

Westphal, Christian

**Working Paper**

## The social costs of gun ownership: Spurious regression and unfounded public policy advocacy

MAGKS Joint Discussion Paper Series in Economics, No. 32-2013

**Provided in Cooperation with:**

Faculty of Business Administration and Economics, University of Marburg

*Suggested Citation:* Westphal, Christian (2013) : The social costs of gun ownership: Spurious regression and unfounded public policy advocacy, MAGKS Joint Discussion Paper Series in Economics, No. 32-2013, Philipps-University Marburg, Faculty of Business Administration and Economics, Marburg

This Version is available at:

<https://hdl.handle.net/10419/93516>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**MAGKS**



**Joint Discussion Paper  
Series in Economics**

by the Universities of  
Aachen · Gießen · Göttingen  
Kassel · Marburg · Siegen

ISSN 1867-3678

**No. 32-2013**

**Christian Westphal**

**The Social Costs of Gun Ownership: Spurious Regression  
and Unfounded Public Policy Advocacy**

This paper can be downloaded from  
[http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index\\_html%28magks%29](http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index_html%28magks%29)

Coordination: Bernd Hayo · Philipps-University Marburg  
Faculty of Business Administration and Economics · Universitätsstraße 24, D-35032 Marburg  
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: [hayo@wiwi.uni-marburg.de](mailto:hayo@wiwi.uni-marburg.de)

# The Social Costs of Gun Ownership: Spurious Regression and Unfounded Public Policy Advocacy<sup>a</sup>

Christian Westphal<sup>b</sup>

This version: July 8, 2013

## Abstract

In 2006, a study, published in the *Journal of Public Economics*, employing a panel regression of 200 U.S. counties across 20 years, found a significant elasticity of homicides with respect to firearms ownership. Based on this finding the authors made the public policy recommendation of taxing gun ownership. However that study fell prey to the ratio fallacy, a trap known since 1896. All the “explanatory power” (goodness-of-fit-wise and significance-wise) of the original analysis was due to regional and intertemporal differences and population being explained by itself. When the ratio fallacy is accounted for, all authors’ results can no longer be found. This is illustrated in this paper using a balanced panel from the data for 1980 to 2004. My findings are robust to (i) alternative specifications not subject to the ratio problem, (ii) using only data from 1980 to 1999 as in the original paper, (iii) using an unbalanced panel for 1980 to either 1999 or 2004, (iv) applying weighting as done by the original authors and (v) using data aggregated at the state level.

*JEL Classifications:* C51; H21; I18; K42

*Keywords:* Gun Ownership; Social Costs; Ratio Fallacy; Spurious Regression

---

<sup>a</sup>Thanks to participants of a Brown Bag Seminar held at Philipps-University Marburg and to participants of a Seminar on replication studies held at University of Göttingen both in which some good thoughts on the problem were had. Thanks to Bernd Hayo for contributing the ‘moving averages’ solution for the noisy proxy mentioned in section 7 and to Florian Neumeier for contributing the growth model used in section 6.4.

<sup>b</sup>University of Marburg, Faculty of Business Administration and Economics, Department of Statistics, christian.westphal@westphal.de, westphal@staff.uni-marburg.de

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>An overview</b>	<b>2</b>
2.1	Summary . . . . .	2
2.2	Criticism . . . . .	3
<b>3</b>	<b>Data Acquisition</b>	<b>4</b>
3.1	Data Sources and Extraction . . . . .	4
3.2	Resulting Dataset . . . . .	6
3.3	Comparison of Descriptives . . . . .	8
<b>4</b>	<b>Regression Analysis</b>	<b>9</b>
4.1	Model . . . . .	9
4.2	Data for Analysis . . . . .	11
4.3	Confirming results . . . . .	11
4.4	Estimation on First Differences . . . . .	13
<b>5</b>	<b>Discussion</b>	<b>13</b>
5.1	Ratio Fallacy . . . . .	13
5.2	First Differences . . . . .	14
5.3	But <i>E95</i> is Not Population . . . . .	15
5.4	Nonsense Regression Between Time Series . . . . .	16
5.5	Misspecification of the Original Model . . . . .	16
5.5.1	Testing for Misspecification . . . . .	16
5.5.2	Theoretical Bias . . . . .	18
<b>6</b>	<b>Alternative specifications</b>	<b>19</b>
6.1	Controlling for Population . . . . .	19
6.2	Algebraic Transformation . . . . .	21
6.3	Risk Model . . . . .	21
6.4	Growth Model . . . . .	22
6.5	Numerators and Denominators . . . . .	23
<b>7</b>	<b>Conclusion</b>	<b>24</b>

## 1 Introduction

The association between guns and crime has been and continues to be a topic of intense debate in society at large and among social scientists. The debate intensified after Lott and Mustard (1997) published results showing that crime declined following the passing of shall-issue laws<sup>1</sup> for concealed carry handgun licenses. This finding sparked a furious academic debate across disciplines. Political scientists, legal scholars, criminologists, economists and scientists working in medical fields all added their voices to the discussion. From my perspective, the most noteworthy (for data and methods used, as well as results) econometric studies on guns and crime appearing after Lott and Mustard (1997) include Ludwig (1998); Duggan (2001); Leenaars and Lester (2001); Cook and Ludwig (2003, 2006); Cook et al. (2007) and Leigh and Neill (2010). Some works in this area bear strongly worded titles that rather clearly reflect their authors' perspectives: "Shooting Down More Guns, Less Crime" (Ayres and Donohue 2003b), "The Final Bullet in the Body of the More Guns Less Crime Hypothesis" (Donohue 2003) and "The Latest Misfires in Support of the More Guns, Less Crime Hypothesis" (Ayres and Donohue 2003a). These titles illustrate the intensity of the debate, which is also evident in Lott (2010: Chapter 7). Academic research results became increasingly important in the public and legal arenas. This can be seen in Fox and McDowall (2008: III.1.A), which begins with the bold statement

"There is A Proven Correlation between the Availability of Handguns and Incidents of Violence."

and then goes on to draw on findings from Duggan (2001). In Fox and McDowall (III.1.C 2008), a result from (Cook et al. 2007: Section 4) is used to bolster their argument. Eventually, more refined econometric methods were applied to the issue. For example, Cook and Ludwig (2006), following the lead of Duggan (2001), apply advanced methods to very detailed data, taking into consideration many empirical problems.

I chose to revisit Cook and Ludwig (2006) (C&L hereafter) due to its rigor and the detailed description of the data sources from a preceding working paper (Cook and Ludwig 2004). The original objective was to address specialized econometric problems, such as the noisy proxy used and truncation of the data due to the logarithmic model, and to also possibly confirm the results with five more years of data. In this attempt, I made a surprising discovery: C&L ignored a statistical property of their data (ratios) leading to spurious results in regression analysis. Even more surprising is that this pitfall has been known about for more

---

<sup>1</sup>A "shall-issue" law forces a state to issue concealed carry licenses to any applicant. No reasons need to be given by the applicant; as long as he does not have any convictions or mental disorders the license must be issued.

than a century (Pearson 1896). This statistical property is the only reason C&L arrived at their result, based on which they advocated taxing gun ownership.

To make educated and welfare-maximizing decisions, public planners often rely on scientific findings. If these findings are biased or spurious, any public policy based on them may not have its intended effect and in the worst case could actually be harmful. To some, it may be “obvious” that externalities are imposed upon others by firearm possession. But even the “obvious” should be backed up by evidence if public policies, not to mention public funds, are going to be directed toward the issue and, unfortunately, C&L’s results are not appropriate for this purpose. Their results are a statistical artifact of a well-known problem in regression analysis. This is my main finding and it is demonstrated in detail below. This work thus contributes to keeping spurious results out of the public policy debate. Furthermore the problem is illustrated in enough detail that other researchers may be alerted to this easy-to-miss problem and thus avoid it in their own empirical analyses.

This paper is organized as follows: C&L’s original study is summarized and put into scientific context in Section 2. Section 3 describes the acquisition of the data necessary to repeat their analysis, and makes my analysis replicable by other researchers. Indeed, only those readers interested in such a replication need read Subsections 3.1 and 3.2. The results from Cook and Ludwig (2006) are repeated in Sections 3.3 and 4.3, with a sharp twist in the results in Section 4.4 nullifying C&L’s original conclusion. Section 5 then shows how the results from Cook and Ludwig (2006) are spurious (mostly) due to ignoring the ratio fallacy – a problem known since Pearson’s (1896) work and worked out in detail by Kronmal (1993). This finding is confirmed via a battery of robustness checks in Section 6, some of which also present possible fixes for the earlier model specification. None of the models yield significance for the parameter of interest.

## 2 “The Social Costs of Gun Ownership”

### 2.1 Summary

Cook and Ludwig (2006) appears to be a rigorous analysis of the relationship between guns and crime. The authors use the advanced method of panel analysis and analyze a comprehensive data set covering 200 U.S. counties and 20 years. The results are presented in a clear fashion and a specific public policy recommendation is made.

*Framework:* The analysis assumes that gun ownership may impose externalities on society (Cook and Ludwig 2006: 379–380), specifically that more guns may result in more homicides.

*Measures:* Due to a lack of administrative data on gun ownership, a proxy

is used. That proxy is the fraction of suicides committed with a firearm (Cook and Ludwig 2006: 380), that is, “firearm suicides” divided by “suicides” ( $FSS$  or  $FS/S$ ). This proxy

- (a) seems reasonable as in a society with zero guns,  $FSS$  will be zero, and in a society where everyone has access to guns, all else equal, it very likely will achieve its maximum value.
- (b) is confirmed to function in the way intended by other studies (Azrael, Cook and Miller 2004; Kleck 2004), at least for the cross-section.
- (c) is supported by evidence that with increased availability of firearms the number of shooting suicides increases (see Klieve, Barnes and De Leo 2009; Klieve, Sveticic and De Leo 2009; Leenaars et al. 2003; Kellermann et al. 1992).
- (d) continues to be valid if
  - i. other methods of suicide are substituted by firearms suicide (as suggested by Klieve, Barnes and De Leo 2009) with increased availability of guns, or
  - ii. the availability of guns increases the overall number of suicides through suicides by shooting (for which there is some evidence; see Leenaars et al. 2003; Klieve, Sveticic and De Leo 2009).<sup>2</sup>

For homicides, the numbers of homicides on the county level are used.

*Data:* The data used are a panel across the 200 U.S. counties with the largest populations (measured in 1990) and for the period 1980 to 1999. Statistics on population size and number of homicides and suicides, as well as some sociodemographic controls for each county, are available. From this information, a panel of ratios is computed with the appropriate numerators and denominators.

*Methods:* The panel of ratios is analyzed by a two-way (individual and time) fixed effects panel model on the logarithms. A variant of the estimating equation with a full description of the variables used can be found in Section 4.1. Different model specifications (different levels of aggregation, different sets of controls) are compared.

*Main Result:* From the logarithmic model an *elasticity* of the homicide rate with respect to the firearms ownership measure significant at the 5% level and on the order of 0.10 is estimated. From this result, an appropriate tax on gun ownership is calculated to be in the range of USD 100 to USD 1,800 depending on the local levels of gun ownership and homicides.

## 2.2 Criticism

Moody and Marvell (2010) note that C&L switch their sets of controls in the “crime equation” between Cook and Ludwig (2006), Cook and Ludwig (2002),

---

<sup>2</sup>I can find no literature discussing the case for no or a negative correlation between the number of firearms and the number of suicides (or suicides by shooting).

and Cook and Ludwig (2003) without giving any reasons for doing so. Kleck (2009) has several criticisms, including (a) that C&L's method of dealing with causal dependence is overly simple, (b) that the *FSS* proxy may not be valid for measuring trends in gun ownership, and, similar to Moody and Marvell's argument, (c) that the controls used are arbitrarily chosen and that some possible necessary controls are missing from the model. This last criticism is valid, but many sociodemographic controls can be substituted for each other and therefore I do not consider such a change in control variables – possibly attributable to data availability – to detract from the value of a study; also the fixed effects model will be able to capture any unobserved variables that do not change over time. Causality remains a problem and causal relationships have to be interpreted with care. Indeed, finding an *association* but stopping short of calling it *causality* seems prudent for the topic of guns and crime. Once an association is found, of course, it is worth trying to discover causal relationships.

## 3 Data Acquisition

### 3.1 Data Sources and Extraction

The aggregated data set used in Cook and Ludwig (2006) is not published and the authors chose not to share it with me. I thus acquired the data from the primary sources given in Cook and Ludwig (2004: Appendix 3). This allowed me to include five more years of data. The four data sources used are:

1. *CDC Wonder*: I used the CDC Wonder database<sup>3</sup> to obtain the *yearly population figure* for the counties. This database was also used to select the *200 counties with the largest population in 1990*<sup>4</sup> and the geographic *FIPS codes valid in 2009*. CDC Wonder cannot be used to extract statistics on number of homicides, firearm homicides, suicides, or firearm suicides as those numbers are suppressed in the later years for many of the 200 counties.
2. *Mortality Detail Data*: Cook and Ludwig (2004: 45) list the exact data sources used. These are ICPSR study data sets<sup>5</sup> 07632, 06798, 06799, 02201, 02392, 02702, 03085, 03306, 03473, 04640, 20540 and 20623, in chronological order. Each of these data sets contains micro data for approximately 2 million deaths in the United States for one year. No geographic codes are available

---

<sup>3</sup>United States Department of Health and Human Services (2010).

<sup>4</sup>The set of selected counties does not change if the 1990 census population from United States Census Bureau (1990) is used instead.

<sup>5</sup>United States Department of Health and Human Services. Centers for Disease Control and Prevention. National Center for Health Statistics (2010); United States Department of Health and Human Services. National Center for Health Statistics (1997, 2008a, b, 2009a, b, 2008c, 2007a, b, c, d, e)



for 2005,<sup>6</sup> and later data was not available to me.

3. *Sociodemographic Controls*: C&L had access to ICPSR study dataset 06054.<sup>7</sup> This data set was not available to me. I used the 1980, 1990, and 2000 censuses<sup>8</sup> directly, with the code also available from Westphal (2013).
  - (a) For the *1980 census*, data were extracted from the summary tape file 3C. All data aggregated above or below county level were dropped. From Table 1, total population and rural population (used to compute urban population) are obtained. Table 10 gives the number of households. Table 12, cell 2 contains the number of blacks. Table 15, sum of cells 1–3 (males) and 28–30 (females) gives the population younger than five years (used to compute population aged five years and older). Table 20, cells 5–7 contains the number of female-headed households identifiable from the census data.<sup>9</sup> Table 34 contains the number of people who have been living in the same house for five years or more.
  - (b) Summary tape file 3A was used for the *1990 census*. The total population is taken from Table P3; the respective number of households from P5. P6 gives the numbers of rural inhabitants. P8 contains the number of blacks. The sum of P13's first three cells yields the population below the age of five years. P17, in cells 10 and 11, gives the number of female household heads identifiable from this census. P43 contains the number of people who have lived in the same house for five years or more.
  - (c) Summary file 3 for the *2000 census* – downloadable at the county level from the American Factfinder application<sup>10</sup> – in Table P005, cell VD01 contains the figure for total population. The same table, cell VD05 gives rural population. Table P006, cell VD03 yields the number of blacks. Table P008, cells VD03 – VD07 (males) and VD42 – VD46 (females), are the numbers of persons below five years of age. Table P009, cell VD21 contains the number of female-headed households identifiable from this census. Table P014, cell VD01 contains the number of households.
4. *Other Crime Data*: The FBI's Uniform Crime Reports are available via ICPSR study datasets<sup>11</sup> 08703, 08714, 09252, 09119, 09335, 09573, 09785, 06036,

---

<sup>6</sup>United States Department of Health and Human Services. National Center for Health Statistics (2008d).

<sup>7</sup>United States Department of Commerce. Bureau of the Census (1993).

<sup>8</sup>United States Census Bureau (1980, 1990, 2000). C&L did not name the 1980 census as a data source in Cook and Ludwig (2004), but it is mentioned in Cook and Ludwig (2006: Table 1). Relying only on the 1990 and 2000 censuses does not change any of the findings discussed below.

<sup>9</sup>The 2010 census gives the number of all female-headed households; it is remarkably higher than the number identifiable from the 1980, 1990 and 2000 censuses.

<sup>10</sup>For detailed download procedures see Westphal (2013).

<sup>11</sup>United States Department of Justice. Federal Bureau of Investigation (2006a, b, r, 2005, 2006s, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q).

06316, 06669, 06850, 02389, 02764, 02910, 03167, 03451, 03721, 04009, 04360 and 04466. These contain reported crime numbers aggregated at the county level. Study dataset 06545<sup>12</sup> for the 1993 Uniform Crime Report data was not available for download at the time of writing.

For data extraction, I used R<sup>13</sup> with Grothendieck (2011), the latter very conveniently allowing SQL operations on R data frames. For the *mortality detail data*, the strategy is to read in the tab-separated file or fixed-width file for one year, and then drop all deaths occurring in counties not on the list of the 200 counties mentioned above. Next, count homicides, firearm homicides, suicides, and firearm suicides<sup>14</sup> – coded by either ICD9 or ICD10 – by county using SQL `count()`.<sup>15</sup> For the *control variables* and the *other crime data*, the data are already aggregated at the county level. The controls have to be interpolated/extrapolated<sup>16</sup> between/from the census years. The 2010 census was not used: the definition of female-household head was changed for that census and the number of people living in the same house for the last five years is missing from the 2010 summary file.

Different geographical coding schemes are found in the data: NCHS<sup>17</sup> coding and FIPS<sup>18</sup> coding. NCHS coding changes with each census and FIPS coding changes as counties are renamed or restructured. Changes relevant to the 200 largest counties between 1979 and 2004 are shown in Tables 1 and 2. These changes may lead to mismatched assignment of values if ignored during data extraction; thus each data source and each year had to be individually checked for such changes. There are many potential sources of error here. Detailed instructions and all code used can be found on my personal website (Westphal 2013) for individual use and critical review. To replicate Cook and Ludwig (2006), the five New York City counties are aggregated into one artificial county.<sup>19</sup>

### 3.2 Resulting Dataset

The resulting dataset shows 24 variables for  $K = 196$  counties in  $T = 26$  years (1979-2004). These variables include *five index variables*, namely, *year* as a time

<sup>12</sup>United States Department of Justice. Federal Bureau of Investigation (2008).

<sup>13</sup>R Development Core Team (2011).

<sup>14</sup>This includes homicides and suicides by explosives as these are not distinguishable in ICD9 coding.

<sup>15</sup>The numbers extracted this way were confirmed using the Stata files provided by ICPSR and repeating the data extraction using independent Stata code, also available from Westphal (2013).

<sup>16</sup>I used linear interpolation.

<sup>17</sup>National Center for Health Statistics.

<sup>18</sup>Federal Information Processing Standard.

<sup>19</sup>I assigned a “FIPS” code of 36998 to that artificial county.

State	County	NCHS 1970	NCHS 1980	FIPS 6-5 2009
Missouri	St. Louis City	26096	26097	29510
Missouri	St. Louis County	26095	26096	29189
Nevada	Clarc County	29002	29003	32003
New York	Kings County	33029	33029	36047
New York	Queens County	33029	33029	36081
New York	New York County	33029	33029	36061
New York	Bronx County	33029	33029	36005
New York	Richmond County	33029	33029	36085
Virginia	Norfolk City	47369	47088	51710
Virginia	Virginia Beach City	47402	47127	51810
Virginia	Fairfax County	47087	47040	51059

Table 1: NCHS code changes between 1970 and 1980 according to <http://www.nber.org/mortality/errata.txt>

Change notice	Year	Affects (out of 200 largest counties in 1990)
2	1992	none
3	1995	none
4	1999	none
5	1999	Dade County, FL changed its name to Miami-Dade County, FL. FIPS changed from 12025 to 12086.
6	2002	Parts of Adams County, CO and Jefferson County, CO now are part of Broomfield County, CO.
7	2001	none
8	2007	none
9	2008	none
10	2008	none

Table 2: FIPS 6-5 changes affecting the 200 largest counties in 1990 according to <http://www.census.gov/geo/www/ansi/changenotes.html>

index, and *nchs* and *fips*, and the tuple of *state* and *county* as interchangeable individual identifiers for the counties.

There are population numbers (per county, per year):

- (1) *pop*: the population number from United States Department of Health and Human Services (2010),
- (2) *total*: the (interpolated) population number from the censuses,
- (3) *UCRpop*: the population number from the FBI's Uniform Crime Report,
- (4) *total5plus*: the (interpolated) population number from the census for persons of five years and older,
- (5) *households*: the (interpolated) number of households from the censuses,
- (6) *deaths*: the number of all deaths (not used).

There are four numbers involving homicides and suicides (variables named by the respective ICD9 code):

- (1) *E96*: homicides,
- (2) *E965*: firearm homicides,
- (3) *E95*: suicides,
- (4) *E955*: firearm suicides.

The remaining nine variables are control variables:

- (1) *resid5Yago*: number of residents not having moved in the last five years,
- (2) *rural*: number of residents living in rural areas,
- (3) *black*: number of black residents,
- (4) *fhh*: number of female household heads,
- (5) *UCRmurder*: number of murders from the UCR (not used),
- (6) *UCRrape*: number of rapes from the UCR (not used),
- (7) *UCRrobbery*: number of robberies from the UCR,
- (8) *UCRassault*: number of assaults from the UCR (not used),
- (9) *UCRburglary*: number of burglaries from the UCR.

I then used these numbers to calculate the percentages with the appropriate denominator. Usually, the denominator is *pop* except for the ratio of people not having moved in the last five years (denominator is *total5plus*) and the ratio of female household heads, where the denominator is the number of *households*. Switching the denominator to either *total* or *UCRpop* changes the results only marginally from those reported below; correlation between *pop* and *total* is  $> 0.999$  and correlation between *pop* and *UCRpop* is  $> 0.989$ .

### 3.3 Comparison of Descriptives

There are detailed descriptives in Cook and Ludwig (2006: 382 and Table 1). I compare my data to those aggregates. The values computed from my dataset are found in Tables 3 and 4, with the values from the original article in parentheses. To avoid comparing different time periods, I restricted the comparison of

descriptives to 1980–1999, the years used in the original study.

%	0	10	20	30	40	50	60	70	80	90	100
Quantile	4	28	35	41	47	55	65	80	106	156	1156
	(NA)	(27)	(NA)	(NA)	(NA)	(52)	(NA)	(NA)	(NA)	(142)	(NA)

Table 3: Quantiles of number of suicides for the 200 largest counties over 1980 to 1999, values from Cook and Ludwig (2006) in parentheses

C&L (p. 383) give all values “weighted by county population”. This does not make any sense for values whose denominator is not county population. Therefore, I added a column *weight* to Table 4. This column allows comparing my data to those of C&L while at the same time giving the correct descriptives. Indiscriminately weighting by county population will not result in the sample mean if the variable does not have county population as a denominator. For example, for the average number of suicides per county in

$$\sum_{k=1}^K \sum_{t=1}^T \frac{E95_{k,t} pop_{k,t}}{KT \overline{pop}} \neq \overline{E95}, \quad (1)$$

where, from now on,  $k$  is the county index and  $t$  is the time index.

Note that the descriptives from Cook and Ludwig (2006: Table 1) and those from my dataset are very similar. Differences may be due to slightly different data sources,<sup>20</sup> a slightly different set of observations used for computation,<sup>21</sup> or, possibly, revised data.

## 4 Regression Analysis

### 4.1 Model

C&L used the logarithmic two-way fixed effects model with  $t = 1, 2, \dots, T$  indicating the years and  $k = 1, 2, \dots, K$  indicating the counties:

$$\ln Y_{k,t} = \beta_1 \ln FSS_{k,t-1} + X_{k,t} \beta_2 + d_k + d_t + \varepsilon_{k,t}. \quad (2)$$

$X_{k,t}$  contains the logarithmic values of the ratios used as controls, namely, (i) burglary rate, (ii) robbery rate, (iii) percentage black, (iv) percentage urban, (v) percentage 5+ year residents and (vi) percentage female-headed households. They include the proxy  $FSS = E955/E95$  lagged by one year to circumvent possible reverse causation, i.e., people buying guns because of a higher homicide rate. The

<sup>20</sup>Remember, I had no access to the ICPSR study files on the censuses.

<sup>21</sup>I do not know if the values from Cook and Ludwig (2006) are calculated on the full data set or on their selection of 3,822 observations used for the regression analysis.

	weight	Full sample (largest 200)	Bottom quartile 1980 FSS	Top quartile 1980 FSS
<i>Full period (1980-1999)</i>				
FSS	<i>E95</i>	52.47	35.80	66.38
	<i>pop</i>	49.98	34.54	66.18
	<i>pop</i>	(49.9)	(34.6)	(66.9)
Homicide rate per 100'000	<i>pop</i>	11.30	11.00	14.27
	<i>pop</i>	(11.0)	(10.9)	(14.4)
Gun homicide rate per 100'000	<i>pop</i>	7.46	6.92	9.93
	<i>pop</i>	(7.3)	(6.9)	(10.1)
% Urban	<i>pop</i>	93.68	95.14	92.66
	<i>pop</i>	(92.6)	(94.7)	(91.8)
% Black	<i>pop</i>	14.33	16.75	18.64
	<i>pop</i>	(14.0)	(13.5)	(19.5)
% not moved in last 5 years	<i>total5plus</i>	57.76	59.43	48.35
% Female household head	<i>households</i>	17.36	18.60	16.40
	<i>pop</i>	17.18	18.40	16.28
	<i>pop</i>	(18.0)	(20.1)	(18.5)
Burglary rate	<i>pop</i>	1339	1218	1643
Robbery rate	<i>pop</i>	319	415	291
Avg. # Suicides per county	1	83.43	80.69	78.97
	<i>pop</i>	194.17	193.50	123.30
	<i>pop</i>	(195.8)	(192.5)	(120.0)
<i>FSS in selected years</i>				
1980	<i>E95</i>	49.91	29.28	72.72
	<i>pop</i>	48.01	29.20	73.11
	<i>pop</i>	(48.0)	(29.2)	(73.3)
1990	<i>E95</i>	54.93	37.87	68.36
	<i>pop</i>	52.57	36.78	68.48
	<i>pop</i>	(52.8)	(37.2)	(69.1)
1999	<i>E95</i>	50.56	35.71	60.01
	<i>pop</i>	48.18	34.93	59.81
	<i>pop</i>	(48.0)	(34.9)	(59.8)

Table 4: Descriptive statistics for county data revisited, values from Cook and Ludwig (2006: Table 1) in parentheses

dependent variable  $Y$  is the homicide rate  $E96/pop$ . Results for this and those in the following sections are qualitatively the same when the rate of firearms homicides  $E965/pop$  is used. Results may be found at Westphal (2013). In their model, they include another constant term,  $\beta_0$ , but as we know this is either caught in the time and county dummies, or the model is overspecified, or we need to impose a restriction on one set of dummies<sup>22</sup>, so I do not include it in Equation (2).

## 4.2 Data for Analysis

Ratios taking a value of zero have to be excluded from the analysis as their logarithm is  $-\infty$ . There are several ways of excluding observations containing a ratio of zero: unbalance the panel or remove counties or years (whichever is less costly) in order to keep the panel balanced. For the remainder of this article I present results from a balanced panel. All observations from 1993 are removed as 1993 UCR data<sup>23</sup> were not available at the time of writing. I also removed all counties having a zero in either of the ratios' numerators (see Table 4). Results for this and those presented in the following sections are qualitatively the same and numerically close when using (various subsets of) the unbalanced panel.<sup>24</sup> The resulting balanced panel is 25 years long (1979 to 2004 without 1993) and 142 counties wide, i.e. there are 3,550 observations. Descriptive statistics do not differ much from those set out in Table 4. From those values, rates, logarithms of rates, and lagged values for  $FSS$  are added to the dataset; 3,408 values remain for 1980 to 2004 after balancing on the lags and differences.

## 4.3 Confirming results

Estimating Equation (2) from this balanced panel yields the estimation output in Table 5, column labeled "Equation (2)". The results are only slightly different from the results in Cook and Ludwig (2006: Table 2, final column). The sample used is different (five more years and balanced) and I did not apply weighting on the panel model, as this is rarely done in the econometric literature. The lack of efficiency can be dealt with after estimation by applying Driscoll and Kraay (1998) via Croissant and Millo's (2008) `vcovSCC` function.<sup>25</sup> Contrary to Cook

---

<sup>22</sup>An example would be  $\sum_k d_k = 0$ .

<sup>23</sup>United States Department of Justice. Federal Bureau of Investigation (2008).

<sup>24</sup>For the unbalanced panel weighting as used by C&L is needed to achieve significance. All results may be found at Westphal (2013).

<sup>25</sup>By applying weighting to account for heteroscedasticity (Cook and Ludwig 2006: 382) and calculating standard errors that are robust to heteroscedasticity (Cook and Ludwig 2006: 382), C&L basically "double correct" for heteroscedasticity. I did not find any econometric literature on this approach; however their weighting may be viewed as easily justifiable "importance weighting".

Years	1980-1999		1980-2004		
	Estimates by C&L $N = 3822$	Equation (2) fixed effects $N = 3408$		Equation (3) first differences $N = 3266$	
Explanatories	Coef.	Coef.	SE	Coef.	p-val
$\ln FSS_{t-1}$	0.086*	0.074*	0.029	0.011	0.67
$\ln$ robbery rate	0.149***	0.111**	0.042	0.028	0.38
$\ln$ burglary rate	0.226***	0.125***	0.033	0.020	0.67
$\ln$ % black	0.278 <sup>†</sup>	0.141**	0.050	-0.071	0.80
$\ln$ % urban	-0.537***	-0.490**	0.160	-0.319	0.77
$\ln$ % same house 5 yr ago	-0.690	-0.467*	0.235	-0.045	0.96
$\ln$ % female headed house	-0.303	0.650***	0.186	0.914	0.31
$R^2_{within}$	NA	0.060		0.001	

Table 5: Estimation output, <sup>†</sup> :  $p < 0.10$ , \* :  $p < 0.05$ , \*\* :  $p < 0.01$ , \*\*\* :  $p < 0.001$ , robust standard errors according to Driscoll and Kraay (1998) computed with `vcovSCC` from Croissant and Millo (2008),  $R^2$  adjusted

and Ludwig (2006: Table 3, model 3), who needed weighting to achieve significance on  $\beta_1$ , significance on the balanced panel is achieved without weighting.<sup>26</sup> This may be due to the large errors in the proxy, as noted by Cook and Ludwig (2006: 382), which bias the coefficient towards zero. Balancing the panel by excluding zeroes favors counties with less error in the proxy, that is, larger counties in terms of population, which therefore, all else equal, have a lesser chance of producing zeroes, are favored by the balanced panel.

The within  $R^2$  reported in Table 5 is magnitudes smaller than the  $R^2$  of around 0.9 reported by Cook and Ludwig (2006: Table 2) for all their models. They reported the  $R^2$  from the least squares dummy variable estimation. That measure includes the fit from the dummies. The within  $R^2$  only includes the fit from the ratios in  $X_{k,t}$  and  $FSS_{k,t-1}$ . This tells us that much of the variation comes from regional and/or intertemporal differences, caught by the dummies. The coefficient on the female household heads changes sign between the original study and my estimation, but this does not affect the arguments in Sections 4.4, 5 or 6. For  $\beta_1$ , we can say the significant positive result from Cook and Ludwig (2006) is confirmed by my estimation.

<sup>26</sup>Notably despite having used the same data sources I cannot exactly replicate the results from Cook and Ludwig (2006). The following data have been updated since their work, but exactly what changes were made is not known: United States Department of Health and Human Services. Centers for Disease Control and Prevention. National Center for Health Statistics (2010: data sets 27, 28, 29).



#### 4.4 Estimation on First Differences

Model (2) can be reformulated on the first differences, as is well known from reading any econometric textbook on panel analysis (e.g., Wooldridge 2002: Section 10.6). The individual fixed effects disappear from the model, the time fixed effects are transformed to the differences between the time fixed effects, and the errors are transformed.<sup>27</sup> The coefficients on the variables of interest remain the same mathematically, as can be seen from Equation (3).

$$\begin{aligned}\Delta \ln Y_{k,t} &= \ln Y_{k,t} - \ln Y_{k,t-1} \\ \Delta \ln Y_{k,t} &= \beta_1 \Delta \ln FSS_{k,t-1} + \Delta X_{k,t} \beta_2 + \delta_t + \nu_{k,t}\end{aligned}\quad (3)$$

Therefore, when estimating Equation (3) we would expect similar results in size and significance as achieved from estimating Equation (2). Looking at Table 5 reveals that *all significance has disappeared from the model*. This will be discussed in the next section.

## 5 Discussion

### 5.1 Ratio Fallacy

To understand what happens when we estimate the first difference model, the estimating equation (2) needs to be written out in full. In a first step we obtain:

$$\begin{aligned}\ln HomR_{k,t} &= \beta_1 \ln FSS_{k,t-1} + \beta_{2,1} \ln BurgR_{k,t} + \beta_{2,2} \ln RobR_{k,t} \\ &+ \beta_{2,3} \ln BlackR_{k,t} + \beta_{2,4} \ln UrbR_{k,t} + \beta_{2,5} \ln Resid5R_{k,t} \\ &+ \beta_{2,6} \ln FHHR_{k,t} + d_k + d_t + \varepsilon_{k,t}\end{aligned}\quad (4)$$

This equation still contains ratios, so it has to be written on the counts, yielding

$$\begin{aligned}\ln E96_{k,t} - \ln \mathbf{pop}_{k,t} &= \beta_1 (\ln E955_{k,t} - \ln \mathbf{E95}_{k,t}) \\ &+ \beta_{2,1} (\ln burglaries_{k,t} - \ln \mathbf{pop}_{k,t}) + \beta_{2,2} (\ln robberies_{k,t} - \ln \mathbf{pop}_{k,t}) \\ &+ \beta_{2,3} (\ln blacks_{k,t} - \ln \mathbf{pop}_{k,t}) + \beta_{2,4} (\ln urbans_{k,t} - \ln \mathbf{pop}_{k,t}) \\ &+ \beta_{2,5} (\ln 5yearResidents_{k,t} - \ln \mathbf{pop5plus}_{k,t}) \\ &+ \beta_{2,6} (\ln fhh_{k,t} - \ln \mathbf{households}_{k,t}) + d_k + d_t + \varepsilon_{k,t}.\end{aligned}\quad (5)$$

One of the left-hand summands  $-\ln \mathbf{pop}_{k,t}$  repeats itself multiple times on the right-hand side. Basically, this model explains “population plus homicides” on the left-hand side by six different “population + something” terms on the right-hand side. Population is a perfect correlate of itself,<sup>28</sup> so as long as the added

<sup>27</sup>For when this is beneficial, see Wooldridge (2002: Section 10.7).

<sup>28</sup>It is nearly perfectly correlated with the population of those five years and older and the number of households;  $r > 0.99$  for those variables.

values do not exhibit too much orthogonal variation to population itself, it *will* be able to explain itself. This is a variant of the ratio fallacy, first discovered by Pearson (1896) and discussed in detail by Kronmal (1993),<sup>29</sup> which here appears disguised in a logarithmic model. This fallacy can be seen more clearly in the logarithms of ratios, as now the variable responsible for the spurious results is linear in the terms on both sides. The *FSS* denominator is the number of all suicides (*E95*), i.e., it is not population. Why this does not affect the argument is shown in Section 5.3.

## 5.2 First Differences

I now demonstrate what happens when we compute the first differences of any of the ratios' logarithms using the example of the left-hand side of Equation (6) to understand why significance vanishes here.

$$\begin{aligned}
 \Delta Y_{k,t} &= \Delta(\ln E96_{k,t} - \ln pop_{k,t}) \\
 &= (\ln E96_{k,t} - \ln pop_{k,t}) - (\ln E96_{k,t-1} - \ln pop_{k,t-1}) \\
 &= \ln E96_{k,t} - \ln pop_{k,t} - \ln E96_{k,t-1} + \ln pop_{k,t-1} \\
 &= \ln \frac{E96_{k,t}}{E96_{k,t-1}} - \underbrace{\ln \frac{pop_{k,t}}{pop_{k,t-1}}}_{=\Delta \ln pop_{k,t-1} \approx 0}
 \end{aligned} \tag{6}$$

Population changes relatively the slowest over time compared to the other values. Therefore, the second fraction is always very close to 1, meaning that the logarithm is always very close to 0. This is shown in Table 6 when comparing (i) the values on the diagonal and (ii) their (rescaled) squared deviations from zero (see the bottom row): the diagonal tells us  $\Delta \ln pop$  has much less variation than any other value and the bottom shows us that the values on average are much closer to zero than all other values. Around sixty percent of the variance in the value based on population  $\Delta \ln pop_{k,t}$  is variance between counties.<sup>30</sup> Close to 100% of all variance in all other values in Table 6 is variance within counties (over time).<sup>31</sup> Together with Table 6 this shows that all other values vary much more strongly over time than does the value based on population. Relative to the other values, the logarithm of the growth rate of the population can be considered constant, as it is depicted in Equation (6). Also, for the right-hand side term

<sup>29</sup>Further ample warning about these specifications is given in the methodological literature: Kuh and Meyer (1955); Madansky (1964); Belsley (1972); Casson (1973).

<sup>30</sup>Computed by analysis of variance decomposition of variance: *within*-county variance is variance over time, *between*-county variance is variance between counties.

<sup>31</sup>Actually, an analysis of variance decomposition shows negative between variance for *E95*, *E955*, and *E96*, which is rare but numerically possible and evidence for very low between variance.

of interest  $\beta_1(\Delta \ln E955_{k,t-1} - \Delta \ln E95_{k,t-1})$ , the term from the numerator is double the mean squared distance from zero and double the variance than the term from the denominator. This means that in *this specific data set*, taking the first differences *at least partially* removes the numbers causing spurious correlations between ratios. This does *not* mean, however, that taking first differences will solve this problem any time in any data set. Here, it basically removes population

	$\Delta \ln pop$	$\Delta \ln E96$	$\Delta \ln E95_{t-1}$	$\Delta \ln E955_{t-1}$
$\Delta \ln pop$	1.00	0.92	0.69	0.47
$\Delta \ln E96$	0.92	450.92	-0.22	2.81
$\Delta \ln E95_{t-1}$	0.69	-0.22	172.24	171.21
$\Delta \ln E955_{t-1}$	0.47	2.81	171.21	356.52
rescaled mean				
sum of squares	1.00	289.19	111.32	230.42

Table 6: Covariance matrix rescaled by  $s_{\Delta \ln pop}^2 = 0.0002409870$ , mean sums of squares rescaled by  $K^{-1}T^{-1} \sum_k \sum_t (\Delta \ln pop_{k,t})^2 = 0.0003727771$

from all the terms and only the (growth rates) of the numerators remain in the model after taking first differences. Once population is removed from both sides, the right-hand side is no longer able to explain the left hand-side.

### 5.3 But E95 is Not Population

One could now argue  $E95_{k,t-1}$  is not population and therefore the results from Cook and Ludwig (2006) *are not* due to the ratio fallacy. When we look at the correlation matrix (Table 7) we immediately see that the correlation between suicides and population is far superior to any other correlation between the left-hand side and the right-hand side, at least in regard to the four variables shown in Table 7. Auxiliary panel regression results in Table 8 support the claim that it

	$\ln pop$	$\ln E96$	$\ln E95_{t-1}$	$\ln E955_{t-1}$
$\ln pop$	1.000	0.676	0.868	0.685
$\ln E96$	0.676	1.000	0.731	0.705
$\ln E95_{t-1}$	0.868	0.731	1.000	0.902
$\ln E955_{t-1}$	0.685	0.705	0.902	1.000

Table 7: Correlation matrix of population and different deaths

is E95 driving the results of the coefficient on the FSS proxy.

Dependent variable	Exogenous variable				$R^2_{within}$
	$\ln E95_{k,t-1}$		$\ln E955_{k,t-1}$		
$\ln pop_{k,t}$	0.24***	(0.047)	-0.05***	(0.010)	0.15
$\ln E96_{k,t}$	0.20***	(0.056)	0.03	(0.029)	0.02

Table 8: Auxiliary two-way fixed effects panel regressions illustrating the explanatory power of the *FSS* denominator in the model, \* :  $p < 0.05$ , \*\* :  $p < 0.01$ , \*\*\* :  $p < 0.001$ , robust standard errors according to Driscoll and Kraay (1998) computed with *vcovSCC* from Croissant and Millo (2008) in parentheses,  $R^2$  adjusted

## 5.4 Nonsense Regression Between Time Series

Regression between time series is known to produce spurious results in the following settings: trending or auto correlated time series (Granger and Newbold 1974),  $I(1)$  processes without drift (Phillips 1986),  $I(1)$  processes with further stationary regressors (Hassler 1996), stationary AR processes (Granger, Hyung and Jeon 2001), random walks with and without drift for fixed effects panel models (Entorf 1997), time-varying means (Hassler 2003), and stationary processes around linear trends (Kim, Lee and Newbold 2004), as well as in fixed effects (or first differences) estimations with weak variation in the time series (Choi 2011). It seems unlikely that none of these situations occurred in the original analysis, and thus there may be more sources for spurious results than just the ratio problem. As noted by C&L themselves (p. 383), there are heterogeneous trends for the dependent variable between counties. This is illustrated in Figure 1. Clearly, a single time dummy is incapable of detrending heterogeneous trends across counties. Therefore, not all trends will be accounted for in C&L's original model. Many of those time-series-related problems are automatically dealt with when taking first differences, while the single time dummy from model (2) is not able to detrend heterogeneous county trends.

## 5.5 Misspecification of the Original Model

### 5.5.1 Testing for Misspecification

We can look at the problem in C&L's model from the perspective of linear model theory. When we write out Equation (2) and rearrange the right-hand side of

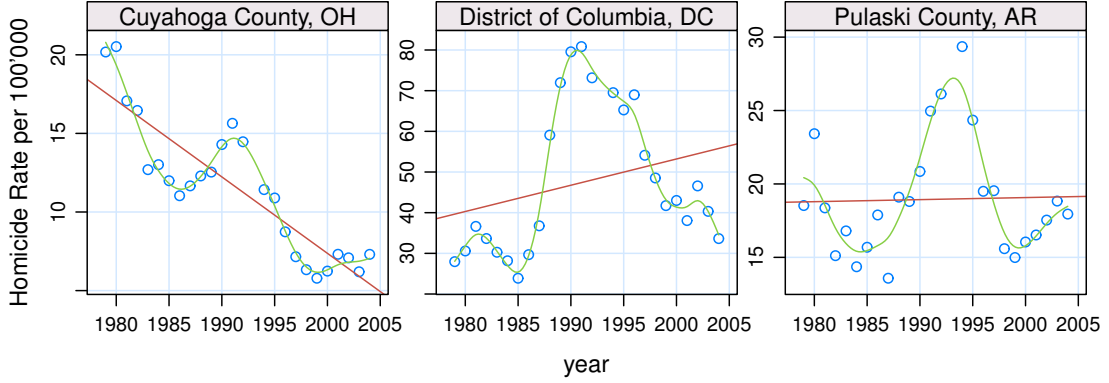


Figure 1: Illustration of heterogeneous time trends of the homicide rate between counties.

Equation (7), we obtain

$$\begin{aligned}
 \ln HomR_{k,t} &= \beta_1 (\ln E955_{k,t-1} - \ln E95_{k,t-1}) \\
 &+ \beta_{2,1} (\ln burglaries_{k,t} - \ln pop_{k,t}) \dots \\
 &= \beta_{E955} \ln E955_{k,t-1} + \beta_{E95} \ln E95_{k,t-1} + \beta_{2,1} \ln burglaries_{k,t} + \dots \\
 &+ \underbrace{(-\beta_{2,1} - \beta_{2,2} - \beta_{2,3} - \beta_{2,4} - \beta_{2,5} - \beta_{2,6})}_{=\beta_{pop}} \ln pop_{k,t} + \varepsilon_{k,t}.
 \end{aligned} \tag{7}$$

We see a linear restriction of

$$\begin{bmatrix} \beta_{E955} + \beta_{E95} \\ \beta_{2,1} + \beta_{2,2} + \beta_{2,3} + \beta_{2,4} + \beta_{2,5} + \beta_{2,6} + \beta_{pop} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{8}$$

which is rejected with a p-value of  $5.372 \times 10^{-14}$ . The number of households and the population of those five years and older are substituted for by population, as these are nearly perfect correlates.<sup>32</sup> Therefore, the original model seems to be misspecified. When estimation is performed on the first differences, the same restriction as in Equation (8) has to hold. In this case, the null hypothesis is not rejected (p-value of 0.90). This is further evidence that the differentiated model has *in this case* taken the ratio problem out of the data. However, this does not have to be the case for any dataset.

<sup>32</sup>Using the original denominators and testing all four linear hypotheses leads to an even more significant rejection of the null hypothesis.

### 5.5.2 Theoretical Bias

For a simple univariate linear model<sup>33</sup> on logarithms of ratios with a common denominator

$$\ln y_j - \ln z_j = b_0 + b_1(\ln x_j - \ln z_j) + \varepsilon_j \quad (9)$$

$$\Leftrightarrow \ln y_j - \ln z_j = b_0 + b_x \ln x_j + b_z \ln z_j + \varepsilon_j \text{ with } b_x = -b_z \quad (10)$$

with  $j = 1, 2, \dots, J$  and a linear restriction of  $b_x + b_z = 0$  in Equation (10), the bias of the estimator of  $b_x = b_1$  can be computed. The bias is known (see Judge et al. 1985: 53, Eq. (3.2.6)) to be

$$(G'G)^{-1}R'[R(G'G)^{-1}R']^{-1}(r - Rb) \quad (11)$$

where  $G = [1_J \ \ln X \ \ln Z]$  is the usual matrix of independent variables in the least squares model and  $1_J$ ,  $\ln X$  and  $\ln Z$  are column vectors of 1s, the  $\ln x_j$ , and the  $\ln z_j$ .  $R$  and  $r$  describe the linear restriction

$$\underbrace{[0 \ 1 \ 1]}_{=R} \times \underbrace{\begin{bmatrix} b_0 \\ b_x \\ b_z \end{bmatrix}}_{=b} = \underbrace{[0]}_{=r}. \quad (12)$$

Using

$$(G'G)^{-1} = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{1,2} & c_{2,2} & c_{2,3} \\ c_{1,3} & c_{2,3} & c_{3,3} \end{bmatrix} \quad (13)$$

the bias for  $\hat{b}_x$  given  $b_x = 0$  can then be calculated:

$$\text{bias}(\hat{b}_x | b_x = 0) = -\frac{c_{2,2} + c_{2,3}}{c_{2,2} + 2c_{2,3} + c_{3,3}} b_z. \quad (14)$$

Considering all  $c_{v,w}$  have the same denominator  $\det(G'G)$ , the fraction in Equation (14) becomes

$$\frac{J \sum z^{*2} - (\sum z^*)^2 + J \sum x^* z^* - \sum x^* \sum z^*}{J \sum z^{*2} - (\sum z^*)^2 + 2(J \sum x^* z^* - \sum x^* \sum z^*) + J \sum x^{*2} - (\sum x^*)^2} \quad (15)$$

<sup>33</sup>I use different symbols here so as not to mislead the reader into thinking this is the same model as Equation (2). Furthermore the result can be directly applied to panel estimation. This can be seen from the appropriate transformations (e.g. Baltagi 2008: Sections 2.2, 3.2).

with  $z_j^* = \ln z_j$ ,  $x_j^* = \ln x_j$ , and where each sum is on all  $j$  and the index has been omitted for readability. This, expanded by  $J^{-2}$  collapses to

$$F = -\frac{s_{z^*}^2 + s_{x^*,z^*}}{s_{z^*}^2 + 2s_{x^*,z^*} + s_{x^*}^2} \quad (16)$$

giving the overall expected bias of  $b_x = b_1$  as  $F b_z$  solely depending on the variances and covariance of  $\ln X$ ,  $\ln Z$  and the true  $b_z$ . This result is well in accordance with an upward biased estimate of  $\beta_1$  in C&L's original model. Furthermore the implications are quite strong: If neither  $X$  nor  $Z$  contribute to the outcome  $Y$  and all variables are uncorrelated, then  $b_z$  takes a value of  $-1$  and from estimating the restricted Model (9) we expect a positive estimate for  $b_1$ .

## 6 Alternative specifications

### 6.1 Controlling for Population

The first method for removing the spuriousness from C&L's model is given by Krommal (1993: 390): include the inverse of the deflating variable as an explanatory variable on the right-hand side. In a logarithmic model, this means we can just add the logarithm of the deflating variable.<sup>34</sup> The model then becomes<sup>35</sup>

$$\ln Y_{k,t} = \beta_0 \ln pop_{k,t} + \beta_1 \ln FSS_{k,t-1} + X_{k,t} \beta_2 + d_k + d_t + \varepsilon_{k,t}. \quad (17)$$

The estimation result from this model is set out in the column labeled "Eq. (17)" in Table 9. Across the board, significance weakens considerably and completely disappears from the ratios computed from the best correlates of population in the numerator. This specification takes out the linear restriction on  $\ln pop$  known from Equations (7) and (8). The linear restriction in (8) is now less restrictive:

$$\beta_{\ln E955_{k,t-1}} + \beta_{\ln E95_{k,t-1}} = 0. \quad (18)$$

The p-value for this null hypothesis is 0.002195. This still does not account for possibly spurious results due to time-series effects. It is likely that not all the left-hand side time series in the panel can be detrended by a single time dummy. Sections 6.2, 6.3, 6.4, and 6.5 discuss further model specifications.

<sup>34</sup>I know of only one other study on this topic that explicitly addresses this technique: Kleck and Patterson (1993).

<sup>35</sup>I use  $\varepsilon_{k,t}$  for the error term multiple times in this section; however, I do not assume it to be identically distributed for all models. Also I reuse  $\beta$  for coefficients with different interpretations. These are not identical across models.

Underlying explanatories	Model			
	Eq. (17) controlling	Eq. (19) rearranged	Eq. (24) risk	Eq. (26) growth
population	-0.48 (0.000)	1.00 (restr.)	-0.000 (0.592)	1.079 (0.198)
<i>suicide measures</i>				
$FSS_{t-1}$	0.05 (0.082)	0.02 (0.516)	NA	NA
other suicides $_{t-1}$	NA	NA	0.120 (0.196)	-0.012 (0.784)
firearms suicides $_{t-1}$	NA	NA	0.017 (0.781)	-0.001 (0.979)
<i>control variables</i>				
robberies	0.13 (0.002)	0.16 (0.001)	0.002 (0.295)	0.028 (0.547)
burglaries	0.10 (0.005)	0.08 (0.076)	0.005 (0.000)	0.020 (0.677)
blacks	0.19 (0.000)	0.24 (0.000)	0.001 (0.011)	-0.067 (0.841)
urbans	-0.18 (0.314)	0.15 (0.477)	0.000 (0.238)	-0.210 (0.854)
same house 5 years ago	-0.11 (0.645)	0.29 (0.231)	-0.000 (0.287)	-0.349 (0.603)
female headed house	0.33 (0.089)	-0.02 (0.928)	-0.002 (0.220)	0.410 (0.371)
$R^2_{within}$	0.075	0.058	0.200	0.002

Table 9: Robustness of the null result to different model specifications eliminating the ratio fallacy. Cells give coefficient estimates and corresponding p-values in parentheses. Robust standard errors according to Driscoll and Kraay (1998) were computed with `vcovSCC` from Croissant and Millo (2008) for significance levels,  $R^2$  adjusted



## 6.2 Algebraic Transformation of the Estimating Equation

Similar to the approach in Section 6.1 one may transform Equation (5) by adding  $\ln pop_{k,t}$  on both sides. This results in

$$\ln E96_{k,t} = \ln pop_{k,t} + \beta_1 \ln FSS_{k,t-1} + X_{k,t} \beta_2 + d_k + d_t + \varepsilon_{k,t}. \quad (19)$$

$\ln pop_{k,t}$  now has a fixed coefficient of  $\beta_{pop} = 1$ . This technique is known from the Poisson regression model for the analysis of rates.<sup>36</sup> Any spurious correlation between the left-hand side and the right-hand side due to population appearing on both sides is no longer possible. Results are reported in the column labeled “Eq. (19)” in Table 9. There is no significance on  $FSS_{t-1}$ . Testing solely for the restriction  $\beta_{pop} = 1$  rejects the null hypotheses.<sup>37</sup> Therefore the model appears to be misspecified. Time series problems are not accounted for.

## 6.3 Risk Model

Duggan (2003: 48–50) proposes a model<sup>38</sup> for explaining individual  $i$ 's suicide decision<sup>39</sup>

$$\Pr(\text{Suicide}_i) = \alpha + X_i \theta + \gamma \text{Gun}_i + \lambda_i + \varepsilon_i \quad (20)$$

with  $X_i$  being individual observable controls,  $\text{Gun}_i$  a dummy for gun ownership, and  $\lambda_i$  individual  $i$ 's unobserved individual propensity to commit suicide. Assume that a gun owner chooses suicide by firearm with a probability  $> 0$ . Then, as long as  $\lambda_i$  is not negatively correlated with gun ownership and as long as  $\gamma \geq 0$ , gun ownership will be associated with a higher probability of committing suicide at all or just with a higher probability of committing suicide by firearm.<sup>40</sup> For the limiting case of zero correlation between  $\text{Gun}_i$  and  $\lambda_i$  the relative risk of a gun owner becomes:<sup>41</sup>

$$RR_{\text{Gun}} = \frac{\alpha + X_i \theta + \gamma \text{Gun}_i + \varepsilon_i}{\alpha + X_i \theta + \varepsilon_i} \geq 1 \quad (21)$$

and the expected number of (firearm) suicides for a population of size  $pop$  will be

$$\mathcal{E}[E95] = pop \cdot (\alpha + \bar{X} \theta + \gamma \cdot \Pr(\text{Gun})), \quad (22)$$

$$\mathcal{E}[E955] = pop \cdot \Pr(\text{GunSuic}|\text{Suicide}, \text{Gun}) \cdot \Pr(\text{Suicide}|\text{Gun}) \cdot \Pr(\text{Gun}) \quad (23)$$

<sup>36</sup>See Osgood (2000: 23–27).

<sup>37</sup>p-value of  $8.22 \times 10^{-12}$ .

<sup>38</sup>I slightly deviate from Duggan's notation without changing the model to keep my formulas simpler.

<sup>39</sup>This is a linear probability model. The parameters must satisfy the requirement of  $0 \leq \Pr(\text{Suicide}_i) \leq 1 \forall i$ . The following argument holds for other monotonous link functions as well.

<sup>40</sup> $\partial \Pr(\text{GunSuic}_i|\text{Suicide}_i) / \partial \text{Gun}_i > 0$ .

<sup>41</sup>There is no evidence in the literature contradicting these assumptions.

under the simplifying assumption that only gun owners are able to commit suicide by firearm. When the  $X_i$  are dummies,  $pop \cdot \bar{X}$  will become count data for those dummies (that is, numbers of people with certain characteristics). The relative risk notion in Equation (21) gives a very clear interpretation to the coefficients in this model.

I go in the opposite direction, and start at the macro level by proposing a “risk model” (coefficients do not match the identically named coefficients in equations (20), (21) and (22)) for *homicides* that can be estimated solely on differences in the counts, thus removing the potential for spuriousness due to time series:

$$\Delta E96_{k,t} = \beta_0 \Delta pop_{k,t} + \beta_{1,1} \Delta E955_{k,t-1} + \beta_{1,2} \Delta E95_{k,t-1} + \Delta X_{k,t} \beta_2 + \delta_t + \varepsilon_{k,t}, \quad (24)$$

where  $X_{k,t}$  now contains the numerators’ values of the control ratios used by C&L.  $E955_{k,t-1}$  now is the gun proxy – which, given Duggan’s (2003) model, should be positively correlated to the number of gun owners – and the number of non-firearm suicides is used as an additional control. Then  $\beta_{1,1}$  should be positive if crime increases with more guns. Let us say  $\beta_0$  is one person’s baseline risk of becoming a victim/committing a homicide. Now attribute an additional risk to each gun owner,<sup>42</sup> then a relation of

$$\frac{\beta_0 + \beta_{1,1}}{\beta_0} \sim RR_{gunowner}, \quad (25)$$

exists, given the number of firearm suicides is somehow linked to the number of gun owners. Results are reported in the column labeled “Eq. (24)” of Table 9. There is no significance on the variable of interest ( $E955$ ). Significance on the other variables must not be over interpreted. For example, for burglaries, it might just mean there are around 170 times as many burglaries as homicides. This model is susceptible to criticism for obvious heteroscedasticity across counties with different levels of population. Standardization would be helpful. Also multicollinearity might be an issue, as all numbers used are part of the population and therefore are in  $pop_{k,t}$ .

## 6.4 Growth Model

A way to standardize without using ratios is to use growth rates. Putting the (logarithm of) the growth rate of homicides on the left-hand side yields the following model:

$$\ln \frac{E96_{k,t}}{E96_{k,t-1}} = \beta_0 \ln \frac{pop_{k,t}}{pop_{k,t-1}} + \beta_1 \ln \frac{E955_{k,t-1}}{E955_{k,t-2}} + X_{k,t} \beta_2 + \varepsilon_{k,t}, \quad (26)$$

<sup>42</sup>Either as a victim or as the perpetrator or by imposing an externality upon the remaining population.

where  $X_{k,t}$  contains the log growth rates for the controls. As in the risk model, non-firearm suicides can be added as a control. Results are reported in the column labeled “Eq. (26)” in Table 9; no significance is observed.<sup>43</sup>

## 6.5 Numerators and Denominators

To check where the explanatory power in C&L’s model comes from, a comparison of the following models seems appropriate:

$$\ln pop_{k,t} = \beta_1 \ln E95_{k,t-1} + d_k + d_t + \varepsilon_{k,t} \quad (27)$$

$$\begin{aligned} \ln pop_{k,t} = & \beta_1 \ln E95_{k,t-1} + \beta_{2,1} \ln burglaries_{k,t} + \beta_{2,2} \ln robberies_{k,t} \\ & + \beta_{2,3} \ln blacks_{k,t} + \beta_{2,4} \ln urbans_{k,t} + \beta_{2,5} \ln 5yearResidents_{k,t} \\ & + \beta_{2,6} \ln fh_{k,t} + d_k + d_t + \varepsilon_{k,t}. \end{aligned} \quad (28)$$

$$\ln E96_{k,t} = \beta_1 \ln E95_{k,t-1} + \beta_3 \ln pop_{k,t} + d_k + d_t + \varepsilon_{k,t} \quad (29)$$

$$\begin{aligned} \ln E96_{k,t} = & \beta_1 \ln E95_{k,t-1} + \beta_{2,1} \ln burglaries_{k,t} + \beta_{2,2} \ln robberies_{k,t} \\ & + \beta_{2,3} \ln blacks_{k,t} + \beta_{2,4} \ln urbans_{k,t} + \beta_{2,5} \ln 5yearResidents_{k,t} \\ & + \beta_{2,6} \ln fh_{k,t} + d_k + d_t + \varepsilon_{k,t}. \end{aligned} \quad (30)$$

These models allow cross-checking whether the right-hand side numerators actually explain the left-hand side numerator as intended, or whether some other mechanism is driving the results. In Equation (27) there is only one right-hand side term. Due to nearly perfect correlation with  $pop_{k,t}^{-1}$ , I removed  $pop5plus_{k,t}^{-1}$  and  $households_{k,t}^{-1}$  from the model. Including  $pop_{k,t}^{-1}$  on the right-hand side is obviously ridiculous. The results are given in Table 10. The comparison of within  $R^2$ s

		Model			
		lhs: <i>pop</i>		lhs: <i>E96</i>	
		Equation (27)	Equation (28)	Equation (29)	Equation (30)
$R^2_{within}$		0.1492	0.9542	0.0322	0.0818
		Modifications of equation (28)			
		lhs: <i>pop</i>		lhs: <i>E96/pop</i>	
		only numerators	ratios	only numerators	ratios
$R^2_{within}$	rhs	0.9542	0.2790	0.0810	0.0635

Table 10: Diagnostics for finding the source of “explanatory power” in the original model,  $R^2$  adjusted

clearly shows that Equations (27) and (28) each display a larger coefficient of determination than either Equation (29) or Equation (30). The very high within  $R^2$

<sup>43</sup>For state-level data, this model yields a negative coefficient on the order of 0.3 on the gun proxy, significant at the 5% level. This is the only setting showing this result.

of 0.9542 for Equation (28) reduces to 0.0810 when the left-hand side is changed to  $E96/pop$ , to 0.2790 when the right-hand side is changed to ratios, and, finally, to 0.06 when lhs and rhs are changed to the full model from Equation (2) (Table 10, second row). Thus, from a goodness-of-fit point of view, the only thing C&L's full model does, is add a lot of noise to Equation (28). This adds additional support to the already strong theoretical and quantitative argument in Section 5 that the results of the original analysis were driven by the ratio fallacy. The coefficients of these estimations have no useful interpretation; this was purely an illustrative exercise to show what is driving the results in the original paper.

## 7 Conclusion

Other aspects of C&L's analysis could be addressed. Namely, (i) their data suffer from truncation for observations with zeroes in the numerators<sup>44</sup> and (ii) the FSS proxy is very noisy for smaller counties: Imagine a county in year  $t - 1$  having 20 suicides, one with a gun, and in  $t$  20 suicides, eight with guns.<sup>45</sup> Surely gun ownership did not increase proportional to this increase in FSS. This problem could be addressed by using moving averages.<sup>46</sup> However, as their results are purely an effect of a technical property of the data, further criticism of the study seems unwarranted. Which model should they have chosen? This remains an open question but given that none yields significance or anything close to it for the parameter of interest, finding the "correct" model becomes somewhat of a moot point. Of much more relevance, especially when it comes to the sensitive topic of gun control, is to discover how many studies on this topic have ignored the ratio problem?

## References

- Ayres, Ian, and John J Donohue.** 2003a. "The Latest Misfires in Support of the 'More Guns, Less Crime' Hypothesis." *Stanford Law Review*, 55: 1371.
- Ayres, Ian, and John J Donohue.** 2003b. "Shooting Down the "More Guns, Less Crime" Hypothesis." *Stanford Law Review*, 1193–1312.
- Azrael, Deborah, Philip J Cook, and Matthew Miller.** 2004. "State and Local Prevalence of Firearms Ownership Measurement, Structure, and Trends." *Journal of Quantitative Criminology*, 20(1): 43–62.

<sup>44</sup>This can be a severe problem for consistency as discussed in Hayashi (2000: Sections 5.3, 8.2).

<sup>45</sup>This is the case for Union County, NJ and  $t = 1998$ .

<sup>46</sup>I did, in fact, try this and all my findings remain robust to using moving averages; results, data, and code available upon request.

- Baltagi, Badi.** 2008. *Econometric Analysis of Panel Data*. Wiley.
- Belsley, David A.** 1972. "Specification with Deflated Variables and Specious Spurious Correlation." *Econometrica*, 923–927.
- Casson, MC.** 1973. "Linear regression with error in the deflating variable." *Econometrica*, 751–759.
- Choi, In.** 2011. "Spurious Fixed Effects Regression." *Oxford Bulletin of Economics and Statistics*.
- Cook, Philip J, and Jens Ludwig.** 2002. "The effects of Gun Prevalence on Burglary: Deterrence vs Inducement." *NBER Working Paper Series*, 8926.
- Cook, Philip J, and Jens Ludwig.** 2003. "Guns and Burglary." *Evaluating Gun Policy: Effects on Crime and Violence*, 74–106.
- Cook, Philip J, and Jens Ludwig.** 2006. "The social costs of gun ownership." *Journal of Public Economics*, 90: 379–391.
- Cook, Philip J, Jens Ludwig, Sudhir Venkatesh, and Anthony A Braga.** 2007. "Underground Gun Markets." *The Economic Journal*, 117(524): F558–F588.
- Cook, Phillip J, and Jens Ludwig.** 2004. "The social costs of gun ownership." *NBER Working Paper Series*, 10736.
- Croissant, Yves, and Giovanni Millo.** 2008. "Panel Data Econometrics in R: The plm Package." *Journal of Statistical Software*, 27(2).
- Donohue, John J.** 2003. "The Final Bullet in the Body of the More Guns, Less Crime Hypothesis." *Criminology & Public Policy*, 2(3): 397–410.
- Driscoll, John C, and Aart C Kraay.** 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics*, 80(4): 549–560.
- Duggan, Mark.** 2001. "More Guns, More Crime." *The Journal of Political Economy*, 109(5): 1086–1114.
- Duggan, Mark.** 2003. "Guns and Suicide." *Evaluating Gun Policy: Effects on Crime and Violence*, 41–73.
- Entorf, Horst.** 1997. "Random walks with drifts: Nonsense regression and spurious fixed-effect estimation." *Journal of Econometrics*, 80(2): 287–296.

- Fox, James Alan, and David McDowall.** 2008. "Brief of Professors of Criminal Justice As Amici Curiae in Support of Petitioners." *Submitted to the Supreme Court of the United States in District of Columbia and Adrian M. Fenty, Mayor of the District of Columbia v. Dick Anthony Heller*, No. 07-290, Jan. 11, 2008.
- Granger, Clive WJ, and P Newbold.** 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics*, 2(2): 111–120.
- Granger, Clive WJ, Namwon Hyung, and Yongil Jeon.** 2001. "Spurious regressions with stationary series." *Applied Economics*, 33(7): 899–904.
- Grothendieck, G.** 2011. "sqldf: Perform SQL Selects on R Data Frames." R package version 0.4-6.1.
- Hassler, Uwe.** 1996. "Spurious regressions when stationary regressors are included." *Economics Letters*, 50(1): 25–31.
- Hassler, Uwe.** 2003. "Nonsense regressions due to neglected time-varying means." *Statistical Papers*, 44(2): 169–182.
- Hayashi, Fumio.** 2000. *Econometrics*. Princeton University Press.
- Judge, George G, WE Griffiths, R Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee.** 1985. *The Theory and Practice of Econometrics*. New York:Wiley.
- Kellermann, Arthur L, Frederick P Rivara, Grant Somes, Donald T Reay, Jerry Francisco, Joyce Gillentine Banton, Janice Prodzinski, Corinne Fligner, and Bela B Hackman.** 1992. "Suicide in the home in relation to gun ownership." *New England Journal of Medicine*, 327(7): 467–472.
- Kim, Tae-Hwan, Young-Sook Lee, and Paul Newbold.** 2004. "Spurious regressions with stationary processes around linear trends." *Economics Letters*, 83(2): 257–262.
- Kleck, Gary.** 2004. "Measures of Gun Ownership Levels for Macro-level Crime and Violence Research." *Journal of Research in Crime and Delinquency*, 41(1): 3–36.
- Kleck, Gary.** 2009. "How Not to Study the Effect of Gun Levels on Violence Rates." *Journal on Firearms & Public Policy*, 21: 65.
- Kleck, Gary, and E Britt Patterson.** 1993. "The Impact of Gun Control and Gun Ownership Levels on Violence Rates." *Journal of Quantitative Criminology*, 9(3): 249–287.

- Klieve, Helen, Jerneja Sveticic, and Diego De Leo.** 2009. "Who uses firearms as a means of suicide? A population study exploring firearm accessibility and method choice." *BMC Medicine*, 7(1): 52.
- Klieve, Helen, Michael Barnes, and Diego De Leo.** 2009. "Controlling firearms use in Australia: has the 1996 gun law reform produced the decrease in rates of suicide with this method?" *Social Psychiatry and Psychiatric Epidemiology*, 44(4): 285–292.
- Kronmal, Richard A.** 1993. "Spurious Correlation and the Fallacy of the Ratio Standard Revisited." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 156(3): 379–392.
- Kuh, Edwin, and John R Meyer.** 1955. "Correlation and Regression Estimates when the Data are Ratios." *Econometrica*, 400–416.
- Leenaars, Antoon A, and David Lester.** 2001. "The impact of gun control (Bill C-51) on homicide in Canada." *Journal of Criminal Justice*, 29(4): 287–294.
- Leenaars, Antoon A, Ferenc Moksony, David Lester, and Susanne Wenckstern.** 2003. "The Impact of Gun Control (Bill C-51) on Suicide in Canada." *Death Studies*, 27(2): 103–124.
- Leigh, Andrew, and Christine Neill.** 2010. "Do Gun Buybacks Save Lives? Evidence from Panel Data." *American Law and Economics Review*, 12(2): 462–508.
- Lott, John R.** 2010. *More Guns, Less Crime: Understanding Crime and Gun Control Laws*. . 3rd ed., University of Chicago Press.
- Lott, John R, and David B Mustard.** 1997. "Crime, Deterrence, and Right-to-carry Concealed Handguns." *The Journal of Legal Studies*, 26(1): 1–68.
- Ludwig, Jens.** 1998. "Concealed-Gun-Carrying Laws and Violent Crime: Evidence from State Panel Data." *International Review of Law and Economics*, 18(3): 239–254.
- Madansky, Albert.** 1964. "Spurious Correlation Due to Deflating Variables." *Econometrica*, 652–655.
- Moody, Carlisle E, and Thomas B Marvell.** 2010. "On the Choice of Control Variables in the Crime Equation." *Oxford Bulletin of Economics and Statistics*, 72(5): 696–715.
- Osgood, D. Wayne.** 2000. "Poisson-Based Regression Analysis of Aggregate Crime Rates." *Journal of Quantitative Criminology*, 16(1): 21–43.

- Pearson, K.** 1896. "Mathematical Contributions to the Theory of Evolution. – On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs." *Proceedings of the Royal Society of London*, 60(359-367): 489–498.
- Phillips, Peter CB.** 1986. "Understanding spurious regressions in econometrics." *Journal of Econometrics*, 33(3): 311–340.
- R Development Core Team.** 2011. "R: A Language and Environment for Statistical Computing." Vienna, Austria, R Foundation for Statistical Computing, ISBN 3-900051-07-0.
- United States Census Bureau.** 1980. "1980 Census of Population and Housing, Summary Tape File 3A."
- United States Census Bureau.** 1990. "1990 Census of Population and Housing, Summary Tape File 3A."
- United States Census Bureau.** 2000. "2000 Census of Population and Housing."
- United States Department of Commerce. Bureau of the Census.** 1993. *Census of Population and Housing, 1990 [United States]: Summary Tape File 3C*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06054.v1>.
- United States Department of Health and Human Services.** 2010. "Compressed Mortality File (CMF) on CDC WONDER Online Database."
- United States Department of Health and Human Services. Centers for Disease Control and Prevention. National Center for Health Statistics.** 2010. *Mortality Detail Files, 1968-1991*. Inter-University Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR07632.v4>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 1997. *Mortality Detail File, 1992*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06798.v1>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2007a. *Multiple Cause of Death, 1998*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR03306.v2>.



- United States Department of Health and Human Services. National Center for Health Statistics.** 2007b. *Multiple Cause of Death, 1999*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR03473.v2>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2007c. *Multiple Cause of Death Public Use Files, 2000-2002*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR04640.v1>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2007d. *Multiple Cause of Death Public Use Files, 2003*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR20540.v1>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2007e. *Multiple Cause of Death Public Use Files, 2004*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR20623.v1>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2008a. *Multiple Cause of Death, 1993*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06799.v1>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2008b. *Multiple Cause of Death, 1994*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR02201.v2>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2008c. *Multiple Cause of Death, 1997*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR03085.v2>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2008d. *Multiple Cause of Death Public Use Files, 2005*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR22040.v1>.
- United States Department of Health and Human Services. National Center for Health Statistics.** 2009a. *Multiple Cause of Death, 1995*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR02392.v2>.

- United States Department of Health and Human Services. National Center for Health Statistics.** 2009b. *Multiple Cause of Death, 1996*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR02702.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2005. *Uniform Crime Reports: County Level Arrest and Offense Data, 1986*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR09119.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006a. *Uniform Crime Reporting Program Data [United States]: County Level Arrest and Offenses Data, 1977-1983*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR08703.v1>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006b. *Uniform Crime Reporting Program Data [United States]: County Level Arrest and Offenses Data, 1984*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR08714.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006c. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1989*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR09573.v1>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006d. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1990*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR09785.v1>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006e. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1991*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06036.v1>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006f. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1992*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06316.v1>.

- United States Department of Justice. Federal Bureau of Investigation.** 2006g. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1994.* Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06669.v3>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006h. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1995.* Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06850.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006i. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1996.* Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR02389.v3>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006j. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1997.* Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR02764.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006k. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1998.* Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR02910.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006l. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1999.* Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR03167.v4>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006m. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 2000.* Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR03451.v4>.

- United States Department of Justice. Federal Bureau of Investigation.** 2006n. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 2001*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR03721.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006o. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 2002*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR04009.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006p. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 2003*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR04360.v2>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006q. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 2004*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR04466.v1>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006r. *Uniform Crime Reports: County Level Detailed Arrest and Offense Data, 1985 and 1987*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR09252.v1>.
- United States Department of Justice. Federal Bureau of Investigation.** 2006s. *Uniform Crime Reports: County Level Detailed Arrest and Offense Data, 1988*. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR09335.v1>.
- United States Department of Justice. Federal Bureau of Investigation.** 2008. *Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1993*. Inter-University Consortium for Political and Social Research (ICPSR) [distributor]. <http://dx.doi.org/10.3886/ICPSR06545.v1>.
- Westphal, Christian.** 2013. "Personal Website." <http://www.westphal.de>.
- Wooldridge, Jeffrey M.** 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.