$ when $1/(r + 1/2) < \alpha < 1/r$ for any $0 < r < 1$. To solve this problem in a more general setting, we propose, by means of the extreme value theory, asymptotically normal estimators of L-functionals for heavy-tailed distributions for which $\sigma^2(J) = \infty$.

The remainder of this paper is organized as follows. Section 2 is devoted to a brief introduction on the domain of attraction of α-stable laws. In Section 3 we define, via the extreme value approach, a new asymptotically normal estimator of L-functionals and state our main results. Applications to trimmed L-moments, risk measures, and related quantities are given in Section 4. All proofs are deferred to Section 5.

2. Domain of Attraction of α-Stable Laws

A df is said to belong to the domain of attraction of a stable law with stability index $0 < \alpha \leq 2$, notation: $F \in D(\alpha)$, if there exist two real sequences $A_n > 0$ and C_n such that

$$A_n^{-1}\left(\sum_{i=1}^{n} X_i - C_n\right) \xrightarrow{\mathfrak{D}} S_\alpha(\sigma, \delta, \mu), \quad \text{as } n \longrightarrow \infty, \tag{2.1}$$

where $S_\alpha(\sigma, \delta, \mu)$ is a stable distribution with parameters $0 < \alpha \leq 2$, $-1 \leq \delta \leq +1$, $\sigma > 0$ and $-\infty < \mu < +\infty$ (see, e.g., Samorodnitsky and Taqqu [32]). This class of distributions was introduced by Lévy during his investigations of the behavior of sums of independent random variables in the early 1920s [33]. $S_\alpha(\sigma, \delta, \mu)$ is a rich class of probability distributions that allow skewness and thickness of tails and have many important mathematical properties. As shown in early work by Mandelbrot (1963) and Fama [34], it is a good candidate to accommodate heavy-tailed financial series and produces measures of risk based on the tails of distributions, such as the Value-at-Risk. They also have been proposed as models for many types of physical and economic systems, for more details see Weron [35]. This class of distributions have nice heavy-tail properties. More precisely, if we denote by $G(x) := \mathbb{P}(|X| \leq x) = F(x) - F(-x), x > 0$, the df of $Z := |X|$, then the tail behavior of $F \in D(\alpha)$, for $0 < \alpha < 2$, may be described by the following

(i) The tail $1 - G$ is regularly varying at infinity with index $-\alpha$. That is

$$\lim_{t\to\infty} \frac{(1 - G(xt))}{(1 - G(t))} = x^{-\alpha}, \quad \text{for any } x > 0. \tag{2.2}$$

(ii) There exists $0 \leq p \leq 1$ such that

$$\lim_{x\to\infty} \frac{1 - F(x)}{1 - G(x)} = p, \qquad \lim_{x\to\infty} \frac{F(-x)}{1 - G(x)} = 1 - p =: q. \tag{2.3}$$

Let, for $0 < s < 1, K(s) := \inf\{x > 0 : G(x) \geq s\}$ be the quantile function pertaining to G and $Q_1(s) := \max(-Q(1 - s), 0)$ and $Q_2(s) := \max(Q(s), 0)$. Then Proposition A.3 in a work by Csörgő et al. [36] says that the set of conditions above is equivalent to the following.

(i′) $K(1 - \cdot)$ is regularly varying at 0 with index $-1/\alpha$. That

$$\lim_{s\downarrow 0} \frac{K(1 - xs)}{K(1 - s)} = x^{-1/\alpha}, \quad \text{for any } x > 0. \tag{2.4}$$

(ii′) There exists $0 \leq p \leq 1$ such that

$$\lim_{s\downarrow 0} \frac{Q_1(1 - s)}{K(1 - s)} = p^{1/\alpha}, \qquad \lim_{s\downarrow 0} \frac{Q_2(1 - s)}{K(1 - s)} = (1 - p)^{1/\alpha} =: q^{1/\alpha}. \tag{2.5}$$

Our framework is a second-order condition that specifies the rate of convergence in statement (i′). There exists a function A, not changing sign near zero, such that

$$\lim_{s\downarrow 0} (A(s))^{-1} \left(\frac{K(1 - xs)}{K(1 - s)} - x^{-1/\alpha} \right) = x^{-1/\alpha} \frac{x^{\varrho} - 1}{\varrho}, \quad \text{for any } x > 0, \tag{2.6}$$

where $\varrho \leq 0$ is the second-order parameter. If $\varrho = 0$, interpret $(x^{\varrho} - 1)/\varrho$ as $\log x$. The second-order condition for heavy-tailed distributions has been introduced by de Haan and Stadtmüller [37].

3. Estimating L-Functionals When $F \in D(\alpha)$

3.1. Extreme Quantile Estimation

The right and left extreme quantiles of small enough level t of df F, respectively, are two reals x_R and x_L defined by $1 - F(x_R) = t$ and $F(x_L) = t$, that is, $x_R = Q(1 - t)$ and $x_L = Q(t)$.

The estimation of extreme quantiles for heavy-tailed distributions has got a great deal of interest, see for instance Weissman [38], Dekkers and de Haan [39], Matthys and Beirlant [40] and Gomes et al. [41]. Next, we introduce one of the most popular quantile estimators. Let $k = k_n$ and $\ell = \ell_n$ be sequences of integers (called trimming sequences) satisfying $1 < k <$ n, $1 < \ell < n$, $k \to \infty$, $\ell \to \infty$, $k/n \to 0$ and $\ell/n \to 0$, as $n \to \infty$. Weissman's estimators of extreme quantiles x_R and x_L are defined, respectively, by

$$\begin{aligned} \widehat{x}_L &= \widehat{Q}_L(t) := \left(\frac{k}{n}\right)^{1/\alpha_L} X_{k,n} t^{-1/\alpha_L}, \quad \text{as } t \downarrow 0, \\ \widehat{x}_R &= \widehat{Q}_R(1-t) := \left(\frac{\ell}{n}\right)^{1/\alpha_R} X_{n-\ell,n} t^{-1/\alpha_R}, \quad \text{as } t \downarrow 0, \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} \widehat{\alpha}_L &= \widehat{\alpha}_L(k) := \left(\frac{1}{k}\sum_{i=1}^{k} \log^+(-X_{i,n}) - \log^+(-X_{k,n})\right)^{-1}, \\ \widehat{\alpha}_R &= \widehat{\alpha}_R(\ell) := \left(\frac{1}{\ell}\sum_{i=1}^{\ell} \log^+(X_{n-i+1,n}) - \log^+(X_{n-\ell,n})\right)^{-1} \end{aligned} \tag{3.2}$$

are two forms of Hill's estimator [42] for the stability index α which could also be estimated, using the order statistics $Z_{1,n} \leq \cdots \leq Z_{n,n}$ associated to a sample $(Z_1, \ldots, Z_n)$ from Z, as follows:

$$\widehat{\alpha} = \widehat{\alpha}(m) := \left(\frac{1}{m}\sum_{i=1}^{m} \log^+(Z_{n-i+1,n}) - \log^+(Z_{n-m,n})\right)^{-1}, \tag{3.3}$$

with $\log^+ u := \max(0, \log u)$ and $m = m_n$ being an intermediate sequence fulfilling the same conditions as k and ℓ. Hill's estimator has been thoroughly studied, improved, and even generalized to any real-valued tail index. Its weak consistency was established by Mason [43] assuming only that the underlying distribution is regularly varying at infinity. The almost sure convergence was proved by Deheuvels et al. [44] and more recently by Necir [45]. The asymptotic normality has been investigated, under various conditions on the distribution tail, by numerous workers like, for instance, Csörgő and Mason [46], Beirlant and Teugels [47], and Dekkers et al. [48].

3.2. A Discussion on the Sample Fractions k and ℓ

Extreme value-based estimators rely essentially on the numbers k and ℓ of lower- and upper-order statistics used in estimate computation. Estimators $\widehat{\alpha}_L$ and $\widehat{\alpha}_R$ have, in general, substantial variances for small values of k and ℓ and considerable biases for large values of k and ℓ. Therefore, one has to look for optimal values for k and ℓ, which balance between these two vices.

Numerically, there exist several procedures for the thorny issue of selecting the numbers of order statistics appropriate for obtaining good estimates of the stability index α;

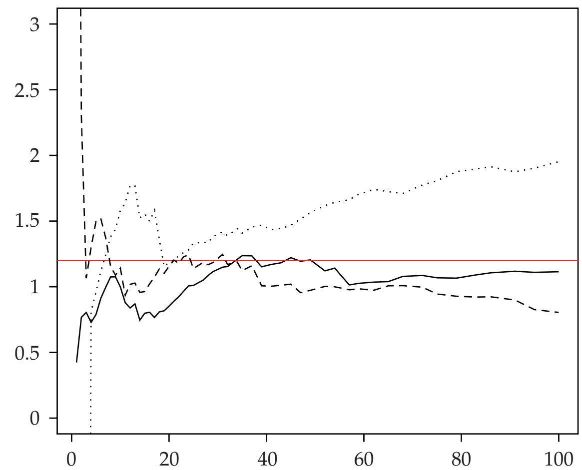


Figure 1: Plots of Hill estimators, as functions of the number of extreme statistics, $\widehat{\alpha}$ (solid line), $\widehat{\alpha}_R$ (dashed line), and $\widehat{\alpha}_L$ (dotted line) of the characteristic exponent α of a stable distribution skewed to the right, based on 1000 observations with 50 replications. The horizontal line represents the true value of $\alpha = 1.2$.

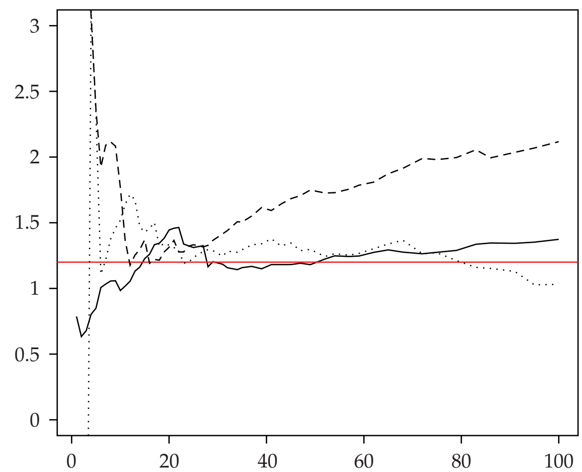


Figure 2: Plots of Hill estimators, as functions of the number of extreme statistics, $\widehat{\alpha}$ (solid line), $\widehat{\alpha}_R$ (dashed line), and $\widehat{\alpha}_L$ (dotted line) of the characteristic exponent α of a stable distribution skewed to the left, based on 1000 observations with 50 replications. The horizontal line represents the true value of $\alpha = 1.2$.

see, for example, Dekkers and de Haan [49], Drees and Kaufmann [50], Danielsson et al. [51], Cheng and Peng [52] and Neves and Alves [53]. Graphically, the behaviors of $\widehat{\alpha}_L$, $\widehat{\alpha}_R$, and $\widehat{\alpha}$ as functions of k, ℓ and m, respectively, are illustrated by Figures 1, 2, and 3 drawn by means of the statistical software R [54]. According to Figure 1, $\widehat{\alpha}_R$ is much more suitable than $\widehat{\alpha}_L$ when estimating the stability index of a distribution which is skewed to the right ($\delta > 0$) whereas Figure 2 shows that $\widehat{\alpha}_L$ is much more reliable than $\widehat{\alpha}_R$ when the distribution is skewed to the left ($\delta < 0$). In the case where the distribution is symmetric ($\delta = 0$), both estimators seem to be equally good as seen in Figure 3. Finally, it is worth noting that, regardless of the distribution skewness, estimator $\widehat{\alpha}$, based on the top statistics pertaining to the absolute value of X, works well and gives good estimates for the characteristic exponent α.

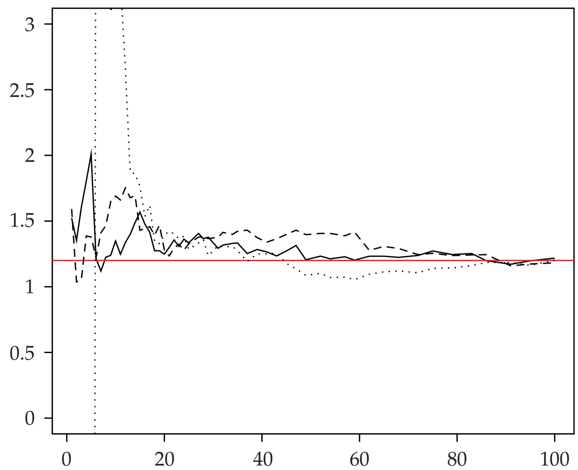


Figure 3: Plots of Hill estimators, as functions of the number of extreme statistics, $\widehat{\alpha}$ (solid line), $\widehat{\alpha}_R$ (dashed line) and $\widehat{\alpha}_L$ (dotted line) of the characteristic exponent α of a symmetric stable distribution, based on 1000 observations with 50 replications. The horizontal line represents the true value of $\alpha = 1.2$.

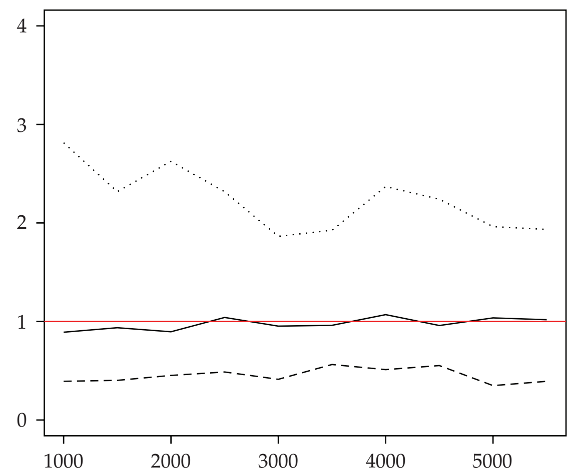


Figure 4: Plots of the ratios of the numbers of extreme statistics, as functions of the sample size, for a stable symmetric distribution $S_{1.2}(1, 0, 0)$ (solid line), a stable distribution skewed to the right $S_{1.2}(1, 0.5, 0)$ (dashed line) and a stable distribution skewed to the left $S_{1.2}(1, -0.5, 0)$ (dotted line).

It is clear that, in general, there is no reason for the trimming sequences k and ℓ to be equal. We assume that there exists a positive real constant θ such that $\ell/k \to \theta$ as $n \to \infty$. If the distribution is symmetric, the value of θ is equal to 1; otherwise, it is less or greater than 1 depending on the sign of the distribution skewness. For an illustration, see Figure 4 where we plot the ratio θ for several increasing sample sizes.

3.3. Some Regularity Assumptions on J

For application needs, the following regularity assumptions on function J are required:

(H1) J is differentiable on $(0,1)$,

(H2) $\lambda := \lim_{s\downarrow 0} \dfrac{J(1-s)}{J(s)} < \infty$,

(H3) both $J(s)$ and $J(1-s)$ are regularly varying at zero with common index $\beta \in \mathbb{R}$.

(H4) there exists a function $a(\cdot)$ not changing sign near zero such that

$$\lim_{t\downarrow 0} \frac{J(xt)/J(t) - x^{\beta}}{a(t)} = x^{\beta}\frac{x^{\omega}-1}{\omega}, \quad \text{for any } x > 0, \tag{3.4}$$

where $\omega \leq 0$ is the second-order *parameter.*

The three remarks below give more motivations to this paper.

Remark 3.1. Assumption (H3) has already been used by Mason and Shorack [55] to establish the asymptotic normality of trimmed L-statistics. Condition (H4) is just a refinement of (H3) called the second order condition that is required for quantile function K in (2.6).

Remark 3.2. Assumptions (H1)–(H4) are satisfied by all weight functions $(J_i)_{i=2,4}$ with $(\beta,\lambda) = (1,\pm 1)$ (see Section 4.1) and by function J_r in (1.9) with $(\beta,\lambda) = (r-1,-1)$. These two examples show that the constants β and λ may be positive or negative depending on application needs.

Remark 3.3. L-functionals $L(J)$ exist for any $0 < \alpha < 2$ and $\beta \in \mathbb{R}$ such that $1/\alpha - \beta < 1$. However, Lemma 5.4 below shows that for $1/\alpha - \beta > 1/2$ we have $\sigma^2(J) = \infty$. Then, recall (1.3); whenever $1/2 < \alpha < 2/3$, the trimmed L-moments exist however $\sigma^2(J_i) = \infty, i = 1,\ldots,4$. Likewise, recall (1.9); whenever $1/(r+1/2) < \alpha < 1/r$, the two-sided deviation $\Delta_r(X)$ exists while $\sigma^2(J_r) = \infty$.

3.4. Defining the Estimator and Main Results

We now have all the necessary tools to introduce our estimator of $L(J)$, given in (1.1), when $F \in D(\alpha)$ with $0 < \alpha < 2$. Let $k = k_n$ and $\ell = \ell_n$ be sequences of integers satisfying $1 < k < n$, $1 < \ell < n$, $k \to \infty$, $\ell \to \infty$, $k/n \to 0$, $\ell/n \to 0$, and the additional condition $\ell/k \to \theta < \infty$ as $n \to \infty$. First, we must note that since $1 + \beta - 1/\alpha > 0$ (see Remark 3.3) and since both $\widehat{\alpha}_L$ and $\widehat{\alpha}_R$ are consistent estimators of α (see, Mason [43]), then we have for all large n

$$\mathbb{P}\left(1+\beta-\frac{1}{\widehat{\alpha}_L} > 0\right) = \mathbb{P}\left(1+\beta-\frac{1}{\widehat{\alpha}_R} > 0\right) = 1 + o(1). \tag{3.5}$$

Observe now that $L(J)$ defined in (1.1) may be split in three integrals as follows:

$$L(J) = \int_0^{k/n} J(t)Q(t)dt + \int_{k/n}^{1-\ell/n} J(t)Q(t)dt + \int_{1-\ell/n}^{1} J(t)Q(t)dt =: T_{L,n} + T_{M,n} + T_{R,n}. \tag{3.6}$$

Substituting $\widehat{Q}_L(t)$ and $\widehat{Q}_R(1-t)$ for $Q(t)$ and $Q(1-t)$ in $T_{L,n}$ and $T_{R,n}$, respectively and making use of assumption (H3) and (3.5) yield that for all large n

$$\begin{aligned}
\int_0^{k/n} J(t)\widehat{Q}_L(t)dt &= \left(\frac{k}{n}\right)^{1/\alpha_L} X_{k,n}\int_0^{k/n} t^{-1/\alpha_L} J(t)dt = (1+o(1))\frac{(k/n)J(k/n)}{1+\beta-1/\widehat{\alpha}_L}X_{k,n}, \\
\int_0^{\ell/n} J(1-t)\widehat{Q}_R(1-t)dt &= \left(\frac{\ell}{n}\right)^{1/\alpha_R} X_{n-\ell,n}\int_0^{\ell/n} t^{-1/\alpha_R} J(1-t)dt \\
&= (1+o(1))\frac{(\ell/n)J(1-\ell/n)}{1+\beta-1/\widehat{\alpha}_R}X_{n-\ell,n}.
\end{aligned} \tag{3.7}$$

Hence we may estimate $T_{L,n}$ and $T_{R,n}$ by

$$\widehat{T}_{L,n} := \frac{(k/n)J(k/n)}{1+\beta-1/\widehat{\alpha}_L}X_{k,n}, \qquad \widehat{T}_{R,n} := \frac{(\ell/n)J(1-\ell/n)}{1+\beta-1/\widehat{\alpha}_R}X_{n-\ell,n}, \tag{3.8}$$

respectively. As an estimator of $T_{M,n}$ we take the sample one that is

$$\widehat{T}_{M,n} := \int_{k/n}^{1-\ell/n} J(t)Q_n(t)dt = \sum_{i=k+1}^{n-\ell} a_{i,n}X_{i,n}, \tag{3.9}$$

with the same constants $a_{i,n}$ as those in (1.11). Thus, the final form of our estimator is

$$\widehat{L}_{k,\ell}(J) = \frac{(k/n)J(k/n)}{1+\beta-1/\widehat{\alpha}_L}X_{k,n} + \sum_{i=k+1}^{n-\ell} a_{i,n}X_{i,n} + \frac{(\ell/n)J(1-\ell/n)}{1+\beta-1/\widehat{\alpha}_R}X_{n-\ell,n}. \tag{3.10}$$

A universal estimator of $L(J)$ may be summarized by

$$\widehat{L}_n^*(J) = \widehat{L}_{k,\ell}(J)\mathbb{I}\left(\sigma^2(J)=\infty\right) + \widehat{L}_n(J)\mathbb{I}\left(\sigma^2(J)<\infty\right), \tag{3.11}$$

where $\widehat{L}_n(J)$ is as in (1.11). More precisely

$$\widehat{L}_n^*(J) = \widehat{L}_{k,\ell}(J)\mathbb{I}(A(\alpha,\beta)) + \widehat{L}_n(J)\mathbb{I}\left(\overline{A}(\alpha,\beta)\right), \tag{3.12}$$

where $A(\alpha,\beta) := \{(\alpha,\beta)\in(0,2)\times\mathbb{R} : 1/2 < 1/\alpha-\beta < 1\}$ and $\overline{A}(\alpha,\beta)$ is its complementary in $(0,2)\times\mathbb{R}$.

Note that for the particular case $\ell = k$ and $J = 1$ the asymptotic normality of the trimmed mean $\widehat{T}_{M,n}$ has been established in Theorem 1 of Csörgő et al. [56]. The following theorem gives the asymptotic normality of $\widehat{T}_{M,n}$ for more general trimming sequences k and ℓ and weighting function J. For convenience, we set, for any $0 < x < 1/2$ and $0 < y < 1/2$,

$$\sigma^2(x, y; J) := \int_x^{1-y} \int_x^{1-y} (\min(s,t) - st) J(s) J(t) dQ(s) Q(t) < \infty, \tag{3.13}$$

and let $\sigma_n^2(J) := \sigma^2(k/n, \ell/n; J)$.

Theorem 3.4. *Assume that $F \in D(\alpha)$ with $0 < \alpha < 2$. For any measurable function J satisfying assumption (H3) with index $\beta \in \mathbb{R}$ such that $0 < 1/\alpha - \beta < 1$ and for any sequences of integers k and ℓ such that$1 < k < n$, $1 < \ell < n$, $k \to \infty$, $\ell \to \infty$, $k/n \to 0$, and $\ell/n \to 0$, as $n \to \infty$, there exists a probability space (Ω, A, P) carrying the sequence $X_1, X_2, \ldots$ and a sequence of Brownian bridges $\{B_n(s), 0 \le s \le 1, n = 1, 2, \ldots\}$ such that one has for all large n*

$$\frac{\sqrt{n}\left(\widehat{T}_{M,n} - T_{M,n}\right)}{\sigma_n(J)} = -\frac{\int_{k/n}^{1-\ell/n} J(s) B_n(s) ds}{\sigma_n(J)} + o_p(1), \tag{3.14}$$

and therefore

$$\frac{\sqrt{n}\left(\widehat{T}_{M,n} - T_{M,n}\right)}{\sigma_n(J)} \xrightarrow{\mathfrak{D}} \mathcal{N}(0,1) \quad \text{as } n \longrightarrow \infty. \tag{3.15}$$

The asymptotic normality of our estimator is established in the following theorem.

Theorem 3.5. *Assume that $F \in D(\alpha)$ with $0 < \alpha < 2$. For any measurable function J satisfying assumptions (H1)–(H4) with index $\beta \in \mathbb{R}$ such that $1/2 < 1/\alpha - \beta < 1$, and for any sequences of integers k and ℓ such that $1 < k < n$, $1 < \ell < n$, $k \to \infty$, $\ell \to \infty$, $k/n \to 0$, $\ell/n \to 0$, $\ell/k \to \theta < \infty$, and $\sqrt{k} a(k/n) A(k/n) \to 0$ as $n \to \infty$, one has*

$$\frac{\sqrt{n}\left(\widehat{L}_{k,\ell}(J) - L(J)\right)}{\sigma_n(J)} \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, \sigma_0^2\right), \quad \text{as } n \longrightarrow \infty, \tag{3.16}$$

where

$$\sigma_0^2 = \sigma_0^2(\alpha, \beta) := (\alpha\beta + 1)(2\alpha\beta + 2 - \alpha) \times \left(\frac{2\alpha^2 + (\beta\alpha - 1)^2 + 2\alpha(\beta\alpha - 1)}{2\left((1+\beta)\alpha - 1\right)^4} + \frac{1}{(1+\beta)\alpha - 1} \right) + 1. \tag{3.17}$$

The following corollary is more practical than Theorem 3.5 as it directly provides confidence bounds for $L(J)$.

Corollary 3.6. *Under the assumptions of Theorem 3.5 one has*

$$\frac{\sqrt{n}\left(\widehat{L}_{k,\ell}(J) - L(J)\right)}{(\ell/n)^{1/2} J(1-\ell/n) X_{n-\ell,n}} \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, V^2\right), \quad \text{as } n \longrightarrow \infty, \tag{3.18}$$

where

$$\begin{aligned} V^2 = V^2(\alpha,\beta,\lambda,\theta,p) :=& \left(1 + \lambda^{-2}\left(\frac{q}{p}\right)^{-2/\alpha} \theta^{-2\beta+2/\alpha-1}\right) \\ &\times \left(\frac{2\alpha^2 + (\beta\alpha-1)^2 + 2\alpha(\beta\alpha-1)}{2((1+\beta)\alpha-1)^4} + \frac{1}{(1+\beta)\alpha-1}\right) + 1, \end{aligned} \tag{3.19}$$

with (p,q) *as in statement (ii) of Section 2 and* (λ,β) *as in assumptions (H2)-(H3) of Section 3.*

3.5. Computing Confidence Bounds for $L(J)$

The form of the asymptotic variance V^2 in (3.20) suggests that, in order to construct confidence intervals for $L(J)$, an estimate of p is needed as well. Using the intermediate order statistic $Z_{n-m,n}$, de Haan and Pereira [57] proposed the following consistent estimator for p :

$$\widehat{p}_n = \widehat{p}_n(m) := \frac{1}{m}\sum_{i=1}^{n} \mathbb{I}\{X_i > Z_{n-m,n}\}, \tag{3.20}$$

where $m = m_n$ is a sequence of integers satisfying $1 < m < n$, $m \to \infty$, and $m/n \to 0$, as $n \to \infty$ (the same as that used in (3.3)).

Let J be a given weight function satisfying (H1)–(H4) with fixed constants β and λ. Suppose that, for n large enough, we have a realization $(x_1,\ldots,x_n)$ of a sample $(X_1,\ldots,X_n)$ from rv X with df F fulfilling all assumptions of Theorem 3.5. The $(1-\varsigma)$-confidence intervals for $L(J)$ will be obtained via the following steps.

Step 1. Select the optimal numbers k^*, ℓ^*, and m^* of lower- and upper-order statistics used in (3.2) and (3.3).

Step 2. Determine $X_{k^*,n}, X_{n-\ell^*,n}, J(k^*/n), J(1-\ell^*/n)$, and $\theta^* := \ell^*/k^*$.

Step 3. Compute, using (3.2), $\widehat{\alpha}_L^* := \widehat{\alpha}_L(k^*)$ and $\widehat{\alpha}_R^* := \widehat{\alpha}_R(\ell^*)$. Then deduce, by (3.10), the estimate $\widehat{L}_{k^*,\ell^*}(J)$.

Step 4. Use (3.3) and (3.20) to compute $\widehat{\alpha}^* := \widehat{\alpha}(m^*)$ and $\widehat{p}_n^* := \widehat{p}_n(m^*)$. Then deduce, by (3.19), the asymptotic standard deviation

$$V^* := \sqrt{V^2(\widehat{\alpha}^*,\beta,\lambda,\theta^*,\widehat{p}_n^*)}. \tag{3.21}$$

Finally, the lower and upper $(1-\varsigma)$-confidence bounds for $L(J)$, respectively, will be

$$\begin{aligned}&\widehat{L}_{k^*,\ell^*}(J) - z_{\varsigma/2}\frac{\sqrt{\ell^*}V^* X_{n-\ell^*,n} J(1-\ell^*/n)}{n},\\&\widehat{L}_{k^*,\ell^*}(J) + z_{\varsigma/2}\frac{\sqrt{\ell^*}V^* X_{n-\ell^*,n} J(1-\ell^*/n)}{n},\end{aligned} \tag{3.22}$$

where $z_{\varsigma/2}$ is the $(1-\varsigma/2)$ quantile of the standard normal distribution $\mathcal{N}(0,1)$ with $0<\varsigma<1$.

4. Applications

4.1. TL-Skewness and TL-Kurtosis When $F \in D(\alpha)$

When the distribution mean EX exists, the skewness and kurtosis coefficients are, respectively, defined by $L_1 := \mu_3/\mu_2^{3/2}$ and $L_2 := \mu_4/\mu_2^2$ with $\mu_k := E(X-EX)^k$, $k = 2, 3$, and 4 being the centered moments of the distribution. They play an important role in distribution classification, fitting models, and parameter estimation, but they are sensitive to the behavior of the distribution extreme tails and may not exist for some distributions such as the Cauchy distribution. Alternative measures of skewness and kurtosis have been proposed; see, for instance, Groeneveld [58] and Hosking [3]. Recently, Elamir and Seheult [4] have used the trimmed L-moments to introduce new parameters called TL-skewness and TL-kurtosis that are more robust against extreme values. For example, when the trimming parameter equals one, the TL-skewness and TL-kurtosis measures are, respectively, defined by

$$\upsilon_1 := \frac{m_3}{m_2}, \qquad \upsilon_2 := \frac{m_4}{m_2}, \tag{4.1}$$

where $m_i, i = 2,3,4$, are the trimmed L-moments defined in Section 1. The corresponding weight functions of (1.3) are defined as follows:

$$\begin{aligned}J_2(s) &:= 6s(1-s)(2s-1),\\J_3(s) &:= \frac{20}{3}s(1-s)\left(5s^2-5s+1\right),\\J_4(s) &:= \frac{15}{2}s(1-s)\left(14s^3-21s^2+9s-1\right).\end{aligned} \tag{4.2}$$

If we suppose that $F \in D(\alpha)$ with $1/2 < \alpha < 2/3$, then, in view of the results above, asymptotically normal estimators for υ_1 and υ_2 will be, respectively,

$$\widehat{\upsilon}_1 := \frac{\widehat{m}_3}{\widehat{m}_2}, \qquad \widehat{\upsilon}_2 := \frac{\widehat{m}_4}{\widehat{m}_2}, \tag{4.3}$$

where

$$\widehat{m}_i = \begin{cases} \dfrac{-6(k/n)^2}{2-1/\widehat{\alpha}_L} X_{k,n} + \displaystyle\sum_{j=k+1}^{n-\ell} a_{j,n}^{(i)} X_{j,n} + \dfrac{6(\ell/n)^2}{2-1/\widehat{\alpha}_R} X_{n-\ell,n}, & \text{for } i = 2, \\ \dfrac{20(k/n)^2}{3(2-1/\widehat{\alpha}_L)} X_{k,n} + \displaystyle\sum_{j=k+1}^{n-\ell} a_{j,n}^{(i)} X_{j,n} + \dfrac{20(\ell/n)^2}{3(2-1/\widehat{\alpha}_R)} X_{n-\ell,n}, & \text{for } i = 3, \\ \dfrac{-15(k/n)^2}{2(2-1/\widehat{\alpha}_L)} X_{k,n} + \displaystyle\sum_{j=k+1}^{n-\ell} a_{j,n}^{(i)} X_{j,n} + \dfrac{15(\ell/n)^2}{2(2-1/\widehat{\alpha}_R)} X_{n-\ell,n}, & \text{for } i = 4, \end{cases} \tag{4.4}$$

with $a_{j,n}^{(i)} := \int_{(j-1)/n}^{j/n} J_i(s)ds$, $i = 2, 3, 4$, and $j = 1, \dots, n$.

Theorem 4.1. *Assume that $F \in D(\alpha)$ with $1/2 < \alpha < 2/3$. For any sequences of integers k and ℓ such that $1 < k < n$, $1 < \ell < n$, $k \to \infty$, $\ell \to \infty$, $k/n \to 0$, $\ell/n \to 0$, $\ell/k \to \theta < \infty$, and $\sqrt{k}a(k/n)A(k/n) \to 0$ as $n \to \infty$, one has, respectively, as $n \to \infty$,*

$$\begin{aligned} \frac{\sqrt{n}(\widehat{v}_1 - v_1)}{(\ell/n)^{3/2} X_{n-\ell,n}} &\xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, V_1^2\right), \\ \frac{\sqrt{n}(\widehat{v}_2 - v_2)}{(\ell/n)^{3/2} X_{n-\ell,n}} &\xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, V_2^2\right), \end{aligned} \tag{4.5}$$

where

$$V_1^2 := \frac{36}{m_2^2}\left(1 - \frac{9m_3}{10m_2}\right)^2 \sigma^{*2}, \qquad V_2^2 := \frac{225}{4m_2^2}\left(1 - \frac{4m_4}{5m_2}\right)^2 \sigma^{*2}, \tag{4.6}$$

with

$$\sigma^{*2} := \left(1 + (q/p)^{-2/\alpha}\theta^{2/\alpha - 3}\right) \times \left(\frac{2\alpha^2 + (\alpha-1)^2 + 2\alpha(\alpha-1)}{2(2\alpha-1)^4} + \frac{1}{2\alpha - 1}\right) + 1. \tag{4.7}$$

4.2. Risk Two-Sided Deviation When $F \in D(\alpha)$

Recall that the risk two-sided deviation is defined by

$$\Delta_r(X) := \int_0^1 J_r(s) Q(s) ds, \quad 0 < r < 1, \tag{4.8}$$

where

$$J_r(s) := \frac{r}{2} \frac{s^{1-r} - (1-s)^{1-r}}{s^{1-r}(1-s)^{1-r}}, \quad 0 < s < 1. \tag{4.9}$$

An asymptotically normal estimator for $\Delta_r(X)$, when $1/(r+1/2) < \alpha < 1/r$, is

$$\widehat{\Delta}_r(X) = -\frac{r(k/n)^r}{2r - 4/\widehat{\alpha}_L} X_{k,n} + \sum_{j=k+1}^{n-\ell} a_{j,n}^{(r)} X_{j,n} + \frac{r(\ell/n)^r}{2r - 4/\widehat{\alpha}_R} X_{n-\ell,n}, \tag{4.10}$$

where

$$a_{j,n}^{(r)} = \frac{1}{2}\left[\left(1 - \frac{i}{n}\right)^{-r} - \left(1 - \frac{i-1}{n}\right)^{-r} - \left(\frac{i}{n}\right)^{-r} + \left(\frac{i-1}{n}\right)^{-r}\right], \tag{4.11}$$

$j = 1, \ldots, n$.

Theorem 4.2. *Assume that $F \in D(\alpha)$ with $0 < \alpha < 2$ such that $1/(r+1/2) < \alpha < 1/r$, for any $0 < r < 1$. Then, for any sequences of integers k and ℓ such that$1 < k < n$, $1 < \ell < n$, $k \to \infty, \ell \to \infty$, $k/n \to 0$, $\ell/n \to 0, \ell/k \to \theta < \infty$, and $\sqrt{k}a(k/n)A(k/n) \to 0$ as $n \to \infty$, one has, as $n \to \infty$,*

$$\frac{\sqrt{n}\left(\widehat{\Delta}_r(X) - \Delta_r(X)\right)}{(\ell/n)^{r-1/2} X_{n-\ell,n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, V_r^2\right), \tag{4.12}$$

where

$$V_r^2 := \frac{r^2}{4}\left(1 + \left(\frac{q}{p}\right)^{-2/\alpha} \theta^{2/\alpha - 2r + 1}\right) \times \left(\frac{2\alpha^2 + (r\alpha - \alpha - 1)^2 + 2\alpha(r\alpha - \alpha - 1)}{2(r\alpha - 1)^4} + \frac{1}{r\alpha - 1}\right) + 1. \tag{4.13}$$

5. Proofs

First we begin by the following three technical lemmas.

Lemma 5.1. *Let f_1 and f_2 be two continuous functions defined on $(0,1)$ and regularly varying at zero with respective indices $\kappa > 0$ and $-\tau < 0$ such that $\kappa < \tau$. Suppose that f_1 is differentiable at zero, then*

$$\lim_{x \downarrow 0} \frac{\int_x^{1/2} f_1(s)\, df_2(s)}{f_1(x) f_2(x)} = \frac{\tau}{\kappa - \tau}. \tag{5.1}$$

Lemma 5.2. *Under the assumptions of Theorem 3.5, one has*

$$\lim_{n \to \infty} \frac{\int_{k/n}^{1-\ell/n} (s(1-s))^{1/2-\nu} J(s)\, dQ(s)}{(k/n)^{1/2-\nu} J(k/n) Q(k/n)} = \frac{1 + \lambda \theta^{1/2-\nu+\beta-1/\alpha} (q/p)^{1/\alpha}}{\alpha(1/2 - \nu + \beta) - 1}, \tag{5.2}$$

for any $0 < \nu < 1/4$.

Lemma 5.3. *For any* $0 < x < 1/2$ *and* $0 < y < 1/2$ *one has*

$$\sigma^2(x,y;J) = xc^2(x) + yc^2(1-y) + \int_x^{1-y} c^2(t)dt - \left(xc(x) + yc(1-y) + \int_x^{1-y} c(t)dt\right)^2, \tag{5.3}$$

where $c(s) := \int_{1/2}^{s} J(t)dQ(t)$, $0 < s < 1/2$.

Lemma 5.4. *Under the assumptions of Theorem 3.5, one has*

$$\lim_{n\to\infty} \frac{(k/n)J^2(k/n)Q^2(k/n)}{\sigma_n^2(J)} = w^2,$$
$$\lim_{n\to\infty} \frac{(\ell/n)J^2(1-\ell/n)Q^2(1-\ell/n)}{\sigma_n^2(J)} = \lambda^2\left(\frac{q}{p}\right)^{2/\alpha}\theta^{2\beta-2/\alpha+1}w^2, \tag{5.4}$$

where

$$w^2 := \frac{(\alpha\beta+1)(2\alpha\beta+2-\alpha)}{2\left(1+\lambda^2(q/p)^{2/\alpha}\theta^{2\beta-2/\alpha+1}\right)}. \tag{5.5}$$

Proof of Lemma 5.1. Let f_1' denote the derivative of f_1. Applying integration by parts, we get, for any $0 < x < 1/2$,

$$\int_x^{1/2} f_1(s)df_2(s) = f_1\left(\frac{1}{2}\right)f_2\left(\frac{1}{2}\right) - f_1(x)f_2(x) - \int_x^{1/2} f_1'(s)f_2(s)ds. \tag{5.6}$$

Since the product f_1f_2 is regularly varying at zero with index $-\tau+\kappa < 0$, then $f_1(x)f_2(x) \to 0$ as $x \downarrow 0$. Therefore

$$\lim_{x\downarrow 0}\frac{\int_x^{1/2} f_1(s)df_2(s)}{f_1(x)f_2(x)} = -1 - \lim_{x\downarrow 0}\frac{\int_x^{1/2} f_1'(s)f_2(s)ds}{f_1(x)f_2(x)}. \tag{5.7}$$

By using Karamata's representation (see, e.g., Seneta [59]), it is easy to show that

$$xf_1'(x) = \kappa(1+o(1))f_1(x), \quad \text{as } x \downarrow 0. \tag{5.8}$$

Hence

$$\lim_{x\downarrow 0}\frac{\int_x^{1/2} f_1(s)df_2(s)}{f_1(x)f_2(x)} = -1 - \kappa\lim_{x\downarrow 0}\frac{\int_x^{1/2} f_1'(s)f_2(s)ds}{xf_1'(x)f_2(x)}. \tag{5.9}$$

It is clear that (5.8) implies that f_1' is regularly varying at zero with index $\kappa - 1$; therefore $f_1' f_2$ is regularly varying with index $-\tau + \kappa - 1 < 0$. Then, Theorem 1.2.1 by de Haan [60, page 15] yields

$$\lim_{x \downarrow 0} - \frac{\int_x^{1/2} f_1'(s) f_2(s) ds}{x f_1'(x) f_2(x)} = \frac{1}{\kappa - \tau}. \tag{5.10}$$

This completes the proof of Lemma 5.1. □

Proof of Lemma 5.2. We have

$$\begin{aligned} I_n &:= \int_{k/n}^{1-\ell/n} (s(1-s))^{1/2-\nu} J(s) dQ(s) = \int_{k/n}^{1/2} (s(1-s))^{1/2-\nu} J(s) dQ(s) \\ &- \int_{1-\ell/n}^{1/2} (s(1-s))^{1/2-\nu} J(1-s) dQ(1-s) =: I_{1n} - I_{2n}. \end{aligned} \tag{5.11}$$

By taking, in Lemma 5.1, $f_1(s) = (s(1-s))^{1/2-\nu} J(s)$ and $f_2(s) = Q(s)$ with $\kappa = 1/2 - \nu + \beta, \tau = 1/\alpha$, and $x = k/n$, we get

$$\lim_{n \to \infty} \frac{I_{1n}}{(k/n)^{1/2-\nu} J(k/n) Q(k/n)} = \frac{1/\alpha}{1/2 - \nu + \beta - 1/\alpha}. \tag{5.12}$$

Likewise if we take $f_1(s) = (s(1-s))^{1/2-\nu} J(1-s)$ and $f_2(s) = Q(1-s)$ with $\kappa = 1/2 - \nu + \beta, \tau = 1/\alpha$, and $x = \ell/n$, we have

$$\lim_{n \to \infty} \frac{I_{2n}}{(\ell/n)^{1/2-\nu} J(1-\ell/n) Q(1-\ell/n)} = \frac{1/\alpha}{1/2 - \nu + \beta - 1/\alpha}. \tag{5.13}$$

Note that statement (ii′) of Section 2 implies that

$$\lim_{s \downarrow 0} Q(1-s)/Q(s) = -\left(\frac{q}{p}\right)^{1/\alpha}. \tag{5.14}$$

The last two relations, together with assumption (H2) and the regular variation of $Q(1-s)$, imply that

$$\lim_{n \to \infty} \frac{I_{2n}}{(k/n)^{1/2-\nu} J(k/n) Q(k/n)} = -\frac{(q/p)^{1/\alpha} \lambda \theta^{1/2-\nu+\beta-1/\alpha}/\alpha}{1/2 - \nu + \beta - 1/\alpha}. \tag{5.15}$$

This achieves the proof of Lemma 5.2. □

Proof of Lemma 5.3. We will use similar techniques to those used by Csörgő et al. [36, Proposition A.2]. For any $0 < s < 1/2$, we set

$$W_{x,y}(t) := \begin{cases} c(1-y) & \text{for } 1-y \le t < 1, \\ c(t) & \text{for } x < t < 1-y, \\ c(x) & \text{for } 0 < t \le x. \end{cases} \tag{5.16}$$

Then $\sigma^2(x, y; J)$ may be rewritten into

$$\sigma^2(x, y; J) = \int_0^1 W_{x,y}^2(s)ds - \left(\int_0^1 W_{x,y}(s)ds\right)^2, \tag{5.17}$$

and the result of Lemma 5.3 follows immediately. □

Proof of Lemma 5.4. From Lemma 5.3 we may write

$$\frac{\sigma_n^2(J)}{(k/n)J^2(k/n)Q^2(k/n)} = T_{n1} + T_{n2} + T_{n3} + T_{n4}, \tag{5.18}$$

where

$$\begin{gathered} T_{n1} := \frac{(k/n)c^2(k/n)}{(k/n)J^2(k/n)Q^2(k/n)}, \qquad T_{n2} := \frac{(\ell/n)c^2(1-\ell/n)}{(k/n)J^2(k/n)Q^2(k/n)}, \\ T_{n3} := \frac{\int_{k/n}^{1-\ell/n} c^2(t)dt}{(k/n)J^2(k/n)Q^2(k/n)}, \\ T_{n4} := \frac{\left((k/n)c(k/n) + (\ell/n)c(1-\ell/n) + \int_{k/n}^{1-\ell/n} c(t)dt\right)^2}{(k/n)J^2(k/n)Q^2(k/n)}. \end{gathered} \tag{5.19}$$

By the same arguments as in the proof of Lemma 5.2, we infer that

$$\begin{gathered} \lim_{n\to\infty} \frac{c(k/n)}{J(k/n)Q(k/n)} = \frac{1}{\alpha\beta - 1}, \\ \lim_{n\to\infty} \frac{c(1-\ell/n)}{J(k/n)Q(k/n)} = \frac{\lambda(q/p)^{1/\alpha}\theta^{\beta-1/\alpha}}{\alpha\beta - 1}. \end{gathered} \tag{5.20}$$

Therefore

$$\lim_{n\to\infty} T_{n1} = \frac{1}{(\alpha\beta-1)^2}, \qquad \lim_{n\to\infty} T_{n2} = \frac{\lambda^2(q/p)^{2/\alpha}\theta^{2\beta-2/\alpha+1}}{(\alpha\beta-1)^2}. \tag{5.21}$$

Next, we consider the third term T_{n3} which may be rewritten into

$$T_{n3} = \frac{\int_{k/n}^{1/2} c^2(t)dt}{(k/n)J^2(k/n)Q^2(k/n)} + \frac{\int_{1/2}^{1-\ell/n} c^2(t)dt}{(k/n)J^2(k/n)Q^2(k/n)}. \tag{5.22}$$

Observe that

$$\frac{\int_{k/n}^{1/2} c^2(t)dt}{(k/n)J^2(k/n)Q^2(k/n)} = \left(\frac{c(k/n)}{J(k/n)Q(k/n)}\right)^2 \frac{\int_{k/n}^{1/2} c^2(t)dt}{(k/n)c^2(k/n)}. \tag{5.23}$$

It is easy to verify that function $c^2(\cdot)$ is regularly varying at zero with index $2(\beta - 1/\alpha)$. Thus, by Theorem 1.2.1 by de Haan [60] we have

$$\lim_{n\to\infty} \frac{\int_{k/n}^{1/2} c^2(t)dt}{(k/n)c^2(k/n)} = \frac{\alpha}{2 - 2\alpha\beta - \alpha}. \tag{5.24}$$

Hence

$$\lim_{n\to\infty} \frac{\int_{k/n}^{1/2} c^2(t)dt}{(k/n)J^2(k/n)Q^2(k/n)} = \frac{\alpha}{(\alpha\beta - 1)^2(2 - 2\alpha\beta - \alpha)}. \tag{5.25}$$

By similar arguments we show that

$$\lim_{n\to\infty} \frac{\int_{1/2}^{1-\ell/n} c^2(t)dt}{(k/n)J^2(k/n)Q^2(k/n)} = \frac{\alpha(q/p)^{2/\alpha}\lambda^2\theta^{2\beta-2/\alpha+1}}{(\alpha\beta - 1)^2(2 - 2\alpha\beta - \alpha)}. \tag{5.26}$$

Therefore

$$\lim_{n\to\infty} T_{n3} = \frac{1 + \alpha(q/p)^{2/\alpha}\lambda^2\theta^{2\beta-2/\alpha+1}}{(\alpha\beta - 1)^2(2 - 2\alpha\beta - \alpha)}. \tag{5.27}$$

By analogous techniques we show that $T_{n4} \to 0$ as $n \to \infty$; we omit details. Summing up the three limits of $T_{ni}, i = 1,2,3$, achieves the proof of the first part of Lemma 5.4. As for the second assertion of the lemma, we apply a similar procedure. □

5.1. Proof of Theorem 3.4

Csörgő et al. [36] have constructed a probability space (Ω, A, P) carrying an infinite sequence $\xi_1, \xi_2, \ldots$ of independent rv's uniformly distributed on $(0,1)$ and a sequence of Brownian bridges $\{B_n(s), 0 \le s \le 1, n = 1,2,\ldots\}$ such that, for the empirical process,

$$\varphi_n(s) := n^{1/2}\{\Gamma_n(s) - s\}, \quad 0 \le s \le 1, \tag{5.28}$$

where $\Gamma_n(\cdot)$ is the uniform empirical df pertaining to the sample $(\xi_1,\ldots,\xi_n)$; we have for any $0 \le \nu < 1/4$ and for all large n

$$\sup_{1/n \le s \le 1-1/n} \frac{\left|\varphi_n(s) - B_n(s)\right|}{(s(1-s))^{1/2-\nu}} = O_p\left(n^{-\nu}\right). \tag{5.29}$$

For each $n \ge 1$, let $\xi_{1,n} \le \cdots \le \xi_{n,n}$ denote the order statistics corresponding to $(\xi_1,\ldots,\xi_n)$. Note that for each n, the random vector $(Q(\xi_{1,n}),\ldots,Q(\xi_{n,n}))$ has the same distribution as $(X_{1,n},\ldots,X_{n,n})$. Therefore, for $1 \le i \le n$, we shall use the rv's $Q(\xi_{i,n})$ to represent the rv's $X_{i,n}$, and without loss of generality, we shall be working, in all the following proofs, on the probability space above. According to this convention, the term $\widehat{T}_{M,n}$ defined in (3.9) may be rewritten into

$$\widehat{T}_{M,n} = \int_{\xi_{k,n}}^{\xi_{n-\ell,n}} Q(s)\, d\Psi(\Gamma_n(s)), \tag{5.30}$$

where $\Psi(s) := \int_0^s J(t)dt$. Integrating by parts yields

$$\frac{n^{1/2}\left(\widehat{T}_{M,n} - T_{M,n}\right)}{\sigma_n(J)} = \Delta_{1,n} + \Delta_{2,n} + \Delta_{3,n}, \tag{5.31}$$

where

$$\begin{aligned}
\Delta_{1,n} &:= -\frac{n^{1/2}\int_{k/n}^{1-\ell/n}\{\Psi(\Gamma_n(s)) - \Psi(s)\}dQ(s)}{\sigma_n(J)}, \\
\Delta_{2,n} &:= \frac{n^{1/2}\int_{k/n}^{\xi_{k,n}}\{\Psi(\Gamma_n(s)) - \Psi(k/n)\}dQ(s)}{\sigma_n(J)}, \\
\Delta_{3,n} &:= \frac{n^{1/2}\int_{\xi_{n-\ell,n}}^{1-\ell/n}\{\Psi(\Gamma_n(s)) - \Psi(1-\ell/n)\}dQ(s)}{\sigma_n(J)}.
\end{aligned} \tag{5.32}$$

Next, we show that

$$\Delta_{1,n} \xrightarrow{\mathfrak{D}} \mathcal{N}(0,1) \quad \text{as } n \longrightarrow \infty, \tag{5.33}$$

$$\Delta_{i,n} \xrightarrow{p} 0 \quad \text{as } n \longrightarrow \infty \text{ for } i = 2,3. \tag{5.34}$$

Making use of the mean-value theorem, we have for each n

$$\Psi(\Gamma_n(s)) - \Psi(s) = (\Gamma_n(s) - s)J(\vartheta_n(s)), \tag{5.35}$$

where $\{\vartheta_n(s)\}_{n\geq 1}$ is a sequence of rv's with values in the open interval of endpoints $s \in (0,1)$ and $\Gamma_n(s)$. Therefore

$$\Delta_{1,n} = \frac{-\int_{k/n}^{1-\ell/n} \varphi_n(s) J(\vartheta_n(s)) dQ(s)}{\sigma_n(J)}. \tag{5.36}$$

This may be rewritten into

$$\begin{aligned}\Delta_{1,n} &= -\frac{\int_{k/n}^{1-\ell/n} \varphi_n(s) J(s) dQ(s)}{\sigma_n(J)} - \frac{\int_{k/n}^{1-\ell/n} \varphi_n(s) J(s)\{J(\vartheta_n(s)) - J(s)\} dQ(s)}{\sigma_n(J)} \\ &=: \Delta_{1,n}^{*} + \Delta_{1,n}^{**}.\end{aligned} \tag{5.37}$$

Note that

$$\begin{aligned}&\frac{\int_{k/n}^{1-\ell/n} \left|\varphi_n(s) - B_n(s)\right| J(s) dQ(s)}{\sigma_n(J)} \\ &\leq \sup_{k/n \leq s \leq 1-\ell/n} \frac{\left|\varphi_n(s) - B_n(s)\right|}{(s(1-s))^{1/2-\nu}} \int_{k/n}^{1-\ell/n} (s(1-s))^{1/2-\nu} |J(s)| dQ(s)/\sigma_n(J),\end{aligned} \tag{5.38}$$

for $0 < \nu < 1/4$, which by (5.29) is equal to

$$\frac{O_p(n^{-\nu}) \int_{k/n}^{1-\ell/n} (s(1-s))^{1/2-\nu} |J(s)| dQ(s)}{\sigma_n(J)}. \tag{5.39}$$

Since we have, from Lemmas 5.2 and 5.3,

$$\left(\frac{n}{k}\right)^{\nu} \int_{k/n}^{1-\ell/n} (s(1-s))^{1/2-\nu} |J(s)| dQ(s)/\sigma_n(J) = O(1), \quad \text{as } n \longrightarrow \infty, \tag{5.40}$$

then the right-hand side of the last inequality is equal to $O_p(k^{-\nu})$ which in turn tends to zero as $n \to \infty$. This implies that as $n \to \infty$

$$\Delta_{1,n}^{*} = -\frac{\int_{k/n}^{1-\ell/n} B_n(s) J(s) dQ(s)}{\sigma_n(J)} + o_p(1). \tag{5.41}$$

Next, we show that $\Delta_{1,n}^{**} = o_p(1)$. Indeed, function J is differentiable on $(0,1)$; then by the mean-value theorem, there exists a sequence $\{\vartheta_n^{*}(s)\}_{n\geq 1}$ of rv's with values in the open interval of endpoints $s \in (0,1)$ and $\vartheta_n(s)$ such that for each n we have

$$\Delta_{1,n}^{**} = \frac{\int_{k/n}^{1-\ell/n} \varphi_n(s) J(s) \{\vartheta_n(s) - s\} J'(\vartheta_n^{*}(s)) dQ(s)}{\sigma_n(J)}. \tag{5.42}$$

From inequalities (3.9) and (3.10) by Mason and Shorack [55], we infer that, for any $0 < \rho < 1$, there exists $0 < M_\rho < \infty$ such that for all large n we have

$$\left| J'(\vartheta_n^*(s)) \right| \leq \frac{M_\rho |J(s)|}{(s(1-s))}, \tag{5.43}$$

for any $0 < s \leq 1/2$. On the other hand, we have for any $0 < s < 1$

$$|\vartheta_n(s) - s| \leq |\Gamma_n(s) - s|. \tag{5.44}$$

Therefore

$$\left| \Delta_{1,n}^{**} \right| \leq \frac{M_\rho n^{-1/2} \int_{k/n}^{1/2} \left(\left| \varphi_n(s) \right|^2 |J(s)| / s(1-s) \right) dQ(s)}{\sigma_n(J)}. \tag{5.45}$$

This implies, since, for each $n \geq 1$, $E|\varphi_n(s)|^2 < s(1-s)$, that

$$E\left| \Delta_{1,n}^{**} \right| \leq \frac{M_\rho n^{-1/2} \int_{k/n}^{1/2} |J(s)| dQ(s)}{\sigma_n(J)}, \tag{5.46}$$

which tends to zero as $n \to \infty$.

Next, we consider the term $\Delta_{2,n}$ which may be rewritten into

$$\Delta_{2,n} = \frac{n^{1/2} \int_{k/n}^{\xi_{k,n}} \{ \Psi(\Gamma_n(s)) - \Psi(s) \} dQ(s)}{\sigma_n(J)} + \frac{n^{1/2} \int_{k/n}^{\xi_{k,n}} \{ \Psi(s) - \Psi(k/n) \} dQ(s)}{\sigma_n(J)}. \tag{5.47}$$

Making use of the mean-value theorem, we get

$$\Delta_{2,n} = \frac{\int_{k/n}^{\xi_{k,n}} \varphi_n(s) J(\mu_n(s)) dQ(s)}{\sigma_n(J)} + \frac{n^{1/2} \int_{k/n}^{\xi_{k,n}} (s - k/n) J(s_n^*) dQ(s)}{\sigma_n(J)}, \tag{5.48}$$

where $\mu_n(s)$ is a sequence of rv's with values in the open interval of endpoints $s \in (k/n, \xi_{k,n})$ and $\Gamma_n(s)$ and s_n^* a sequence of rv's with values in the open interval of endpoints $s \in (k/n, \xi_{k,n})$ and k/n. Again we may rewrite $\Delta_{2,n}$ into

$$\begin{aligned}
\Delta_{2,n} = {} & \frac{\int_{k/n}^{\xi_{k,n}} \varphi_n(s) (J(\mu_n(s)) - J(s)) dQ(s)}{\sigma_n(J)} + \frac{\int_{k/n}^{\xi_{k,n}} \varphi_n(s) J(s) dQ(s)}{\sigma_n(J)} \\
& + n^{1/2} J\left(\frac{k}{n}\right) \int_{k/n}^{\xi_{k,n}} \left(s - \frac{k}{n} \right) \left(\frac{J(s_n^*)}{J(k/n)} - 1 \right) dQ(s) / \sigma_n(J) \\
& + n^{1/2} J\left(\frac{k}{n}\right) \int_{k/n}^{\xi_{k,n}} \left(s - \frac{k}{n} \right) dQ(s) / \sigma_n(J).
\end{aligned} \tag{5.49}$$

Recall that, as $n \to \infty$, both k and ℓ tend to infinity with $k/n \to 0$ and $\ell/n \to 0$. This implies that

$$\frac{n}{k^{1/2}}\left(\xi_{k,n} - \frac{k}{n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \longrightarrow \infty, \tag{5.50}$$

$$\frac{n}{l^{1/2}}\left(\xi_{n-l,n} - 1 + \frac{l}{n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \longrightarrow \infty, \tag{5.51}$$

(see, e.g., Balkema and de Haan [61, page 18]). Next, we use similar arguments to those used in the proof of Theorem 1 by Csörgő et al. [56]. For any $0 < c < \infty$ write

$$\begin{aligned}
\Delta_{2,n}^{(1)}(c) &:= \frac{\int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)} |\varphi_n(s)| |J(\mu_n(s)) - J(s)| dQ(s)}{\sigma_n(J)} \\
&\quad + \frac{\int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)} |J(s)| |\varphi_n(s)| dQ(s)}{\sigma_n(J)}, \\
\Delta_{2,n}^{(2)}(c) &:= \frac{\int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)} n^{1/2} |s - k/n| J(s_n^*) dQ(s)}{\sigma_n(J)} \\
&\quad + \frac{n^{1/2} J(k/n) \int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)} (s - k/n)(J(s_n^*)/J(k/n) - 1) dQ(s)}{\sigma_n(J)}.
\end{aligned} \tag{5.52}$$

Notice that by (5.49)

$$\lim_{c\to\infty} \liminf_{n\to\infty} \mathbb{P}\left\{ |\Delta_{2,n}| \le \Delta_{2,n}^{(1)}(c) + \Delta_{2,n}^{(2)}(c) \right\} \ge \lim_{c\to\infty} \liminf_{n\to\infty} \mathbb{P}\left\{ \left| \xi_{k,n} - \frac{k}{n} \right| \le c\frac{k^{1/2}}{n} \right\}. \tag{5.53}$$

In view of (5.50), this last quantity equals 1. Therefore to establish (5.34) for $i = 2$, it suffices to show that for each $0 < c < \infty$

$$\Delta_{2,n}^{(1)}(c) \xrightarrow{p} 0, \quad \Delta_{2,n}^{(2)}(c) \longrightarrow 0 \quad \text{as } n \longrightarrow \infty. \tag{5.54}$$

By the mean-value theorem, there exists $\{\mu_n^*(s)\}_{n\ge1}$ a sequence of rv's with values in the open interval of endpoints s and $\mu_n(s)$ such that for each n we have

$$J(\mu_n(s)) - J(s) = (\mu_n(s) - s) J'(\mu_n^*(s)). \tag{5.55}$$

Since $|\mu_n(s) - s| \le |\Gamma_n(s) - s|$, then by inequality (5.43) we infer that, for any $0 < \rho < 1$, there exists $0 < M'_\rho < \infty$ such that for all large n we have

$$\left|J(\mu_n(s)) - J(s)\right| \le \frac{M'_\rho n^{-1/2}\left|\varphi_n(s)\right||J(s)|}{(s(1-s))}. \tag{5.56}$$

This implies that the first term in $\Delta_{2,n}^{(1)}(c)$ is less than or equal to

$$M'_\rho n^{-1/2}\int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)}\left(\frac{\varphi_n^2(s)|J(s)|}{(s(1-s))}\right)dQ(s)/\sigma_n(J)\ . \tag{5.57}$$

Since $E(\varphi_n^2(s)) \le s(1-s)$, then the expected value of the previous quantity is less than or equal to

$$\frac{M'_\rho n^{-1/2}\int_{(1-c)k/n}^{(1+c)k/n}|J(s)|dQ(s)}{\sigma_n(J)}. \tag{5.58}$$

Likewise the expected value of the second term in $\Delta_{2,\mathrm{n}}^{(1)}(c)$ is less than or equal to

$$\frac{\int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)}(s(1-s))^{1/2}|J(s)|dQ(s)}{\sigma_n(J)} \le \frac{\left(k/n+c(k^{1/2}/n)\right)^{1/2}\int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)}|J(s)|dQ(s)}{\sigma_n(J)}. \tag{5.59}$$

Fix $1 < \varepsilon < \infty$. It is readily verified that, for all large n, the quantity (5.58) is less than

$$\frac{M'_\rho n^{-1/2}\int_{k/\varepsilon n}^{\varepsilon(k/n)}|J(s)|dQ(s)}{\sigma_n(J)}, \tag{5.60}$$

and the right-hand side of (5.59) is less than

$$2\left(\frac{k}{n}\right)^{1/2}\int_{k/\varepsilon n}^{\varepsilon(k/n)}|J(s)|dQ(s)/\sigma_n(J). \tag{5.61}$$

Therefore for all large n

$$E\left(\Delta_{2,n}^{(1)}(c)\right) \le \left(2 + M'_\rho k^{-1/2}\right)\left(\frac{k}{n}\right)^{1/2}\int_{k/\varepsilon n}^{\varepsilon(k/n)}|J(s)|dQ(s)/\sigma_n(J). \tag{5.62}$$

By routine manipulations, as in the proofs of Lemmas 5.1 and 5.2 (we omit details), we easily show that

$$\lim_{n\to\infty}\left(\frac{k}{n}\right)^{1/2}\int_{k/\varepsilon n}^{\varepsilon(k/n)}|J(s)|dQ(s)/\sigma_n(J) = \frac{w}{\alpha\beta-1}\left(\varepsilon^{1/\alpha-\beta} - \varepsilon^{\beta-1/\alpha}\right). \tag{5.63}$$

Since $k \to \infty$, then for any $0 < c < \infty$

$$\limsup_{n\to\infty} E\left(\Delta_{2,n}^{(1)}(c)\right) \leq 2\left(\varepsilon^{1/\alpha-\beta} - \varepsilon^{\beta-1/\alpha}\right)w, \tag{5.64}$$

for any fixed $1 < \varepsilon < \infty$. This implies that for all $0 < c < \infty$

$$\lim_{n\to\infty} E\left(\Delta_{2,n}^{(1)}(c)\right) = 0. \tag{5.65}$$

Therefore, by Markov inequality, we have the first result of (5.34).

Now, consider the term $\Delta_{2,n}^{(2)}(c)$. Observe that ns^*/k is a sequence of rv's with values in the open interval of endpoints $n\xi_{k,n}/k$ and 1. On the other hand, (5.49) implies that $n\xi_{k,n}/k \xrightarrow{p} 1$ as $n \to \infty$. Hence $ns^*/k \xrightarrow{p} 1$ as well. Then, it is readily checked that, in view of relation (3.9) by Mason and Shorack [55], we have

$$\frac{\sup_{s\in H_n} J(s_n^*)}{J(k/n)} = O_p(1) \quad \text{as } n \longrightarrow \infty. \tag{5.66}$$

Therefore

$$\Delta_{2,n}^{(2)}(c) = O_p(1) J\left(\frac{k}{n}\right) \int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)} n^{1/2}\left|s-\frac{k}{n}\right| dQ(s)/\sigma_n(J). \tag{5.67}$$

Observe that

$$\int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)} n^{1/2}\left|s-\frac{k}{n}\right| dQ(s) \leq c\left(\frac{k}{n}\right)^{1/2} \int_{k/n-c(k^{1/2}/n)}^{k/n+c(k^{1/2}/n)} dQ(s). \tag{5.68}$$

Hence for all large n

$$\Delta_{2,n}^{(2)}(c) = O_p(1) J\left(\frac{k}{n}\right) \int_{k/\varepsilon n}^{\varepsilon(k/n)} dQ(s)/\sigma_n(J), \tag{5.69}$$

for any fixed $1 < \varepsilon < \infty$, we have

$$\Delta_{2,n}^{(2)}(c) = O_p(1)\left(\varepsilon^{1/\alpha-\beta} - \varepsilon^{\beta-1/\alpha}\right)w. \tag{5.70}$$

this means that $\Delta_{2,n}^{(2)}(c) \to 0$ as $n \to \infty$. By the same arguments and making use of (5.51) we show that $\Delta_{3,n} \xrightarrow{p} 0$ as $n \to \infty$ (we omit details), which achieves the proof of Theorem 3.4.

5.2. Proof of Theorem 3.5

Recall (3.8), (3.9), and (3.10) and write

$$\widehat{L}_{k,\ell}(J) - L(J) = \left(\widehat{T}_{L,n} - T_{L,n}\right) + \left(\widehat{T}_{M,n} - T_{M,n}\right) + \left(\widehat{T}_{R,n} - T_{R,n}\right). \tag{5.71}$$

It is easy to verify that

$$\widehat{T}_{R,n} - T_{R,n} = \frac{(k/n)J(k/n)X_{k,n}\widehat{\alpha}_L}{(1+\beta)\widehat{\alpha}_L - 1} - \int_0^{k/n} J(s)Q(t)dt = S_{n1}^L + S_{n2}^L + S_{n3}^L, \tag{5.72}$$

where

$$\begin{aligned} S_{n1}^L &:= \left(\frac{k}{n}\right)J\left(\frac{k}{n}\right)X_{k,n}\left\{\frac{\widehat{\alpha}_L}{(1+\beta)\widehat{\alpha}_L - 1} - \frac{\alpha}{(1+\beta)\alpha - 1}\right\}, \\ S_{n2}^L &:= \frac{\alpha(k/n)Q(k/n)J(k/n)}{(1+\beta)\alpha - 1}\left\{\frac{X_{k,n}}{Q(k/n)} - 1\right\}, \\ S_{n3}^L &:= \frac{\alpha(k/n)J(k/n)Q(k/n)}{(1+\beta)\alpha - 1} - \int_0^{k/n} J(s)Q(t)dt. \end{aligned} \tag{5.73}$$

Likewise we have

$$\widehat{T}_{R,n} - T_{R,n} = \frac{(\ell/n)J(1-\ell/n)X_{n-\ell,n}\widehat{\alpha}_R}{(1+\beta)\widehat{\alpha}_R - 1} - \int_0^{\ell/n} J(s)Q(1-t)dt = S_{n1}^R + S_{n2}^R + S_{n3}^R, \tag{5.74}$$

where

$$\begin{aligned} S_{n1}^R &:= (k/n)J(k/n)X_{k,n}\left\{\frac{\widehat{\alpha}_L}{(1+\beta)\widehat{\alpha}_L - 1} - \frac{\alpha}{(1+\beta)\alpha - 1}\right\}, \\ S_{n2}^R &:= \frac{\alpha(\ell/n)Q(1-\ell/n)J(1-\ell/n)}{(1+\beta)\alpha - 1}\left\{\frac{X_{n-\ell,n}}{Q(1-\ell/n)} - 1\right\}, \\ S_{n3}^R &:= \frac{\alpha(\ell/n)Q(1-\ell/n)J(1-\ell/n)}{(1+\beta)\alpha - 1} - \int_0^{\ell/n} J(s)Q(1-t)dt. \end{aligned} \tag{5.75}$$

It is readily checked that S_{n1}^R may be rewritten into

$$S_{n1}^L = \frac{\widehat{\alpha}_L\alpha(k/n)J(k/n)X_{k,n}}{((1+\beta)\widehat{\alpha}_L - 1)((1+\beta)\alpha - 1)}\left(\frac{1}{\widehat{\alpha}_L} - \frac{1}{\alpha}\right). \tag{5.76}$$

Since $\widehat{\alpha}_L$ is a consistent estimator of α, then for all large n

$$S_{n1}^L = (1 + o_p(1)) \frac{\alpha^2 (k/n) J(k/n) X_{k,n}}{((1+\beta)\alpha - 1)^2} \left(\frac{1}{\widehat{\alpha}_L} - \frac{1}{\alpha} \right). \tag{5.77}$$

In view of Theorems 2.3 and 2.4 of Csörgő and Mason [46], Peng [30], and Necir et al. [17] has been shown that under the second-order condition (2.6) and for all large n

$$\begin{gathered}
\sqrt{k}\alpha \left(\frac{1}{\widehat{\alpha}_L} - \frac{1}{\alpha} \right) = -\sqrt{\frac{n}{k}} B_n \left(\frac{k}{n} \right) + \sqrt{\frac{n}{k}} \int_0^{k/n} \frac{B_n(s)}{s} ds + o_p(1), \\
\sqrt{k} \left(\frac{X_{k,n}}{Q(k/n)} - 1 \right) = -\alpha^{-1} \sqrt{\frac{n}{k}} B_n \left(\frac{k}{n} \right) + o_p(1), \\
\frac{X_{k,n}}{Q} \left(\frac{k}{n} \right) = 1 + o_p(1),
\end{gathered} \tag{5.78}$$

where $\{B_n(s), 0 \le s \le 1, n = 1, 2, \ldots\}$ is the sequence of Brownian bridges defined in Theorem 3.4. This implies that for all large n

$$\begin{aligned}
S_{n1}^L &= (1 + o_p(1)) \frac{\alpha (k^{1/2}/n) J(k/n) Q(k/n)}{((1+\beta)\alpha - 1)^2} \left\{ -\sqrt{\frac{n}{k}} B_n \left(\frac{k}{n} \right) + \sqrt{\frac{n}{k}} \int_0^{k/n} \frac{B_n(s)}{s} ds + o_p(1) \right\}, \\
S_{n2}^L &= \frac{\alpha (k^{1/2}/n) Q(k/n) J(k/n)}{(1+\beta)\alpha - 1} \left\{ -\sqrt{\frac{n}{k}} B_n \left(\frac{k}{n} \right) + o_p(1) \right\}.
\end{aligned} \tag{5.79}$$

Then, in view of Lemma 5.3, we get for all large n

$$\begin{aligned}
\frac{\sqrt{n}(S_{n1}^L + S_{n2}^L)}{\sigma_n(J)} = &-\frac{\alpha w}{((1+\beta)\alpha - 1)^2} \left\{ -\sqrt{\frac{n}{k}} B_n \left(\frac{k}{n} \right) + \sqrt{\frac{n}{k}} \int_0^{k/n} \frac{B_n(s)}{s} ds \right\} \\
&- \frac{w}{(1+\beta)\alpha - 1} \sqrt{\frac{n}{k}} B_n \left(\frac{k}{n} \right) + o_p(1).
\end{aligned} \tag{5.80}$$

By the same arguments (we omit details), we show that for all large n

$$\begin{aligned}
\frac{\sqrt{n}(S_{n1}^R + S_{n2}^R)}{\sigma_n(J)} = &\frac{\alpha w_R}{((1+\beta)\alpha - 1)^2} \left\{ \sqrt{\frac{n}{\ell}} B_n \left(1 - \frac{\ell}{n} \right) - \sqrt{\frac{n}{\ell}} \int_{1-\ell/n}^1 \frac{B_n(s)}{1-s} ds \right\} \\
&+ \frac{w_R}{(1+\beta)\alpha - 1} \left\{ -\sqrt{\frac{n}{\ell}} B_n \left(1 - \frac{\ell}{n} \right) \right\} + o_p(1),
\end{aligned} \tag{5.81}$$

where $w_R := |\lambda|(q/p)^{1/\alpha}\theta^{\beta-1/\alpha+1/2}w$. Similar arguments as those used in the proof of Theorem 1 by Necir et al. [17] yield that

$$\frac{\sqrt{n}S_{n3}^{R}}{\sigma_n(J)} = \frac{\sqrt{n}S_{n3}^{R}}{\sigma_n}(J) = o(1) \quad \text{as } n \longrightarrow \infty. \tag{5.82}$$

Then, by (5.80), (5.81), and (5.82) we get

$$\begin{aligned}
\frac{\sqrt{n}\left(\widehat{L}_{k,\ell}(J) - L(J)\right)}{\sigma_n(J)} = & -\frac{\alpha w}{((1+\beta)\alpha-1)^2}\left\{-\sqrt{\frac{n}{k}}B_n\left(\frac{k}{n}\right) + \sqrt{\frac{n}{k}}\int_0^{k/n}\frac{B_n(s)}{s}ds\right\} \\
& - \frac{w}{(1+\beta)\alpha-1}\sqrt{\frac{n}{k}}B_n\left(\frac{k}{n}\right) + o_p(1) - \frac{\int_{k/n}^{1-\ell/n}J(s)B_n(s)ds}{\sigma_n(J)} \\
& + \frac{\alpha w_R}{((1+\beta)\alpha-1)^2}\left\{\sqrt{\frac{n}{\ell}}B_n\left(1-\frac{\ell}{n}\right) - \sqrt{\frac{n}{\ell}}\int_{1-\ell/n}^{1}\frac{B_n(s)}{1-s}ds\right\} \\
& + \frac{w_R}{(1+\beta)\alpha-1}\left\{-\sqrt{\frac{n}{\ell}}B_n\left(1-\frac{\ell}{n}\right)\right\} + o_p(1).
\end{aligned} \tag{5.83}$$

The asymptotic variance of $\sqrt{n}(\widehat{L}_{k,\ell}(J) - L(J))/\sigma_n(J)$ will be computed by

$$\begin{aligned}
\sigma_0^2 = \lim_{n\to\infty}\Bigg\{ & w^2\frac{\alpha^2}{((1+\beta)\alpha-1)^4}\frac{n}{k}\int_0^{k/n}ds\int_0^{k/n}\frac{\min(s,t)-st}{st}dt \\
& + w^2\frac{(1-\beta\alpha)^2}{((1-\beta)\alpha-1)^4}\frac{n}{k}\frac{k}{n}\left(1-\frac{k}{n}\right) + \frac{\int_{k/n}^{1-\ell/n}dc(s)\int_{k/n}^{1-\ell/n}(\min(s,t)-st)dc(s)}{\sigma_n(J)} \\
& + w_R^2\frac{(1-\beta\alpha)^2}{((1-\beta)\alpha-1)^4}\frac{n}{\ell}\frac{\ell}{n}\left(1-\frac{\ell}{n}\right) \\
& + w_R^2\frac{\alpha^2}{((1+\beta)\alpha-1)^4}\frac{n}{\ell}\int_{1-\ell/n}^{1}ds\int_{1-\ell/n}^{1}\frac{\min(s,t)-st}{st}dt \\
& - 2w^2\frac{\alpha(1-\beta\alpha)}{((1-\beta)\alpha-1)^4}\frac{n}{k}\int_0^{k/n}\frac{t-(k/n)t}{t}dt \\
& + 2w\frac{\alpha}{((1+\beta)\alpha-1)^2}\sqrt{\frac{n}{k}}\int_0^{k/n}ds\int_{k/n}^{1-\ell/n}\frac{s-st}{s}dc(t)/\sigma_n(J)
\end{aligned}$$

$$
\begin{aligned}
&-2ww_R\frac{\alpha(1-\beta\alpha)}{((1+\beta)\alpha-1)^4}\sqrt{\frac{n}{k}}\sqrt{\frac{n}{\ell}}\int_0^{k/n}\frac{s-(1-\ell/n)s}{s}ds\\
&+2ww_R\frac{\alpha^2}{((1+\beta)\alpha-1)^4}\sqrt{\frac{n}{k}}\sqrt{\frac{n}{\ell}}\int_0^{k/n}ds\int_{1-\ell/n}^{1}\frac{\min(s,t)-st}{st}dt\\
&-2w\left(\frac{(1-\beta\alpha)}{((1+\beta)\alpha-1)^2}\right)\sqrt{\frac{n}{k}}\int_{k/n}^{1-\ell/n}\left(\frac{k}{n}-s\left(\frac{k}{n}\right)\right)dc(t)/\sigma_n(J)\\
&-2w\left(\frac{(1-\beta\alpha)}{((1+\beta)\alpha-1)^2}\right)\sqrt{\frac{n}{k}}\int_{k/n}^{1-\ell/n}\left(\frac{k}{n}-s\left(\frac{k}{n}\right)\right)dc(t)/\sigma_n(J)\\
&+2ww_R\frac{(1-\beta\alpha)^2}{((1+\beta)\alpha-1)^4}\sqrt{\frac{n}{k}}\sqrt{\frac{n}{\ell}}\left(\frac{k}{n}-\frac{k}{n}\left(1-\frac{\ell}{n}\right)\right)\\
&-2ww_R\frac{\alpha(1-\beta\alpha)}{((1+\beta)\alpha-1)^4}\int_{1-\ell/n}^{1}\frac{k/n-(k/n)s}{1-s}ds\\
&-2w_R\left(\frac{(1-\beta\alpha)}{((1+\beta)\alpha-1)^2}\right)\sqrt{\frac{n}{\ell}}\int_{k/n}^{1-\ell/n}\left(s-s\left(1-\frac{\ell}{n}\right)\right)dc(t)/\sigma_n(J)\\
&+2w_R\left(\frac{\alpha}{((1+\beta)\alpha-1)^2}\right)\sqrt{\frac{n}{\ell}}\int_{1-\ell/n}^{1}ds\int_{k/n}^{1-\ell/n}\frac{(t-st)}{(1-s)}dc(t)/\sigma_n(J)\\
&-2w_R^2\frac{\alpha(1-\beta\alpha)}{((1+\beta)\alpha-1)^4}\left(\frac{n}{\ell}\right)\int_{1-\ell/n}^{1}\frac{(1-\ell/n)-(1-\ell/n)s}{1-s}ds\Bigg\}.
\end{aligned}
\tag{5.84}
$$

After calculation we get

$$
\begin{aligned}
\sigma_0^2 &= w^2\frac{2\alpha^2}{((1+\beta)\alpha-1)^4}+w^2\frac{(1-\beta\alpha)^2}{((1+\beta)\alpha-1)^4}+1+w_R^2\frac{(1-\beta\alpha)^2}{((1+\beta)\alpha-1)^4}+w_R^2\frac{2\alpha^2}{((1+\beta)\alpha-1)^4}\\
&\quad-2w^2\frac{\alpha(1-\beta\alpha)}{((1+\beta)\alpha-1)^4}+2w\frac{\alpha}{((1+\beta)\alpha-1)^2}w\\
&\quad-2w\frac{(1-\beta\alpha)}{((1+\beta)\alpha-1)^2}w+2w_R\frac{\alpha}{((1+\beta)\alpha-1)^2}w_R\\
&\quad-2w_R\frac{(1-\beta\alpha)}{((1+\beta)\alpha-1)^2}w_R-w_R^2\frac{2\alpha}{((1+\beta)\alpha-1)^4}\\
&=\left(w^2+w_R^2\right)\left[\frac{2\alpha^2+(\beta\alpha-1)^2+2\alpha(\beta\alpha-1)}{((1+\beta)\alpha-1)^4}+\frac{2}{(1+\beta)\alpha-1}\right]+1.
\end{aligned}
\tag{5.85}
$$

Finally, it is easy to verify that

$$w^2 + w_R^2 = \frac{(\alpha\beta + 1)(2\alpha\beta + 2 - \alpha)}{2}. \tag{5.86}$$

This completes the proof of Theorem 3.5.

5.3. Proof of the Corollary

Straightforward by combining Theorem 3.5 and Lemma 5.4, we omit details.

5.4. Proof of Theorem 4.1

We will only present details for the proof concerning the first part of Theorem 4.1. The proof for the second part is very similar. For convenience we set

$$\Delta m_2 := \widehat{m}_2 - m_2, \qquad \Delta m_3 := \widehat{m}_3 - m_3. \tag{5.87}$$

Then we have

$$\widehat{v}_1 - v_1 = \frac{\widehat{m}_3}{\widehat{m}_2} - \frac{m_3}{m_2} = \frac{m_2 \Delta m_3 - m_3 \Delta m_2}{\widehat{m}_2 m_2}. \tag{5.88}$$

Since $\widehat{m}_2$ is consistent estimator of m_2, then for all large n

$$\widehat{v}_1 - v_1 = (1 + o_P(1))\left(\frac{\Delta m_3}{m_2} - \frac{m_3 \Delta m_2}{m_2^2}\right), \tag{5.89}$$

and therefore

$$\frac{\sqrt{n}(\widehat{v}_1 - v_1)}{\sigma_n(J_3)} = \frac{1}{m_2}(C_{1n} + C_{2n}) + o_p(1), \tag{5.90}$$

where

$$C_{1n} := \frac{\sqrt{n}\Delta m_3}{\sigma_n(J_3)}, \qquad C_{2n} := -\frac{m_3\sqrt{n}\Delta m_2}{m_2 \sigma_n(J_3)}. \tag{5.91}$$

In view of (5.83) we may write that for all large n

$$
\begin{aligned}
C_{1n} = &-\frac{\alpha w_1}{(2\alpha-1)^2}\left\{-\sqrt{\frac{n}{k}}B_n\left(\frac{k}{n}\right)+\sqrt{\frac{n}{k}}\int_0^{k/n}\frac{B_n(s)}{s}ds\right\} \\
&-\frac{w_1}{2\alpha-1}\sqrt{\frac{n}{k}}B_n\left(\frac{k}{n}\right)+o_p(1)-\frac{\int_{k/n}^{1-\ell/n}J(s)B_n(s)ds}{\sigma_n(J)} \\
&+\frac{\alpha w_{R,1}}{(2\alpha-1)^2}\left\{\sqrt{\frac{n}{\ell}}B_n\left(1-\frac{\ell}{n}\right)-\sqrt{\frac{n}{\ell}}\int_{1-\ell/n}^{1}\frac{B_n(s)}{1-s}ds\right\} \\
&+\frac{w_{R,1}}{2\alpha-1}\left\{-\sqrt{\frac{n}{\ell}}B_n\left(1-\frac{\ell}{n}\right)\right\}+o_p(1),
\end{aligned}
\tag{5.92}
$$

and, by Lemma 5.4, we infer that $\sigma_n(J_3)/\sigma_n(J_2) \to 10/9$ as $n \to \infty$,

$$
\begin{aligned}
-\frac{10m_2}{9m_3}C_{2n} = &-\frac{\alpha w_1}{(2\alpha-1)^2}\left\{-\sqrt{\frac{n}{k}}B_n\left(\frac{k}{n}\right)+\sqrt{\frac{n}{k}}\int_0^{k/n}\frac{B_n(s)}{s}ds\right\} \\
&-\frac{w_1}{2\alpha-1}\sqrt{\frac{n}{k}}B_n\left(\frac{k}{n}\right)+o_p(1)-\frac{\int_{k/n}^{1-\ell/n}J(s)B_n(s)ds}{\sigma_n(J)} \\
&+\frac{\alpha w_{R,1}}{((1+\beta)\alpha-1)^2}\left\{\sqrt{\frac{n}{\ell}}B_n\left(1-\frac{\ell}{n}\right)-\sqrt{\frac{n}{\ell}}\int_{1-\ell/n}^{1}\frac{B_n(s)}{1-s}ds\right\} \\
&+\frac{w_{R,1}}{2\alpha-1}\left\{-\sqrt{\frac{n}{\ell}}B_n\left(1-\frac{\ell}{n}\right)\right\}+o_p(1),
\end{aligned}
\tag{5.93}
$$

with

$$
w_1 := \frac{(\alpha+1)(2\alpha+2-\alpha)}{2\left(1+(q/p)^{2/\alpha}\theta^{3-2/\alpha}\right)}, \qquad w_{R,1} := (q/p)^{1/\alpha}\theta^{3/2-1/\alpha}w_1. \tag{5.94}
$$

By the same arguments as the proof of Theorem 3.5 we infer that

$$
\frac{\sqrt{n}(\widehat{v}_1-v_1)}{(\ell/n)^{3/2}X_{n-\ell,n}} \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0,V_1^2\right) \quad \text{as } n \longrightarrow \infty. \tag{5.95}
$$

This achieves the proof of Theorem 4.1.

5.5. Proof of Theorem 4.2

Theorem 4.2 is just an application of Theorem 3.5, we omit details.

Acknowledgments

The authors thank anonymous referees for their insightful comments and valuable suggestions. The first author is indebted to Ričardas Zitikis for his constructive advice and also for having introduced him (the first author) to the Actuarial Research Group (ARG) in London, Ontario.

References

[1] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, NY, USA, 1980.

[2] H. Chernoff, J. L. Gastwirth, and M. V. Johns, "Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation," *Annals of Mathematical Statistics*, vol. 38, pp. 52–72, 1967.

[3] J. R. M. Hosking, "*L*-moments: analysis and estimation of distributions using linear combinations of order statistics," *Journal of the Royal Statistical Society Serie B*, vol. 52, pp. 105–124, 1990.

[4] E. A. H. Elamir and A. H. Seheult, "Trimmed *L*-moments," *Computational Statistics and Data Analysis*, vol. 43, no. 3, pp. 299–314, 2003.

[5] J. R. M. Hosking, "Some theory and practical uses of trimmed *L*-moments," *Journal of Statistical Planning and Inference*, vol. 137, no. 9, pp. 3024–3039, 2007.

[6] D. Denneberg, *Non-Additive Measure and Integral*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[7] S. S. Wang, "Insurance pricing and increased limits ratemaking by proportional hazards transforms," *Insurance: Mathematics & Economics*, vol. 17, no. 1, pp. 43–54, 1995.

[8] S. Wang, "An actuarial index of the right-tail risk," *North American Actuarial Journal*, vol. 2, no. 2, pp. 88–101, 1998.

[9] B. L. Jones and R. Zitikis, "Empirical estimation of risk measures and related quantities," *North American Actuarial Journal*, vol. 7, no. 4, pp. 44–54, 2003.

[10] B. L. Jones and R. Zitikis, "Testing for the order of risk measures: an application of *L*-statistics in actuarial science," *Metron*, vol. 63, pp. 193–211, 2005.

[11] B. L. Jones and R. Zitikis, "Risk measures, distortion parameters, and their empirical estimation," *Insurance: Mathematics and Economics*, vol. 41, no. 2, pp. 279–297, 2007.

[12] V. Brazauskas, B. L. Jones, M. L. Puri, and R. Zitikis, "Nested *L*-statistics and their use in comparing the riskiness of portfolios," *Scandinavian Actuarial Journal*, vol. 2007, pp. 162–179, 2007.

[13] V. Brazauskas, B. L. Jones, and R. Zitikis, "Robust fitting of claim severity distributions and the method of trimmed moments," *Journal of Statistical Planning and Inference*, vol. 139, no. 6, pp. 2028–2043, 2009.

[14] F. Greselin, M. L. Puri, and R. Zitikis, "*L*-functions, processes, and statistics in measuring economic inequality and actuarial risks," *Statistics and Its Interface*, vol. 2, pp. 227–245, 2009.

[15] S. S. Wang, "Ordering of risks under PH-transforms," *Insurance: Mathematics and Economics*, vol. 18, no. 2, pp. 109–114, 1996.

[16] S. S. Wang, "A class of distortion operators for pricing financial and insurance risks," *Journal of Risk and Insurance*, vol. 67, no. 1, pp. 15–36, 2000.

[17] A. Necir, D. Meraghni, and F. Meddi, "Statistical estimate of the proportional hazard premium of loss," *Scandinavian Actuarial Journal*, vol. 3, pp. 147–161, 2007.

[18] A. Necir and D. Meraghni, "Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts," *Insurance: Mathematics and Economics*, vol. 45, no. 1, pp. 49–58, 2009.

[19] G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York, NY, USA, 1986.

[20] P. J. Bickel, "Some contributions to the theory of order statistics," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 575–591, University of California Press, Berkeley, Calif, USA, 1967.

[21] G. R. Shorack, "Asymptotic Normality of Linear Combinations of Function of Order Statistics," *Annals of Mathematical Statistics*, vol. 40, pp. 2041–2050, 1969.

[22] G. R. Shorack, "Functions of order statistics," *Annals of Mathematical Statistics*, vol. 43, pp. 412–427, 1972.

[23] S. M. Stigler, "Functions of order statistics," *Annals of Mathematical Statistics*, vol. 43, pp. 412–427, 1969.
[24] S. M. Stigler, "Linear functions of order statistics with smooth weight functions," *Annals of Statistics*, vol. 2, pp. 676–693, 1974.
[25] F. H. Ruymgaart and M. C. A. Van Zuijlen, "Asymptotic normality of linear combinations of functions of order statistics in the non-i.i.d. case," *Indagationes Mathematicae*, vol. 39, no. 5, pp. 432–447, 1977.
[26] P. K. Sen, "An invariance principle for linear combinations of order statistics," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 42, no. 4, pp. 327–340, 1978.
[27] D. D. Boos, "A differential for L-statistics," *Annals of Statistics*, vol. 7, no. 5, pp. 955–959, 1979.
[28] D. M. Mason, "Asymptotic normality of linear combinations of order statistics with a smooth score function," *Annals of Statistics*, vol. 9, pp. 899–908, 1981.
[29] K. Singh, "On asymptotic representation and approximation to normality of L-statistics. I," *Sankhya Series A*, vol. 43, no. 1, pp. 67–83, 1981.
[30] L. Peng, "Estimating the mean of a heavy tailed distribution," *Statistics and Probability Letters*, vol. 52, no. 3, pp. 255–264, 2001.
[31] L. Peng, "Empirical-likelihood-based confidence interval for the mean with a heavy-tailed distribution," *Annals of Statistics*, vol. 32, no. 3, pp. 1192–1214, 2004.
[32] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, NY, USA, 1994.
[33] P. Lévy, *Calcul des Probabilités*, Gauthier-Villars, Paris, France, 1925.
[34] E. F. Fama, "The Behavior of stock market prices," *Journal of Business*, vol. 38, pp. 34–105, 1965.
[35] R. Weron, "Levy-stable distributions revisited: tail index $>$ 2 does not exclude the Levy-stable regime," *International Journal of Modern Physics C*, vol. 12, no. 2, pp. 209–223, 2001.
[36] M. Csörgő, S. Csörgő, L. Horváth, and D. M. Mason, "Normal and stable convergence of integral functions of the empirical distribution function," *Annals of Probability*, vol. 14, pp. 86–118, 1986.
[37] L. de Haan and U. Stadtmüller, "Generalized regular variation of second order," *Journal of the Australian Mathematical Society*, vol. 61, no. 3, pp. 381–395, 1996.
[38] I. Weissman, "Estimation of parameters and large quantiles based on the k largest observations," *Journal of American Statistical Association*, vol. 73, pp. 812–815, 1978.
[39] A. L. M. Dekkers and L. de Haan, "On the estimation of the extreme-value index and large quantile estimation," *Annals of Statistics*, vol. 17, pp. 1795–1832, 1989.
[40] G. Matthys and J. Beirlant, "Estimating the extreme value index and high quantiles with exponential regression models," *Statistica Sinica*, vol. 13, no. 3, pp. 853–880, 2003.
[41] M. I. Gomes, F. Figueiredo, and S. Mendonça, "Asymptotically best linear unbiased tail estimators under a second-order regular variation condition," *Journal of Statistical Planning and Inference*, vol. 134, pp. 409–433, 2005.
[42] B. M. Hill, "A simple approach to inference about the tail of a distribution," *Annals of Statistics*, vol. 3, pp. 1136–1174, 1975.
[43] D. M. Mason, "Laws of the large numbers for sums of extreme values," *Annals of Probability*, vol. 10, pp. 754–764, 1982.
[44] P. Deheuvels, E. Haeusler, and D. M. Mason, "Almost sure convergence of the Hill estimator," *Mathematical Proceedings of the Cambridge Philosophy Society*, vol. 104, pp. 371–381, 1988.
[45] A. Necir, "A functional law of the iterated logarithm for kernel-type estimators of the tail index," *Journal of Statistical Planning and Inference*, vol. 136, no. 3, pp. 780–802, 2006.
[46] M. Csörgő and D. M. Mason, "Central limit theorems for sums of extreme values," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 98, pp. 547–558, 1985.
[47] J. Beirlant and J. Teugels, "Asymptotic normality of Hill's estimator," in *Extreme Value Theory (Oberwolfach, 1987)*, vol. 51 of *Lecture Notes in Statistics*, pp. 148–155, Springer, New York, NY, USA, 1989.
[48] A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan, "A moment estimator for the index of an extreme-value distribution," *Annals of Statistics*, vol. 17, pp. 1833–1855, 1989.
[49] A. L. M. Dekkers and L. Dehaan, "Optimal choice of sample fraction in extreme-value estimation," *Journal of Multivariate Analysis*, vol. 47, no. 2, pp. 173–195, 1993.
[50] H. Drees and E. Kaufmann, "Selecting the optimal sample fraction in univariate extreme value estimation," *Stochastic Processes and Their Applications*, vol. 75, no. 2, pp. 149–172, 1998.
[51] J. Danielsson, L. de Haan, L. Peng, and C. G. de Vries, "Using a bootstrap method to choose the sample fraction in tail estimation," *Journal of Multivariate Analysis*, vol. 76, no. 2, pp. 226–248, 2001.
[52] S. Cheng and L. Peng, "Confidence intervals for the tail index," *Bernoulli*, vol. 7, no. 5, pp. 751–760, 2001.

[53] C. Neves and M. I. F. Alves, "Reiss and Thomas' automatic selection of the number of extremes," *Computational Statistics and Data Analysis*, vol. 47, no. 4, pp. 689–704, 2004.

[54] R. Ihaka and R. Gentleman, "R: a language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 299–314, 1996.

[55] D. M. Mason and G. R. Shorack, "Necessary and sufficient conditions for asymptotic normality of trimmed L-statistics," *Annals of Probability*, vol. 20, pp. 1779–1804, 1992.

[56] S. Csörgő, L. Horváth, and D. M. Mason, "What portion of the sample makes a partial sum asymptotically stable or normal?" *Probability Theory and Related Fields*, vol. 72, no. 1, pp. 1–16, 1986.

[57] L. de Haan and T. T. Pereira, "Estimating the index of a stable distribution," *Statistics and Probability Letters*, vol. 41, no. 1, pp. 39–55, 1999.

[58] R. A. Groeneveld, "A class of quantile measures for kurtosis," *American Statistician*, vol. 52, no. 4, pp. 325–329, 1998.

[59] E. Seneta, *Regularly Varying Functions*, vol. 508 of *Lecture Notes in Mathematics*, Springer, Berlin, Germany, 1976.

[60] L. de Haan, *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*, vol. 32 of *Mathematical Center Tracts*, Center for Mathematics and Computer Science, Amsterdam, The Netherlands, 1970.

[61] A. A. Balkema and L. de Haan, "Limit laws for order statistics," in *Limit Theorems of Probability Theory (Colloquium, Keszthely, 1974)*, vol. 11 of *Colloquia Mathematica Societatis Janos Bolyai*, pp. 17–22, North-Holland, Amsterdam, The Netherlands, 1975.

[62] M. Csörgő and D. M. Mason, "On the asymptotic distribution of weighted uniform empirical and quantile processes in the middle and on the tails," *Stochastic Processes and their Applications*, vol. 21, no. 1, pp. 119–132, 1985.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 965672, 14 pages
doi:10.1155/2010/965672

Research Article

POT-Based Estimation of the Renewal Function of Interoccurrence Times of Heavy-Tailed Risks

Abdelhakim Necir,[1] Abdelaziz Rassoul,[2] and Djamel Meraghni[1]

[1] *Laboratory of Applied Mathematics, University Mohamed Khider, P.O. Box 145, Biskra 07000, Algeria*
[2] *Ecole Nationale Superieure d'Hydraulique, Guerouaou, BP 31, Blida 09000, Algeria*

Correspondence should be addressed to Abdelaziz Rassoul, rsl_aziz@yahoo.fr

Received 13 October 2009; Accepted 24 February 2010

Academic Editor: Ričardas Zitikis

Copyright © 2010 Abdelhakim Necir et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Making use of the peaks over threshold (POT) estimation method, we propose a semiparametric estimator for the renewal function of interoccurrence times of heavy-tailed insurance claims with infinite variance. We prove that the proposed estimator is consistent and asymptotically normal, and we carry out a simulation study to compare its finite-sample behavior with respect to the nonparametric one. Our results provide actuaries with confidence bounds for the renewal function of dangerous risks.

1. Introduction

Let $X_1, X_2, \ldots$ be independent and identically distributed (iid) positive random variables (rvs), representing claim interoccurrence times of an insurance risk, with common distribution function (df) F having finite mean μ and variance σ^2. Let

$$\mathbb{S}_m := \begin{cases} X_1 + \cdots + X_m, & m = 1, 2, \ldots, \\ 0, & m = 0 \end{cases} \tag{1.1}$$

be the claim occurrence times, and define the number of claims recorded over the time interval $[0, t]$ by

$$\mathbb{N}(t) := \max\{m \geq 0,\ \mathbb{S}_m \leq t\}. \tag{1.2}$$

The corresponding renewal function is defined by

$$\mathbb{R}(t) := E[\mathbb{N}(t)] = \sum_{k=1}^{\infty} F^{(k)}(t), \quad t > 0, \tag{1.3}$$

where $F^{(k)}$ is the k-fold convolution of F for $k \geq 1$.

The renewal theory has proved to be a powerful tool in stochastic modeling in a wide variety of applications such as reliability theory, where a renewal process is used to model the successive repairs of a failed machine (see [1]), risk theory, where a renewal process is used to model the successive occurrences of risks (see [2, 3]), inventory theory, where a renewal process is used to model the successive times between demand points (see [4]), manpower planning, where a renewal process is used to model the sequence of resignations from a given job (see [5]), and warranty analysis, where a renewal process is used to model the successive purchases of a new item following the expiry of a free-replacement warranty (see [6]). Therefore, the need for renewal function estimates seems more than pressing in many practical problems. For a summary of renewal theory, one refers to Feller [7], Asmussen [8], and Resnick [9].

Statistical estimation of the renewal function has been considered in several ways. Using a nonparametric approach, Frees [10] introduced two estimators based on the empirical counterparts of F and $F^{(k)}$ by suitably truncating the sum in (1.3). Zhao and Subba Rao [11] proposed an estimation method based on the kernel estimate of the density and the renewal equation. A histogram-type estimator, resembling to the second estimator of Frees, was given by Markovich and Krieger [12].

When $E[X^2] = \infty$, Sgibnev [13] gave an asymptotic approximation of (1.3) as follows:

$$\mathbb{R}(t) - \frac{t}{\mu} \sim \frac{1}{E[X^2]} \int_0^t \left(\int_y^{\infty} \overline{F}(x) dx \right) dy, \tag{1.4}$$

with $\overline{F} := 1 - F$ being the tail of F.

By replacing F by its empirical counterpart F_n in (1.4), Bebbington et al. [14] recently proposed a nonparametric estimator for $\mathbb{R}(t)$ in the case where F is of infinite variance, given by

$$\widetilde{\mathbb{R}}_n(t) := \frac{t}{\widetilde{\mu}} + \frac{1}{\widetilde{\mu}_2} \int_0^t \left(\int_y^{\infty} \overline{F}_n(x) dx \right) dy, \tag{1.5}$$

where $\widetilde{\mu}$ and $\widetilde{\mu}_2$, respectively, represent the first and second sample moments of F. Their main result says that whenever F belongs to the domain of attraction of a stable law S_α with $1/2 < \alpha < 1$ (see, e.g., [15]), the df of $\widetilde{\mathbb{R}}_n(t)$ converges, for suitable normalizing constants, to S_α. This result provides confidence bounds for $\mathbb{R}(t)$ with respect to the quantiles of S_α.

In general, people prefer estimators having simple formulas and carrying some kind of asymptotic normality property in order to facilitate confidence interval construction. From this point of view, the estimator $\widetilde{\mathbb{R}}_n(t)$ may not be as satisfactory to the users as it should be. Then an alternative estimator to $\widetilde{\mathbb{R}}_n(t)$ would be more useful in practice. Our task is to use the extreme value theory tools to construct such an alternative estimator.

Indeed, an important class of models having infinite second-order moments is the set of heavy-tailed distributions (e.g., Pareto, Burr, Student, etc.). A df F is said to be heavy-tailed with tail index $\xi > 0$ if

$$\overline{F}(x) = cx^{-1/\xi}\left(1 + x^{-\delta}\mathbb{L}(x)\right), \quad \text{as } x \longrightarrow \infty, \tag{1.6}$$

for $\xi \in (0,1)$, $\delta > 0$, and some real constant c, with $\mathbb{L}$ a slowly varying function at infinity, that is, $\mathbb{L}(tx)/\mathbb{L}(x) \to 1$ as $x \to \infty$ for any $t > 0$. For details on these functions, see Chapter 0 in Resnick [16] or Seneta [17]. Notice that when $\xi \in (1/2,1)$ we have $\mu < \infty$ and $E[X^2] = \infty$. In this case, an asymptotic approximation of the renewal function $\mathbb{R}(t)$ is given in (1.4).

Prior to Sgibnev [13], Teugels [18] obtained an approximation of $\mathbb{R}(t)$ when F is heavy-tailed with tail index $\xi \in (1/2,1)$:

$$\mathbb{R}(t) - \frac{t}{\mu} \sim \frac{\xi^2 t^2 \overline{F}(t)}{\mu^2(1-\xi)(2\xi - 1)}, \quad \text{as } t \longrightarrow \infty. \tag{1.7}$$

Extreme value theory allows for an accurate modeling of the tails of any unknown distribution, making the (semiparametric) statistical inference more accurate for heavy-tailed distributions. Indeed, the semiparametric approach permits extrapolating beyond the largest value of a given sample while the nonparametric one does not since the empirical df vanishes outside the sample. This represents a big handicap for those dealing with heavy-tailed data.

Extreme value theory has two aspects. The first one consists in approximating the tail distribution by the generalized extreme value (GEV) distribution, thanks to Fisher-Tippett theorem (see [19, 20]). The second aspect (commonly known as POT method) is based on Balkema-de Haan result which says that the distribution of the excesses over a fixed threshold is approximated by the generalized Pareto distribution (GPD) (see [21, 22]). Those interested in extreme value theory and its applications are referred to the textbooks of de Haan and Ferriera [23] and Embrechts et al. [24]. In our situation, we have a fixed threshold equal to the horizon $t = t_n$ (see Section 3). Therefore, the POT method would be the appropriate choice to derive an estimator for $\mathbb{R}(t)$ by exploiting the heavy-tail property of df F used in approximation (1.4). The asymptotic normality of our estimator is established under suitable assumptions.

The remainder of the paper is organized as follows. In Section 2, we introduce the GPD approximation, mostly known as the POT method. A new estimator of the renewal function $\mathbb{R}(t)$ is proposed in Section 3, along with two main results on its limiting behavior. Section 4 is devoted to a simulation study. The proofs are postponed until Section 5.

2. GPD Approximation

The distribution of the excesses, over a "fixed" threshold t, pertaining to df F is defined by

$$F_t(y) := P(X_1 - t \le y \mid X_1 > t), \quad \text{for } y > 0. \tag{2.1}$$

It is shown, in Balkema and de Haan [21] and Pickands [22], that F_t is approximated by a generalized Pareto distribution (GPD) function $\mathbb{G}_{\xi,\beta}$ with shape parameter $\xi \in \mathbb{R}$ and scale parameter $\beta = \beta(t) > 0$, in the following sense:

$$\sup_{y>0} \left| F_t(y) - \mathbb{G}_{\xi,\beta}(y) \right| = O\left(t^{-\delta} \mathbb{L}(t) \right), \quad \text{as } t \longrightarrow \infty, \tag{2.2}$$

where $t^{-\delta}\mathbb{L}(t) \rightarrow 0$ as $t \rightarrow \infty$ for any $\delta > 0$. The GPD function $\mathbb{G}_{\xi,\beta}$ is a two-parameter df defined by

$$\mathbb{G}_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \xi \dfrac{y}{\beta} \right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\dfrac{y}{\beta} \right), & \xi = 0, \end{cases} \tag{2.3}$$

for $0 \leq y < \infty$ if $\xi \geq 0$ and $0 \leq y < -\beta/\xi$ if $\xi < 0$.

Let $Y_1, \ldots, Y_N$ be iid rvs with exact GPD $\mathbb{G}_{\xi,\beta}$. It is well known by standard arguments (see, e.g., [25, Chapter 9]) that there exists, with probability 1 as N tends to infinity, a local maximum $(\widehat{\xi}_N, \widehat{\beta}_N)$ for the Log-Likelihood of $\mathbb{G}_{\xi,\beta}$'s density based on the sample $(Y_1, \ldots, Y_N)$. In this case, by Theorem 3.7 page 447 in the work of Lehmann and Casella [26], we infer that $\widehat{\xi}_N$ and $\widehat{\beta}_N$ are consistent estimators of ξ and β. Moreover, these estimators are asymptotically normal provided that $\xi > -1/2$. The extension to $\xi \leq -1/2$ was investigated by Smith [27].

Suppose now that $Y_1, \ldots, Y_N$ are drawn not from $\mathbb{G}_{\xi,\beta}$, but from F_t. In view of the asymptotic approximation (2.2), Smith [27] has proposed estimates for (ξ, β) via the Maximum Likelihood approach. The obtained estimators $(\widehat{\xi}_N, \widehat{\beta}_N)$ are solutions of the following system:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^{N} \log\left(1 + \xi \frac{y_i}{\beta} \right) &= \xi, \\ \frac{1}{N} \sum_{i=1}^{N} \frac{y_i/\beta}{1 + y_i/\beta} &= \frac{1}{1+\xi}, \end{aligned} \tag{2.4}$$

where $(y_1, \ldots, y_N)$ is a realization of $(Y_1, \ldots, Y_N)$.

Letting $t = t_N \rightarrow \infty$ as $N \rightarrow \infty$ and $\beta_N = t_N \xi$ and making use of (2.2), Smith [28] established, in Theorem 3.2, the asymptotic normality of $(\widehat{\xi}_N, \widehat{\beta}_N)$ as follows:

$$\sqrt{N} \begin{pmatrix} \dfrac{\widehat{\beta}_N}{\beta_N} - 1 \\ \widehat{\xi}_N - \xi \end{pmatrix} \xrightarrow{\mathfrak{D}} \mathcal{N}_2\left(0, \mathbb{Q}^{-1} \right) \quad \text{as } N \longrightarrow \infty, \tag{2.5}$$

where

$$\mathbb{Q}^{-1} = (1+\xi)\begin{pmatrix} 2 & -1 \\ -1 & 1+\xi \end{pmatrix}, \tag{2.6}$$

provided that $\sqrt{N} t_N^{-\delta} \mathbb{L}(t_N) \to 0$ as $N \to \infty$ and $x \mapsto x^{-\delta}\mathbb{L}(x)$ is nonincreasing near infinity. In the case $\sqrt{N} t_N^{-\delta} \mathbb{L}(t_N) \nrightarrow 0$, the limiting distribution in (2.5) is biased. Here $\xrightarrow{\mathfrak{D}}$ denotes convergence in distribution and $\mathcal{N}_2(\omega, \Sigma)$ stands for the bivariate normal distribution with mean vector ω and covariance matrix Σ.

3. Estimating the Renewal Function in Infinite Time

Since we are interested in the renewal function in infinite time, we must assume that time t is large enough and for asymptotic considerations, we will assume that t depends on the sample size n. That is, $t = t_n$, with $t_n \to \infty$ as $n \to \infty$. Relation (1.7) suggests that in order to construct an estimator of $\mathbb{R}(t_n)$, we need to estimate μ, ξ and $\overline{F}(t_n)$. Let $n = n(t)$ be the number of X_is, which are observed on horizon t_n and denoted by

$$N_{t_n} := \operatorname{card}(\{X_i > t_n : 1 \le i \le n\}), \tag{3.1}$$

the number of exceedances over t_n, with $\operatorname{card}(K)$ being the cardinality of set K. Notice that N_{t_n} is a binomial rv with parameters n and $p_n := \overline{F}(t_n)$ for which the natural estimator is $\widehat{p}_n := N_t/n$.

Select, from the sample $(X_1, \ldots, X_n)$, only those observations $X_{i_1}, \ldots, X_{i_{N_{t_n}}}$ that exceed t_n. The N_t excesses

$$E_{j:n} := X_{i_j} - t_n, \quad j = 1, \ldots, N_{t_n} \tag{3.2}$$

are iid rvs with common df F_{t_n}. As seen in Section 2, the maximum likelihood estimators $(\widehat{\xi}_n, \widehat{\beta}_n)$ are solutions of the following system:

$$\begin{gathered} \frac{1}{v_n}\sum_{j1}^{v_n} \log\left(1 + \xi \frac{e_{j:n}}{\beta}\right) = \xi, \\ \frac{1}{v}\sum_{j=1}^{v_n} \frac{e_{j:n}/\beta}{1 + e_{j:n}/\beta} = \frac{1}{1+\xi}, \end{gathered} \tag{3.3}$$

where v_n is an observation of N_{t_n} and the vector $(e_{1:n}, \ldots, e_{v_n:n})$ a realization of $(E_{1:n}, \ldots, E_{N_{t_n}:n})$. Regarding the distribution mean $\mu = E[X_1]$, we know that, for $\xi \in (0, 1/2]$, X_1 has finite variance and therefore μ could naturally be estimated by the sample mean $\overline{X} := n^{-1} S_n$ which, by the Central Limit Theorem (CLT), is asymptotically normal. Whereas for $\xi \in (1/2, 1)$, X_1 has infinite variance, in which case the CLT is no longer valid. This case is frequently met in real insurance data (see, e.g., [29]). Using the GPD approximation, Johansson [30] has proposed an alternative estimator for $\mu = \int_0^\infty x dF(x)$. For each $n \geq 1$, we write μ as the sum of two components:

$$\mu_n^* := \int_0^{t_n} x dF(x), \qquad \tau_n := \int_{t_n}^\infty x dF(x) = -\int_0^\infty (t_n + s) d\overline{F}(t_n + s). \tag{3.4}$$

Johansson [30] defined his estimator of μ, by estimating both $F(x)$ and $\overline{F}(t_n + s)$, as follows:

$$\widehat{\mu}_n^{(J)} := \int_0^{t_n} x dF_n(x) - \int_0^\infty (t_n + s) d\widehat{\overline{F}}(t_n + s), \tag{3.5}$$

where F_n is the empirical df based on the sample $(X_1, \ldots, X_n)$ and $\widehat{\overline{F}}(t_n + s)$ is an estimate of $\overline{F}(t_n + s)$ obtained from the relation

$$\overline{F}_{t_n}(s) = \frac{\overline{F}(t_n + s)}{\overline{F}(t_n)}, \quad s > 0, \tag{3.6}$$

which implies that $\overline{F}(t_n + s) = p_n \overline{F}_{t_n}(s),\ s > 0$. Approximation (2.2) motivates us to estimate $\overline{F}_{t_n}(s)$ by $\widehat{\overline{F}}_{t_n}(s) := \overline{\mathbb{G}}_{\widehat{\xi}_n, \widehat{\beta}_n}(s),\ s > 0$. Hence, an estimate of $\overline{F}(t_n + s)$ is

$$\widehat{\overline{F}}(t_n + s) := \widehat{p}_n \overline{\mathbb{G}}_{\widehat{\xi}_n, \widehat{\beta}_n}(s), \quad s > 0. \tag{3.7}$$

By integrating (3.5), we get

$$\begin{aligned} \widehat{\mu}_n^{(J)} &= \frac{1}{n} \sum_{i=1}^n X_i \mathbf{1}_{\{X_i \leq t_n\}} + \widehat{p}_n \left(t_n + \frac{\widehat{\beta}_n}{1 - \widehat{\xi}_n} \right) \\ &=: \widehat{\mu}_n^* + \widehat{\tau}_n, \end{aligned} \tag{3.8}$$

with $\widehat{\xi}_n \in (0, 1)$ with large probability. Here, $\mathbf{1}_K$ denotes the indicator function of set K. Respectively, substituting $\widehat{\mu}_n^{(J)}$, $\widehat{\xi}_n$, and $\widehat{p}_n$ for μ, ξ and $\overline{F}(t_n)$ in (1.7) yields the following estimator for the renewal function $\mathbb{R}(t_n)$

$$\widehat{\mathbb{R}}_n(t_n) := \frac{t_n}{\widehat{\mu}_n^{(J)}} + \frac{\widehat{\xi}_n^2 t_n^2 \widehat{p}_n}{\widehat{\mu}_n^{(J)2} \left(1 - \widehat{\xi}_n\right)\left(2\widehat{\xi}_n - 1\right)}. \tag{3.9}$$

The asymptotic behavior of $\widehat{\mathbb{R}}_n(t_n)$ is given by the following two theorems.

Theorem 3.1. *Let F be a df fulfilling* (1.6) *with $\xi \in (1/2, 1)$. Suppose that $\mathbb{L}$ is locally bounded in $[x_0, +\infty)$ for $x_0 \geq 0$ and $x \mapsto x^{-\delta}\mathbb{L}(x)$ is nonincreasing near infinity, for some $\delta > 0$. Then, for any $t_n = O(n^{\alpha\xi/4})$ with $\alpha \in (0,1)$, one has*

$$\widehat{\mathbb{R}}_n(t_n) - \mathbb{R}(t_n) = O_{\mathbb{P}}\left(n^{(\alpha/2)(\xi-1/4)-1/2}\right), \quad \text{as } n \longrightarrow \infty. \tag{3.10}$$

Theorem 3.2. *Let F be as in Theorem 3.1. Then for any $t_n = O(n^{\alpha\xi/4})$ with $\alpha \in (4/(1+2\xi\delta), 1)$, we have*

$$\frac{\sqrt{n}}{s_n t_n}\left(\widehat{\mathbb{R}}_n(t_n) - \mathbb{R}(t_n)\right) \xrightarrow{\mathfrak{D}} \mathcal{N}(0,1), \quad \text{as } n \longrightarrow \infty, \tag{3.11}$$

where

$$\begin{aligned} s_n^2 := \theta_1^2 &+ \frac{p_n(1-p_n)}{\gamma_n^2}\left(\theta_2 + \theta_1\left(t_n + \frac{\beta_n}{1-\xi}\right)\right)^2 + \frac{p_n}{\gamma_n^2}\left(\theta_3 + \frac{\theta_1 p_n \beta_n}{(1-\xi)^2}\right)^2 \\ &+ \frac{\theta_1^2 \beta_n^2 p_n^3}{\gamma_n^2(1-\xi)^2} - \frac{\theta_1 \beta_n p_n^2}{\gamma_n^2}\left(\theta_3 + \frac{\theta_1 p_n \beta_n}{(1-\xi)^2}\right) \end{aligned} \tag{3.12}$$

with

$$\begin{aligned} \theta_1 &:= -\frac{1}{\mu^2} - \frac{2\xi^2 t_n p_n}{\mu^3(1-\xi)(2\xi-1)}, \\ \theta_2 &:= \frac{\xi^2 t_n}{\mu^2(1-\xi)(2\xi-1)}, \\ \theta_3 &:= \frac{t_n p_n}{\mu^2(1-\xi)(2\xi-1)}\left(2\xi + \frac{4\xi^3 - 3\xi^2}{(1-\xi)(2\xi-1)}\right), \end{aligned} \tag{3.13}$$

$p_n := \overline{F}(t_n)$, $\beta_n := t_n\xi$, *and* $\gamma_n^2 := \mathrm{Var}(X_1 \mathbf{1}_{\{X_1 \leq t_n\}})$.

4. Simulation Study

In this section, we carry out a simulation study (by means of the statistical software **R**, see [31]) to illustrate the performance of our estimation procedure, through its application to sets of samples taken from two distinct Pareto distributions $F(x) = 1 - x^{-1/\xi}$, $x > 1$ (with tail indices $\xi = 3/4$ and $\xi = 2/3$). We fix the threshold at 4, which is a value above the intermediate statistic corresponding to the optimal fraction of upper-order statistics in each sample. The latter is obtained by applying the algorithm of Cheng and Peng [32]. For each sample size, we generate 200 independent replicates. Our overall results are then taken as the empirical means of the values in the 200 repetitions.

A comparison with the nonparametric estimator is done as well. In the graphical illustration, we plot both estimators versus the sample size ranging from 1000 to 20000.

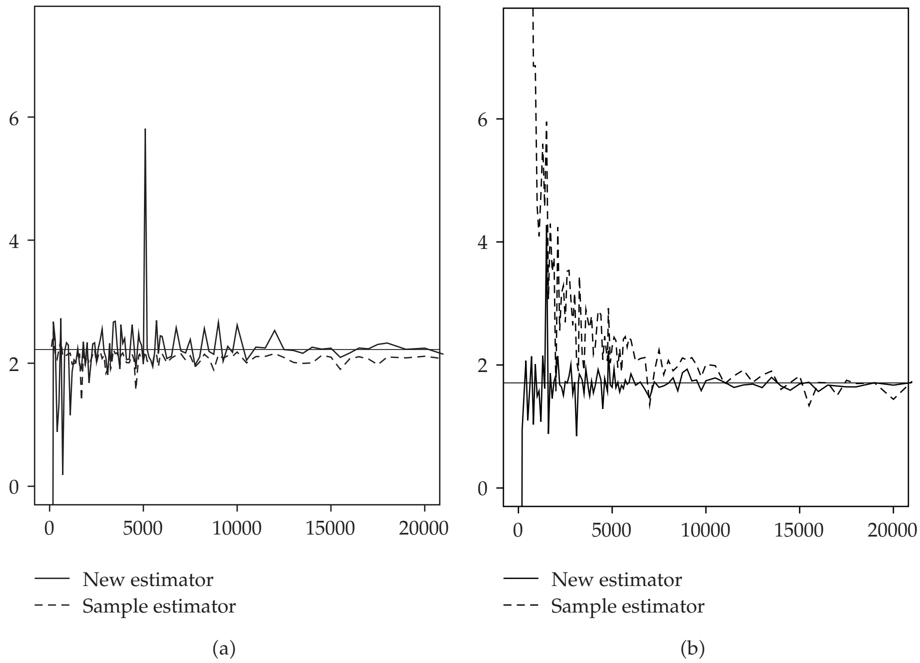


Figure 1: Plots of the new and sample estimators of the renewal function, of interoccurrence times of Pareto-distributed claims with tail indices 2/3 (a) and 3/4 (b), versus the sample size. The horizontal line represents the true value of the renewal function $\mathbb{R}(t)$ evaluated at $t = 4$.

Table 1: Semiparametric and nonparametric estimates of the renewal function of interoccurrence times of Pareto-distributed claims with shape parameter 3/4. Simulations are repeated 200 times for different sample sizes.

	True value $R = 1.708$					
Sample size	Semiparametric $\widehat{R}$			Nonparametric $\widetilde{R}$		
	Mean	Bias	RMSE	Mean	Bias	RMSE
1000	1.696	−0.013	0.250	2.141	0.433	0.553
2000	1.719	0.011	0.183	1.908	0.199	0.288
5000	1.705	−0.003	0.119	1.686	−0.022	0.168

Figure 1 clearly shows that the new estimator is consistent and that it is always better than the nonparametric one. For the numerical investigation, we take samples of sizes 1000, 2000 and 5000. In each case, we compute the semiparametric estimate $\widehat{R}$ as well as the nonparametric estimate $\widetilde{R}$. We also provide the bias and the root mean squared error (rmse).

The results are summarized in Tables 1 and 2 for $\xi = 3/4$ and $\xi = 2/3$ respectively. We notice that, regardless of the tail index value and the sample size, the semiparametric estimation procedure is more accurate than the nonparametric one.

5. Proofs

The following tools will be instrumental for our needs.

Table 2: Semiparametric and nonparametric estimates of the renewal function of interoccurence times of Pareto-distributed claims with shape parameter 2/3. Simulations are repeated 200 times for different sample sizes.

Sample size	True value $R = 2.222$					
	Semiparametric $\widehat{R}$			Nonparametric $\widetilde{R}$		
	Mean	Bias	RMSE	Mean	Bias	RMSE
1000	2.265	0.042	0.185	2.416	0.193	0.229
2000	2.247	0.024	0.157	2.054	−0.167	0.223
5000	2.223	0.001	0.129	2.073	−0.149	0.192

Proposition 5.1. *Let F be a df fulfilling* (1.6) *with $\xi \in (1/2, 1)$, $\delta > 0$, and some real c. Suppose that $\mathbb{L}$ is locally bounded in $[x_0, +\infty)$ for $x_0 \geq 0$. Then for n large enough and for any $t_n = O(n^{\alpha\xi/4})$, $\alpha \in (0,1)$, one has*

$$\begin{gathered}
p_n = c(1 + o(1))n^{-\alpha/4}, \\
\gamma_n^2 = O\left(n^{(\alpha/2)(\xi-1/2)}\right), \\
s_n^2 = O\left(n^{(\alpha/2)(\xi-1/2)}\right), \\
\sqrt{np_n}t_n^{-\delta}\mathbb{L}(t_n) = O\left(n^{-\alpha/8-\alpha\xi\delta/4+1/2}\right),
\end{gathered} \tag{5.1}$$

where p_n, γ_n^2, and s_n^2 are those defined in Theorem 3.2.

Lemma 5.2. *Under the assumptions of Theorem 3.2, one has, for any real numbers u_1, u_2, u_3 and u_4,*

$$\begin{aligned}
&E\left[\exp\left\{iu_1\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\mu}_n^* - \mu_n^*\right) + i\sqrt{np_n}(u_2, u_3)\begin{pmatrix}\dfrac{\widehat{\beta}_n}{\beta_n} - 1 \\ \widehat{\xi}_n - \xi\end{pmatrix} + iu_4\frac{\sqrt{n}(\widehat{p}_n - p_n)}{\sqrt{p_n(1-p_n)}}\right\}\right] \\
&\quad\longrightarrow \exp\left\{-\frac{u_1^2}{2} - \frac{1}{2}(u_2, u_3)\mathbb{Q}^{-1}\begin{pmatrix}u_2 \\ u_3\end{pmatrix} - \frac{u_4^2}{2}\right\}, \quad \text{as } n \longrightarrow \infty,
\end{aligned} \tag{5.2}$$

where $i^2 = -1$.

Proof of the Proposition. We will only prove the second result, the other ones are straightforward from (1.6). Let $x_0 > 0$ be such that $\overline{F}(x) = cx^{-1/\xi}\,(1 + x^{-\delta}\mathbb{L}(x))$, for $x > x_0$. Then for n large enough, we have

$$E\left[X_1\mathbf{1}_{\{X_1 \leq t_n\}}\right] = \int_0^{t_n} x dF(x) = \int_0^{x_0} x dF(x) + \int_{x_0}^{t_n} x dF(x). \tag{5.3}$$

Recall that $\mu < \infty$, hence $\int_0^{x_0} x dF(x) < \infty$. Making use of the proposition assumptions, we get $E[X_1\mathbf{1}_{\{X_1 \leq t_n\}}] = O(1)$ and $E[X_1^2\mathbf{1}_{\{X_1 \leq t_n\}}] = O(t_n^{2-1/\xi})$ and therefore $\gamma_n^2 = O(n^{\alpha/2(\xi/2-1)})$. □

Proof of Lemma 5.2. See Johansson [30]. □

Proof of Theorem 3.1. We may readily check that for all large n,

$$\frac{\left(\widehat{\mathbb{R}}_n(t_n)-\mathbb{R}_n(t_n)\right)}{t_n}\sim A_n+B_n+C_n, \tag{5.4}$$

where

$$\begin{aligned} A_n &:= \left(-\frac{1}{\widehat{\mu}_n^{(J)}\mu}-\frac{\xi^2 t_n p_n\left(\widehat{\mu}_n^{(J)}+\mu\right)}{\widehat{\mu}_n^{(J)2}\mu^2(1-\xi)(2\xi-1)}\right)\left(\widehat{\mu}_n^{(J)}-\mu\right),\\ B_n &:= \frac{\widehat{\xi}_n^2 t_n}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\xi}_n\right)\left(2\widehat{\xi}_n-1\right)}\left(\widehat{p}_n-p_n\right),\\ C_n &:= \frac{t_n p_n}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\xi}_n\right)\left(2\widehat{\xi}_n-1\right)}\times\left(\widehat{\xi}_n+\xi+\frac{2\xi^2\left(\widehat{\xi}_n+\xi\right)-3\xi^2}{(1-\xi)(2\xi-1)}\right)\left(\widehat{\xi}_n-\xi\right). \end{aligned} \tag{5.5}$$

Johansson [30] proved that there exists a bounded sequence k_n such that

$$\widehat{\mu}_n^{(J)}-\mu=O_{\mathbb{P}}\left(\gamma_n\sqrt{\frac{k_n}{n}}\right), \tag{5.6}$$

hence $\widehat{\mu}_n^{(J)}-\mu=O_{\mathbb{P}}(n^{(\alpha/4)(\xi-1/2)-1/2})$. The first result of the proposition yields that

$$t_n p_n\left(\widehat{\mu}_n^{(J)}-\mu\right)=O_{\mathbb{P}}\left(n^{(\alpha/4)(2\xi-3/2)-1/2}\right). \tag{5.7}$$

Since $(\alpha/4)(2\xi-3/2)-1/2<0$, then $t_n p_n\ (\widehat{\mu}_n^{(J)}-\mu)=o_{\mathbb{P}}(1)$. On the other hand, by the CLT we have

$$\widehat{p}_n-p_n=O_{\mathbb{P}}\left(\sqrt{\frac{p_n}{n}}\right), \tag{5.8}$$

then $t_n(\widehat{p}_n-p_n)=O_{\mathbb{P}}(n^{(\alpha/4)(\xi-1/2)-1/2})=o_{\mathbb{P}}(1)$. On the other hand, Smith [28], yields

$$\widehat{\xi}_n-\xi=O_{\mathbb{P}}t_n^{-\delta}\mathbb{L}(t_n), \tag{5.9}$$

it follows that, $\widehat{\xi}_n^2 t_n \ (\widehat{p}_n - p_n) = O_{\mathbb{P}}(n^{(\alpha/4)(\xi(1-2\delta)-1/2)-1/2}) = o_{\mathbb{P}}(1)$, therefore

$$\begin{gathered}
\frac{\widehat{\xi}_n^2 t_n (\widehat{p}_n - p_n)}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\xi}_n\right)\left(2\widehat{\xi}_n - 1\right)} = o_{\mathbb{P}}(1), \\
t_n p_n \left(\widehat{\xi}_n - \xi\right) = O_{\mathbb{P}}\left(n^{(\alpha/4)(\xi(1-\delta)-1)}\right) = o_{\mathbb{P}}(1), \\
\widehat{\xi}_n t_n p_n \left(\widehat{\xi}_n - \xi\right) = O_{\mathbb{P}}\left(n^{(\alpha/4)(\xi(1-2\delta)-1)}\right) = o_{\mathbb{P}}(1), \\
p_n \left(\widehat{\xi}_n - \xi\right) = O_{\mathbb{P}}\left(n^{(-\alpha/4)((1+\xi\delta))}\right) = o_{\mathbb{P}}(1).
\end{gathered} \tag{5.10}$$

Thus,

$$\begin{gathered}
\frac{\left(\widehat{\xi}_n + \xi\right)}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\xi}_n\right)\left(2\widehat{\xi}_n - 1\right)} t_n p_n \left(\widehat{\xi}_n - \xi\right) \xrightarrow{\mathbb{P}} 0, \\
\frac{t_n p_n \left(2\xi^2\left(\widehat{\xi}_n + \xi\right) - 3\xi^2\right)}{\widehat{\mu}_n^{(J)2}\left(1-\widehat{\xi}_n\right)\left(2\widehat{\xi}_n - 1\right)(1-\xi)(2\xi-1)} \left(\widehat{\xi}_n - \xi\right) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \longrightarrow \infty.
\end{gathered} \tag{5.11}$$

Therefore for all large n, we get $\widehat{\mathbb{R}}(t_n) - \mathbb{R}(t_n) = O_{\mathbb{P}}(n^{(\alpha/2)(\xi-1/4)-1/2})$, as sought. □

Proof of Theorem 3.2. From the proof of Theorem 3.1, for all large n, it is easy to verify that

$$\begin{aligned}
\frac{\left(\widehat{\mathbb{R}}_n(t_n) - \mathbb{R}_n(t_n)\right)}{t_n} = {} & \theta_1 (1 + o_{\mathbb{P}}(1)) \left(\widehat{\mu}_n^{(J)} - \mu\right) \\
& + \theta_2 (1 + o_{\mathbb{P}}(1)) (\widehat{p}_n - p_n) \\
& + \theta_3 (1 + o_{\mathbb{P}}(1)) \left(\widehat{\xi}_n - \xi\right),
\end{aligned} \tag{5.12}$$

where

$$\begin{aligned}
\theta_1 &= -\frac{1}{\mu^2} - \frac{2\xi^2 t_n p_n}{\mu^3 (1-\xi)(2\xi-1)}, \\
\theta_2 &= \frac{\xi^2 t_n}{\mu^2 (1-\xi)(2\xi-1)}, \\
\theta_3 &= \frac{t_n p_n}{\mu^2 (1-\xi)(2\xi-1)} \left(2\xi + \frac{4\xi^3 - 3\xi^2}{(1-\xi)(2\xi-1)}\right).
\end{aligned} \tag{5.13}$$

Multiplying by $\sqrt{n}/\gamma_n$ and using the proposition and the lemma together with the continuous mapping theorem, we find that

$$\begin{aligned}\frac{\sqrt{n}}{\gamma_n t_n}\left(\widehat{\mathbb{R}}_n(t_n)-\mathbb{R}_n(t_n)\right) &= \theta_1(1+o_{\mathbb{P}}(1))\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\mu}_n^{(J)}-\mu\right)\\ &\quad+\theta_2(1+o_{\mathbb{P}}(1))\frac{\sqrt{n}}{\gamma_n}(\widehat{p}_n-p_n)\\ &\quad+\theta_3(1+o_{\mathbb{P}}(1))\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\xi}_n-\xi\right).\end{aligned}\tag{5.14}$$

On the other hand, from Johansson [30], we have for all large n

$$\begin{aligned}\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\mu}_n^{(J)}-\mu\right) &= \frac{\sqrt{n}}{\gamma_n}(\widehat{\mu}_n^*-\mu_n^*)+\left(t_n+\frac{\beta_n}{1-\xi_n}\right)\frac{\sqrt{n}}{\gamma_n}(\widehat{p}_n-p_n)\\ &\quad+\frac{p_n\beta_n}{(1-\xi)^2}\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\xi}_n-\xi\right)+\frac{p_n}{1-\xi}\frac{\sqrt{n}}{\gamma_n}\left(\widehat{\beta}_n-\beta_n\right)+o_{\mathbb{P}}(1).\end{aligned}\tag{5.15}$$

This enables us to rewrite $(\sqrt{n}/\gamma_n t_n)(\widehat{\mathbb{R}}_n(t_n)-\mathbb{R}_n(t_n))$ into

$$\begin{aligned}&\theta_1\frac{\sqrt{n}}{\gamma_n}(\widehat{\mu}_n^*-\mu_n^*)+\frac{\sqrt{p_n(1-p_n)}}{\gamma_n}\left(\theta_2+\theta_1\left(t_n+\frac{\beta_n}{1-\xi}\right)\right)\frac{\sqrt{n}(\widehat{p}_n-p_n)}{\sqrt{p_n(1-p_n)}}\\ &\quad+\theta_1\frac{\beta_n p_n\sqrt{p_n}}{\gamma_n(1-\xi)}\sqrt{\frac{n}{p_n}}\left(\frac{\widehat{\beta}_n}{\beta_n}-1\right)\\ &\quad+\frac{\sqrt{p_n}}{\gamma_n}\left(\theta_3+\theta_1\frac{p_n\beta_n}{(1-\xi)^2}\right)\sqrt{\frac{n}{p_n}}\left(\widehat{\xi}_n-\xi\right)\ +o_{\mathbb{P}}(1),\\ &\qquad\mathbb{Q}^{-1}=(1+\xi)\begin{pmatrix}2 & -1\\ -1 & 1+\xi\end{pmatrix}.\end{aligned}\tag{5.16}$$

In view of Lemma 5.2, we infer that for all large n, the previous quantity is

$$\begin{aligned}&\theta_1\mathcal{W}_1+\frac{\sqrt{p_n(1-p_n)}}{\gamma_n}\left(\theta_2+\theta_1\left(t_n+\frac{\beta_n}{1-\xi}\right)\right)\mathcal{W}_2\\ &\quad+\frac{\sqrt{2(1+\xi)}\theta_1\beta_n p_n\sqrt{p_n}}{\gamma_n(1-\xi)}\mathcal{W}_3+\frac{(1+\xi)\sqrt{p_n}}{\gamma_n}\left(\theta_3+\frac{\theta_1 p_n\beta_n}{(1-\xi)^2}\right)\mathcal{W}_4+o_{\mathbb{P}}(1),\end{aligned}\tag{5.17}$$

where $(\mathcal{W}_i)_{i=1,4}$ are standard normal rvs with $E[W_iW_j] = 0$ for every $i, j = 1, \ldots, 4$ with $i \neq j$, except for

$$\begin{aligned}
E[W_3W_4] &= E\left[\sqrt{2(1+\xi)}\sqrt{\frac{n}{p_n}}\left(\frac{\widehat{\beta}_n}{\beta_n}-1\right)(1+\xi)\sqrt{\frac{n}{p_n}}\left(\widehat{\xi}_n-\xi\right)\right] \\
&= \sqrt{2(1+\xi)}(1+\xi)E\left[\sqrt{\frac{n}{p_n}}\left(\frac{\widehat{\beta}_n}{\beta_n}-1\right)\sqrt{\frac{n}{p_n}}\left(\widehat{\xi}_n-\xi\right)\right] \\
&= -\sqrt{2(1+\xi)}(1+\xi)^2.
\end{aligned} \tag{5.18}$$

Therefore, the rv $(\sqrt{n}/\gamma_n t_n)(\widehat{\mathbb{R}}_n(t_n) - \mathbb{R}_n(t_n))$ is Gaussian with mean zero with asymptotic variance

$$\begin{aligned}
K_n^2 := \theta_1^2 &+ \frac{p_n(1-p_n)}{\gamma_n^2}\left(\theta_2+\theta_1\left(t_n+\frac{\beta_n}{1-\xi}\right)\right)^2 + \frac{2(1+\xi)\theta_1^2\beta_n^2p_n^3}{\gamma_n^2(1-\xi)^2} \\
&+ \frac{(1+\xi)^2p_n}{\gamma_n^2}\left(\theta_3+\frac{\theta_1p_n\beta_n}{(1-\xi)^2}\right)^2 - \frac{2\theta_1\beta_np_n^2(1+\xi)^4}{(1-\xi)\gamma_n^2}\left(\theta_3+\frac{\theta_1p_n\beta_n}{(1-\xi)^2}\right) + o_{\mathbb{P}}(1).
\end{aligned} \tag{5.19}$$

Observe now that $K_n^2 = s_n^2 + o_{\mathbb{P}}(1)$, where s_n^2 is that in (3.12), this completes the proof of Theorem 3.2. □

6. Conclusion

In this paper, we have proposed a new estimator for the renewal function of heavy-tailed claim interoccurence times, via a semiparametric approach. Our considerations are based on one aspect of the extreme value theory, namely, the POT method. We have proved that our estimator is consistent and asymptotically normal. Moreover, simulations show that it is more accurate than the nonparametric estimator given by Bebbington et al. [14].

Acknowledgment

The authors are grateful to the referees whose suggestions led to an improvement of the paper.

References

[1] D. R. Cox, *Renewal Theory*, Methuen, London, UK, 1962.
[2] P. Embrechts, M. Maejima, and E. Omey, "Some limit theorems for generalized renewal measures," *Journal of the London Mathematical Society*, vol. 31, no. 1, pp. 184–192, 1985.
[3] P. Embrechts, M. Maejima, and J. Teugels, "Asymptotic behavior of compound distributions," *Astin Bulletin*, vol. 15, pp. 45–48, 1985.
[4] S. M. Ross, *Stochastic Processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 1983.

[5] D. J. Bartholomew and A. F. Forbes, *Statistical Techniques for Man-Power Planning*, Wiley, Chichester, UK, 1979.

[6] W. R. Blischke and E. M. Scheuer, “Applications of renewal theory in analysis of the free-replacement warranty,” *Naval Research Logistics Quarterly*, vol. 28, no. 2, pp. 193–205, 1981.

[7] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, Wiley, New York, NY, USA, 2nd edition, 1971.

[8] S. Asmussen, *Applied Probability and Queues*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Chichester, UK, 1987.

[9] S. Resnick, *Adventures in Stochastic Processes*, Birkhäuser, Boston, Mass, USA, 1992.

[10] E. W. Frees, “Nonparametric renewal function estimation,” *The Annals of Statistics*, vol. 14, no. 4, pp. 1366–1378, 1986.

[11] Q. Zhao and S. Subba Rao, “Nonparametric renewal function estimation based on estimated densities,” *Asia-Pacific Journal of Operational Research*, vol. 14, no. 1, pp. 115–126, 1997.

[12] N. M. Markovich and U. R. Krieger, “Nonparametric estimation of the renewal function by empirical data,” *Stochastic Models*, vol. 22, no. 2, pp. 175–199, 2006.

[13] M. S. Sgibnev, “On the renewal theorem in the case of infinite variance,” *Sibirskiĭ Matematicheskiĭ Zhurnal*, vol. 22, no. 5, pp. 178–189, 1981.

[14] M. Bebbington, Y. Davydov, and R. Zitikis, “Estimating the renewal function when the second moment is infinite,” *Stochastic Models*, vol. 23, no. 1, pp. 27–48, 2007.

[15] V. M. Zolotarev, *One-Dimensional Stable Distributions*, vol. 65 of *Translations of Mathematical Monographs*, American Mathematical Society, Providence, RI, USA, 1986.

[16] S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes*, vol. 4 of *Applied Probability. A Series of the Applied Probability Trust*, Springer, New York, NY, USA, 1987.

[17] E. Seneta, *Regularly Varying Functions*, Lecture Notes in Mathematics, Vol. 508, Springer, Berlin, Germany, 1976.

[18] J. L. Teugels, “Renewal theorems when the first or the second moment is infinite,” *Annals of Mathematical Statistics*, vol. 39, pp. 1210–1219, 1968.

[19] R. A. Fisher and L. H. C. Tippett, “Limiting forms of the frequency distribution of the largest or smallest member of a sample,” *Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 180–190, 1928.

[20] B. Gnedenko, “Sur la distribution limite du terme maximum d’une série aléatoire,” *Annals of Mathematics*, vol. 44, pp. 423–453, 1943.

[21] A. A. Balkema and L. de Haan, “Residual life time at great age,” *Annals of Probability*, vol. 2, pp. 792–804, 1974.

[22] J. Pickands III, “Statistical inference using extreme order statistics,” *The Annals of Statistics*, vol. 3, pp. 119–131, 1975.

[23] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering, Springer, New York, NY, USA, 2006.

[24] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, vol. 33 of *Applications of Mathematics*, Springer, Berlin, Germany, 1997.

[25] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman and Hall, London, UK, 1974.

[26] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Texts in Statistics, Springer, New York, NY, USA, 2nd edition, 1998.

[27] R. L. Smith, “Maximum likelihood estimation in a class of nonregular cases,” *Biometrika*, vol. 72, no. 1, pp. 67–90, 1985.

[28] R. L. Smith, “Estimating tails of probability distributions,” *The Annals of Statistics*, vol. 15, no. 3, pp. 1174–1207, 1987.

[29] J. Beierlant, G. Matthys, and G. Dierckx, “Heavy-tailed distributions and rating,” *Astin Bulletin*, vol. 31, no. 1, pp. 37–58, 2001.

[30] J. Johansson, “Estimating the mean of heavy-tailed distributions,” *Extremes*, vol. 6, no. 2, pp. 91–109, 2003.

[31] R. Ihaka and R. Gentleman, “R: a language for data analysis and graphics,” *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 299–314, 1996.

[32] S. Cheng and L. Peng, “Confidence intervals for the tail index,” *Bernoulli*, vol. 7, no. 5, pp. 751–760, 2001.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 596839, 17 pages
doi:10.1155/2010/596839

Research Article

Estimating the Conditional Tail Expectation in the Case of Heavy-Tailed Losses

Abdelhakim Necir,[1] Abdelaziz Rassoul,[2] and Ričardas Zitikis[3]

[1] *Laboratory of Applied Mathematics, Mohamed Khider University of Biskra, 07000, Algeria*
[2] *Ecole Nationale Superieure d'Hydraulique, Guerouaou, BP 31, Blida, 09000, Algeria*
[3] *Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON, Canada N6A5B7*

Correspondence should be addressed to Ričardas Zitikis, zitikis@stats.uwo.ca

Received 21 October 2009; Accepted 20 January 2010

Academic Editor: Edward Furman

Copyright © 2010 Abdelhakim Necir et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The conditional tail expectation (CTE) is an important actuarial risk measure and a useful tool in financial risk assessment. Under the classical assumption that the second moment of the loss variable is finite, the asymptotic normality of the nonparametric CTE estimator has already been established in the literature. The noted result, however, is not applicable when the loss variable follows any distribution with infinite second moment, which is a frequent situation in practice. With a help of extreme-value methodology, in this paper, we offer a solution to the problem by suggesting a new CTE estimator, which is applicable when losses have finite means but infinite variances.

1. Introduction

One of the most important actuarial risk measures is the conditional tail expectation (CTE) (see, e.g., [1]), which is the average amount of loss given that the loss exceeds a specified quantile. Hence, the CTE provides a measure of the capital needed due to the exposure to the loss, and thus serves as a risk measure. Not surprisingly, therefore, the CTE continues to receive increased attention in the actuarial and financial literature, where we also find its numerous extensions and generalizations (see, e.g., [2–8], and references therein). We next present basic notation and definitions.

Let X be a loss random variable with cumulative distribution function (cdf) F. Usually, the cdf F is assumed to be continuous and defined on the entire real line, with negative loss interpreted as gain. We also assume the continuity of F throughout the present paper. The CTE of the risk or loss X is then defined, for every $t \in (0,1)$, by

$$\mathrm{CTE}_F(t) = \mathbf{E}[X \mid X > Q(t)], \tag{1.1}$$

where $Q(t) = \inf\{x : F(x) \geq t\}$ is the quantile function corresponding to the cdf F. Since the cdf F is continuous, we easily check that

$$\mathrm{CTE}_F(t) = \frac{1}{1-t}\int_t^1 Q(s)ds. \tag{1.2}$$

Naturally, the CTE is unknown since the cdf F is unknown. Hence, it is desirable to establish statistical inferential results such as confidence intervals for $\mathrm{CTE}_F(t)$ with specified confidence levels and margins of error. We shall next show how to accomplish this task, initially assuming the classical moment assumption $\mathbf{E}[X^2] < \infty$. Namely, suppose that we have independent random variables $X_1, X_2, \ldots$, each with the cdf F, and let $X_{1:n} < \cdots < X_{n:n}$ denote the order statistics of $X_1, \ldots, X_n$. It is natural to define an empirical estimator of $\mathrm{CTE}_F(t)$ by the formula

$$\widehat{\mathrm{CTE}}_n(t) = \frac{1}{1-t}\int_t^1 Q_n(s)ds, \tag{1.3}$$

where $Q_n(s)$ is the empirical quantile function, which is equal to the ith order statistic $X_{i:n}$ for all $s \in ((i-1)/n, i/n]$, and for all $i = 1, \ldots, n$. The asymptotic behavior of the estimator $\widehat{\mathrm{CTE}}_n(t)$ has been studied by Brazauskas et al. [9], and we next formulate their most relevant result for our paper as a theorem.

Theorem 1.1. *Assume that* $\mathbf{E}[X^2] < \infty$*. Then for every* $t \in (0,1)$*, we have the asymptotic normality statement*

$$\sqrt{n}\left(\widehat{\mathrm{CTE}}_n(t) - \mathrm{CTE}_F(t)\right)(1-t) \longrightarrow_d \mathcal{N}\left(0, \sigma^2(t)\right), \tag{1.4}$$

when $n \to \infty$*, where the asymptotic variance* $\sigma^2(t)$ *is given by the formula*

$$\sigma^2(t) = \int_t^1\int_t^1 (\min(x,y) - xy)dQ(x)dQ(y). \tag{1.5}$$

The assumption $\mathbf{E}[X^2] < \infty$ is, however, quite restrictive as the following example shows. Suppose that F is the Pareto cdf with index $\gamma > 0$, that is, $1 - F(x) = x^{-1/\gamma}$ for all $x \geq 1$. Let us focus on the case $\gamma < 1$, because when $\gamma \geq 1$, then $\mathrm{CTE}_F(t) = +\infty$ for every $t \in (0,1)$. Theorem 1.1 covers only the values $\gamma \in (0, 1/2)$ in view of the assumption $\mathbf{E}[X^2] < \infty$. When $\gamma \in [1/2, 1)$, we have $\mathbf{E}[X^2] = \infty$ but, nevertheless, $\mathrm{CTE}_F(t)$ is well defined and finite since $\mathbf{E}[X] < \infty$. Analogous remarks hold for other distributions with Pareto-like tails, an we shall indeed work with such general distributions in this paper.

Namely, recall that the cdf F is regularly varying at infinity with index $(-1/\gamma) < 0$ if

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma} \tag{1.6}$$

for every $x > 0$. This class includes a number of popular distributions such as Pareto, generalized Pareto, Burr, Fréchet, Student, and so forth, which are known to be appropriate models for fitting large insurance claims, fluctuations of prices, log-returns, and so forth (see, e.g., [10]). In the remainder of this paper, therefore, we restrict ourselves to this class of distributions. For more information on the topic and, generally, on extreme value models and their manifold applications, we refer to the monographs by Beirlant et al. [11], Castillo et al. [12], de Haan and Ferreira [13], Resnick [14].

The rest of the paper is organized as follows. In Section 2 we construct an alternative, called "new", CTE estimator by utilizing an extreme value approach. In Section 3 we establish the asymptotic normality of the new CTE estimator and illustrate its performance with a little simulation study. The main result, which is Theorem 3.1 stated in Section 3, is proved in Section 4.

2. Construction of a New CTE Estimator

We have already noted that the "old" estimator $\widehat{\mathrm{CTE}}_n(t)$ does not yield the asymptotic normality (in the classical sense) beyond the condition $\mathbf{E}[X^2] < \infty$. Indeed, this follows by setting $t = 0$, in which case $\widehat{\mathrm{CTE}}_n(t)$ becomes the sample mean of $X_1, \ldots, X_n$, and thus the asymptotic normality of $\widehat{\mathrm{CTE}}_n(0)$ is equivalent to the classical Central Limit Theorem (CLT). Similar arguments show that the finite second moment is necessary for having the asymptotic normality (in the classical sense) of $\widehat{\mathrm{CTE}}_n(t)$ at any fixed "level" $t \in (0,1)$. Indeed, note that the asymptotic variance $\sigma^2(t)$ in Theorem 1.1 is finite only if $\mathbf{E}[X^2] < \infty$.

For this reason, we next construct an alternative CTE estimator, which takes into account different asymptotic properties of moderate and high quantiles in the case of heavy-tailed distributions. Hence, from now on we assume that $\gamma \in (1/2, 1)$. Before indulging ourselves into construction details, we first formulate the new CTE estimator:

$$\widetilde{\mathrm{CTE}}_n(t) = \frac{1}{1-t}\int_t^{1-k/n} Q_n(s)\,ds + \frac{kX_{n-k,n}}{n(1-t)\left(1-\widehat{\gamma}\right)}, \tag{2.1}$$

where we use the simplest yet useful and powerful Hill's [15] estimator

$$\widehat{\gamma}_n = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1:n} - \log X_{n-k:n} \tag{2.2}$$

of the tail index $\gamma \in (1/2, 1)$. Integers $k = k_n \in \{1, \ldots, n\}$ are such that $k \to \infty$ and $k/n \to 0$ when $n \to \infty$, and we note at the outset that their choices present a challenging task. In Figures 1 and 2, we illustrate the performance of the new estimator $\widetilde{\mathrm{CTE}}_n(t)$ with respect to the sample size $n \geq 1$, with the integers $k = k_n$ chosen according to the method proposed by Cheng and Peng [16]. Note that when t increases through the values 0.25, 0.50, 0.75, and 0.90

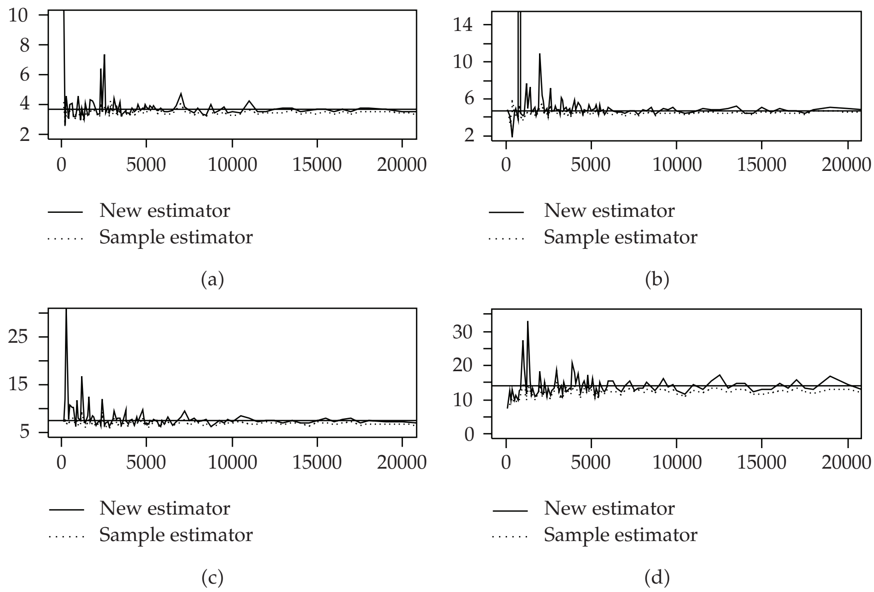


Figure 1: Values of the CTE estimator $\widetilde{\mathrm{CTE}}_n(t)$ (vertical axis) versus sample sizes n (horizontal axis) evaluated at the levels $t = 0.25$, $t = 0.50$, $t = 0.75$, and $t = 0.90$ (panels (a)–(d), resp.) in the Pareto case with the tail index $\gamma = 2/3$.

(panels (a)–(d), resp.), the vertical axes of the panels also increase, which reflects the fact that the larger the t gets, the more erratic the "new" and "old" estimators become. Note also that the empirical (i.e., "old") estimator underestimates the theoretical $\mathrm{CTE}_F(t)$, which is a well known phenomenon (see [17]).

We have based the construction of $\widetilde{\mathrm{CTE}}_n(t)$ on the recognition that one should estimate moderate and high quantiles differently when the underlying distribution is heavy-tailed. For this, we first recall that the high quantile q_s is, by definition, equal to $Q(1-s)$ for sufficiently small s. For an estimation theory of high quantiles in the case of heavy-tailed distributions we refer to, for example, Weissman [18], Dekkers and de Haan [19], Matthys and Beirlant [20], Gomes et al. [21], and references therein. We shall use the Weissman estimator

$$\widetilde{q}_s = \left(\frac{k}{n}\right)^{\widehat{\gamma}} X_{n-k:n} s^{-\widehat{\gamma}}, \quad 0 < s < \frac{k}{n}, \tag{2.3}$$

of the high quantile q_s. Then we write $\mathrm{CTE}_F(t)$ as the sum $\mathrm{CTE}_{1,n}(t) + \mathrm{CTE}_{2,n}(t)$ with the two summands defined together with their respective empirical estimators $\widetilde{\mathrm{CTE}}_{1,n}(t)$ and $\widetilde{\mathrm{CTE}}_{2,n}(t)$ as follows:

$$\begin{aligned}
\mathrm{CTE}_{1,n}(t) &= \frac{1}{1-t}\int_t^{1-k/n} Q(s)ds \approx \frac{1}{1-t}\int_t^{1-k/n} Q_n(s)ds = \widetilde{\mathrm{CTE}}_{1,n}(t), \\
\mathrm{CTE}_{2,n}(t) &= \frac{1}{1-t}\int_{1-k/n}^{1} Q(s)ds \approx \frac{1}{1-t}\int_{1-k/n}^{1} \widetilde{q}_{1-s}ds = \widetilde{\mathrm{CTE}}_{2,n}(t).
\end{aligned} \tag{2.4}$$

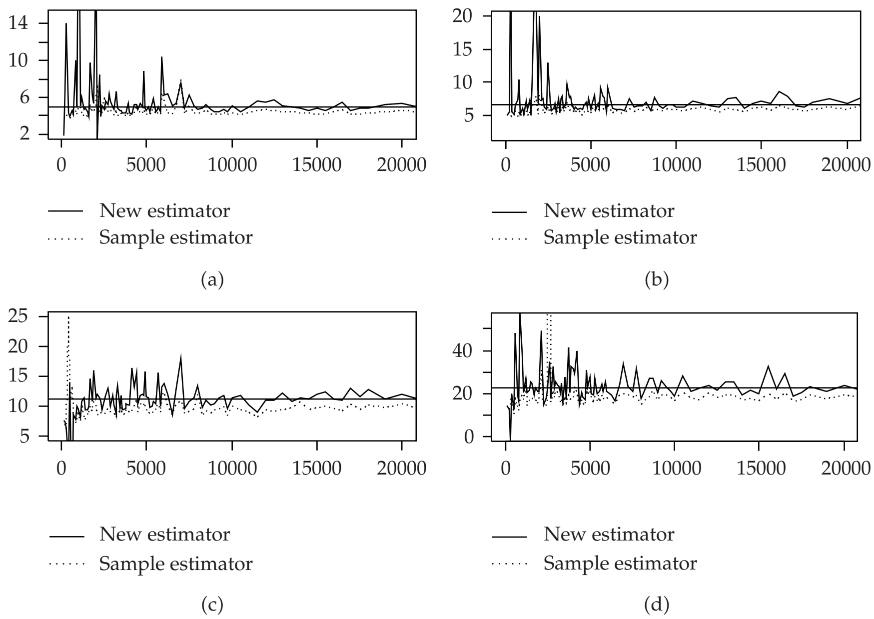


Figure 2: Values of the CTE estimator $\widetilde{\mathrm{CTE}}_n(t)$ (vertical axis) versus sample sizes n (horizontal axis) evaluated at the levels $t = 0.25$, $t = 0.50$, $t = 0.75$, and $t = 0.90$ (panels (a)–(d), resp.) in the Pareto case with the tail index $\gamma = 3/4$.

Simple integration gives the formula

$$\widetilde{\mathrm{CTE}}_{2,n}(t) = \frac{kX_{n-k,n}}{n(1-\widehat{\gamma})(1-t)}. \tag{2.5}$$

Consequently, the sum $\widetilde{\mathrm{CTE}}_{1,n}(t) + \widetilde{\mathrm{CTE}}_{2,n}(t)$ is an estimator of $\mathrm{CTE}_F(t)$, and this is exactly the estimator $\widetilde{\mathrm{CTE}}_n(t)$ introduced above. We shall investigate asymptotic normality of the new estimator in the next section, accompanied with an illustrative simulation study.

3. Main Theorem and Its Practical Implementation

We start this section by noting that Hill's estimator $\widehat{\gamma}_n$ has been thoroughly studied, improved, and generalized in the literature. For example, weak consistency of $\widehat{\gamma}_n$ has been established by Mason [22] assuming only that the underlying distribution is regularly varying at infinity. Asymptotic normality of $\widehat{\gamma}$ has been investigated under various conditions by a number of researchers, including Csörgő and Mason [23], Beirlant and Teugels [24], Dekkers et al. [25], see also references therein.

The main theoretical result of this paper, which is Theorem 3.1 below, establishes asymptotic normality of the new CTE estimator $\widetilde{\mathrm{CTE}}_n(t)$. To formulate the theorem, we need to introduce an assumption that ensures the asymptotic normality of Hill's estimator $\widehat{\gamma}_n$. Namely, the cdf F satisfies the generalized second-order regular variation condition with second-order parameter $\rho \leq 0$ (see [26, 27]) if there exists a function $a(t)$ which does not change its sign in a neighbourhood of infinity and is such that, for every $x > 0$,

$$\lim_{t\to\infty} \frac{1}{a(t)} \left(\frac{1 - F(tx)}{1 - F(t)} - x^{-1/\gamma} \right) = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\rho/\gamma}. \tag{3.1}$$

When $\rho = 0$, then the ratio on the right-hand side of (3.1) is interpreted as $\log x$. For statistical inference concerning the second-order parameter ρ, we refer, for example, to Peng and Qi [28], Gomes et al. [21], Gomes and Pestana [29]. Furthermore, in the formulation of Theorem 3.1, we shall also use the function $A(z) = \gamma^2 a(\mathbb{U}(z))$, where $\mathbb{U}(z) = Q(1 - 1/z)$.

Theorem 3.1. *Assume that the cdf F satisfies condition* (3.1) *with $\gamma \in (1/2, 1)$. Then for any sequence of integers $k = k_n \to \infty$ such that $k/n \to 0$ and $k^{1/2}A(n/k) \to 0$ when $n \to \infty$, we have that for any fixed $t \in (0, 1)$,*

$$\frac{\sqrt{n}\left(\widetilde{\mathrm{CTE}}_n(t) - \mathrm{CTE}_F(t)\right)(1 - t)}{(k/n)^{1/2} X_{n-k:n}} \longrightarrow_d \mathcal{N}\left(0, \sigma_\gamma^2\right), \tag{3.2}$$

where the asymptotic variance σ_γ^2 is given by the formula

$$\sigma_\gamma^2 = \frac{\gamma^4}{(1-\gamma)^4 (2\gamma - 1)}. \tag{3.3}$$

The asymptotic variance σ_γ^2 does not depend on t, unlike the variance in Theorem 1.1. This is not surprising because the heaviness of the right-most tail of F makes the asymptotic behaviour of $\int_t^1 (Q_n(s) - Q(s))ds$ "heavier" than the classical CLT-type behaviour of $\int_0^t (Q_n(s) - Q(s))ds$, for any fixed t. This in turn implies that under the conditions of Theorem 3.1, statement (3.2) is equivalent to the same statement in the case $t = 0$. The latter statement concerns estimating the mean $\mathbf{E}[X]$ of a heavy-tailed distribution. Therefore, we can view Theorem 3.1 as a consequence of Peng [30], and at the same time we can view results of Peng [30] as a consequence of Theorem 3.1 by setting $t = 0$ in it. Despite this equivalence, in Section 4 we give a proof of Theorem 3.1 for the sake of completeness. Our proof, however, is crucially based on a powerful technique called the Vervaat process (see [31–33], for details and references).

To discuss practical implementation of Theorem 3.1, we first fix a significance level $\varsigma \in (0, 1)$ and use the classical notation $z_{\varsigma/2}$ for the $(1 - \varsigma/2)$-quantile of the standard normal distribution $\mathcal{N}(0, 1)$. Given a realization of the random variables $X_1, \ldots, X_n$ (e.g., claim amounts), which follow a cdf F satisfying the conditions of Theorem 3.1, we construct a level $1 - \varsigma$ confidence interval for $\mathrm{CTE}_F(t)$ as follows. First, we choose an appropriate number k of extreme values. Since Hill's estimator has in general a substantial variance for small k and a considerable bias for large k, we search for a k that balances between the two shortcomings, which is indeed a well-known hurdle when estimating the tail index. To resolve this issue, several procedures have been suggested in the literature, and we refer to, for example, Dekkers and de Haan [34], Drees and Kaufmann [35], Danielsson et al. [36], Cheng and Peng [16], Neves and Fraga Alves [37], Gomes et al. [38], and references therein.

Table 1: Point estimates $\widetilde{\mathrm{CTE}}_n(t)$ and 95% confidence intervals for $\mathrm{CTE}_F(t)$ when $\gamma = 2/3$.

$t = 0.75$				$\mathrm{CTE}_F(t) = 7.005$				
n	k^*	$\widetilde{\mathrm{CTE}}_n(t)$	error	rmse	lower	upper	cover	length
1000	054	6.876	0.045	0.303	6.356	7.397	0.839	1.041
2000	100	6.831	0.025	0.231	6.463	7.199	0.882	0.736
5000	219	7.119	0.016	0.194	6.881	7.357	0.895	0.476
$t = 0.90$				$\mathrm{CTE}_F(t) = 12.533$				
n	k^*	$\widetilde{\mathrm{CTE}}_n(t)$	error	rmse	lower	upper	cover	length
1000	054	12.753	0.017	0.534	12.241	13.269	0.847	1.028
2000	100	12.487	0.003	0.294	12.137	12.838	0.841	0.701
5000	219	12.461	0.005	0.236	12.246	12.676	0.887	0.430

Table 2: Point estimates $\widetilde{\mathrm{CTE}}_n(t)$ and 95% confidence intervals for $\mathrm{CTE}_F(t)$ when $\gamma = 3/4$.

$t = 0.75$				$\mathrm{CTE}_F(t) = 9.719$				
n	k^*	$\widetilde{\mathrm{CTE}}_n(t)$	error	rmse	lower	upper	cover	length
1000	051	9.543	0.018	0.582	8.589	9.543	0.854	0.954
2000	104	9.808	0.009	0.466	9.150	10.466	0.888	1.316
5000	222	9.789	0.007	0.410	9.363	10.215	0.915	0.852
$t = 0.90$				$\mathrm{CTE}_F(t) = 18.494$				
n	k^*	$\widetilde{\mathrm{CTE}}_n(t)$	error	rmse	lower	upper	cover	length
1000	051	18.199	0.015	0.989	17.437	18.960	0.874	1.523
2000	104	18.696	0.011	0.858	18.052	19.340	0.895	1.288
5000	222	18.541	0.002	0.798	18.092	18.990	0.925	0.898

In our current study, we employ the method of Cheng and Peng [16] for an appropriate value k^* of the "parameter" k. Having computed Hill's estimator and consequently determined $X_{n-k^*:n}$, we then compute the corresponding values of $\widetilde{\mathrm{CTE}}_n(t)$ and $\sigma^2_{\widehat{\gamma}_n}$, and denote them by $\widetilde{\mathrm{CTE}}^*_n(t)$ and $\sigma^{2*}_{\widehat{\gamma}_n}$, respectively. Finally, using Theorem 3.1 we arrive at the following $(1-\varsigma)$-confidence interval for $\mathrm{CTE}_F(t)$:

$$\widetilde{\mathrm{CTE}}^*_n(t) \pm z_{\varsigma/2} \frac{(k^*/n)^{1/2} X_{n-k^*:n}\, \sigma^{2*}_{\widehat{\gamma}_n}}{(1-t)\sqrt{n}}. \tag{3.4}$$

To illustrate the performance of this confidence interval, we have carried out a small-scale simulation study based on the Pareto cdf $F(x) = 1 - x^{-1/\gamma}$, $x \geq 1$, with the tail index γ set to $2/3$ and $3/4$, and the level t set to 0.75 and 0.90. We have generated 200 independent replicates of three samples of sizes $n = 1000$, 2000, and 5000. For every simulated sample, we have obtained estimates $\widetilde{\mathrm{CTE}}_n(t)$. Then we have calculated the arithmetic averages over the values from the 200 repetitions, with the absolute error (error) and root mean squared error (rmse) of the new estimator $\widetilde{C}_n(t)$ reported in Table 1 ($\gamma = 2/3$) and Table 2 ($\gamma = 3/4$). In the tables, we have also reported 95%-confidence intervals (3.4) with their lower and upper bounds, coverage probabilities, and lengths.

We note emphatically that the above coverage probabilities and lengths of confidence intervals can be improved by employing more precise but, naturally, considerably more complex estimators of the tail index. Such estimators are described in the monographs by Beirlant et al. [11], Castillo et al. [12], de Haan and Ferreira [13], and Resnick [14]. Since the publication of these monographs, numerous journal articles have appeared on the topic. Our aim in this paper, however, is to present a simple yet useful result that highlights how much Actuarial Science and developments in Mathematical Statistics, Probability, and Stochastic Processes are interrelated, and thus benefit from each other.

4. Proof of Theorem 3.1

We start the proof of Theorem 3.1 with the decomposition

$$\left(\widetilde{\mathrm{CTE}}_n(t) - \mathrm{CTE}_F(t)\right)(1-t) = A_{n,1}(t) + A_{n,2}, \tag{4.1}$$

where

$$\begin{aligned} A_{n,1}(t) &= \int_t^{1-k/n} (Q_n(s) - Q(s))ds, \\ A_{n,2} &= \frac{k/n}{1-\widehat{\gamma}_n} X_{n-k:n} - \int_{1-k/n}^1 Q(s)ds. \end{aligned} \tag{4.2}$$

We shall show below that there are Brownian bridges B_n such that

$$\frac{\sqrt{n}A_{n,1}(t)}{(k/n)^{1/2}Q(1-k/n)} = -\frac{\int_0^{1-k/n} B_n(s)dQ(s)}{(k/n)^{1/2}Q(1-k/n)} + o_{\mathbf{P}}(1), \tag{4.3}$$

$$\begin{aligned} \frac{\sqrt{n}A_{n,2}}{(k/n)^{1/2}Q(1-k/n)} &= \frac{\gamma^2}{(1-\gamma)^2}\sqrt{\frac{n}{k}}\, B_n\left(1-\frac{k}{n}\right) \\ &\quad - \frac{\gamma}{(1-\gamma)^2}\sqrt{\frac{n}{k}} \int_{1-k/n}^1 \frac{B_n(s)}{1-s}ds + o_{\mathbf{P}}(1). \end{aligned} \tag{4.4}$$

Assuming for the time being that statements (4.3) and (4.4) hold, we next complete the proof of Theorem 3.1. To simplify the presentation, we use the following notation:

$$\begin{aligned} W_{1,n} &= -\frac{\int_0^{1-k/n} B_n(s)dQ(s)}{(k/n)^{1/2}Q(1-k/n)}, \\ W_{2,n} &= \frac{\gamma^2}{(1-\gamma)^2}\sqrt{\frac{n}{k}}B_n\left(1-\frac{k}{n}\right), \\ W_{3,n} &= -\frac{\gamma}{(1-\gamma)^2}\sqrt{\frac{n}{k}} \int_{1-k/n}^1 \frac{B_n(s)}{1-s}ds. \end{aligned} \tag{4.5}$$

Hence, we have the asymptotic representation

$$\frac{\sqrt{n}\left(\widetilde{\mathrm{CTE}}_n(t)-\mathrm{CTE}_F(t)\right)(1-t)}{(k/n)^{1/2}Q(1-k/n)}=W_{1,n}+W_{2,n}+W_{3,n}+o_{\mathbf{P}}(1). \tag{4.6}$$

The sum $W_{1,n}+W_{2,n}+W_{3,n}$ is a centered Gaussian random variable. To calculate its asymptotic variance, we establish the following limits:

$$\begin{gathered}\mathbf{E}\left[W_{1,n}^2\right]\longrightarrow\frac{2\gamma}{2\gamma-1},\qquad\mathbf{E}\left[W_{2,n}^2\right]\longrightarrow\frac{\gamma^4}{(1-\gamma)^4},\qquad\mathbf{E}\left[W_{3,n}^2\right]\longrightarrow\frac{2\gamma^2}{(1-\gamma)^4},\\ 2\mathbf{E}[W_{1,n}W_{2,n}]\longrightarrow\frac{-2\gamma^2}{(1-\gamma)^2},\qquad 2\mathbf{E}[W_{1,n}W_{3,n}]\longrightarrow\frac{2\gamma}{(1-\gamma)^2},\\ 2\mathbf{E}[W_{2,n}W_{3,n}]\longrightarrow\frac{-2\gamma^3}{(1-\gamma)^4}.\end{gathered} \tag{4.7}$$

Summing up the right-hand sides of the above six limits, we obtain σ_γ^2, whose expression in terms of the parameter γ is given in Theorem 3.1. Finally, since $X_{n-k:n}/Q(1-k/n)$ converges in probability to 1 (see, e.g., the proof of Corollary in [39]), the classical Sultsky's lemma completes the proof of Theorem 3.1. Of course, we are still left to verify statements (4.3) and (4.4), which make the contents of the following two subsections.

4.1. Proof of Statement (4.3)

If Q were continuously differentiable, then statement (4.3) would follow easily from the proof of Theorem 2 in [39]. We do not assume differentiability of Q and thus a new proof is required, which is crucially based on the Vervaat process (see [31–33], and references therein)

$$V_n(t)=\int_0^t(Q_n(s)-Q(s))ds+\int_{-\infty}^{Q(t)}(F_n(x)-F(x))dx. \tag{4.8}$$

Hence, for every t such that $0<t<1-k/n$, which is satisfied for all sufficiently large n since t is fixed, we have that

$$\begin{aligned}A_{n,1}(t)&=\int_0^{1-k/n}(Q_n(s)-Q(s))ds-\int_0^t(Q_n(s)-Q(s))ds\\&=-\int_{Q(t)}^{Q(1-k/n)}(F_n(x)-F(x))dx+V_n\left(1-\frac{k}{n}\right)-V_n(t).\end{aligned} \tag{4.9}$$

It is well known (see [31–33]) that $V_n(t)$ is nonnegative and does not exceed $-(F_n(Q(t))-t)(Q_n(t)-Q(t))$. Since the cdf F is continuous by assumption, we therefore have that

$$\sqrt{n}V_n(t)\le|e_n(t)||Q_n(t)-Q(t)|, \tag{4.10}$$

where $e_n(t)$ is the uniform empirical process $\sqrt{n}(F_n(Q(t)) - F(Q(t)))$, which for large n looks like the Brownian bridge $B_n(t)$. Note also that with the just introduced notation e_n, the integral on the right-hand side of (4.9) is equal to $\int_{Q(t)}^{Q(1-k/n)} e_n(F(x))dx$. Hence,

$$\begin{aligned}\frac{\sqrt{n}A_{n,1}(t)}{(k/n)^{1/2}Q(1-k/n)} &= -\frac{\int_{Q(t)}^{Q(1-k/n)} e_n(F(x))dx}{(k/n)^{1/2}Q(1-k/n)} \\ &\quad + O_{\mathbf{P}}(1)\frac{|e_n(1-k/n)||Q_n(1-k/n) - Q(1-k/n)|}{(k/n)^{1/2}Q(1-k/n)} \\ &\quad + O_{\mathbf{P}}(1)\frac{|e_n(t)||Q_n(t) - Q(t)|}{(k/n)^{1/2}Q(1-k/n)}.\end{aligned} \tag{4.11}$$

We shall next replace the empirical process e_n by an appropriate Brownian bridge B_n in the first integral on the right-hand side of (4.11) with an error term of magnitude $o_{\mathbf{P}}(1)$, and we shall also show that the second and third summands on the right-hand side of (4.11) are of the order $o_{\mathbf{P}}(1)$. The replacement of e_n by B_n can be accomplished using, for example, Corollary 2.1 on page 48 of Csörgő et al. [40], which states that on an appropriately constructed probability space and for any $0 \le \nu < 1/4$, we have that

$$\sup_{1/n \le s \le 1-1/n} \frac{|e_n(s) - B_n(s)|}{s^{1/2-\nu}(1-s)^{1/2-\nu}} = O_{\mathbf{P}}(n^{-\nu}). \tag{4.12}$$

This result is applicable in the current situation since we can always place our original problem into the required probability space, because our main results are "in probability". Furthermore, since $Q(t) \le x \le Q(1-k/n)$, we have that $t \le F(x) \le 1-k/n$. Hence, statement (4.12) implies that

$$-\frac{\int_{Q(t)}^{Q(1-k/n)} e_n(F(x))dx}{(k/n)^{1/2}Q(1-k/n)} = -\frac{\int_{Q(t)}^{Q(1-k/n)} B_n(F(x))dx}{(k/n)^{1/2}Q(1-k/n)} + O_{\mathbf{P}}(1)\frac{\int_{Q(t)}^{Q(1-k/n)} (1-F(x))^{1/2-\nu}dx}{n^{\nu}(k/n)^{1/2}Q(1-k/n)}. \tag{4.13}$$

Changing the variables of integration and using the property $(k/n)^{1/2}Q(1-k/n) \to \infty$ when $n \to \infty$, we obtain that

$$-\frac{\int_{Q(t)}^{Q(1-k/n)} B_n(F(x))dx}{(k/n)^{1/2}Q(1-k/n)} = -\frac{\int_0^{1-k/n} B_n(s)dQ(s)}{(k/n)^{1/2}Q(1-k/n)} + o_{\mathbf{P}}(1). \tag{4.14}$$

The main term on the right-hand side of (4.14) is $W_{1,n}$. We shall next show that the right-most summand of (4.13) converges to 0 when $n \to \infty$.

Changing the variable of integration and then integrating by parts, we obtain the bound

$$\frac{\int_{Q(t)}^{Q(1-k/n)}(1-F(x))^{1/2-\nu}dx}{n^{\nu}(k/n)^{1/2}Q(1-k/n)} \le \frac{(1-s)^{1/2-\nu}Q(s)\Big|_{t}^{1-k/n}}{n^{\nu}(k/n)^{1/2}Q(1-k/n)} + O(1)\frac{\int_{t}^{1-k/n}(1-s)^{-1/2-\nu}Q(s)ds}{n^{\nu}(k/n)^{1/2}Q(1-k/n)}. \tag{4.15}$$

We want to show that the right-hand side of bound (4.15) converges to 0 when $n \to \infty$. For this, we first note that

$$\frac{(1-s)^{1/2-\nu}Q(s)\Big|_{t}^{1-k/n}}{n^{\nu}(k/n)^{1/2}Q(1-k/n)} = \frac{1}{k^{\nu}} - \frac{(1-t)^{1/2-\nu}Q(t)}{n^{\nu}(k/n)^{1/2}Q(1-k/n)} \longrightarrow 0. \tag{4.16}$$

Next, with the notation $\phi(u) = Q(1-u)/u^{1/2+\nu}$, we have that

$$\frac{\int_{t}^{1-k/n}(1-s)^{-1/2-\nu}Q(s)ds}{n^{\nu}(k/n)^{1/2}Q(1-k/n)} = \frac{1}{k^{\nu}}\frac{\int_{k/n}^{1-t}\phi(s)ds}{(k/n)\phi(k/n)} \longrightarrow 0 \tag{4.17}$$

when $n \to \infty$, where the convergence to 0 follows from Result 1 in the Appendix of Necir and Meraghni [39]. Taking statements (4.15)–(4.17) together, we have that the right-most summand of (4.13) converges to 0 when $n \to \infty$.

Consequently, in order to complete the proof of statement (4.3), we are left to show that the second and third summands on the right-hand side of (4.11) are of the order $o_{\mathbf{P}}(1)$. The third summand is of the order $o_{\mathbf{P}}(1)$ because $|e_n(t)||Q_n(t)-Q(t)| = O_{\mathbf{P}}(1)$ and $(k/n)^{1/2}Q(1-k/n) \to \infty$. Hence, we are only left to show that the second summand on the right-hand side of equation (4.11) is of the order $o_{\mathbf{P}}(1)$, for which we shall show that

$$\frac{|e_n(1-k/n)|}{(k/n)^{1/2}}\left|\frac{Q_n(1-k/n)}{Q(1-k/n)} - 1\right| = o_{\mathbf{P}}(1). \tag{4.18}$$

To prove statement (4.18), we first note that

$$\frac{|e_n(1-k/n)|}{(k/n)^{1/2}} \le \frac{|e_n(1-k/n) - B_n(1-k/n)|}{(k/n)^{1/2}} + \frac{|B_n(1-k/n)|}{(k/n)^{1/2}}. \tag{4.19}$$

The first summand on the right-hand side of bound (4.19) is of the order $O_{\mathbf{P}}(1)$ due to statement (4.12) with $\nu = 0$. The second summand on the right-hand side of bound (4.19) is of the order $O_{\mathbf{P}}(1)$ due to a statement on page 49 of Csörgő et al. [40] (see the displayed bound just below statement (2.39) therein). Hence, to complete the proof of statement (4.18), we need to check that

$$\frac{Q_n(1-k/n)}{Q(1-k/n)} = 1 + o_{\mathbf{P}}(1). \tag{4.20}$$

Observe that, for each n, the distribution of $Q_n(1-k/n)$ is the same as that of $Q(E_n^{-1}(1-k/n))$, where E_n^{-1} is the uniform empirical quantile function. Furthermore, the processes $\{1-E_n^{-1}(1-s),\ 0 \le s \le 1\}$ and $\{E_n^{-1}(s),\ 0 \le s \le 1\}$ are equal in distribution. Hence, statement (4.20) is equivalent to

$$\frac{Q(1-E_n^{-1}(k/n))}{Q(1-k/n)} = 1 + o_{\mathbf{P}}(1). \tag{4.21}$$

From the Glivenko-Cantelli theorem we have that $E_n^{-1}(k/n) - k/n \to 0$ almost surely, which also implies that $E_n^{-1}(k/n) \to 0$ since $k/n \to 0$ by our choice of k. Moreover, we know from Theorem 0 and Remark 1 of Wellner [41] that

$$\sup_{1/n \le s \le 1} s^{-1}\left|E_n^{-1}(s) - s\right| = o_{\mathbf{P}}(1), \tag{4.22}$$

from which we conclude that

$$\frac{nE_n^{-1}(k/n)}{k} = 1 + o_{\mathbf{P}}(1). \tag{4.23}$$

Since the function $s \mapsto Q(1-s)$ is slowly varying at zero, using Potter's inequality (see the 5th assertion of Proposition B.1.9 on page 367 of de Haan and Ferreira [13],we obtain that

$$\frac{Q(1-E_n^{-1}(k/n))}{Q(1-k/n)} = (1+o_{\mathbf{P}}(1))\left(\frac{nE_n^{-1}(k/n)}{k}\right)^{-\gamma\pm\theta} \tag{4.24}$$

for any $\theta \in (0,\gamma)$. In view of (4.23), the right-hand side of (4.24) is equal to $1 + o_{\mathbf{P}}(1)$, which implies statement (4.21) and thus finishes the proof of statement (4.3).

4.2. Proof of Statement (4.4)

The proof of statement (4.4) is similar to that of Theorem 2 in Necir et al. [42], though some adjustments are needed since we are now concerned with the CTE risk measure. We therefore present main blocks of the proof together with pinpointed references to Necir et al. [42] for specific technical details.

We start the proof with the function $\mathbb{U}(z) = Q(1-1/z)$ that was already used in the formulation of Theorem 3.1. Hence, if Y is a random variable with the distribution function $G(z) = 1 - 1/z$, $z \ge 1$, then $\mathbb{U}(Y) = Q(G(Y)) =_d X$ because $G(Y)$ is a uniform on the interval $[0,1]$ random variable. Hence,

$$A_{n,2} = \frac{k/n}{1-\widehat{\gamma}_n}\mathbb{U}(Y_{n-k:n}) - \int_0^{k/n}\mathbb{U}\left(\frac{1}{s}\right)ds, \tag{4.25}$$

and so we have

$$\begin{aligned}
\frac{\sqrt{n}A_{n,2}}{(k/n)^{1/2}Q(1-k/n)} &= \sqrt{k}\left(\frac{1}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)} - \frac{(n/k)\int_0^{k/n}\mathbb{U}(1/s)ds}{\mathbb{U}(n/k)}\right) \\
&= \sqrt{k}\left(\frac{1}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)} - \frac{1}{1-\gamma}\right) \\
&\quad + \sqrt{k}\left(\frac{1}{1-\gamma} - \frac{\int_1^{\infty}s^{-2}\mathbb{U}(ns/k)ds}{\mathbb{U}(n/k)}\right).
\end{aligned} \tag{4.26}$$

We next show that the right-most term in (4.26) converges to 0 when $n \to \infty$. For this reason, we first rewrite the term as follows:

$$\sqrt{k}\left(\frac{1}{1-\gamma} - \frac{\int_1^{\infty}s^{-2}\mathbb{U}(ns/k)ds}{\mathbb{U}(n/k)}\right) = -\sqrt{k}\int_1^{\infty}\frac{1}{s^2}\left(\frac{\mathbb{U}(ns/k)}{\mathbb{U}(n/k)} - s^{\gamma}\right)ds. \tag{4.27}$$

The right-hand side of (4.27) converges to 0 (see notes on page 149 of Necir et al. [42]) due to the second-order condition (3.1), which can equivalently be rewritten as

$$\lim_{z\to\infty}\frac{1}{A(z)}\left(\frac{\mathbb{U}(zs)}{\mathbb{U}(z)} - s^{\gamma}\right) = s^{\gamma}\frac{s^{\rho}-1}{\rho} \tag{4.28}$$

for every $s > 0$, where $A(z) = \gamma^2 a(\mathbb{U}(z))$. Note that $\sqrt{k}A(n/k) \to 0$ when $n \to \infty$. Hence, in order to complete the proof of statement (4.4), we need to check that

$$\begin{aligned}
\sqrt{k}\left(\frac{1}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)} - \frac{1}{1-\gamma}\right) &= \frac{\gamma^2}{(1-\gamma)^2}\sqrt{\frac{n}{k}}B_n\left(1-\frac{k}{n}\right) \\
&\quad - \frac{\gamma}{(1-\gamma)^2}\sqrt{\frac{n}{k}}\int_{1-k/n}^{1}\frac{B_n(s)}{1-s}ds + o_{\mathbf{P}}(1).
\end{aligned} \tag{4.29}$$

With Hill's estimator written in the form

$$\widehat{\gamma}_n = \frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{\mathbb{U}(Y_{n-i+1:n})}{\mathbb{U}(Y_{n-k:n})}\right), \tag{4.30}$$

we proceed with the proof of statement (4.29) as follows:

$$\begin{aligned}
&\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)} - 1\right) \\
&\quad = \sqrt{k}\,\frac{1-\gamma}{1-\widehat{\gamma}_n}\left(\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)} - \left(\frac{Y_{n-k:n}}{n/k}\right)^{\gamma}\right) + \sqrt{k}\,\frac{1-\gamma}{1-\widehat{\gamma}_n}\left(\left(\frac{Y_{n-k:n}}{n/k}\right)^{\gamma} - 1\right) + \sqrt{k}\,\frac{\widehat{\gamma}_n - \gamma}{1-\widehat{\gamma}_n}.
\end{aligned} \tag{4.31}$$

Furthermore, we have that

$$\sqrt{k}\,\frac{\widehat{\gamma}_n-\gamma}{1-\widehat{\gamma}_n}=\frac{1}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{\mathbb{U}(Y_{n-i+1:n})}{\mathbb{U}(Y_{n-k:n})}\right)-\gamma\log\left(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right)\right)$$
$$+\frac{\gamma}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right)-1\right). \tag{4.32}$$

Arguments on page 156 of Necir et al. [42] imply that the first term on the right-hand side of (4.32) is of the order $O_{\mathbf{P}}(\sqrt{k}A(Y_{n-k:n}))$, and a note on page 157 of Necir et al. [42] says that $\sqrt{k}\ A(Y_{n-k:n})=o_{\mathbf{P}}(1)$. Hence, the first term on the right-hand side of (4.32) is of the order $o_{\mathbf{P}}(1)$. Analogous considerations using bound (2.5) instead of (2.4) on page 156 of Necir et al. [42] imply that the first term on the right-hand side of (4.31) is of the order $o_{\mathbf{P}}(1)$. Hence, in summary, we have that

$$\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-1\right)=\frac{1-\gamma}{1-\widehat{\gamma}_n}\sqrt{k}\left(\left(\frac{Y_{n-k:n}}{n/k}\right)^{\gamma}-1\right)$$
$$+\frac{\gamma}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}\right)-1\right)+o_{\mathbf{P}}(1). \tag{4.33}$$

We now need to connect the right-hand side of (4.33) with Brownian bridges B_n. To this end, we first convert the Y-based order statistics into U-based (i.e., uniform on $[0,1]$) order statistics. For this we recall that the cdf of Y is G, and thus Y is equal in distribution to $G^{-1}(U)$, which is $1/(1-U)$. Consequently,

$$\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-1\right)=\frac{1-\gamma}{1-\widehat{\gamma}_n}\sqrt{k}\left(\left(\frac{1}{(n/k)(1-U_{n-k:n})}\right)^{\gamma}-1\right)$$
$$+\frac{\gamma}{1-\widehat{\gamma}_n}\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{(1-U_{n-k:n})}{(1-U_{n-i+1:n})}\right)-1\right)+o_{\mathbf{P}}(1). \tag{4.34}$$

Next we choose a sequence of Brownian bridges B_n (see pages 158-159 in [42] and references therein) such that the following two asymptotic representations hold:

$$\sqrt{k}\left(\left(\frac{1}{(n/k)(1-U_{n-k:n})}\right)^{\gamma}-1\right)=-\gamma\sqrt{\frac{n}{k}}B_n\left(1-\frac{k}{n}\right)+o_{\mathbf{P}}(1),$$
$$\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(\log\left(\frac{(1-U_{n-k:n})}{(1-U_{n-i+1:n})}\right)-1\right)=\sqrt{\frac{n}{k}}B_n\left(1-\frac{k}{n}\right) \tag{4.35}$$
$$-\sqrt{\frac{n}{k}}\int_{1-k/n}^{1}\frac{B_n(s)}{1-s}ds+o_{\mathbf{P}}(1).$$

Using these two statements on the right-hand side of (4.34) and also keeping in mind that $\widehat{\gamma}_n$ is a consistent estimator of γ (see [22]), we have that

$$\sqrt{k}\left(\frac{1-\gamma}{1-\widehat{\gamma}_n}\frac{\mathbb{U}(Y_{n-k:n})}{\mathbb{U}(n/k)}-1\right)=\frac{\gamma^2}{1-\gamma}\sqrt{\frac{n}{k}}\,B_n\left(1-\frac{k}{n}\right)-\frac{\gamma}{1-\gamma}\sqrt{\frac{n}{k}}\int_{1-k/n}^{1}\frac{B_n(s)}{1-s}ds+o_{\mathbf{P}}(1). \tag{4.36}$$

Dividing both sides of equation (4.36) by $1-\gamma$, we arrive at (4.29). This completes the proof of statement (4.4) and of Theorem 3.1 as well.

Acknowledgments

Our work on the revision of this paper has been considerably influenced by constructive criticism and suggestions by three anonymous referees and the editor in charge of the manuscript, Edward Furman, and we are indebted to all of them. Results of the paper were first announced at the 44th Actuarial Research Conference at the University of Wisconsin, Madison, Wisconsin, July 30–August 1, 2009. The authors are grateful to participants of this most stimulating conference, organized by the Society of Actuaries, for generous feedback. The research has been partially supported by grants from the Society of Actuaries (SOA) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas, *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, Wiley, Chichester, UK, 2005.

[2] Z. M. Landsman and E. A. Valdez, "Tail conditional expectations for elliptical distributions," *North American Actuarial Journal*, vol. 7, no. 4, pp. 55–71, 2003.

[3] M. R. Hardy and J. L. Wirch, "The iterated CTE: a dynamic risk measure," *North American Actuarial Journal*, vol. 8, no. 4, pp. 62–75, 2004.

[4] J. Cai and H. Li, "Conditional tail expectations for multivariate phase-type distributions," *Journal of Applied Probability*, vol. 42, no. 3, pp. 810–825, 2005.

[5] B. J. Manistre and G. H. Hancock, "Variance of the CTE estimator," *North American Actuarial Journal*, vol. 9, no. 2, pp. 129–156, 2005.

[6] E. Furman and Z. Landsman, "Tail variance premium with applications for elliptical portfolio of risks," *Astin Bulletin*, vol. 36, no. 2, pp. 433–462, 2006.

[7] E. Furman and R. Zitikis, "Weighted risk capital allocations," *Insurance: Mathematics & Economics*, vol. 43, no. 2, pp. 263–269, 2008.

[8] E. Furman and R. Zitikis, "Weighted pricing functionals with applications to insurance: an overview," *North American Actuarial Journal*, vol. 13, pp. 483–496, 2009.

[9] V. Brazauskas, B. L. Jones, M. L. Puri, and R. Zitikis, "Estimating conditional tail expectation with actuarial applications in view," *Journal of Statistical Planning and Inference*, vol. 138, no. 11, pp. 3590–3604, 2008.

[10] J. Beirlant, G. Matthys, and G. Dierckx, "Heavy-tailed distributions and rating," *Astin Bulletin*, vol. 31, no. 1, pp. 37–58, 2001.

[11] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers, *Statistics of Extremes*, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, UK, 2004.

[12] E. Castillo, A. S. Hadi, N. Balakrishnan, and J. M. Sarabia, *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ, USA, 2005.

[13] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering, Springer, New York, NY, USA, 2006.

[14] S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer Series in Operations Research and Financial Engineering, Springer, New York, NY, USA, 2007.

[15] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

[16] S. Cheng and L. Peng, "Confidence intervals for the tail index," *Bernoulli*, vol. 7, no. 5, pp. 751–760, 2001.

[17] J. T. Kim and M. R. Hardy, "Quantifying and correcting the bias in estimated risk measures," *Astin Bulletin*, vol. 37, no. 2, pp. 365–386, 2007.

[18] I. Weissman, "Estimation of parameters and large quantiles based on the k largest observations," *Journal of the American Statistical Association*, vol. 73, no. 364, pp. 812–815, 1978.

[19] A. L. M. Dekkers and L. de Haan, "On the estimation of the extreme-value index and large quantile estimation," *The Annals of Statistics*, vol. 17, no. 4, pp. 1795–1832, 1989.

[20] G. Matthys and J. Beirlant, "Estimating the extreme value index and high quantiles with exponential regression models," *Statistica Sinica*, vol. 13, no. 3, pp. 853–880, 2003.

[21] M. I. Gomes, F. Figueiredo, and S. Mendonça, "Asymptotically best linear unbiased tail estimators under a second-order regular variation condition," *Journal of Statistical Planning and Inference*, vol. 134, no. 2, pp. 409–433, 2005.

[22] D. M. Mason, "Laws of large numbers for sums of extreme values," *The Annals of Probability*, vol. 10, no. 3, pp. 754–764, 1982.

[23] S. Csörgő and D. M. Mason, "Central limit theorems for sums of extreme values," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 98, no. 3, pp. 547–558, 1985.

[24] J. Beirlant and J. L. Teugels, "Asymptotic normality of Hill's estimator," in *Extreme Value Theory (Oberwolfach, 1987)*, J. Hüsler and R.-D. Reiss, Eds., vol. 51 of *Lecture Notes in Statistics*, pp. 148–155, Springer, New York, NY, USA, 1989.

[25] A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan, "A moment estimator for the index of an extreme-value distribution," *The Annals of Statistics*, vol. 17, no. 4, pp. 1833–1855, 1989.

[26] L. de Haan and U. Stadtmüller, "Generalized regular variation of second order," *Australian Mathematical Society Journal Series A*, vol. 61, no. 3, pp. 381–395, 1996.

[27] J. Geluk, L. de Haan, S. Resnick, and C. Stărică, "Second-order regular variation, convolution and the central limit theorem," *Stochastic Processes and Their Applications*, vol. 69, no. 2, pp. 139–159, 1997.

[28] L. Peng and Y. Qi, "Estimating the first- and second-order parameters of a heavy-tailed distribution," *Australian & New Zealand Journal of Statistics*, vol. 46, no. 2, pp. 305–312, 2004.

[29] M. I. Gomes and D. Pestana, "A simple second-order reduced bias' tail index estimator," *Journal of Statistical Computation and Simulation*, vol. 77, no. 5-6, pp. 487–504, 2007.

[30] L. Peng, "Estimating the mean of a heavy tailed distribution," *Statistics & Probability Letters*, vol. 52, no. 3, pp. 255–264, 2001.

[31] R. Zitikis, "The Vervaat process," in *Asymptotic Methods in Probability and Statistics*, B. Szyszkowicz, Ed., pp. 667–694, North-Holland, Amsterdam, The Netherlands, 1998.

[32] Y. Davydov and R. Zitikis, "Generalized Lorenz curves and convexifications of stochastic processes," *Journal of Applied Probability*, vol. 40, no. 4, pp. 906–925, 2003.

[33] Y. Davydov and R. Zitikis, "Convex rearrangements of random elements," in *Asymptotic Methods in Stochastics*, B. Szyszkowicz, Ed., vol. 44 of *Fields Institute Communications*, pp. 141–171, American Mathematical Society, Providence, RI, USA, 2004.

[34] A. L. M. Dekkers and L. de Haan, "Optimal choice of sample fraction in extreme-value estimation," *Journal of Multivariate Analysis*, vol. 47, no. 2, pp. 173–195, 1993.

[35] H. Drees and E. Kaufmann, "Selecting the optimal sample fraction in univariate extreme value estimation," *Stochastic Processes and Their Applications*, vol. 75, no. 2, pp. 149–172, 1998.

[36] J. Danielsson, L. de Haan, L. Peng, and C. G. de Vries, "Using a bootstrap method to choose the sample fraction in tail index estimation," *Journal of Multivariate Analysis*, vol. 76, no. 2, pp. 226–248, 2001.

[37] C. Neves and M. I. Fraga Alves, "Reiss and Thomas' automatic selection of the number of extremes," *Computational Statistics & Data Analysis*, vol. 47, no. 4, pp. 689–704, 2004.

[38] M. I. Gomes, D. Pestana, and F. Caeiro, "A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator," *Statistics & Probability Letters*, vol. 79, no. 3, pp. 295–303, 2009.

[39] A. Necir and D. Meraghni, "Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts," *Insurance: Mathematics & Economics*, vol. 45, no. 1, pp. 49–58, 2009.

[40] M. Csörgő, S. Csörgő, L. Horváth, and D. M. Mason, "Weighted empirical and quantile processes," *The Annals of Probability*, vol. 14, no. 1, pp. 31–85, 1986.

[41] J. A. Wellner, "Limit theorems for the ratio of the empirical distribution function to the true distribution function," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 45, no. 1, pp. 73–88, 1978.

[42] A. Necir, D. Meraghni, and F. Meddi, "Statistical estimate of the proportional hazard premium of loss," *Scandinavian Actuarial Journal*, no. 3, pp. 147–161, 2007.

Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2010, Article ID 543065, 17 pages
doi:10.1155/2010/543065

Research Article

An Analysis of the Influence of Fundamental Values' Estimation Accuracy on Financial Markets

Hiroshi Takahashi

Graduate School of Business Administration, Keio University, 4-1-1 Hiyoshi, Kohoku-ku, Yokohama-city 223-8572, Japan

Correspondence should be addressed to Hiroshi Takahashi, htaka@kbs.keio.ac.jp

Received 14 September 2009; Revised 17 December 2009; Accepted 17 February 2010

Academic Editor: Edward Furman

Copyright © 2010 Hiroshi Takahashi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research analyzed the influence of the differences in the forecast accuracy of fundamental values on the financial market. As a result of intensive experiments in the market, we made the following interesting findings: (1) improvements in forecast accuracy of fundamentalists can contribute to an increase in the number of fundamentalists; (2) certain situations might occur, according to the level of forecast accuracy of fundamentalists, in which fundamentalists and passive management coexist, or in which fundamentalists die out of the market, and furthermore; (3) where a variety of investors exist in the market, improvements in the forecast accuracy could increase the number of fundamentalists more than the number of investors that employ passive investment strategy. These results contribute to clarifying the mechanism of price fluctuations in financial markets and also indicate one of the factors for the low ratio of passive investors in asset management business.

1. Introduction

A growing body of studies regarding asset pricing have been conducted, and many prominent theories have been proposed [1–4]. Along with the advancement of these theories, many arguments regarding securities investment in practical business affairs in finance have been actively discussed. The theory of asset pricing and investment strategy for shares are also currently being discussed with enthusiasm. The accurate valuation of fundamental values of investment grade assets is one of significant interest for those investors that actually make transactions in real financial markets. For example, many of institutional investors have a number of security analysts in their own companies in order to try to evaluate the fundamental values of each security.

Market efficiency is a central hypothesis in traditional asset pricing theories and there has been a large amount of discussion regarding it [5]. For example, in the Capital

Asset Pricing Model (CAPM), which is one of the most popular asset pricing theories, equilibrium asset prices are derived on the assumption of efficient markets and rational investors. CAPM indicates that the optimal investment strategy is to hold market portfolio[1] [2]. Since it is very difficult for investors to get an excess return in an efficient market, it is assumed to be difficult to beat market portfolio even though fundamental values are estimated correctly based on public information [2, 6]. On the other hand, passive investment strategy, which tries to maintain an average return using benchmarks based on market indices, is consistent with traditional asset pricing theories and is considered to be an effective method in efficient markets. On the basis of such arguments, there has been growing interest in passive investment strategy in the practical business affairs of asset management. Many investors employ the passive investment strategy for their investment.[2]

Recently, however, traditional financial theories have been criticized in terms of their explanation power and the validity of their assumptions. Behavioral finance has recently been in the limelight and many reports indicate that deviation from rational decision-making can explain anomalies which cannot be explained with traditional financial theories [7–10]. Generally, investor behavior which is assumed in behavioral finance has complicated rules for decision making compared to decision making based on expected utility maximization. For this reason, in many cases, it is difficult to derive the influence of investor behavior on prices analytically [11]. In order to take such investors behavior into account in analyzing financial markets, we need to introduce a different analytical method.

In the area of computer science, Agent-Based Modeling has been proposed as an effective approach to analyze the relation between microrules and macrobehavior [12]. This is a bottom-up approach that tries to describe macrobehavior of the entire system using local rules. This approach is appropriate for analyzing a multiagent system in which a great number of agents that act autonomously gather together.[3] The agent-based approach is applied in a wide variety of study fields such as engineering and biology, and many reports have been made about analyses adopting this approach in the field of social science [13–17].

In the background of the above-mentioned arguments, the purpose of this research is to clarify the influence of the difference in the forecast accuracy of fundamental values on financial markets by using the agent-based model for analysis. This analysis includes the relationship between fundamentalists that invest based on fundamentals and passive investment strategy. Section 2 describes the model used in this analysis. Section 3 shows the results of the analysis. Section 4 summarizes this paper.

2. Model

A computer simulation of the financial market involving 1000 investors was used as the model for this research. Shares and risk-free assets were the two types of assets used along with the possible transaction methods. Several types of investors exist in the market, each undertaking transactions based on their own stock evaluations. This market was composed of three major stages, (1) generation of corporate earnings, (2) formation of investor forecasts, and (3) setting transaction prices. The market advances through repetition of these stages. The following sections describe negotiable transaction assets, modeling of investor behavior, setting transaction prices, and natural selection rules in the market.

Table 1: List of investor types.

No.	Investor types
1	Fundamentalist
2	Forecasting by past average (most recent 10 days)
3	Forecasting by trend (most recent 10 day)
4	Passive investor

2.1. Negotiable Assets in the Market

This market has risk-free and risk assets. There are risky assets in which all profits gained during each term are distributed to the shareholders. Corporate earnings (y_t) are expressed as $y_t = y_{t-1} \cdot (1+\varepsilon_t)$. However, they are generated according to the process of $\varepsilon_t \sim N(0, \sigma_y^2)$ with share trading being undertaken after the public announcement of profits for the term [18]. Each investor is given common asset holdings at the start of the term with no limit placed on debit and credit transactions (1000 in risk-free assets and 1000 in stocks). Investors adopt the buy-and-hold method for the relevant portfolio as a benchmark[4] to conduct decision-making by using a one-term model.

2.2. Modeling Investor Behavior

Each type of investor handled in this analysis is organized in Table 1.[5] The investors in this market evaluate transaction prices based on their own forecasts for market tendency, taking into consideration both risk and return rates when making decisions. Each investor determines the investment ratio (w_t^i) based on the maximum objective function $(f(w^i t))$ as shown below[6] [19]:

$$f\left(w_t^i\right) = r_{t+1}^{\mathrm{int},i} \cdot w_t^i + r_f \cdot \left(1 - w_t^i\right) - \lambda\left(\sigma_{t-1}^{s,i}\right)^2 \cdot \left(w_t^i\right)^2. \tag{2.1}$$

Here, $r_{t+1}^{\mathrm{int},i}$ and $\sigma_{t-1}^{s,i}$ express the expected rate of return and risk for stocks as estimated by each investor i. r_f indicates the risk-free rate. w_t^i represents the stock investment ratio of the investor i for term t.

The expected rate of return for shares is calculated as follows [19]:

$$r_{t+1}^{\mathrm{int},i} = \frac{1 \cdot c^{-1}\left(\sigma_{t-1}^{s,i}\right)^{-2} \cdot r_{t+1}^{f,i} + 1 \cdot \left(\sigma_{t-1}^{s,i}\right)^{-2} \cdot r_t^{im}}{1 \cdot c^{-1}\left(\sigma_{t-1}^{s,i}\right)^{-2} + 1 \cdot \left(\sigma_{t-1}^{s,i}\right)^{-2}}. \tag{2.2}$$

Here, $r_{t+1}^{f,i} C r_t^{im}$ expresses the expected rate of return, calculated from short-term expected rate of return, and risk and gross current price ratio of stocks, respectively. c is a coefficient that adjusts the dispersion level of the expected rate of return calculated from risk and gross current price ratio of stocks [19].

The short-term expected rate of return ($r_t^{f,i}$) is obtained where ($P_{t+1}^{f,i}, y_{t+1}^{f,i}$) is the equity price and profit forecast for term $t+1$ is estimated by the investor, as shown below:

$$r_{t+1}^{f,i} = \left(\frac{P_{t+1}^{f,i} + y_{t+1}^{f,i}}{P_t} - 1 \right) \cdot \left(1 + \eta_t^i \right). \tag{2.3}$$

The short-term expected rate of return includes the error term ($\eta_t^i \sim N(0, \sigma_n^2)$) reflecting that even investors using the same forecast model vary slightly in their detailed outlook. The stock price ($P_{t+1}^{f,i}$), profit forecast ($y_{t+1}^{f,i}$), and risk estimation methods are described in Section 2.2.2.

The expected rate of return obtained from stock risk and so forth is calculated from stock risk (σ_{t-1}^i), benchmark equity stake (W_{t-1}), investorsf degree of risk avoidance (λ), and risk-free rate (r_f), as shown below [19, 20]:

$$r_t^{im} = 2 \cdot \lambda \cdot \left(\sigma_{t-1}^i \right)^2 \cdot W_{t-1} + r_f. \tag{2.4}$$

2.2.1. Stock Price Forecasting Method

The fundamental value is estimated by using the dividend discount model, which is a well-known model in the field of finance. Fundamentalists estimate the forecasted stock price and forecasted profit from profit for the term (y_t) and the discount rate (δ) as $P_{t+1}^{f,i} = y_t/\delta$, $y_{t+1}^{f,i} = y_t$.

Forecasting based on trends involves forecasting the next term stock prices and profit through extrapolation of the most recent stock value fluctuation trends. The next term stock price and profit is estimated from the most recent trends of stock price fluctuation (a_{t-1}) from time point $t-1$ as $P_{t+1}^{f,i} = P_{t-1} \cdot (1 + a_{t-1})^2$, $y_{t+1}^{f,i} = y_t \cdot (1 + a_{t-1})$.

Forecasting based on past averages involves estimating the next term stock prices and profit based on the most recent average stock value.

2.2.2. Risk Estimation Method

In this analysis, each investor estimates risk from past price fluctuations. Specifically, stock risk is estimated as $\sigma_{t-1}^i = \sigma_{t-1}^h$ (common to each investor). Here, σ_{t-1}^h represents the stock volatility that is calculated from price fluctuation from the most recent 100 terms.

2.3. Determination of Transaction Prices

Transaction prices are determined as the price where stock supply and demand converge ($\sum_{i=1}^{M} (F_t^i w_t^i)/P_t = N$). In this case, the total asset (F_t^i) of investor i is calculated from transaction price (P_t) for term t, profit (y_t) and total assets from the term $t-1$, stock investment ratio (w_{t-1}^i), and risk-free rate (r_f), as $F_t^i = F_{t-1}^i \cdot (w_{t-1}^i \cdot (P_t + y_t)/P_{t-1} + (1 - w_{t-1}^i) \cdot (1 + r_f))$.

2.4. Rules of Natural Selection in the Market

The rules of natural selection can be identified in this market. The driving force behind these rules is cumulative excess profit [21]. The rules of natural selection go through the following two stages: (1) the identification of investors who alter their investment strategy, and (2) the actual alteration of investment strategy [17, 22].

Each investor determines the existence of investment strategy alteration based on the most recent performance of each 5-term period after 25 terms have passed since the beginning of market transactions. The higher the profit rate obtained most recently is, the lesser the possibility of strategy alteration becomes. The lower the profit, the higher the possibility becomes. Specifically, when an investor could not obtain a positive excess profit for the benchmark portfolio profitability, they are likely to alter their investment strategy with the probability below:[7]

$$p_i = \min(1, \max(-100 \cdot r^{\text{cum}}, 0)). \tag{2.5}$$

Here, however, r_i^{cum} is the cumulative excess profitability for the most recent benchmark of investor i. Measurement was conducted for 1 term, 5 terms, and 25 terms, and the cumulative excess profitability was a profitability of one-term conversion.

Regarding the determination of a new investment strategy, an investment strategy that has a high cumulative excess profit for the most recent five terms (forecasting type) is more likely to be selected. Where the strategy of the investor i is z_i and the cumulative excess profit for the most recent five terms is r_i^{cum}, the probability p_i that z_i is selected as a new investment strategy is given as $p_i = e^{(a \cdot r_i^{\text{cum}})} / \sum_{j=1}^{M} e^{(a \cdot r_j^{\text{cum}})}$.[8] Those investors who altered their strategies make investments based on the new strategies after the next step.

3. Analysis Results

First of all, the case where investors make decisions based on past prices in the market is analyzed. Specifically, a market where there are investors that make forecasts based on past price trends and past price averages, as well as fundamentals, is analyzed. Afterwards, the case where there are investors that conduct passive investment strategy in the market is analyzed.

3.1. Where There Exist Investors That Forecast Based on Past Price Fluctuations

First of all, the influence of differences in the forecast accuracy of fundamentals on the market was analyzed. Afterwards, the influence of the difference in the forecast dispersion of investors other than fundamentalists was analyzed.

3.1.1. Influence of the Difference in the Forecast Accuracy of Fundamentals

This section analyzes the influence of the difference in the forecast accuracy of fundamentalists on the market where there exist heterogeneous investors in the market. From the start, there are a similar number of the three types of investors (Table 1: Type 1–3) in the market. In

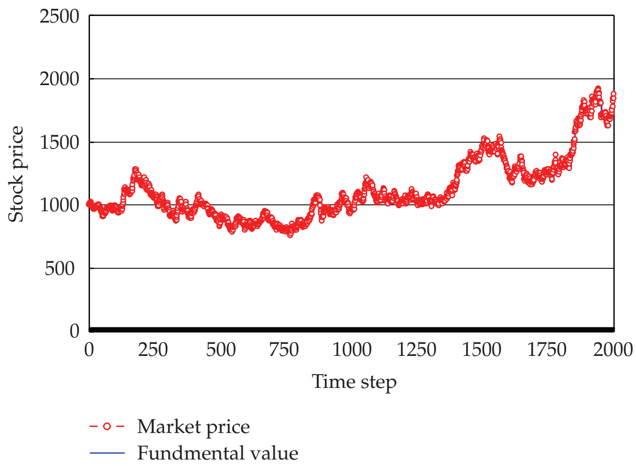


Figure 1: Price transitions ($\sigma_n = 1\%$).

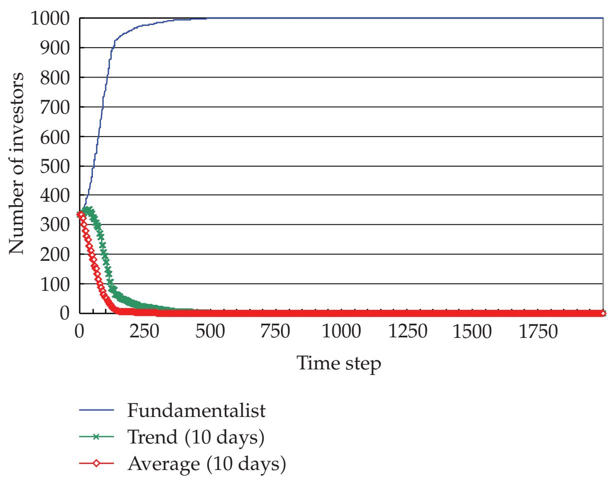


Figure 2: Transition of the number of investors ($\sigma_n = 1\%$).

the beginning, the case where the forecast dispersion of investors (σ_n) is 1% is described, and then the case where the forecast error of fundamentalists differs is described.

Where the Forecast Dispersion of Investors Is 1%

Figures 1 and 2 show the transitions of transaction prices and the number of investors. The typical price transitions obtained in this analysis are shown with respect to the transition of transaction prices. With regard to the transition of the numbers of investors, the average value obtained by conducting the analysis 50 times was used (the same being true in the following analysis). Figure 1 shows that transaction prices are consistent with fundamental values throughout the entire transaction period. Figure 2 confirms that the number of fundamentalists increases as time goes on.

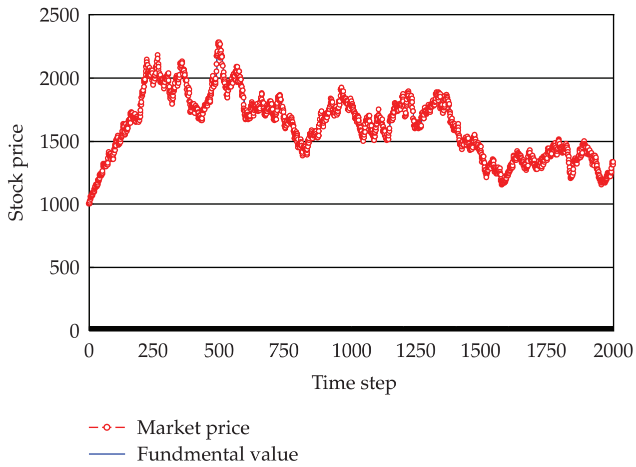


Figure 3: Price transitions ($\sigma_n = 2\%$).

Where the Forecast Error of Fundamentalists Is 2%

Figures 3 and 4 show the transaction prices and the transition of the number of investors where the forecast error of fundamentalists increases. In this analysis, the forecast error of fundamentalists (σ_n) is 2%, and the forecast dispersion of investors other than fundamentalists (σ_n) is 1%.

With respect to the transitions of transaction prices, as shown in Figure 3, it can be confirmed that transaction prices are consistent with fundamental values. Regarding the transition of the number of investors, similarly, the number of fundamentalists increases as time passes. When Figures 2 and 4 are compared regarding the rate of increase in the number of fundamentalists, the rate of increase is significant where fundamentalists have high accuracy in their forecasting. These results show that the better the forecast accuracy of fundamentalists becomes, the greater the rate of increase in the number of fundamentalists. The next section analyzes the case where the forecast error is 0%.[9]

Where the Forecast Error of Fundamentalists Is 0%

Figure 5 shows the transition of the number of investors where the forecast error of fundamentalists (σ_n) is 0%.[10] The forecast dispersion of investors other than fundamentalists (σ_n) is constant at 1%. Where Figure 5 is compared with Figures 2 and 4, the rate of increase of the number of fundamentalists is fastest in the case of Figure 5. When it comes to the rate of increase in the number of fundamentalists, the rate of increase goes up as the forecast accuracy improves. These results show that the better the forecast accuracy becomes, the more likely it is that fundamentalists can survive in the market. This result is consistent with traditional financial theory. The influence of the difference in the forecast accuracy of fundamentalists was analyzed in this analysis. Whether or not fundamentalists can survive in the market likely depends on the influence of the forecast dispersion of other investors. To confirm the influence of other investors, the next section analyzes the influence of forecast dispersion of investors other than fundamentalists.

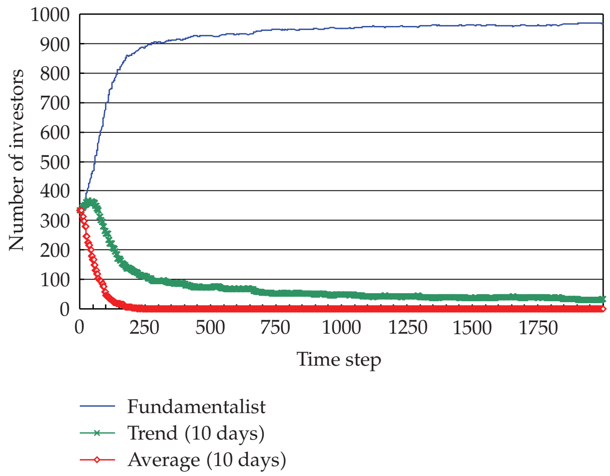


Figure 4: Transition of the number of investors (σ_n = 2%).

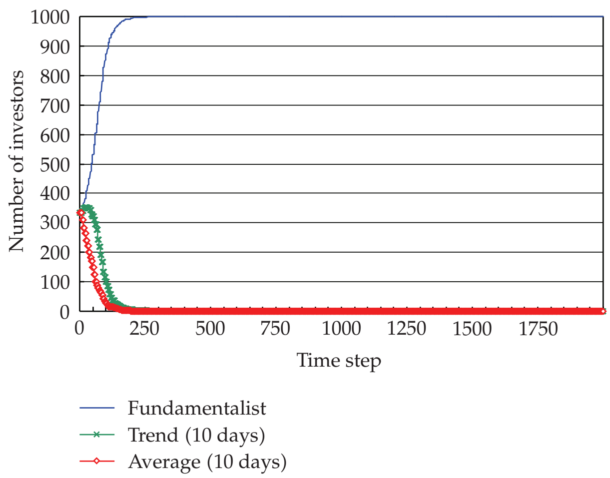


Figure 5: Transition of the number of investors (σ_n = 0%).

3.1.2. Influence of Forecast Dispersion of Other Investors

Here, the influence of forecast dispersion of other investors on the rate of increase in the number of fundamentalists is analyzed. The forecast accuracy of fundamentalists in this analysis (σ_n) is consistent at 1%.

Figures 6 and 7 show the transitions of the number of investors where the forecast dispersion of investors other than fundamentalists is 2% and 3%. As the same with the previous section with respect to the transition of the number of investors, the number of fundamentalists increases as time passes. Furthermore, the rate of increase becomes faster as the dispersion of investors other than fundamentalists becomes significant.[11] These results show that interaction with other investors should be taken into consideration in order to clarify the mechanism of financial markets. In this sense, these results are highly suggestive.[12]

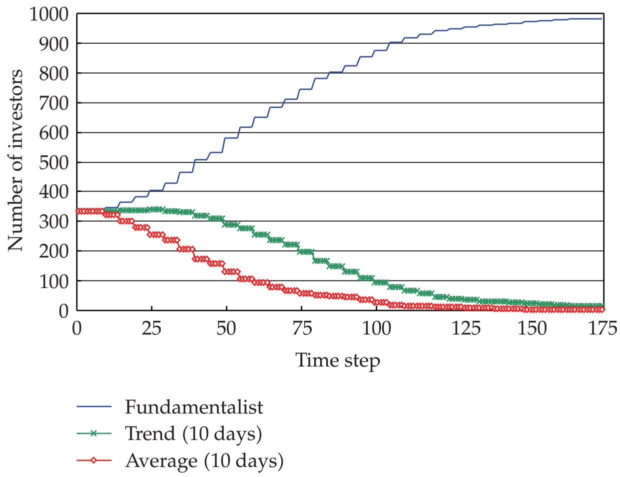


Figure 6: Transition of the number of investors ($\sigma_n = 2\%$).

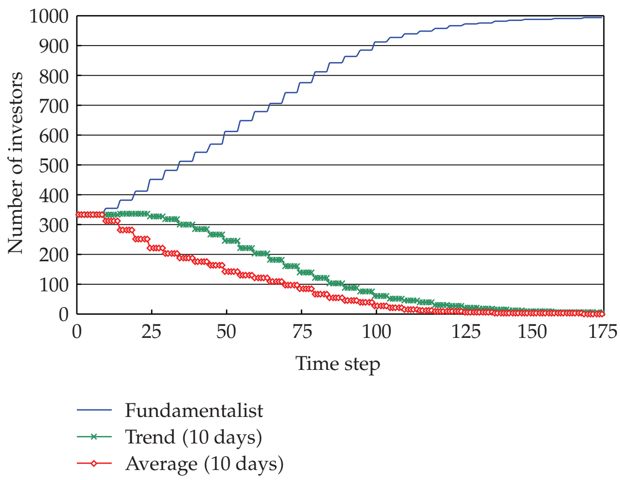


Figure 7: Transition of the number of investors ($\sigma_n = 3\%$).

3.2. Where There Exist Investors That Conduct Passive Investment Strategy

This section analyzes the case where there exist investors that conduct passive investment strategy. First of all, the influence of the difference in the forecast accuracy of fundamentalists is analyzed, and then another analysis considers more practical conditions.

3.2.1. Influence of the Difference in Forecast Accuracy of Fundamentalists

This section analyzes the influence of differences in the forecast accuracy of fundamentalists on the market where there exist investors that conduct passive management. In the early stages, there are a similar number of the four types of investors (Table 1: Type 1–4) in the market. First of all, the case where the forecast dispersion of investors (σ_n) is 1% is described. Afterwards, a case where the forecast error of fundamentalists is 0% is described.

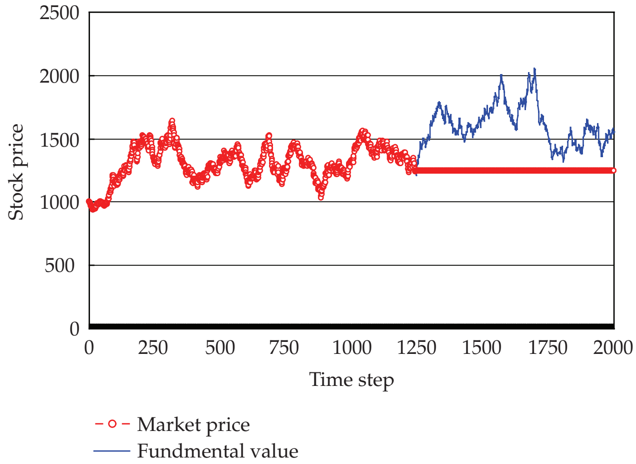


Figure 8: Price transitions ($\sigma_n = 1\%$).

Where the Forecast Dispersion of Investors Is 1%

Figures 8 and 9 show the transitions of transaction prices and the numbers of investors. Figure 9 shows that the number of investors that conduct passive investment strategy increases as time passes, and all the investors are conducting passive investment strategy after a certain period of time.[13] These results support the effectiveness of conducting passive investment strategy from the viewpoint of investment performance, which is consistent with traditional asset pricing theories [2]. Transaction prices, however, do not show fundamental values from around the middle of the transaction period. The same trend can also be confirmed in the case where the forecast accuracy of fundamentalists is 2%. Under the present conditions, the investment behavior of fundamentalists and passive management is the same on average. However, among fundamentalists, the number of investors that conduct passive investment strategy increases due to forecast error of fundamentalists.[14] Where the forecast accuracy of fundamentalists is good, ($\sigma_n = 0$), coexistence of fundamentalists and passive management can be predicted. The next section analyzes the case where the forecast error of fundamentalists (σ_n) is 0%.

Where the Forecast Error of Fundamentalists Is 0%

Figures 10 and 11 show the transitions of transaction prices and the number of investors where the forecast error of fundamentalists (σ_n) is 0%. In this analysis, the forecast dispersion of investors other than fundamentalists (σ_n) is constant at 1%. The price history shows that traded prices are consistent with fundamental values throughout the entire transaction period. Additionally, the transitions of the number of investors show that fundamentalists coexist with those investors that conduct passive investment strategy in the market. Just as with the present conditions, where there exist only two types of investors in the market, those investors that conduct passive investment strategy and fundamentalists, and where the forecast accuracy of fundamentalists is good ($\sigma_n = 0$), as a result, investment behavior of both investors becomes equal. Given this, both types of investors are likely to exist in the market.

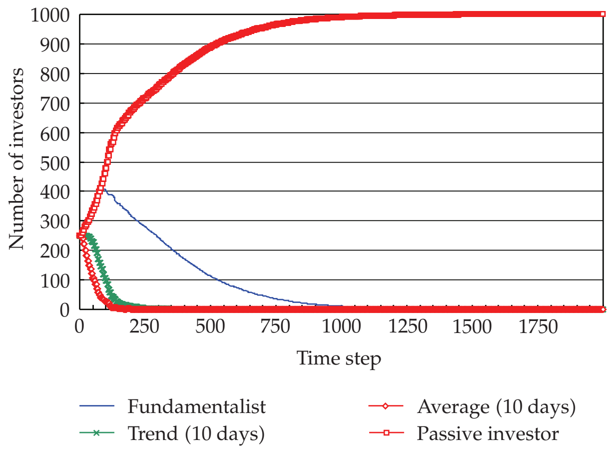


Figure 9: Transition of the number of investors ($\sigma_n = 1\%$).

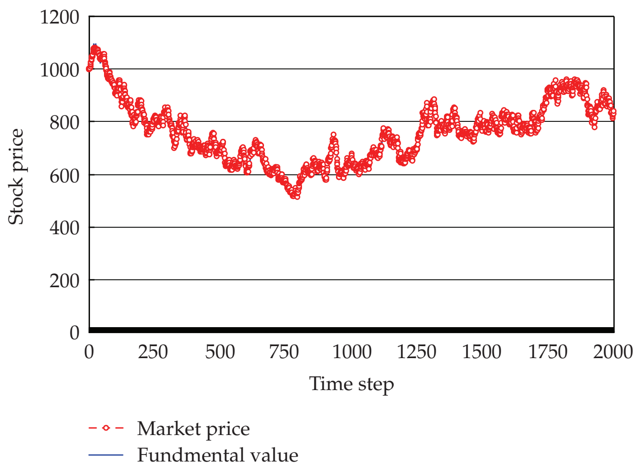


Figure 10: Price transitions ($\sigma_n = 0\%$).

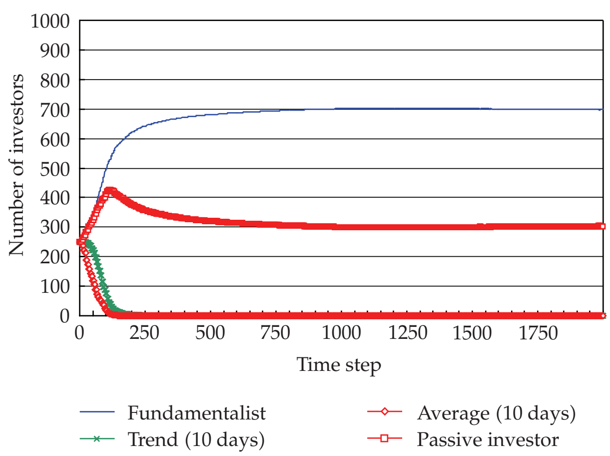


Figure 11: Transition of the number of investors ($\sigma_n = 0\%$).

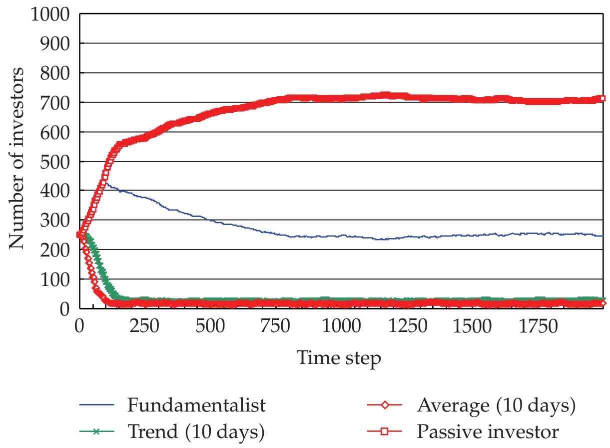


Figure 12: Transition of the number of investors (mutation: 1%, $\sigma_n = 1\%$).

The analysis conducted in this section confirmed that different situations could be generated. For example, fundamentalists and investors who employ passive investment strategy coexisted in the market according to the forecast accuracy level of fundamentalists or fundamentalists could die out in the market. These results suggest that the difference in estimation accuracy of fundamentalists should have a significant impact on the market. Thus, the results obtained in this analysis are very interesting.

3.2.2. Analysis That Considers the Actual Investment Environment

This section conducts analysis under conditions close to that of actual market conditions. In real markets, investors do not always determine their investment strategy based only on past investment performance. This section focuses on how to change investment strategy in order to analyze the case where some investors randomly change investment strategy.[15]

First of all, the case where 1% of investors changes investment strategy in a random manner is analyzed. Afterwards, another case, where there is an increase in the rate of investors in changing investment strategy randomly, is analyzed. In the early stage, there are a similar number of the four types of investors (Table 1: Type 1–4) in the market. The forecast dispersion of investors other than fundamentalists (σ_n) is constant at (1%).

Where a 1% of Investors Randomly Changes Investment Strategy

Figure 12 shows the transitions of the number of investors where the forecast accuracy of fundamentalists (σ_n) is 1%. This shows that fundamentalists and investors that conduct passive investment strategy coexist together in the market.[16] Since investors who randomly change investment strategy exist, fundamentalists and passive investors coexist in the market when fundamentalists' forecasts are not entirely accurate. The existence of a wide variety of investors can make it possible for fundamentalists to obtain more excess earnings through market transactions. As a result, the number of fundamentalists probably increases.[17]

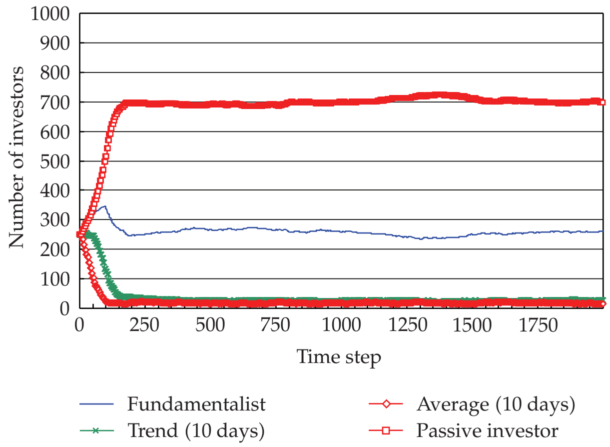


Figure 13: Transition of the number of investors (mutation: 1%, $\sigma_n = 2\%$).

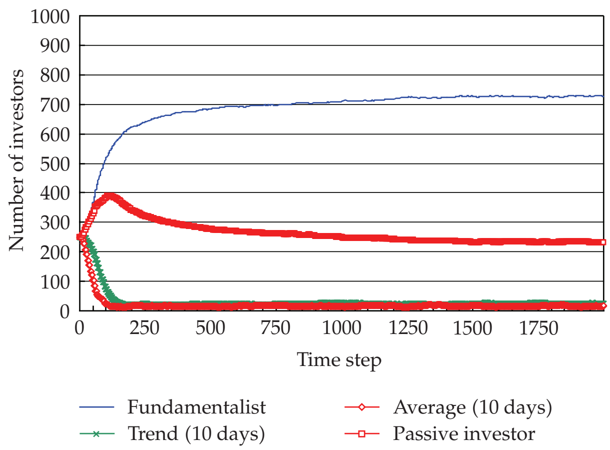


Figure 14: Transition of the number of investors (mutation: 1%, $\sigma_n = 0\%$).

Figures 13 and 14 show the transitions of the number of investors where the forecast accuracy of fundamentalists is 2% and 0%. These results can confirm that fundamentalists and investors that conduct passive investment strategy coexist. In addition, a comparison of Figures 13 and 14 shows that as the estimation accuracy of fundamentalists increases, there is a corresponding increase in the number of fundamentalists over time. These results are interesting. They show that the number of fundamentalists who can survive in the market is significantly influenced by the estimation accuracy level of fundamentalists.

Where the Rate of Investors That Randomly Change Investment Strategy Increases

Figure 15 shows the transitions of the number of investors where the rate of investors that randomly change investment strategy is 2%.[18] If there is an increase in the rate of

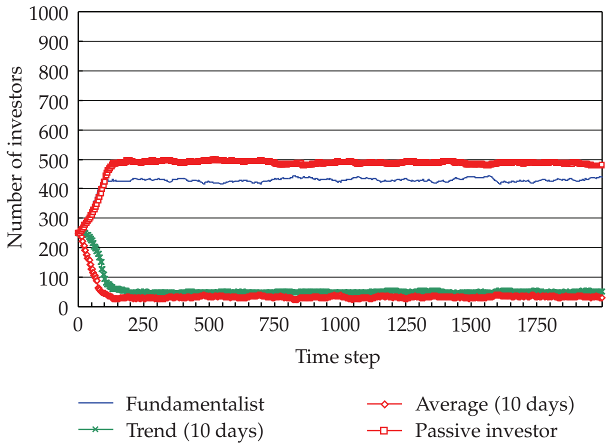


Figure 15: Transition of the number of investors (mutation: 2%, $\sigma_n = 1\%$).

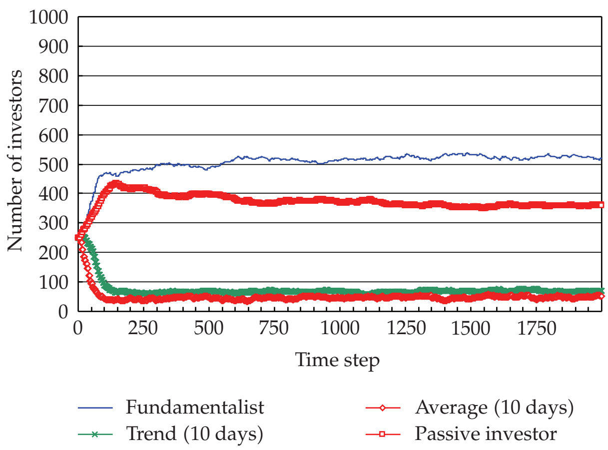


Figure 16: Transition of the number of investors (mutation: 3%, $\sigma_n = 1\%$).

investors who randomly change investment strategy, there is a corresponding increase in the number of fundamentalists.[19] Figure 16 shows the transitions of the number of investors where the rate of investors who randomly change investment strategy is 3%. This shows that the number of fundamentalists increases even further, exceeding the number of investors who conduct passive investment strategy.[20] In real markets, the effectiveness of passive investment strategy has widely been recognized from the viewpoint of practical business affairs as well as the academic standpoint. However, when the entire market is focused on, the rate of investors that adopt passive investment strategy is not always high. These results suggest that the existence of various sources of excess earnings should be included as one of the factors for the low ratio of passive investors. These are interesting results from both business and academic viewpoints.[21]

4. Summary

Using analyses of agent-based model, this research looked at the influence of the difference in the forecast accuracy of fundamental values on financial markets. As a result of this computer-based market analysis, the following findings were made: (1) improvements in the forecast accuracy of fundamentalists can contribute to an increase in the number of fundamentalists; (2) certain situations might occur, according to the level of the forecast accuracy of fundamentalists, in which fundamentalists and passive management coexist, or in which fundamentalists die out of the market, and furthermore; (3) where a variety of investors exist in the market, improvements in forecast accuracy could increase the number of fundamentalists more than the number of investors that conduct passive investment strategy. These results contribute to clarifying the mechanism of price fluctuations in financial markets and also indicate one of the factors for the low ratio of passive investors in real financial markets. At the same time, they indicate that agent-based modeling is effective in conducting analyses in the field of financial studies. The results obtained in this analysis have significant meaning from both an academic and a practical business viewpoint. A more detailed analysis that considers the actual investment environment should be included in future research.

5. List of Parameters

This section lists the major parameters of the financial market designed for this paper. The explanation and value for each parameter is described.

Parameters Abbreviations

M: Number of investors (1000)
N: Number of shares (1000)
F_t^i: Total asset value of investor i for term t ($F_0^i = 2000$: common)
W_t: Ratio of stock in benchmark for term t ($W_0 = 0.5$)
w_t^i: Stock investment rate of investor i for term t ($w_0^i = 0.5$: common)
y_t: Profits generated during term t ($y_0 = 0.5$)
σ_y: Standard deviation of profit fluctuation ($0.2/\sqrt{200}$)
δ: Discount rate for stock (0.1/200)
λ: Degree of investor risk aversion (1.25)
σ_n: Standard deviation of dispersion from short-term expected rate of return on shares (0.01–0.03)
a: Degree of selection pressure (10)
c: Adjustment coefficient (0.01)
r_t^{im}: Expected rate of share return as estimated from risk etc.
σ_t^s: Assessed value of standard deviation of share fluctuation
σ_t^h: Historical volatility of shares
P_t: Transaction prices for term t
$P_t^{f(,i)}$: Forecast value of transaction prices (of investor i) for term t
$y_t^{f(,i)}$: Forecast value of profits (of investor i) for term t
$r^{f(,i)}$: Short-term expected rate of return on shares (of investor i)
a_t: Price trend on stock until term t
r_i^{cum}: Cumulative excess return of investor i for the most recent five terms
p_i: Probability that investorsf who alter their strategy will adopt investor ifs strategy.

Endnotes

1. CAPM is also applied frequently to evaluate the enterprise value in Mergers and Acquisitions (M&A) [23].
2. Passive investment strategy has been well-known in the actual asset management businesses. On the other hand, active investment strategy that tries to obtain excess earnings using investments has been widely prevalent. There also exist investment trust funds that look for their basis of conducting active management in behavioral finance.
3. In the case of a financial market, investors represent agents and a stock market represents a multiagent system [17, 24].
4. Buy-and-hold method is an investment method to hold shares for medium to long term.
5. This analysis covered major types of investor behavior as the analysis object [9].
6. The value of objective function $f(w_t^i)$ depends on the investment ratio (w_t^i). The investor decision-making model here is based on the Black/Litterman model that is used in the practice of securities investment [19, 25].
7. In the actual market, evaluation tends to be conducted according to baseline profit and loss.
8. Selection pressures on an investment strategy become higher as the coefficients' value increases.
9. This is one of the characteristics of agent-based modeling where such an analysis can be conducted.
10. Where the forecast accuracy (σ_n) is 0%, there are no forecast errors by fundamentalists.
11. In other words, these results show that the rate of increase in the number of fundamentalists is influenced by the forecast dispersion of other investors.
12. A detailed analysis of the forecast accuracy of fundamentalists and the forecast dispersion of other investors needs to be carried out in the future.
13. See Takahashi et al. [6] for a detailed analysis of the influence of passive investment strategy on stock markets.
14. Under the present conditions, where the estimation accuracy of fundamentalists (σ_n) is 1% and 2%, all the investors conduct passive investment strategy in either case. In this sense, under the present conditions, the estimation accuracy of fundamentalists (σ_n) does not have any impact on whether or not fundamentalists can survive in the market.
15. Such a mechanism that works can make it possible for investors other than fundamentalists to always exist in the market.
16. In this case, transaction prices are consistent with fundamental values.
17. In the case under discussion here, market transactions consist of funds transferred between fundamentalists and other kinds of investors. The existence of various investors serves to provide the source of excess earnings for fundamentalists. These transactions serve to determine transaction prices and therefore conform to fundamental values.
18. The increase in the rate of investors who randomly change investment strategy means that there are more investors whose investments are based on trends and past averages in the market.
19. In this case, transaction prices are consistent with fundamental values.

20. This paper analyzes the relationship between microbehavior and macrobehavior under conditions where market prices are consistent with fundamental values. Analyzing the market under other conditions, such as when fundamentalists are eliminated from the market, will form part of our future work [17].
21. These results provide a significant suggestion with regard to the meaning of conducting active investment strategy.

References

[1] H. Markowitz, "Portfolio Selection," *Journal of Finance*, vol. 7, pp. 77–91, 1952.

[2] W. F. Sharpe, "Capital asset prices: a theory of market equilibrium under condition of risk," *The Journal of Finance*, vol. 19, pp. 425–442, 1964.

[3] F. Black and M. Scholes, "Pricing of options and corporate liabilities," *Bell Journal of Economics and Management Science*, vol. 4, pp. 141–183, 1973.

[4] J. E. Ingersoll, *Theory of Financial Decision Making*, Rowman & Littlefield, Lanham, Md, USA, 1987.

[5] E. Fama, "Efficient capital markets: a review of theory and empirical work," *Journal of Finance*, vol. 25, pp. 383–417, 1970.

[6] H. Takahashi, S. Takahashi, and T. Terano, "Agent-based modeling to investigate the effects of passive investment strategies in financial markets," in *Social Simulation Technologies: Advances and New Discoveries*, B. Edmonds, C. Hernandes, and K. Troitzsch, Eds., pp. 224–238, Idea Group, 2007, Representing the best of the European Social Simulation Association conferences.

[7] D. Kahneman and A. Tversky, "Prospect theory of decisions under risk," *Econometrica*, vol. 47, pp. 263–291, 1979.

[8] A. Tversky and D. Kahneman, "Advances in prospect theory: cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[9] A. Shleifer, *Inefficient Markets*, Oxford University Press, Oxford, UK, 2000.

[10] R. J. Shiller, *Irrational Exuberance*, Princeton University Press, Princeton, NJ, USA, 2000.

[11] R. Axtell, *Why Agents? On the Varied Motivation for Agent Computing in the Social Sciences*, Working Paper no.17, Brookings Institution Center on Social and Economic Dynamics, 2000.

[12] S. J. Russel and P. Norvig, *Artificial Intelligence*, Prentice-Hall, Upper Saddle River, NJ, USA, 1995.

[13] J. M. Epstein and R. Axtell, *Growing Artificial Societies Social Science from the Bottom Up*, MIT Press, Cambridge, Mass, USA, 1996.

[14] W. B. Arthur, J. H. Holland, B. LeBaron, R. G. Palmer, and P. Taylor, "Asset pricing under endogenous expectations in an artificial stock market," in *The Economy as an Evolving Complex System II*, pp. 15–44, Addison-Wesley, Reading, Mass, USA, 1997.

[15] R. Axelrod, *The Complexity of Cooperation -Agent-Based Model of Competition and Collaboration*, Princeton University Press, Upper Saddle River, NJ, USA, 1997.

[16] L. Tesfatsion, *Agent-Based Computational Economics*, Economics Working Paper, No.1, Iowa Sate University, Iowa, Miss, USA, 2002.

[17] H. Takahashi and T. Terano, "Agent-based approach to investors' behavior and asset price fluctuation in financial markets," *Journal of Artificial Societies and Social Simulation*, vol. 6, no. 3, 2003.

[18] P. O'Brien, "Analysts' forecasts as earnings expectations," *Journal of Accounting and Economics*, pp. 53–83, 1988.

[19] F. Black and R. Litterman, "Global portfolio optimization," *Financial Analysts Journal*, pp. 28–43, 1992.

[20] W. F. Sharpe, "Integrated asset allocation," *Financial Analysts Journal*, pp. 25–32, 1987.

[21] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.

[22] H. Takahashi and T. Terano, "Analysis of micro-macro structure of financial markets via agent-based model: risk management and dynamics of asset pricing," *Electronics and Communications in Japan, Part II*, vol. 87, no. 7, pp. 38–48, 2004.

[23] R. Brealey, S. Myers, and F. Allen, *Principles of Corporate Finance*, McGraw-Hill, New York, NY, USA, 8th edition, 2006.

[24] M. Levy, H. Levy, and S. Solomon, *Microscopic Simulation of Financial Markets*, Academic Press, Boston, Mass, USA, 2000.

[25] L. Martellini and V. Ziemann, "Extending black-litterman analysis beyond the mean-variance framework: an application to hedge fund style active allocation decisions," *Journal of Portfolio Management*, vol. 33, no. 4, pp. 33–45, 2007.