NPS-AS-91-005

NAVAL POSTGRADUATE SCHOOL
Monterey, California

LEARNING CURVE AND RATE ADJUSTMENT
MODELS: AN INVESTIGATION OF BIAS

O. Douglas Moses

February 1991

Prepared for: Naval Sea Systems Command
Cost Estimating and Analysis Division
Washington, DC 20362

NAVAL POSTGRADUATE SCHOOL
Monterey, California

RADM R. W. West, Jr.
Superintendent

Harrison Shull
Provost

This report was prepared in conjunction with research conducted for Naval Sea Systems Command Cost Estimating and Analysis Division and funded by the Naval Postgraduate School.

This report was prepared by:

O. Douglas Moses
Associate Professor
Department of Administrative Sciences

Reviewed by:

David R. Whipple, Chairman
Department of Administrative Sciences

Released by:

Paul J. Marto
Dean of Research

REPORT DOCUMENTATION PAGE

REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b RESTRICTIVE MARKINGS
SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release, Distribution Unlimited
DECLASSIFICATION/DOWNGRADING SCHEDULE	
PERFORMING ORGANIZATION REPORT NUMBER(S): NPS-AS-91-005	5 MONITORING ORGANIZATION REPORT NUMBER(S)
NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School; 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION: Naval Sea Systems Command
ADDRESS (City, State, and ZIP Code): Monterey, CA 93943	7b ADDRESS (City, State, and ZIP Code): Code 017, Washington, DC 20362
NAME OF FUNDING/SPONSORING ORGANIZATION: Naval Postgraduate School; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: O&MN, Direct Funding
ADDRESS (City, State, and ZIP Code): Monterey, CA 93943	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

TITLE (Include Security Classification)
Learning Curve and Rate Adjustment Models: An Investigation of Bias

PERSONAL AUTHOR(S)
O. Douglas Moses

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical Report	FROM ____ TO ____	1991 February	60

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Cost Estimation, Cost Analysis, Learning Curve, Production Rate, Prediction, Forecasting

ABSTRACT (Continue on reverse if necessary and identify by block number)

Learning curve models have gained widespread acceptance as a technique for analyzing and forecasting the cost of items produced from a repetitive process. Considerable research has investigated augmenting the traditional learning curve model with the addition of a production rate variable, creating a rate adjustment model. This study compares the forecasting bias of the learning curve and rate adjustment models. A simulation methodolog is used to vary conditions along seven dimensions. The magnitude and direction of errors i estimating future cost are analyzed and compared under the various simulated conditions, using ANOVA. Overall results indicate that the rate adjustment model is generally unbiased If the cost item being forecast contains any element that is not subject to learning then the traditional learning curve model is consistently biased toward underestimation of future cost. Conditions when the bias is strongest are identified.

DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☑ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: Unclassified
NAME OF RESPONSIBLE INDIVIDUAL: O. Douglas Moses	22b TELEPHONE (Include Area Code): 408-646-3218; 22c OFFICE SYMBOL: AS/Mo

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE: Unclassified

LEARNING CURVE AND RATE ADJUSTMENT MODELS:

AN INVESTIGATION OF BIAS

O. Douglas Moses

Associate Professor
Department of Administrative Science
Naval Postgraduate School
Monterey, CA 93943
408-646-3218

February 1991

This work was prepared for the Cost Estimating and Analysis Division of the Naval Sea Systems Command. Funding was provided by the Naval Postgraduate School.

PREFACE

This report is a companion to an earlier report ("Learning Curves and Rate Adjustment Models: Comparative Prediction Accuracy under Varying Conditions," Naval Postgraduate School Technical Report No. NPS-AS-91-001). Both reports investigate and evaluate two cost estimating approaches commonly used by cost analysts. Both use the same methodology. The earlier report focused on investigating the accuracy of the two approaches; the current report focuses on bias. Readers familiar with the earlier report will find the first 19 pages of this report, describing the methodology, to be quite familiar. For readers unfamiliar with the earlier report, this current report is designed to be self-contained.

LEARNING CURVE AND RATE ADJUSTMENT MODELS:

AN INVESTIGATION OF BIAS

ABSTRACT

Learning curve models have gained widespread acceptance as a technique for analyzing and forecasting the cost of items produced from a repetitive process. Considerable research has investigated augmenting the traditional learning curve model with the addition of a production rate variable, creating a rate adjustment model. This study compares the forecasting bias of the learning curve and rate adjustment models. A simulation methodology is used to vary conditions along seven dimensions. The magnitude and direction of errors in estimating future cost are analyzed and compared under the various simulated conditions, using ANOVA. Overall results indicate that the rate adjustment model is generally unbiased. If the cost item being forecast contains any element that is not subject to learning then the traditional learning curve model is consistently biased toward underestimation of future cost. Conditions when the bias is strongest are identified.

LEARNING CURVE AND RATE ADJUSTMENT MODELS:

AN INVESTIGATION OF BIAS

INTRODUCTION

The problem of cost overruns has consistently plagued the process of acquiring weapons systems by the U. S. Department of Defense. Technical improvements in the conduct of cost estimation and institutional changes in the process of procurement have occurred over the past few decades, but unanticipated cost growth during procurement continues. A cost overrun, by definition, occurs when the actual cost of a program exceeds the estimated cost. There are, in principle, two broad reasons that a cost overrun could occur. Either a) initial cost estimates are fair when made, but subsequently actual costs are poorly managed and controlled; or b) actual costs are well managed, but initial cost estimates were unrealistic. This paper focuses on the latter situation. The paper examines and compares bias in two estimating models used frequently by cost analysts: the learning curve and the rate adjustment model.

Learning curves have gained widespread acceptance as a tool for planning, analyzing, explaining, and predicting the behavior of the unit cost of items produced from a repetitive production process. (See Yelle, 1979, for a review.) Cost estimation techniques for planning the cost of acquiring weapon systems by the Department of Defense, for example, typically consider the role of learning in the estimation process. The premise of learning curve analysis is that cumulative quantity is the primary driver of unit cost. Unit cost is expected to decline as cumulative quantity increases.

There is general acknowledgement that cumulative quantity is not the only factor that influences unit cost and that the simple learning curve is not a fully adequate description of cost behavior. Hence prior research has attempted to augment learning curve models by including additional variables (e.g., Moses, 1990a). Most attention has been focused on the addition of a production rate term.[1] The resulting augmented model is usually referred to as a rate adjustment model.

Conceptually, production rate should be expected to affect unit cost because of the impact of economies of scale. Higher production rates may lead to several related effects: greater specialization of labor, quantity discounts and efficiencies associated with raw materials purchases, and greater use of facilities, permitting fixed overhead costs to be spread over a larger output quantity. Together, these effects work to increase efficiency and reduce production cost (Bemis, 1981; Boger and Liao, 1990; Large, et. al., 1974; Linder and Wilbourn, 1973). However, higher production rate does not guarantee lower cost. When production rate exceeds capacity, such factors as over-time pay, lack of skilled labor, or the need to bring more facilities online may lead to inefficiencies and increased unit cost. In short,

---

[1]One review of the literature pertaining to learning curves (Cheney, 1977) found that 36% of the articles reviewed attempted to augment the learning curve model in some manner by the inclusion of production related variables.

production rate may be associated with both economies and diseconomies of scale.

PRIOR RESEARCH

Numerous studies, using data on actual production cost elements, have been conducted to empirically examine the impact of production rate on unit cost. The broad objective of the research has been to document rate/cost relationships and determine if consideration of production rate leads to improvements in cost explanation or prediction. Results have been inconsistent and general findings inconclusive. Various studies (e.g., Alchian, 1963; Cochran, 1960; Hirsh, 1952; Large, Campbell and Cates, 1976) found little or no significance for rate variables. Other studies did document significant rate/cost relationships (e.g., Bemis, 1981; Cox and Gansler, 1981). Some research found significant results only for particular individual cost elements, such as labor (Smith, 1976), tooling (Levenson, et. al., 1971) or overhead (Large, Hoffmayer, and Kontrovich, 1974). But rate/cost relationships for these same cost elements were not consistently evident in other studies. When significant, estimates of the rate/cost slope varied greatly and the direction of the relationship was sometimes negative and sometimes positive (e.g., Moses, 1990a). In reviewing the existing research on production rate, Smith (1980) concluded that a rate/cost relationship may exist but that the existence, strength and nature of the relationship varies with the item produced and the cost element examined.[2]

The prior research suggests that consideration of production rate sometimes improves cost explanation, but not always. The prior research suggests that a traditional learning curve model sometimes is preferable to a rate adjustment model, but not always. The prior research provides little guidance concerning the circumstances under which explicit incorporation of production rate into a learning curve model is likely to lead to improved explanation or prediction. This issue is important in a number of cost analysis and cost estimation situations. Dorsett (1990), for example, describes the current situation facing military cost estimators who, with the military facing budget reductions and program stretchouts, are required to rapidly develop weapon system acquisition cost estimates under many different quantity profiles. One choice the cost analyst faces is between using a rate adjustment model or a traditional learning model to develop estimates.[3]

---

[2] Several explanations for these varying, inconclusive empirical results can be offered: (a) Varying results are to be expected because rate changes can lead to both economies and diseconomies of scale. (b) Production rate effects are difficult to isolate empirically because of colinearity with cumulative quantity (Gulledge and Womer, 1986). (c) Researchers have usually used inappropriate measures of production rate leading to misspecified models (Boger and Liao, 1990). (d) The impact of a production rate change is dominated by other uncertainties (Large, Hoffmayer, and Kontrovich, 1974), particularly by cumulative quantity (Asher, 1956). Alchian (1963), for example, was unable to find results for rate adjustment models that improved on the traditional learning curve without a rate parameter.

[3] Two other techniques for making cost estimates when production rate changes are also mentioned by Dorsett: curve rotation, which involves an ad hoc upward or downward revision to

Reacting to the inconsistent findings in the literature, Moses (1990b) raised the question of under what circumstances it would be beneficial to incorporate consideration of production rate into a cost estimation problem. The objective of the research was to attempt to identify conditions when a rate adjustment model would outperform the traditional learning curve model (and vice versa). The ability of each model to accurately estimate future cost was assessed under various conditions. Generally findings were that neither model dominated; each was relatively more accurate under certain conditions.

OBJECTIVE OF THE STUDY

One limitation of the Moses study was that accuracy was measured as the absolute difference between estimated and actual cost, without concern for the direction of the difference. When controlling real-world projects, the consequences of errors in estimation typically depend on whether costs are under or over estimated. Underestimation, resulting in cost growth or cost overruns, is typically met with less pleasure than overestimation. Thus the question of model bias toward over or under estimation is of interest.

The objective of this study is to investigate and compare estimation bias for the learning curve and rate adjustment models.

---

the slope of the learning curve, and the use of repricing models (e.g., Balut, 1981; Balut, Gulledge, and Womer, 1989) which adjust learning curve estimates to reflect a greater or lesser application of overhead cost. Dorsett criticized curve rotation for being subjective and leading to a compounding of error when the prediction horizon is not short. He criticized repricing models because they must be plant-specific to be effective.

Does either model exhibit consistent or systematic bias? Are there circumstances where one model may be biased and the other not? Is the bias produced toward underestimation or overestimation of future cost?

RESEARCH APPROACH

Operationally the research questions require an examination of the estimation errors from two competing cost estimation models. The two competing models were as follows:

The traditional learning curve model, which predicts unit cost as a function of cumulative quantity[4]:

$$C_L = aQ^b \qquad (1)$$

where

C_L	=	Unit cost of item at quantity Q (i.e., with learning considered).
Q	=	Cumulative quantity produced.
a	=	Theoretical first unit cost.
b	=	Learning curve exponent (which can be converted to a learning slope by slope = 2^b).

And the most widely used rate adjustment model, which modifies the traditional learning curve model with the addition of a production rate term:

$$C_R = aQ^bR^c \qquad (2)$$

where

[4]Note that this is an incremental unit cost model rather than a cumulative average cost model. Liao (1988) discusses the differences between the two approaches and discusses why the incremental model has become dominant in practice. One reason is that the cumulative model weights early observations more heavily and, in effect, "smooths" away period-to-period changes in average cost.

C_R = Unit cost of item at quantity Q and production rate per period R (i. e., with production rate as well as learning considered).
Q = Cumulative quantity produced.
R = Production rate per period measure.
a = Theoretical first unit cost.
b = Learning curve exponent.
c = Production rate exponent (which can be converted to a production rate slope by slope = 2^c).

A simulation approach was used to address the research questions. In brief, cost series were generated under varying simulated conditions. The learning curve model and the rate adjustment model were separately fit to the cost series to estimate model parameters. The estimated models were then used to separately predict future cost. Actual cost was compared with predicted cost to measure bias. Finally, an analysis (ANOVA) was conducted relating bias (dependent variable) to the simulated conditions (independent variables).

There are three main benefits gained from the simulation approach. First, factors hypothesized to influence bias can be varied over a wider range of conditions than would be encountered in any one (or many) sample(s) of actual cost data. Second, explicit control is achieved over the manipulation of factors. Third, noise caused by factors not explicitly investigated is removed. Hence simulation provides the most efficient way of investigating data containing a wide variety of combinations of the factor levels while controlling for the effects of other factors not explicitly identified.

RESEARCH CHOICES

There were five choices that had to be made in conducting the simulation experiment:

(1) The form of the rate adjustment (RA) model whose performance was to be compared to the learning curve (LC) model.

(2) The functional form of the cost model used to generate the simulated cost data.

(3) The conditions to be varied across simulation treatments.

(4) The cost objective (what cost was to be predicted).

(5) The measure of bias.

Items (1), (2), (4) and (5) deal with methodological issues. Item (3) deals with the various conditions simulated; conditions which may affect the nature and magnitude of bias. Each item will be discussed in turn.

1. The Rate Adjustment Model. Various models, both theoretical and empirical, have been suggested for incorporating production rate into the learning curve (Balut 1981; Balut, Gulledge, and Womer, 1989; Linder and Wilbourn, 1973; Smith, 1980, 1981; Washburn, 1972; Womer, 1979). The models vary with respect to tradeoffs made between theoretical completeness and empirical tractability. Equation 2, described above, was the specific rate adjustment model analyzed in this study, for several reasons: First, it is the most widely used rate adjustment model in the published literature. Second, it is commonly used today in the practice of cost analysis (e.g., Dorsett, 1990). Third, in addition to cost and quantity data (needed to estimate any LC

model), equation 2 requires only production rate data.[5] Thus equation 2 is particularly appropriate for examining the incremental effect of attention to production rate. In short, equation 2 is the most widely applicable and most generally used rate adjustment model.

2. The Cost Generating Function: A "true" cost function for an actual item depends on the item, the firm, the time period and all the varying circumstances surrounding actual production. It is likely that most manufacturers do not "know" the true cost function underlying goods they manufacture. Thus the choice of a cost function to generate simulated cost data is necessarily ad hoc. The objective here was to choose a "generic" cost function which had face validity, which included components (parameters and variables) that were generalizable to all production situations, and which resulted in a unit cost that depended on both learning and production rate factors. The following explanation of the cost function used reflects these concerns.

At the most basic level the cost of any unit is just the sum of the variable cost directly incurred in creating the unit and the share of fixed costs assigned to the unit, where the amount of fixed costs assigned depend on the number of units produced.

---

[5]Other RA models offered in the literature require knowledge of still additional variables. The equation 2 model is particularly applicable in situations where a cost analyst or estimator does not have ready access to or sufficient knowledge about the cost structure and cost drivers of a manufacturer. Examples include the Department of Defense procuring items from government contractors in the private sector, or prime contractors placing orders with subcontractors.

$$UC = VC + \frac{FC}{PQ} \quad (3)$$

where

UC = Unit cost.
VC = Variable cost per unit.
FC = Total fixed costs per period.
PQ = Production quantity per period.

The original concept of "learning" (Wright, 1936) involved the reduction in variable cost per unit expected with increases in cumulative quantity produced. (By definition, fixed costs are assumed to be unaffected by volume or quantity.) To incorporate the effect of learning, variable cost can be expressed as:

$$VC_Q = VC_1(Q^d) \quad (4)$$

where

Q = Cumulative quantity.
VC_Q = Variable cost of the Qth unit.
VC_1 = Variable cost of the first unit.
d = Parameter, the learning index.

Substituting into equation 3:

$$UC_Q = VC_1(Q^d) + \frac{FC}{PQ} \quad (5)$$

Additionally, assume the existence of a "standard" ("benchmark," "normal," "planned") production quantity per period (PQ_s). Standard fixed cost per unit (SFC) at the standard production quantity would be:

$$SFC = \frac{FC}{PQ_s} \quad (6)$$

The production rate (PR) for any period can then be expressed as a ratio of the production quantity to the standard quantity:

$$PR = \frac{PQ}{PQ_s} \qquad (7)$$

The second term of equation (6) can then be rewritten as:

$$\frac{FC}{PQ} = \frac{SFC}{PR} \qquad (8)$$

and equation 5 rewritten as:

$$UC_Q = VC_1(Q^d) + SFC\ (PR^{-1}) \qquad (9)$$

In this formulation it can be seen that total cost per unit is the sum of variable cost per unit (adjusted for learning) plus standard fixed cost per unit (adjusted for production rate). This model incorporates the two factors presumed to impact unit costs that have been most extensively investigated: cumulative quantity (Q) and production rate per period (PR).[6] It is consistent with both the theoretical and empirical literature which sees the primary impact of learning to be on variable costs and the primary impact of production rate to be on the spreading of fixed costs (Smith, 1980). Simulated cost data in this study was generated using equation 9, while varying values for the variables and parameters on the right hand side of the equation to reflect differing conditions.

---

[6]Smith (1980, 1981), for example, used a model similar to equation 9 to explore the effect of different production rates on unit cost. Balut (1981) and Balut, Gulledge and Womer (1989) construct models based on learning and production quantity to assist in "redistributing" overhead and "repricing" unit costs when changes in production rate occur. The Balut and Balut, Gulledge and Womer models differ in that they determine a learning rate for total (not variable) unit cost and then apply an adjustment factor to allow for the impact of varying production quantity on the amount of fixed cost included in total cost.

3. The Simulated Conditions: The general research hypothesis is that the estimation bias of the LC and RA models will depend on the circumstances in which they are used. What conditions might be hypothesized to affect bias? Seven different factors (independent variables) were varied during the simulation. These factors were selected for examination because they have been found to affect the magnitude of model prediction errors in prior research (Smunt, 1986; Moses, 1990b). In the following paragraphs, each factor is discussed. A label for each is provided, along with a discussion of how the factor was operationalized in the simulation. Table 1 summarizes the seven factors.

i) Data History (DATAHIST): The number of data points available to estimate parameters for a model should affect the accuracy of a model. More data available during the model estimation period should be associated with greater accuracy for both the LC and the RA model.[7] The effect of the number of data points on bias however is unclear. If a model is inherently an "incorrect," biased representation of a phenomena, having more data on which to estimate the model parameters will not eliminate the bias.

In the simulation, data history was varied from four to seven to ten data points available for estimating model parameters. This simulates having knowledge of costs and quantities for four, seven or ten production lots. Four is the minimum number of observations

[7] There are, of course, cost/benefit tradeoffs. The marginal benefits of increased prediction accuracy for any model must be weighed against the marginal costs of additional data collection.

TABLE 1

INDEPENDANT VARIABLES

Concept	Label	Levels		
Data History	DATAHIST[1]	4	7	10
Variable Cost Learning Rate	VCRATE	75%	85%	95%
Fixed Cost Burden	BURDEN[2]	15%	33%	50%
Production Rate Trend	PROTREND[3]	Level		Growth
Production Rate Instability/Variance	RATEVAR[4]	.05	.15	.25
Cost Noise/Variance	COSTVAR[5]	.05	.15	.25
Future Production Level	FUTUPROD[6]	Low	Same	High

[1]Number of data points available during the model estimation period; simulates the number of past production lots.

[2]Standard per unit fixed cost as a percentage of cumulative average per unit total cost, during the model estimation period.

[3]A level trend means production at 100% of standard production for each lot during the estimation period. A growth trend means production rate gradually increasing to 100% of standard production during the estimation period. The specific growth pattern depends on the number of production lots in the estimation period, with sequences as follows (expressed as a % of standard): For DATAHIST = 4: 33%, 67%, 100%, 100%. For DATAHIST = 7: 20%, 40%, 60%, 80%, 100%, 100%, 100%. For DATAHIST = 10: 10%, 20%, 35%, 50%, 70%, 90%, 100%, 100%, 100%, 100%.

[4]Coefficient of variation of production rate. (Degree of instability of production rate around the general production rate trend.)

[5]Coefficient of variation of total per unit cost.

[6]"Same" means production rate at 100% of standard for each lot produced within the prediction zone. "Low" means production rate at 50%. "High" means production rate at 150%.

needed to estimate the parameters of the RA model by regression. The simulation focuses on lean data availability both because the effects of marginal changes in data availability should be most pronounced when few observations are available and because many real world applications (e.g., cost analysis of Department of Defense weapon system procurement) occur under lean data conditions.

ii) Variable Cost Learning Rate (VCRATE): In the cost generating function, learning affects total unit cost by affecting variable cost per unit. Past research (Smunt, 1986) has shown that the improvement in prediction accuracy from including a learning parameter in a model (when compared to its absence) depends on the degree of learning that exists in the underlying phenomena being modeled. The association between learning rate and degree of bias however is unclear. In the simulation, variable cost learning rate (reflected in parameter d in equation 9) was varied from 75% to 85% to 95%. Generally, complex products or labor intensive processes tend to experience high rates of learning (70-80%) while simple products or machine-paced processes experience lower (90-100%) rates (Smunt, 1986).[8]

iii) Fixed Cost Burden (BURDEN): In theory (and in the cost function, equation 9) a change in the number of units produced during a period affects unit cost in two ways: First, increasing volume increases cumulative quantity and decreases variable cost

[8]See Conway and Schultz (1959) for further elaboration of factors impacting learning rates.

per unit, due to learning. Second, increasing volume increase the production rate for a period and reduces fixed cost per unit, due to the spreading of total fixed cost over a larger output. Both these effects operate in the same direction; i. e., increasing volume leads to lower per unit cost. This has led some cost analysts to conclude that in practice, it is sufficient to use an LC model, letting the cumulative quantity variable reflect the dual impacts of increased volume. Adding a production rate term to an LC model is seen as empirically unnecessary.

In principle, if fixed cost was zero, cumulative quantity would be sufficient to explain total unit cost and production rate would be irrelevant. But as fixed cost increases as a proportion of total cost, the impact of production rate should become important. This suggests that the relative bias of the LC and RA models may depend on the amount of fixed cost burden assigned to total cost.

Fixed cost burden was simulated by varying the percentage of total unit cost made up of fixed cost.[9] Three percentages were used in the simulation: 15%, 33%, and 50%. The different percentages can be viewed as simulating different degrees of operating

---

[9]Operationally this is a bit complex, since both per unit variable and per unit fixed cost depend on other simulation inputs (cumulative quantity and production rate per period). The process of relating fixed cost to total cost was as follows: First, a cumulative average per unit variable cost for all units produced during the estimation period was determined. Then a standard fixed cost per unit was set relative to the cumulative average per unit variable cost. For example, if standard fixed cost per unit was set equal to cumulative average variable cost per unit, then "on average" fixed cost would comprise 50% of total unit cost during the estimation period. Actual fixed cost per unit may differ from standard fixed cost per unit if the production rate (discussed later) was not at 100% of standard.

leverage, of capital intensiveness, or of plant automation. The 15% level reflects the average fraction of price represented by fixed overhead in the aerospace industry, as estimated at one time by DOD (Balut, 1981).[10] The larger percentages are consistent with the trend toward increased automation (McCullough and Balut, 1986).

iv) Production Rate Trend (PROTREND): When initiating a new product, it is not uncommon for the production rate per period to start low and trend upward to some "normal" level. This may be due both to the need to develop demand for the output or the desire to start with a small production volume, allowing slack for working bugs out of the production process. Alternatively, when a "new" product results from a relatively small modification of an existing product, sufficient customer demand or sufficient confidence in the production process may be assumed and full scale production may be initiated rapidly. In short, two different patterns in production volume may be exhibited early on when introducing a new item: a gradual growing trend toward full scale production or a level trend due to introduction at full scale production volume.

The simulation created two production trends during the model estimation period: "level" and "growth." These represented general trends (but, as will become clear momentarily, variance around the general trend was introduced). The level trend simulated a production rate set at a "standard" 100% each period

[10] In the absence of firm-specific cost data, the Cost Analysis Improvement Group in the Office of the Secretary of Defense treats 15% of the unit price of a defense system as representing fixed cost (Pilling, 1990).

during model estimation. The growth trend simulated production rate climbing gradually to 100%. Details of the trends are in table 1.

v) Production Rate Instability/Variance (RATEVAR): Numerous factors, in addition to the general trend in output discussed above, may operate to cause period-to-period fluctuations in production rate. Manufacturers typically do not have complete control over either demand for output or supply of inputs. Conditions in either market can cause instability in production rate. (Of course, unstable demand, due to the uncertainties of annual budget negotiations, is claimed to be a major cause of cost growth during the acquisition of major weapon systems by the DoD).

Production rate instability was simulated by adding random variance to each period's production rate during the estimation period. The amount of variance ranged from a coefficient of variation of .05 to .15 to .25. For example, if the production trend was level and the coefficient of variation was .05 then "actual" production rates simulated were generated by a normal distribution with mean equal to the standard production rate (100%) and sigma equal to 5%.

vi) Cost Noise/Variance (COSTVAR): From period to period there will be unsystematic, unanticipated, non-recurring, random factors that will impact unit cost. Changes in the cost, type or availability of input resources, temporary increases or decreases in efficiency, and unplanned changes in the production process are all possible causes. Conceptually such unsystematic factors can be thought of as adding random noise to unit cost. While unsystematic variation in cost cannot (by definition) be controlled, it is often possible to characterize different production processes in terms of the degree of unsystematic variation; some processes are simply less well-understood, more uncertain, and less stable than others.

Does bias depend on the stability of the process underlying cost? To investigate this question, random variance was added to the simulated costs generated from the cost function. The amount of variance ranged from a coefficient of variation of .05 to .15 to .25. For example, when the coefficient of variation was .25, then "actual" unit costs simulated were generated by a normal distribution with mean equal to cost from equation 9 and sigma equal to 25%.

vii) Future Production Level (FUTUPROD): Once a model is constructed (from data available during the estimation period), it is to be used to predict future cost. The production rate planned for the future may vary from past levels. Further growth may be planned. Cutbacks may be anticipated. Will the level of the future production rate affect the bias of the LC and RA models? Does one model tend to under (or over) estimate cost if cutbacks in production are anticipated and another if growth is planned? One might expect that inclusion of a rate term might be expected to reduce bias when production rate changes significantly (i. e., either growth or decline in the future period).

In the simulation, future production was set at three levels:

low (50% of standard), same (100% of standard) and high (150% of standard). These simulate conditions of cutting back, maintaining or increasing production relative to the level of production existing at the end of the model estimation period.

4. The Cost Objective: What is to be predicted? Up to this point the stated purpose of the study has been to evaluate bias when predicting future cost. But which future cost? Three alternatives were examined.

i) Next period average unit cost: As the label suggests this is the average per.unit cost of items in the production "lot" manufactured in the first period following the estimation period. Here the total cost of producing the output for the period is simply divided by the output volume to arrive at unit cost. Attention to this cost objective simulates the need to predict near term unit cost.

ii) Total cost over a finite production horizon: The objective here is to predict the total cost of all units produced during a fixed length production horizon. Three periods was used as the length of the production horizon (one production lot produced each period). If the future production rate is low (high) then relatively few (many) units will be produced during the finite production horizon. Attention to this cost objective simulates the need to predict costs over some specific planning period, regardless of the volume to be produced during that planning period.

iii) Total program cost: The objective here is to predict total cost for a specified number of units. If the future production rate is low (high) then relatively more (fewer) periods will be required to manufacture the desired output. The simulation was constructed such that at a low (same, high) level of future production six (three, two) future periods were required to produce the output. Attention to this cost objective simulates the need to predict total cost for a particular production program, regardless of the number of future periods necessary to complete the program.

Examining each of these three cost objectives was deemed necessary to provide a well-rounded investigation of bias. However, the findings were the same across the three cost objectives. In the interest of space, the remainder of this paper will discuss the analysis and results only for the first cost objective, the average cost per unit for the next period's output.

5. <u>The Measure of Bias</u>: A model specific measure of bias (BIAS) was determined separately for each (LC or RA) model as follows:

$$\text{BIAS} = (\text{PUC} - \text{AUC}) \div \text{AUC}$$

where

PUC = Predicted unit cost from either the learning curve or the rate adjustment model.

AUC = Actual unit cost as generated by the cost function.

Positive values for BIAS indicate that a model overestimates actual future cost; negative values indicate underestimation. A model that is unbiased should, on average, produce values for BIAS of zero. BIAS represents the dependent variable in the statistical analysis. The research question then becomes: What factors or conditions explain variance in BIAS?

Figure 1 summarizes the complete simulation process leading up to the determination of BIAS. The simulation was run once for each possible combination of treatments. Given seven factors varied and three possible values for each factor (except for PROTREND which had two), there were 3 x 3 x 3 x 3 x 3 x 3 x 2 = 1458 combinations. Thus the simulation generated 1458 observations and 1458 values for BIAS for each of the two models.[11]

ANALYSIS AND FINDINGS

BIAS was evaluated using analysis of variance (ANOVA) to conduct tests of statistical significance. All main effects and first order (pairwise) interactions were examined. Findings with probability less than .01 were considered significant.

LC Model Bias. Table 2 provides ANOVA results addressing BIAS from the LC model. As shown, four main effects, DATAHIST, BURDEN, PROTREND, and FUTUPROD, are significant, indicating that values for

---

[11]In the simulation, just as in the real practice of cost analysis, it is possible for a model estimated on limited data to be very inaccurate, leading to extreme values for BIAS. If such outlier values were to be used in the subsequent analysis, findings would be driven by the outliers. Screening of the observations for outliers was necessary. During the simulation, if a model produced an BIAS value in excess of 100%, then that value was replaced with 100%. This truncation has the effect of reducing the impact of an outlier on the analysis while still retaining the observation as one that exhibited poor accuracy. Alternative approaches to the outlier problem included deletion instead of truncation and use of a 50% BIAS cutoff rather than the 100% cutoff. Findings were not sensitive to these alternatives.

FIGURE 1. SIMULATION FLOWCHART

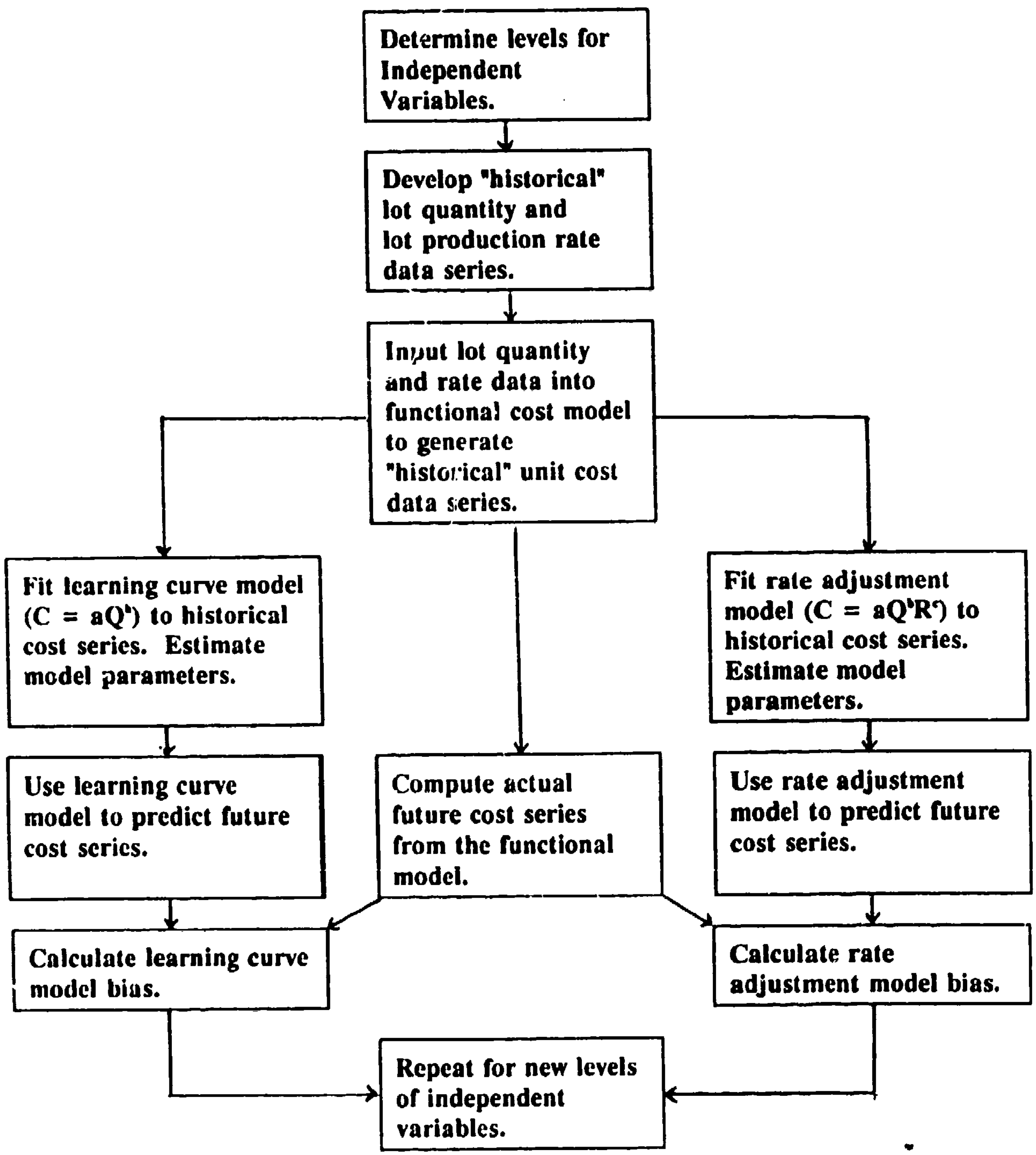


TABLE 2

BIAS FROM LEARNING CURVE MODEL

ANALYSIS OF VARIANCE RESULTS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
Model	85	56.92195	.6697	29.07
Error	1372	31.60432	.0230	PR>F:
Corrected Total	1457	88.52626		.0000

R^2	CV	BIAS MEAN
.6430	140.35	-.1081

SOURCE	DF	ANOVA SS	F VALUE	PR>F
DATAHIST	2	0.2937	6.38	0.0018*
VCRATE	2	0.0085	0.19	0.8311
BURDEN	2	0.3710	8.05	0.0003*
PROTREND	1	4.6998	204.03	0.0001*
RATEVAR	2	0.1167	2.53	0.0797
COSTVAR	2	0.0976	2.12	0.1205
FUTUPROD	2	47.0628	1021.54	0.0000*
DATAHIST*VCRATE	4	0.1184	1.29	0.2737
DATAHIST*BURDEN	4	0.0363	0.39	0.8124
DATAHIST*PROTREND	2	0.1280	2.78	0.0625
DATAHIST*RATEVAR	4	0.0265	0.29	0.8854
DATAHIST*COSTVAR	4	0.1503	1.63	0.1637
DATAHIST*FUTUPROD	4	0.1398	1.52	0.1944
VCRATE*BURDEN	4	0.0506	0.55	0.6990
VCRATE*PROTREND	2	0.0374	0.81	0.4435
VCRATE*RATEVAR	4	0.0623	0.68	0.6083
VCRATE*COSTVAR	4	0.1068	1.16	0.3271
VCRATE*FUTUPROD	4	0.2820	3.06	0.0159
BURDEN*PROTREND	2	0.3131	6.80	0.0012*
BURDEN*RATEVAR	4	0.0282	0.31	0.8738
BURDEN*COSTVAR	4	0.1631	1.77	0.1323
BURDEN*FUTUPROD	4	1.8751	20.35	0.0001*
PROTREND*RATEVAR	2	0.0176	0.38	0.6812
PROTREND*COSTVAR	2	0.0323	0.70	0.4955
PROTREND*FUTUPROD	2	0.3652	7.93	0.0004*
RATEVAR*COSTVAR	4	0.1570	1.70	0.1464
RATEVAR*FUTUPROD	4	0.0949	1.03	0.3902
COSTVAR*FUTUPROD	4	0.0855	0.93	0.4467

BIAS are influenced by these treatment conditions. Table 3 summarizes BIAS values under the various conditions. Some interesting patterns are evident.

First, the overall mean BIAS across all observations is -.108. This means that, on average, the LC produces cost estimates that are about 11% too low.

Second, the mean BIAS for each treatment for every variable of interest, is negative, (with only one exception, when FUTUPROD is "high"). This means that the LC model bias toward underestimation is a consistent, pervasive phenomena. It is not driven by isolated conditions.

Third, in spite of the general tendency toward underestimation, the degree of bias does differ depending on the conditions. The effects of the different conditions perhaps can be best demonstrated by a plot of BIAS values by treatments. Figure 2 shows such a plot, with the (four significant) variables superimposed. In this plot, 1, 2, and 3 on the X-axis reflect low, medium and high values for the independent variables (which are taken from the left, middle and right columns of Table 3). Figure 2 reiterates the point made previously: BIAS is consistently negative (except when FUTUPROD is high). More importantly, trends are evident:

a) Data History: Negative bias, the underestimation of future cost, tends to increase as the number of observations available for estimating model parameters (DATAHIST) increases. At first glance this seems counter-intuitive. Traditional wisdom says that having more data available leads to better parameter estimates and better model forecasts. But that is true only if a model is correctly specified. This issue will be discussed further later.

b) Fixed Cost Burden: Negative bias tends to increase as the proportion of total costs comprised of fixed costs (BURDEN) increases. This result is perhaps not surprising. In the underlying cost phenomena being modeled, learning impacts the incurrence of variable costs, not fixed costs. It is plausible that the LC model would become more biased as fixed costs increase.

c) Past Production Trend: The negative bias is considerably stronger if the rate of production was growing, rather than level during the model estimation period. This is not difficult to explain. An increasing production rate during the model estimation period will result in a steadily declining fixed cost per unit. An LC model will interpret this rate effect as a learning effect, and overestimate the degree of learning actually occurring. Future forecasts of cost will thus be biased downward.

d) Future Production Level: As the production rate, during the period for which costs are being forecast, shifts from "low" to "high", the LC model shifts from strongly underestimating to overestimating cost. In short, there is an inverse relationship between future production level and the bias toward underestimation. This effect is to be expected. Higher (lower) future production will result in lower (higher) fixed cost, and total cost, per unit, creating a tendency toward positive (negative) bias for any cost estimate.

Note that the only time cost is overestimated by the LC model

TABLE 3

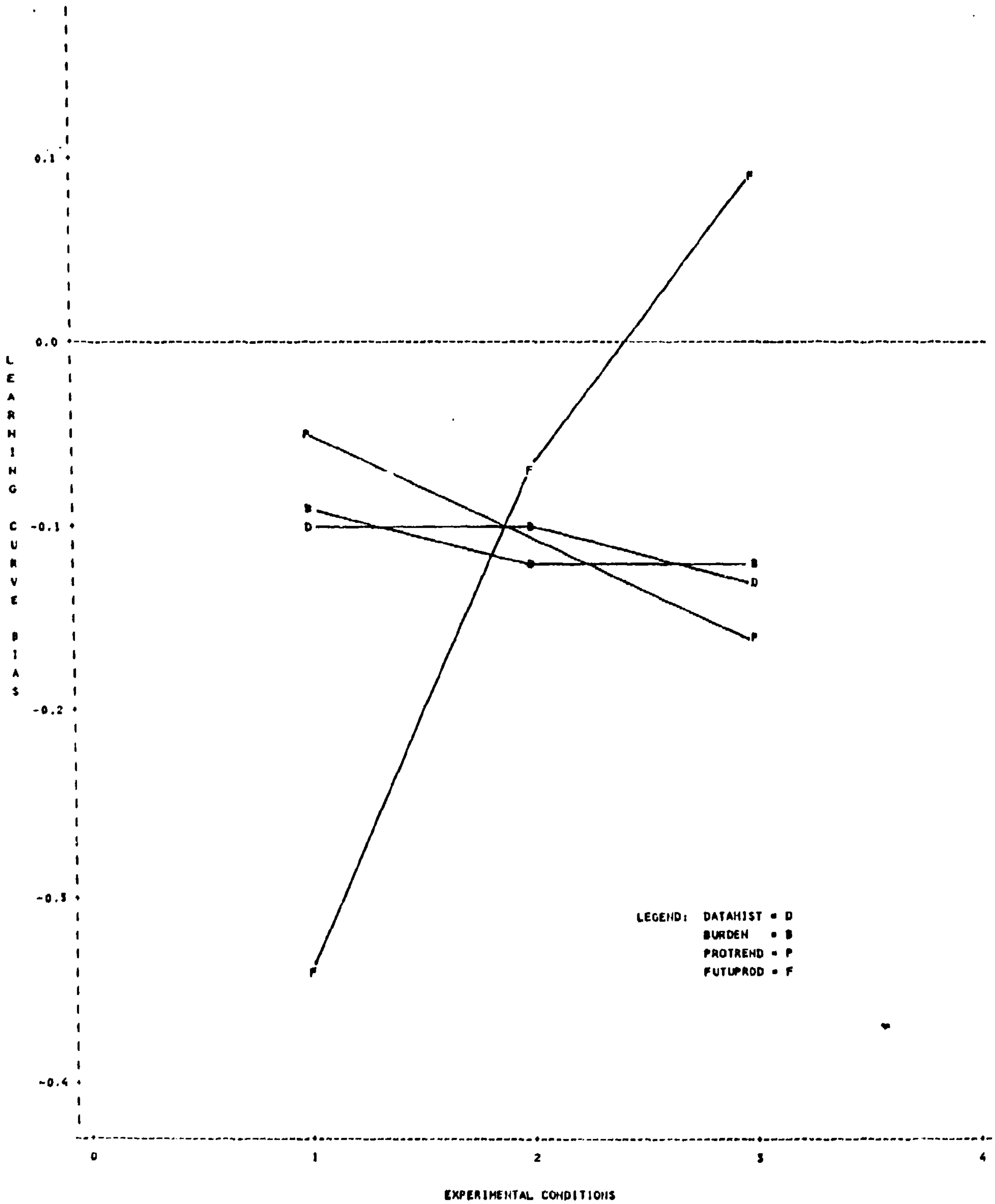
LEARNING CURVE MODEL BIAS
BY MAIN EFFECTS

Independent Variable	BIAS for Each Level		
DATAHIST Value:	4	7	10
BIAS Mean:	-.096	-.100	-.128
VCRATE Value:	75%	85%	95%
BIAS Mean:	-.108	-.111	-.105
BURDEN Value:	15%	33%	50%
BIAS Mean:	-.086	-.119	.120
PROTREND Value:	level	-	growth
BIAS Mean:	-.051	-	-.165
RATEVAR Value:	.05	.15	.25
BIAS Mean:	-.120	-.098	-.107
COSTVAR Value:	.05	.15	.25
BIAS Mean:	-.099	-.106	-.119
FUTUPROD Value	low	same	high
BIAS Mean:	-.344	-.070	.091

Overall Mean: -.108
Range of Group Means: -.344 to .091

FIGURE 2

PLOT OF LEARNING CURVE MODEL BIAS
BY MAIN EFFECTS



is when future production level is high. The LC is still biased toward underestimation, but if the future production level increases enough to reduce per unit fixed cost enough, the tendency toward underestimation is masked by the offsetting tendency toward reduced actual per unit cost.

In addition to these main effects, the Table 2 ANOVA results indicated that pairwise interactions involving BURDEN, PROTREND and FUTUPROD are also significant; not only does BIAS depend on these three variables, it depends on how they interact. These interactions are illustrated in Figures 3, 4 and 5.

Figure 3, the interaction between Fixed Cost Burden and Production Rate Trend, merely reinforces previous findings: Negative bias tends to be greater when burden is higher or when the production rate grows during the model estimation period. The figure just indicates that the combination of these two conditions--high burden coupled with growing production volume--magnifies the negative bias.

Figure 4, the interaction between Fixed Cost Burden and Future Production Level, clearly reinforces the previously noted inverse relationship between future production level and the bias toward underestimation. But findings concerning Burden now appear conditional. High burden increases the tendency toward underestimation, if future production level is low. But high burden increases the tendency toward overestimation when future production level is high. In short, increasing fixed cost burden magnifies the biasing effect--in either direction--caused by shifts in the future production level.

Figure 5 shows the interaction between the production trend during the model estimation period and the future production level during the forecast period. The most interesting observation concerns the two points where BIAS is close to zero. These occur when a) a "level" production trend is coupled with the "same" level in the future forecast period, and b) a "growing" production trend is coupled with a "high" level in the forecast period. Consistency characterizes both situations; the production rate is either consistently level or consistently increasing throughout the joint estimation/forecast periods. In contrast, the greatest bias occurs when a "growing" production trend is coupled with a "low" level in the future forecast period. Here an inconsistent pattern, a shift from increasing to decreasing production rate, causes severe underestimation of cost.

RA Model Bias. Table 4 provides ANOVA results addressing BIAS from the RA model. Table 5 summarizes BIAS values under the various experiment conditions. Two findings are evident. First the overall mean BIAS for all observations is only -.0016. Thus, on average, the RA model exhibits no bias. Second, this absence of bias is evident for all treatments across all variables of interest. There are no significant main effects in the ANOVA results and group means for BIAS in table 5 range only from -.021 to .026. Thus the overall absence of bias is not caused by positive bias under some conditions offsetting negative bias under other conditions. Rather the absence of noticeable bias exists

FIGURE 3

LEARNING CURVE MODEL BIAS

INTERACTION OF FIXED COST BURDEN
AND PRODUCTION RATE TREND

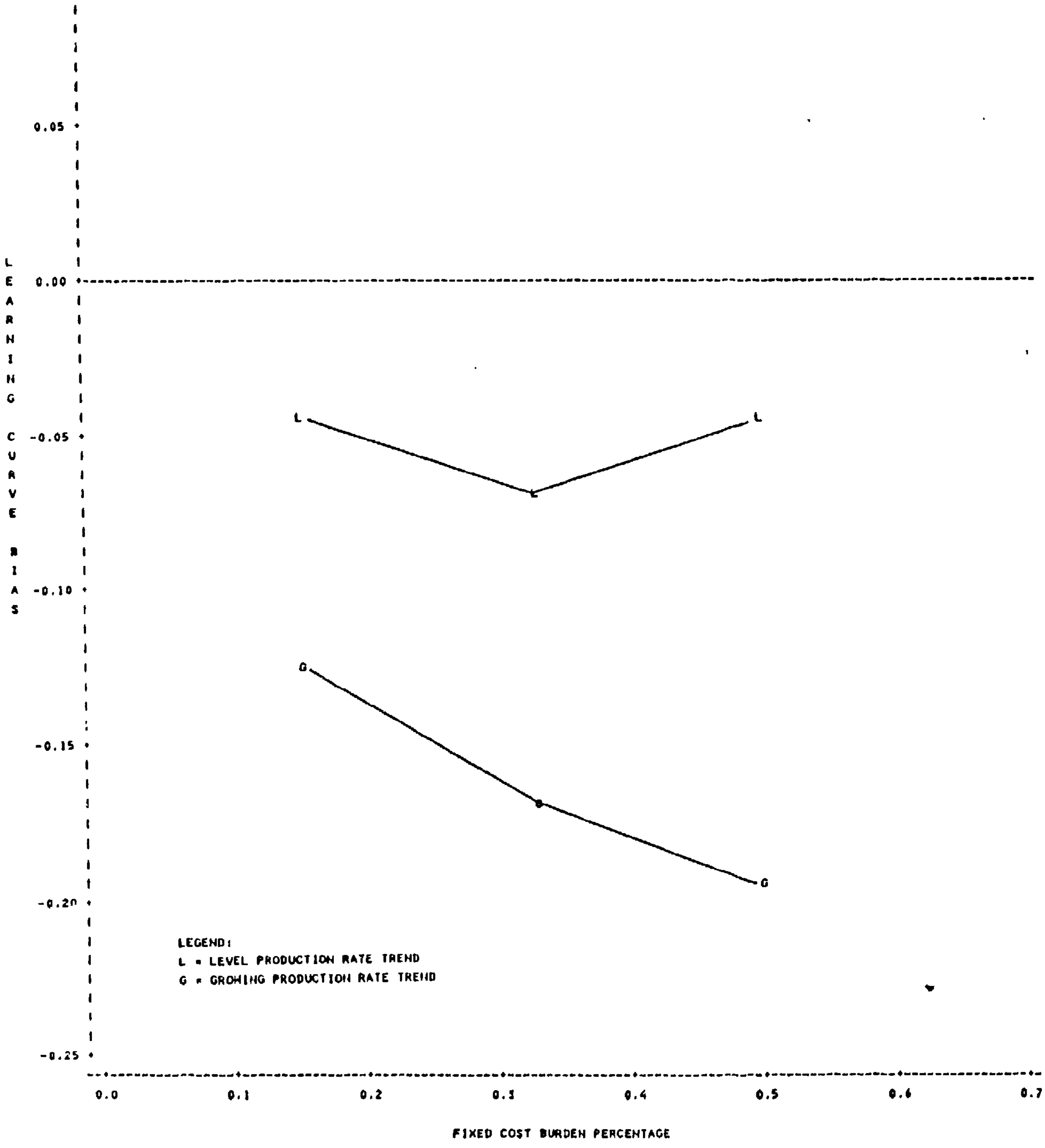


FIGURE 4

LEARNING CURVE MODEL BIAS

INTERACTION OF FIXED COST BURDEN
AND FUTURE PRODUCTION LEVEL

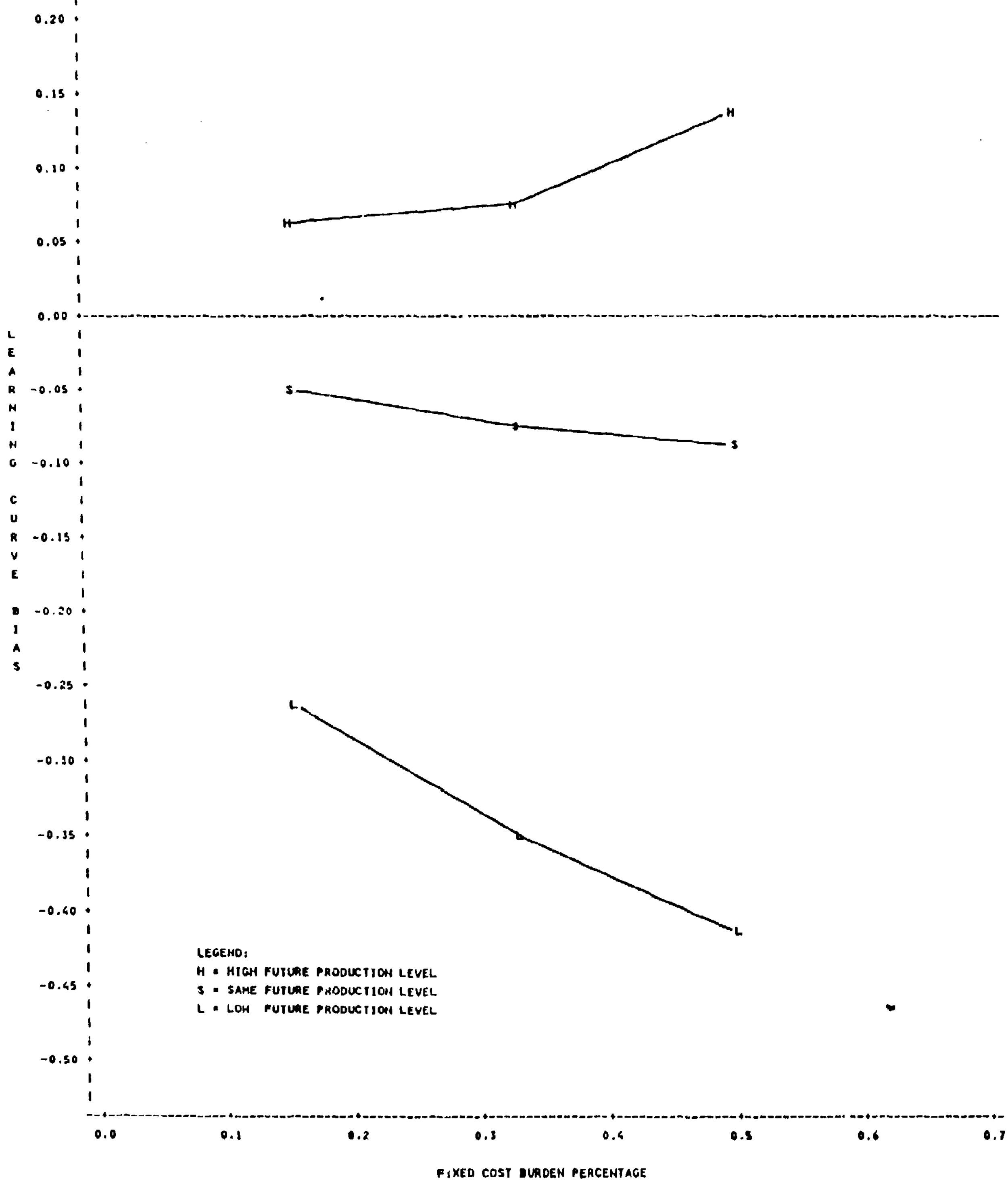


FIGURE 5

LEARNING CURVE MODEL BIAS

INTERACTION OF PRODUCTION RATE TREND
AND FUTURE PRODUCTION LEVEL

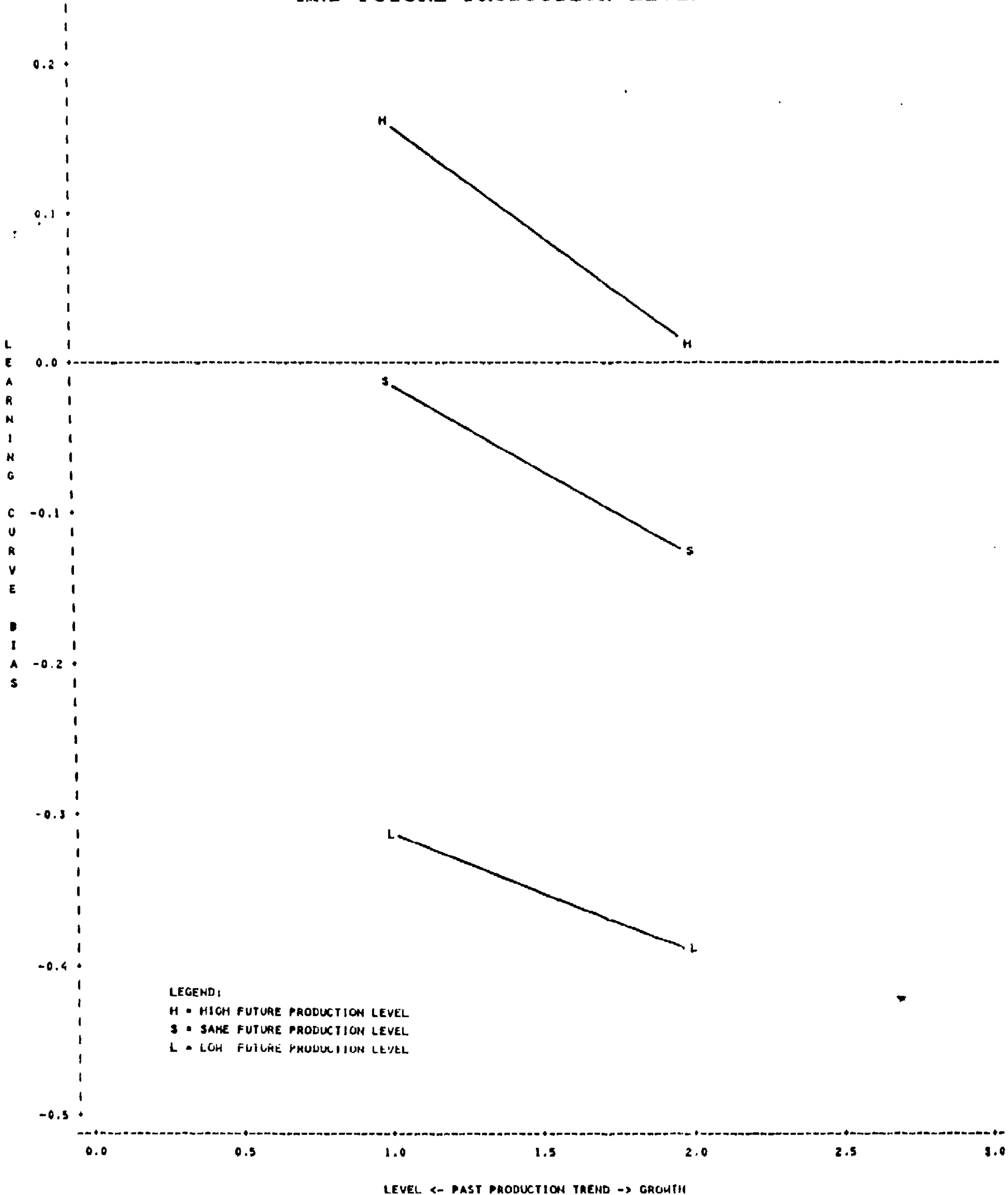


TABLE 4

BIAS FROM RATE ADJUSTMENT MODEL

ANALYSIS OF VARIANCE RESULTS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
Model	85	11.18626	.1316	1.08
Error	1372	166.9451	.1217	PR>F:
Corrected Total	1457	178.1314		.2919

R^2	CV	BIAS MEAN
.0628	21638.82	-.0016

SOURCE	DF	ANOVA SS	F VALUE	PR>F
DATAHIST	2	0.1779	0.73	0.4815
VCRATE	2	0.3435	1.41	0.2441
BURDEN	2	0.0539	0.22	0.8012
PROTREND	1	0.2335	1.92	0.1662
RATEVAR	2	0.2986	1.23	0.2934
COSTVAR	2	0.5567	2.29	0.1019
FUTUPROD	2	0.3965	1.63	0.1964
DATAHIST*VCRATE	4	0.3066	0.63	0.6412
DATAHIST*BURDEN	4	0.0866	0.18	0.9498
DATAHIST*PROTREND	2	0.0972	0.40	0.6706
DATAHIST*RATEVAR	4	0.3802	0.78	0.5373
DATAHIST*COSTVAR	4	0.0617	0.13	0.9727
DATAHIST*FUTUPROD	4	0.2723	0.56	0.6921
VCRATE*BURDEN	4	0.6156	1.26	0.2818
VCRATE*PROTREND	2	0.1873	0.77	0.4633
VCRATE*RATEVAR	4	0.3605	0.74	0.5642
VCRATE*COSTVAR	4	0.1389	0.29	0.8875
VCRATE*FUTUPROD	4	1.3745	2.82	0.0238
BURDEN*PROTREND	2	0.0470	0.19	0.8243
BURDEN*RATEVAR	4	0.3449	0.71	0.5860
BURDEN*COSTVAR	4	0.3527	0.72	0.5751
BURDEN*FUTUPROD	4	0.6125	1.26	0.2844
PROTREND*RATEVAR	2	0.1738	0.71	0.4897
PROTREND*COSTVAR	2	0.2152	0.88	0.4132
PROTREND*FUTUPROD	2	1.1777	4.84	0.0080*
RATEVAR*COSTVAR	4	0.1900	0.39	0.8156
RATEVAR*FUTUPROD	4	1.5652	3.22	0.0122
COSTVAR*FUTUPROD	4	0.5640	1.16	0.3273

TABLE 5

RATE ADJUSTMENT MODEL BIAS
BY MAIN EFFECTS

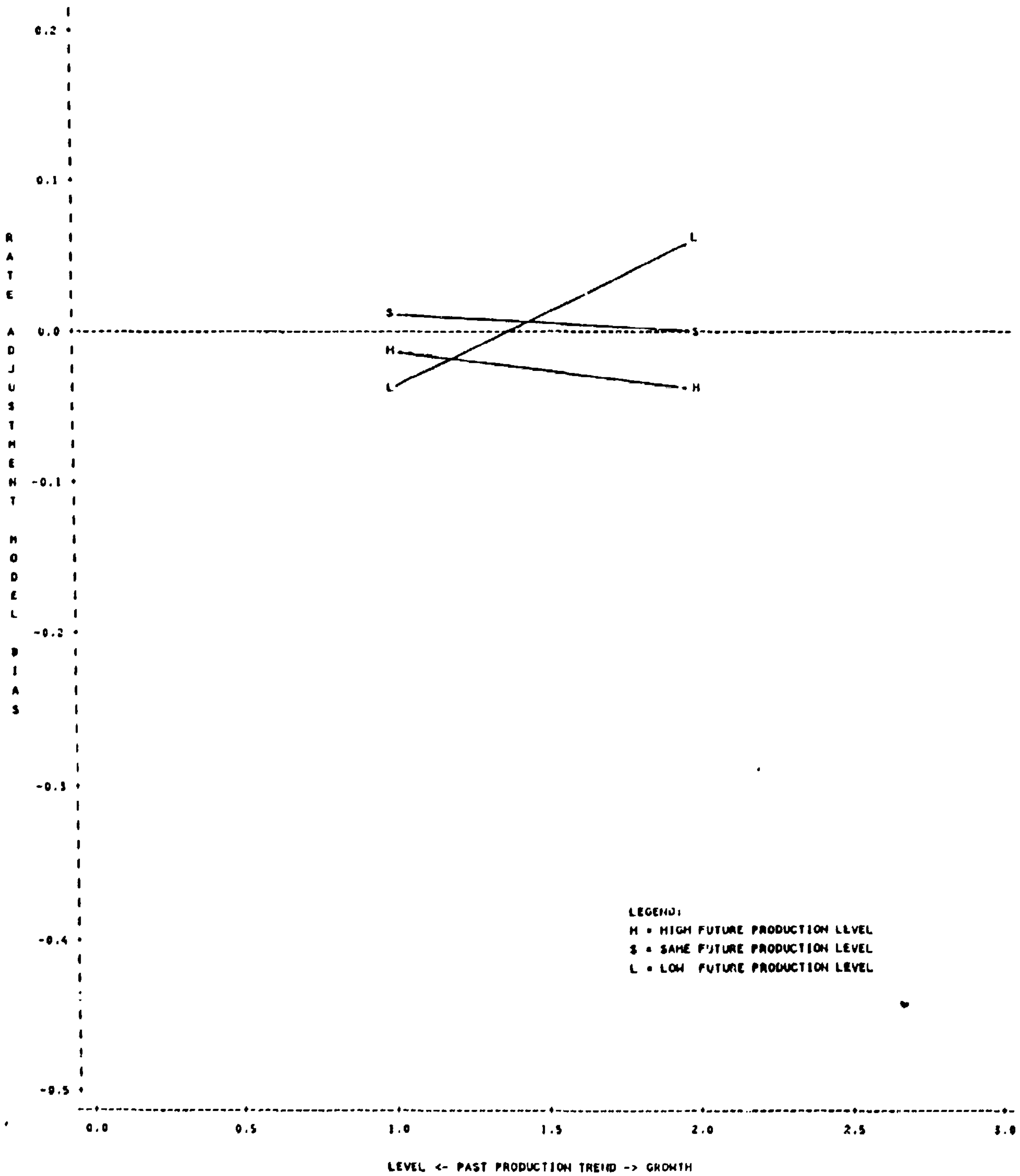
Independent Variable	BIAS		
DATAHIST Value:	4	7	10
BIAS Mean:	.004	.008	-.017
VCRATE Value:	75%	85%	95%
BIAS Mean:	-.021	-.000	.016
BURDEN Value:	15%	33%	50%
BIAS Mean:	-.004	-.008	.007
PROTREND Value:	level	-	growth
BIAS Mean:	-.014	-	.011
RATEVAR Value:	.05	.15	.25
BIAS Mean:	.016	-.002	-.019
COSTVAR Value:	.05	.15	.25
BIAS Mean:	-.019	-.011	.026
FUTUPROD Value	low	same	high
BIAS Mean:	.015	.004	-.024

Overall Mean: -.0016
Range of Group Means: -.021 to .026

FIGURE 6

RATE ADJUSTMENT MODEL BIAS

INTERACTION PRODUCTION RATE TREND
AND FUTURE PRODUCTION LEVEL



across the various treatments.

There is one statistically significant first order interaction in the ANOVA. Figure 6 plots this interaction between Production Rate Trend and Future Production Level. Two points seem noteworthy. First, the greatest bias occurs when a "growing" production trend during the model estimation period is coupled with a "low" production level in the forecast period. So, as with the LC model, a shift from increasing to decreasing production causes bias to occur. Second, in spite of this interaction result being statistically significant, the magnitude of bias evident is far less than with the LC model. In a comparative sense, the RA model still does not appear to create a bias problem.

Additional Analysis of LC Bias: The findings that the degree of bias in the LC model is dependent on PROTREND and FUTUPROD is not completely surprising. Both variables reflect how production rate varies from period to period, and the LC model does not include a rate term.[12]

The findings that LC model bias also depends on DATAHIST and BURDEN merit a bit more attention. To further investigate, some addition simulations were run under "ideal" conditions, where

---

[12]This does not mean the finding is without interest. Many researchers and cost analysts (e.g., Gulledge and Womer, 1986) have noticed that empirically there is often high colinearity between cumulative quantity and production rate. This colinearity has been argued to make production rate a somewhat redundant variable in a model, leading to unreliable parameter estimates when the model is estimated and providing little incremental benefit when the model is used for forecasting future cost. The current findings suggest that one role of a production rate variable in a model is to reduce model bias.

impacts on cost caused by the other variables were suppressed. More specifically equation 9 was used to generate cost series where a) production rate was level during the model estimation period, b) production rate stayed at the same level during the cost forecast period, c) random noise in cost was set at zero, and d) production rate variance was set at zero. Only VCRATE, BURDEN and DATAHIST were varied. Again LC models were fit to the cost series and then estimated future costs were compared with actual future costs.

i) The Concave Curve: Figure 7 shows a log-log plot of residuals (actual minus estimate cost) by quantity for one illustrative situation (where VCRATE = 75%; BURDEN = 50%; DATAHIST = 7). Recall that a central assumption of a learning curve is that cost and quantity are log linear. Figure 7 shows cost as estimated and predicted by the LC model as a horizontal line (abscissa of zero), while the plot of the residuals displays the pattern of actual costs relative to the LC line. Note that actual costs are not log linear with quantity; instead an obvious concave curve is evident. This pattern is not a result of the particular values for VCRATE, BURDEN, and DATAHIST; the same pattern was evident for all other combinations of variable values examined.

The vertical line in the figure separates the seven cost observations used to estimate the LC model, on the left, from three future costs the model is used to predict, on the right. The concavity of the actual cost curve results in each successive actual cost diverging increasingly from the LC model prediction.

FIGURE 7

ESTIMATED COST VERSUS ACTUAL COST

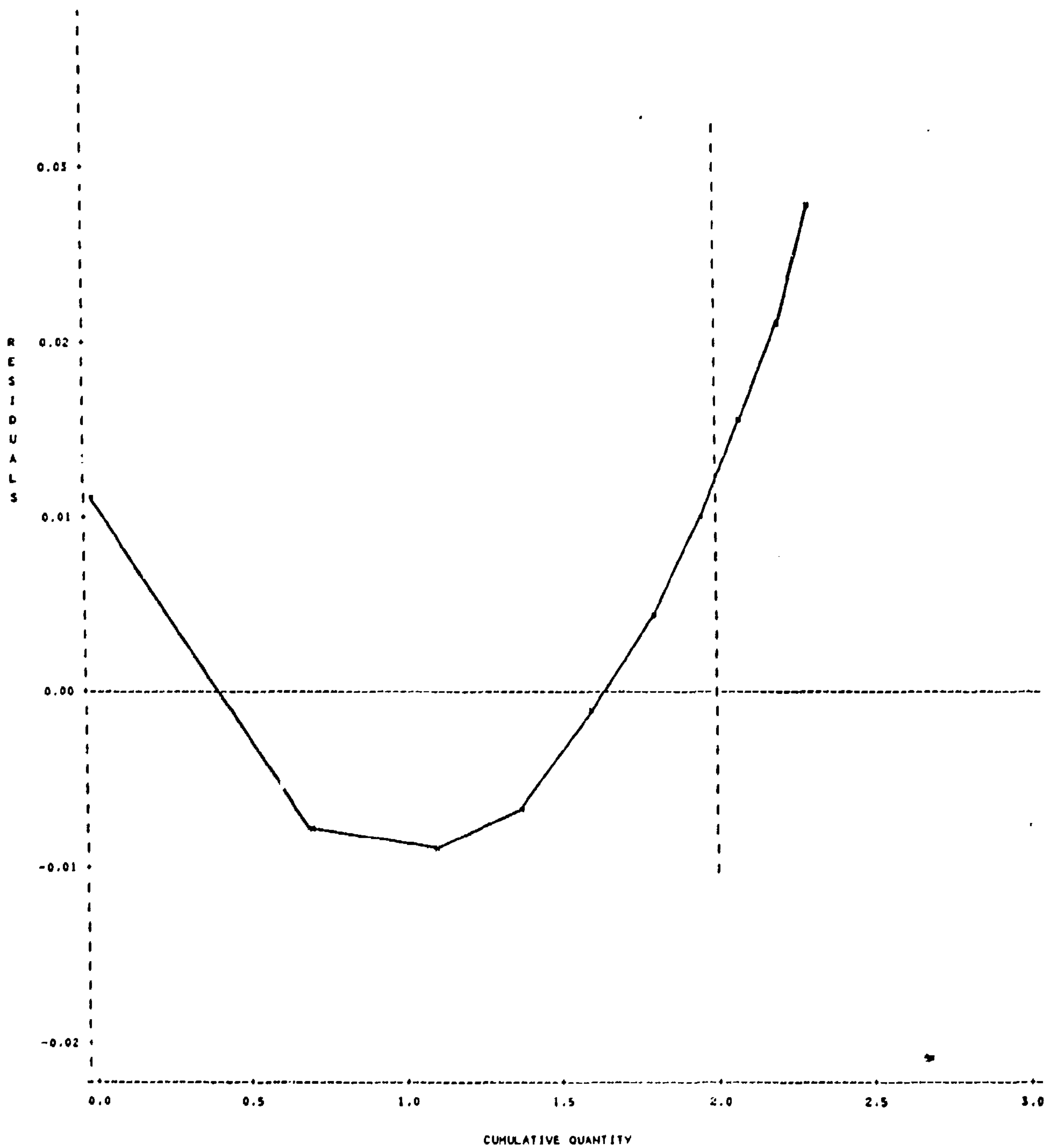


TABLE 6

BIAS PATTERNS FROM THE LC MODEL
(At selected values for BURDEN and VCRATE)

VCRATE	BURDEN	$BIAS_1$	$BIAS_2$	$BIAS_3$	$BIAS_4$
75%	10%	-.00544	-.00758	-.00973	-.01186
	20%	-.00973	-.01348	-.01720	-.02087
	30%	-.01287	-.01772	-.02250	-.02719
	40%	-.01487	-.02037	-.02575	-.03099
	50%	-.01570	-.02138	-.02692	-.03229
	60%	-.01531	-.02076	-.02604	-.03113
	70%	-.01365	-.01843	-.02303	-.02747
	80%	-.01063	-.01429	-.01781	-.02119
85%	10%	-.00176	-.00243	-.00310	-.00376
	20%	-.00315	-.00435	-.00552	-.00668
	30%	-.00418	-.00573	-.00727	-.00877
	40%	-.00482	-.00660	-.00834	-.01005
	50%	-.00507	-.00692	-.00873	-.01050
	60%	-.00492	-.00670	-.00843	-.01012
	70%	-.00436	-.00592	-.00743	-.00891
	80%	-.00336	-.00456	-.00571	-.00683
95%	10%	-.00018	-.00025	-.00032	-.00038
	20%	-.00033	-.00045	-.00056	-.00068
	30%	-.00043	-.00058	-.00074	-.00089
	40%	-.00049	-.00067	-.00084	-.00101
	50%	-.00051	-.00069	-.00088	-.00105
	60%	-.00049	-.00067	-.00084	-.00101
	70%	-.00043	-.00058	-.00073	-.00088
	80%	-.00032	-.00044	-.00056	-.00067

NOTE: DATAHIST = 7. $BIAS_1$ is the bias in forecasting the cost of the first unit produced after the model estimation period; $BIAS_2$ relates to the second unit, etc.

The conclusion to be drawn is that whenever a learning curve is used to model a cost series that includes some fixed cost component (some component that is not subject to learning), then a log linear model is being fit to a log concave phenomena. A systematic bias toward underestimation of future cost is inherent in the LC model.

ii) Bias Patterns: Table 6 lists measures of BIAS for various combinations of BURDEN and VCRATE. The absolute magnitude of the BIAS values is not important; three patterns in the table are. First, reading $BIAS_1$ through $BIAS_4$ values across any row reiterates the pattern exhibited in figure 4. Bias increases when estimating each additional future unit. This suggests that the further into the future the LC model is used to estimate costs, the greater the underestimation will be.

Second, moving from the bottom, to the middle, to the top panel of the table--from VCRATE 95%, to 85%, to 75%--it is clear that BIAS increases. The general pattern suggested is that as the "true" underlying learning rate (of the portion of total cost subject to learning) increases, the tendency of the LC model to underestimate future cost also increases.

Third, read down any column to observe the pattern of BIAS values as BURDEN increases from 10% to [illegible]0% of total cost. Negative bias consistently increases with increases in fixed cost burden--up to a point--then negative bias decreases with further increases in burden. The turn around point for all observations is when burden is 50%. This confirms the finding from the earlier ANOVA test, that bias increases with burden, but indicates that that pattern holds only when fixed cost is less than half of total cost; the pattern is not universal. This reversal is perhaps understandable. Consider the two extremes. If BURDEN = 0%, then all cost would be variable, all cost would be subject to learning, an LC model would be a correct specification of the "true" underlying cost function, and zero bias would result. If BURDEN = 100% then all cost would be fixed, no cost would be subject to learning, an LC model would again be a correct specification of the "true" underlying cost function (which would be a learning curve with slope of zero--no learning), and zero bias would result. Only when costs--some subject to learning, some not--are combined does the bias result. And the bias is at a maximum when the mixture is about fifty-fifty.

iii) Bias and Estimated LC Slope: Recall that the total cost of any unit produced depends on both VCRATE, which determines the learning experienced by the variable cost portion of total cost, and BURDEN, which determines the magnitude of the fixed cost portion of total cost. Given the findings that BIAS depends on both VCRATE and BURDEN raises an interesting practical question. In many circumstances, cost analysts may not have access to detailed cost data and hence may not "know" the values for VCRATE and BURDEN in a real world cost problem being analyzed. In fact, the point of fitting a learning curve to cost data is typically to arrive at a summary description of an unknown cost function. What is observable by the analyst is an <u>estimated</u> learning curve slope for a given observable total cost series. Is there a relationship between estimated LC slope and BIAS? The nature of that relationship is not obvious. Ceteris paribus, as VCRATE become steeper, estimated LC slope will become steeper as well. Given the tendency of BIAS to vary with VCRATE, this suggests that BIAS will increase as estimated LC gets steeper. But, ceteris paribus, as BURDEN increases, estimated LC slope will become more shallow. Given the tendency of BIAS to first increase, then decrease with increases in BURDEN, the relationship between estimated LC slope and BIAS is ambiguous.

Figure 8 plots BIAS against estimated LC slope (generated for combinations of VCRATE, varied from 70% to 95%; BURDEN, varied from 10% to 80%; DATAHIST = 7). Note that the scatter diagram is not tightly clustered along any trend line. In the most general sense, there is no strong relationship between estimated LC slope and bias. But consider the segment of the plot falling within the boundaries formed by the two dotted lines. These represent the boundaries for BIAS when BURDEN is constrained, in this case, to fall between 30-40%. Given that burden is assumed to vary through only a small range, then there is a strong empirical relationship: steeper estimated LC slopes are associated with a greater tendency toward underestimation of cost.

iv) Bias and Data History: Table 7 explores the impact of DATAHIST on BIAS. Here BIAS is measured for cost forecasts from models estimated on n data points, where n is varied from 4 through 10. For each model, BIAS is measured for n + 1, n + 2, etc. Recall from the earlier ANOVA results that bias increased as

DATAHIST increased. This lead to the somewhat counter-intuitive conclusion that LC models get progressively more biased the more observations there are available on which to fit the model. Two patterns in the table confirm this finding but clarify its implications.

First, observe the BIAS values in the diagonal (top left to bottom right) of the table. BIAS consistently increases. The prediction of, say, the 7th cost in a series using an LC model estimated on the first six costs will be more biased than the prediction of the 6th cost using a model estimated on the first five. Bias in predicting the n + 1 cost does increase with n. (This is the same finding as from the ANOVA.)

But observe also the BIAS values in any column. BIAS consistently decreases as DATAHIST increases. The prediction of, say, the 7th cost using a model estimated on the first six costs is less biased than the prediction of that same 7th cost using a model estimated on only the first five. In short, given a task of forecasting a specific given cost, ceteris paribus, it is always beneficial to use as many data points as are available to estimate the LC model.

SUMMARY AND CONCLUSIONS

The central purpose of this study was to examine bias in estimating future cost from two models commonly used in cost estimation. The analysis simulated prediction for both the traditional learning curve and a rate adjustment model, and

FIGURE 8

PLOT OF BIAS VERSUS
ESTIMATED LC SLOPE

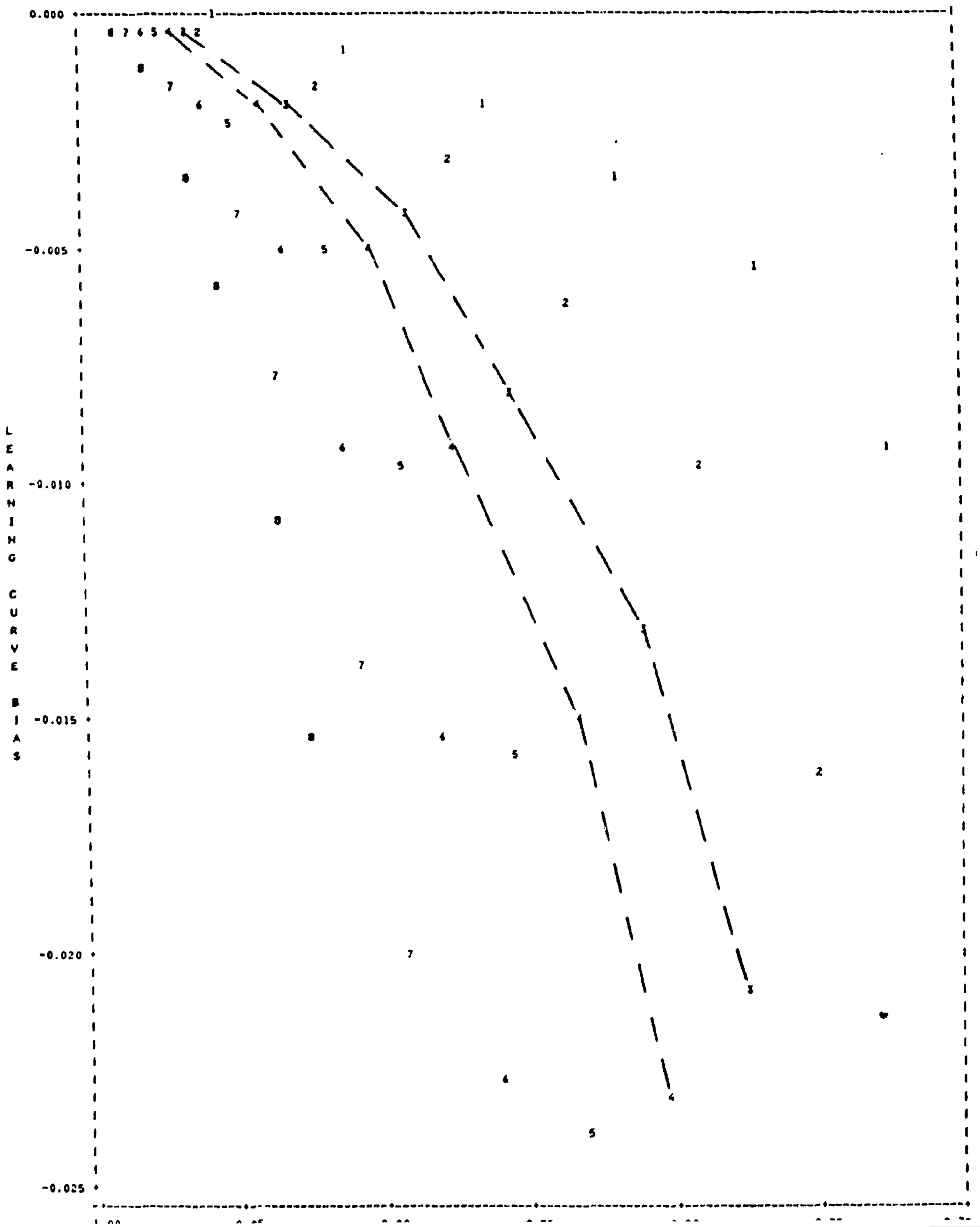


TABLE 7

THE IMPACT OF DATAHIST ON BIAS

DATAHIST	$BIAS_5$	$BIAS_6$	$BIAS_7$	$BIAS_8$	$BIAS_9$	$BIAS_{10}$	$BIAS_{11}$
4	-.0125	-.0204	-.0281	-.0355	-.0425	-.0493	-.0557
5		-.0137	-.0207	-.0275	-.0341	-.0404	-.0464
6			-.0148	-.0211	-.0271	-.0330	-.0387
7				-.0157	-.0214	-.0269	-.0323
8					-.0165	-.0217	-.0268
9						-.0172	-.0220
10							-.0178

NOTE:

VCRATE = 75%
BURDEN = 50%
$BIAS_n$: $BIAS_n$ measures the bias associated with estimating the cost of the nth unit in the cost series.

evaluated bias under varying conditions. The broadest finding was that the rate ljustment model provided cost estimates that were unbiased, while the learning curve model consistently produced estimates that understated actual cost. Most additional findings concerned the conditions related to bias in the learning curve model:

- The cause of the bias is the existence of fixed cost in total cost. The learning curve assumes a log linear relationship between cost and quantity, which does not hold when fixed cost (not subject to learning) is present.

- The bias increases as the proportion of fixed cost in total cost increases--up to the point where fixed cost comprises about 50% of total cost--after that further increases in fixed cost reduce bias. This finding would appear to be relevant given the trend in modern production processes toward increasing automation and hence an increasing fixed component in total cost.

- The degree of bias is affected by the production rate during both the period of model estimation and the period for which costs are forecast. A consistent production rate trend throughout these periods minimizes bias. A shift in production rate trend, particularly to a cutback in volume, magnifies bias. This finding would appear to be relevant to cost estimators analyzing programs where cutbacks are anticipated.

- Assuming the proportional relationship between fixed and variable components of total cost does not vary greatly, bias is greater when the estimated learning curve slope is steeper.

- The bias problem is not diminished as more observations become available to estimate the learning curve. In fact the degree of bias increases as the number of observations increases.

- The degree of bias increases the further into the future predictions are made. Next period cost is somewhat underestimated; cost two periods in the future is underestimated to a greater degree, etc.

Some of the conclusions are a bit ironic. One typically expects to improve forecasting when more data is available for model estimation. The findings here suggest that bias grows worse. One typically expects future costs to decline most rapidly when past costs have exhibited a high rate of learning. The findings here suggests that such circumstances are the ones most likely to result in actual costs higher than forecasted.

Caution should be exercised in drawing direct practice-related implications from these findings. The finding that the rate adjustment model is unbiased while the learning curve is biased does not mean that the rate model should always be preferred to the learning curve model. Bias is only one criteria for evaluating a cost estimation model. Consider accuracy. Evidence indicates that under some circumstances learning curves are more accurate than rate adjustment models (Moses, 1990b). Thus model selection decisions would need to consider (at a minimum) tradeoffs between bias and accuracy. An accurate model with a known bias, which could be adjusted for, would typically be preferable to an inaccurate, unbiased model.

The conclusions of any study must be tempered by any limitations. The most prominent limitation of this study is the use of simulated data. Use of the simulation methodology was justified by the need to create a wide range of treatments and maintain control over extraneous influences. This limitation suggests some directions for future research.

-- Re-analyze the research question while altering aspects of the simulation methodology. For example, are findings sensitive to the cost function assumed?

-- Address the same research question using actual cost and production rate data. Are the same findings evident when using "real-world" data?

Providing confirmation of the findings by tests using alternative approaches would be beneficial.

Additional future research may be directed toward new, but related, research questions.

-- Investigate other competing models or approaches to cost prediction. Perhaps bias can be reduced by using some version of a "moving average" prediction model. Can such a model outperform both the learning curve and the rate adjustment approach? If so, under what circumstances?

-- Investigate tradeoffs between various characteristics of cost estimation models, such as bias versus accuracy.

REFERENCES

1. Alchian, A. (1963), "Reliability of Progress Curves in Airframe Production," Econometrica, Vol. 31, pp. 679-693.

2. Asher, H. (1956), Cost-Quantity Relationships in the Airframe Industry, R-291, RAND Corporation, Santa Monica, CA.

3. Balut, S. (1981), "Redistributing Fixed Overhead Costs," Concepts, Vol. 4, No. 2, pp. 63-72.

4. Balut, S., T. Gulledge, Jr., and N. Womer (1989), "A Method of Repricing Aircraft Procurement," Operations Research, Vol. 37, pp. 255-265.

5. Bemis, J. (1981), "A Model for Examining the Cost Implications of Production Rate," Concepts, Vol. 4, No. 2, pp. 84-94.

6. Boger, D. and S. Liao (1990), "The Effects of Different Production Rate Measures and Cost Structures on Rate Adjustment Models," in W. Greer and D. Nussbaum, editors, Cost Analysis and Estimating Tools and Techniques, Springer-Verlag, New York, 1990, pp. 82-98.

7. Cheney, W. (1977), Strategic Implications of the Experience Curve Effect for Avionics Acquisition by the Department of Defense, Ph. D. Dissertation, Purdue University, West Lafayette, IN.

8. Cochran, E. (1960), "New Concepts of the Learning Curve," Journal of Industrial Engineering," Vol. 11, 317-327.

9. Conway, R. and A. Schultz (1959), "The Manufacturing Progress Function," Journal of Industrial Engineering, 10, pp. 39-53.

10. Cox, L. and J. Gansler (1981), "Evaluating the Impact of Quantity, Rate, and Competition," Concepts, Vol. 4, No. 4, pp. 29-53.

11. Dorsett, J. (1990), "The Impacts of Production Rate on Weapon System Cost," paper presented at the Joint Institute of Cost Analysis/National Estimating Society National Conference, Los Angeles, CA, June 20-22.

12. Gulledge, T. and N. Womer (1986), The Economics of Made-to-Order Production, Springer-Verlag, New York, NY.

13. Hirsch, W. (1952), "Manufacturing Progress Functions," The Review of Economics and Statistics, Vol. 34, pp. 143-155.

14. Large, J., H. Campbell and D. Cates (1976), Parametric Equations for Estimating Aircraft Airframe Costs, R-1693-1-PA&E, RAND Corporation, Santa Monica, CA.

15. Large, J., K. Hoffmayer, and F. Kontrovich (1974), Production Rate and Production Cost, R-1609-PA&E, The RAND Corporation, Santa Monica, CA.

16. Levenson, G., et. al. (1971), Cost Estimating Relationships for Aircraft Airframes, R-761-PR, RAND Corporation, Santa Monica, CA.

17. Liao, S. (1988), "The Learning Curve: Wright's Model vs. Crawford's Model," Issues in Accounting Education, Vol. 3, No. 2, pp. 302-315.

18. Linder, K. and C. Wilbourn (1973), "The Effect of Production Rate on Recurring Missile Costs: A Theoretical Model," Proceedings, Eighth Annual Department of Defense Cost Research Symposium, Airlie VA, compiled by Office of the Comptroller of the Navy, 276-300.

19. McCullough, J. and S. Balut (1986), Defense Contractor Indirect Costs: Trends, 1973-1982, IDA P-1909, Institute for Defense Analysis, Alexandria, VA.

20. Moses, O. (1990a), Extensions to the Learning Curve: An Analysis of Factors Influencing Unit Cost of Weapon Systems, Naval Postgraduate School Technical Report, NPS-54-90-016, Monterey, CA.

21. Moses, O. (1990b), "Learning Curve and Rate Adjustment Models: Comparative Prediction Accuracy under Varying Conditions," Naval Postgraduate School Technical Report, NPS-AS-91-001, Monterey, CA.

22. Pilling, D. (1989), Competition in Defense Procurement, The Brookings Institution, Washington DC, p. 35.

23. Smith, C. (1980), Production Rate and Weapon System Cost: Research Review, Case Studies, and Planning Model, APR080-05, U. S. Army Logistics Management Center, Fort Lee, VA

24. Smith, C. (1981), "Effect of Production Rate on Weapon System Cost," Concepts, Vol. 4, No. 2, pp. 77-83.

25. Smith, L. (1976), An Investigation of Changes in Direct Labor Requirements Resulting From Changes in Airframe Production Rate, Ph. D. dissertation, University of Oregon, Eugene, OR.

26. Smunt, T. (1986), "A Comparison of Learning Curve Analysis and Moving Average Ratio Analysis for Detailed Operational Plan-

ning," Decision Sciences, Vol. 17, No. 4, Fall, pp. 475-494.

27. Washburn, A. (1972), "The Effects of Discounting Profits in the Presence of Learning in the Optimization of Production Rates," AIIE Transactions, 4, pp. 255-313.

28. Wetherill, G. (1986), Regression Analysis with Applications, Chapman and Hall, New York.

29. Womer, N. (1979), "Learning Curves, Production Rate and Program Costs," Management Science, Vol. 25, No. 4, April, pp. 312-319.

30. Wright, T. (1936), "Factors Affecting the Cost of Airplanes," Journal of Aeronautical Sciences, Vol. 3, pp. 122-128.

31. Yelle, L. (1979), "The Learning Curve: Historical Review and Comprehensive Survey," Decisions Sciences, Vol. 10, No. 2, April, pp. 302-328.

Distribution List

Agency	No. of copies
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Dudley Knox Library, Code 52 Naval Postgraduate School Monterey, CA 93943	2
Office of Research Administration Code 012 Naval Postgraduate School Monterey, CA 93943	1
Library, Center for Naval Analyses 4401 Ford Avenue Alexandria, VA 22302-0268	1
Department of Administrative Sciences Library Code AS Naval Postgraduate School Monterey, CA 93943	2
Mr. Michael C. Hammes NAVSEA 017 Department of the Navy Naval Sea Systems Command Washington, DC 20362	4
Professor O. Douglas Moses Code AS/Mo Department of Administrative Sciences Naval Postgraduate School Monterey, CA 93943	30