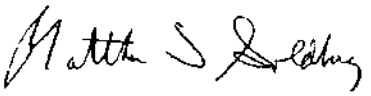
CIM D0006870.A3/1Rev
March 2003

Statistical Methods for Learning Curves and Cost Analysis

Matthew S. Goldberg • Anduin Touw

4825 Mark Center Drive • Alexandria, Virginia 22311-1850

Approved for distribution: March 2003



Matthew S. Goldberg, Director
Cost and Acquisition Team
Resource Analysis Division

This document represents the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the Department of the Navy.

FOREWARD

In 1989, I joined the Cost Analysis and Research Division at the Institute for Defense Analyses (IDA). I spent a very fulfilling 10 years at IDA before moving on to my current position as Director of the Cost and Acquisition Program at the CNA Corporation (the parent organization for the Center for Naval Analyses).

While at IDA, I keenly observed the practice of cost analysis among my colleagues at IDA, as well as at various offices within the Department of Defense and among the contractors who supported them. I was fascinated by statistical techniques, such as lot-midpoint iteration, that were being widely used throughout the military cost-analysis community. Given my own background in econometrics, operations research, and statistics, I sought the theoretical justification for these statistical techniques. Much to my chagrin, I could not find anybody who had asked — much less answered — fundamental questions about the mathematical and statistical properties of lot-midpoint iteration: existence of a solution, uniqueness of the solution, convergence to that solution, and unbiasedness or consistency of the statistical estimators.

It turned out that I was not completely alone in my quest. I experienced something of an epiphany when I encountered David Lee's monograph, *The Cost Analyst's Companion*, which was published by the Logistics Management Institute (LMI) in 1997. Lee addressed many fundamental questions concerning the nature of learning curves, the differences between alternative formulations of the learning curve, and construction of cost-estimating relationships (CERs) using principles from physics and engineering. Lee dealt quite deftly with the *mathematical* underpinnings of learning curves and CERs. Lee also touched on the *statistical* calibration of these models. Although I have great respect for Lee's work, I must say that I found his discussion of the statistical properties of these models far less complete and less satisfying than his discussion of their mathematical properties.

So, late in 1998, I sat down to write David Lee a letter and open up a dialogue on improving the state of statistical practice in military cost analysis. I soon found that my "letter" was chock-full of equations and beginning to look more like a research paper or a journal article. I never did mail the letter, instead redoubling my effort toward writing a research paper for publication in one of the professional journals. By the middle of 1999, I had drafted a 40-page paper on statistical estimation of learning curves and CERs.

Then I experienced a second epiphany — I met Anduin Touw. IDA hired Anduin, a promising young research analyst with a graduate degree in statistics from UCLA and prior work experience at Hughes Space and Communications Corporation. I introduced myself to Anduin, and asked her to peer-review my 40-page paper. Anduin's immediate reaction to my paper was, "If you can't prove all of these results from theory, why not investigate them using Monte Carlo analysis?" Having been hit between the eyes with the obvious, I invited Anduin to actually perform the Monte Carlo analysis, and join me as a co-author on what was now looking more and more like a book. Stephen Balut, Director of IDA's Cost Analysis and Research Division, generously arranged for financial support, and off we went.

Many friends and colleagues read portions of this work or otherwise educated me at various points along the way: Robert Book, Stephen Book, Jino Choi, Henry Eskew, Bruce Harmon, David Hunter, Ted Jaditz, and Philip Lurie. Two colleagues reviewed an early, near-complete draft of this book: Vadim Kutsyy, who had recently completed a Ph.D. in Statistics from the University of Michigan; and Robert Trost, Professor of Economics (and senior econometrician) at George Washington University. Linda Garlet provided editorial assistance on the complete draft that we first submitted for publication.

Anduin and I presented our preliminary findings during two seminars at George Mason University. We extend our thanks to participants in the Statistics Seminar (organized by James Gentle), as well as the Operations Research Seminar (particularly Andrew Loerch and Roman Polyak).

As previously mentioned, financial support was provided by my former Division Director at IDA, Stephen Balut, who also rekindled my long-standing interest in operations research. Financial support was sustained by my current Division Director at the CNA Corporation, Samuel Kleinman, along with the CNA Corporation's Senior Vice-President and Director of Research, David Kelsey.

The idea of publishing the manuscript in the INFORMS *Topics in Operations Research* series was first broached by Thomas Frazier of IDA, who was then the series editor. During the lengthy process of completing the manuscript, the editor's job rotated to Professor Keith Womer of the University of Mississippi, himself one of the leaders in the field of cost analysis. Keith shares my interest in the nexus of econometrics, operations research, and statistics; his review comments and shepherding of the project have been invaluable.

The views that Anduin and I express in this book are solely our own; they do not represent official positions of the Institute for Defense Analyses, the CNA Corporation, the Department of the Navy, the Department of Defense, Hughes Space and Communications Corporation, or Boeing Corporation. Indeed, we persist in some of our views over the objections of a few of our aforementioned colleagues.

Finally, I must thank a pair of feline companions, first Snowy and now Murphy, for keeping my lap warm during many laborious hours at the computer.

M.S.G.
Alexandria, Virginia
March 2003

For me this project began as a result of my sometimes beneficial and sometimes tragic habit of leaping into projects before I fully know the scope or even whether I am welcome. Luckily for me, I found a supportive leader, and although the project grew beyond our initial expectations, it has been well worth the effort. I believe that it has been a great example of why statisticians should venture out to explore fields in which statistics and mathematical models are used, but statisticians are not commonplace. And of why statisticians, not just statistical software, are needed on projects.

I would like to thank Matt for the opportunity to work on such an interesting and fundamental project in cost analysis. I also thank my husband, Brian Jackson, for his support and understanding during my foray into this field. I will always be in debt to Dr. Tony Lin for his advice and insight on this project, on Monte Carlo analysis, and on statistics in general. I would also like to express my appreciation to Dr. Lynne Butler, who guided me into this career, and to my parents, who have always encouraged me to seek out unusual perspectives on and applications of mathematics.

A.E.T.
El Segundo, California
March 2003

This page intentianally left blank.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

In this chapter, we first discuss statistical methods for estimating "cost progress" or "learning." We use these two terms interchangeably to describe a reduction in unit production cost as more items have been cumulatively produced over the course of a manufacturing program. Some older works defined the term "learning" in a much narrower sense, to encompass only the reduction in manufacturing labor hours as workers learn to perform repetitive tasks faster or with fewer errors. Most modern authors have expanded the concept of "learning" to include redesign of the production process itself, perhaps changing the tasks that workers perform or complementing those workers with improved automation. In addition, as a production program unfolds, manufacturers may find cheaper suppliers, or enter into long-term contracts under which they enjoy quantity discounts from suppliers. We retain the older term "learning" without much concern for whether the *source* of the unit cost reduction is confined to production workers performing repetitive tasks, or extends to some other economic or technological factors. We also use the term "learning curve" to describe the mathematical relationship between unit production cost and the cumulative quantity produced.[1]

Next, we turn our attention from the learning curve to the cost-estimating relationship (CER), a regression equation to predict the development or production cost of a system based on performance and technical characteristics such as weight, speed, and composite materials content. We define a class of statistical models known as multiplicative regression models. Many CERs, as well as a particular representation of the learning curve, fall into this class of models. We discuss two specialized statistical techniques for calibrating learning curves. We also discuss several general-purpose statistical techniques that apply to all multiplicative regression models, including CERs as well as learning curves.

We attempt to keep the level of mathematics to a minimum throughout this introductory chapter. Only a few of the equations we display should appear difficult to most readers, and these few we have simply copied into the current chapter without a full

---

[1] The various definitions of "learning" are surveyed in Yelle (1979) and Dutton, Thomas, and Butler (1984). The seminal papers are Asher (1956) and Conway and Schultz (1959).

derivation. In subsequent chapters, we provide the derivations of the few difficult equations. Our intention is for the reader to grasp the major content of our work from the current chapter, and defer the more difficult mathematics until later.

1.1 Data on production lots

Large hardware items are often purchased not as individual units, but rather as *lots*. For example, the U.S. Navy might sign a contract to purchase 1,200 tactical missiles, to be delivered 100 per month over a period of one year. Or, the U.S. Air Force might sign a contract to purchase 36 fighter aircraft to be delivered 3 per month over a period of one year. Further, a production run often extends over several consecutive years. So, expanding on the second example, the U.S. Air Force might purchase 12 aircraft during the first year, 24 aircraft during the second year, 36 aircraft per year during several years of peak production, and finally 12 aircraft during the final year of production. In each of these cases, the units comprising a single year's purchase are considered as one lot.

Two of the fundamental defining features of a lot are the number of units that comprise the lot and the total price of the lot. A perhaps surprising aspect of large hardware purchases is that individual units within each lot are not separately priced. This point may, at first, seem trivial or even wrong. If the initial lot consisting of 12 units of some hardware item costs a total of \$18.75 million, isn't the average cost simply \$18.75 ÷ 12 or \$1.56 million per unit?

The average cost per unit can always be computed by simple division, *for a fixed number of units*. That is not to say, however, that all of the units in the lot are equally costly. If the buyer were to renegotiate the number of units in the lot (the "lot size"), the seller would likely adjust the total lot cost in a non-proportional manner; i.e., adjust the average cost. Figure 1.1 illustrates the situation. The 12 units within the lot exhibit a trend of decreasing unit cost due to learning. Although the average cost of all 12 units is indeed \$1.56 million, the average cost of only the first 8 units is higher — \$1.64 million. Thus, a renegotiation that reduced the lot size from 12 units to only 8 units would yield an increase in the average cost.

The challenge for the data analyst is to deduce the trend in learning, given only data on lot size and total lot cost, but *not* the cost of individual units. One possibility would be to ask the seller for alternative price quotes corresponding to various lot sizes. Abstracting from profit margins that drive a wedge between cost and price, one could

attempt to estimate the learning curve from the knowledge that an 8–unit lot costs $1.64 million per unit, whereas a 12–unit lot costs only $1.56 million per unit.

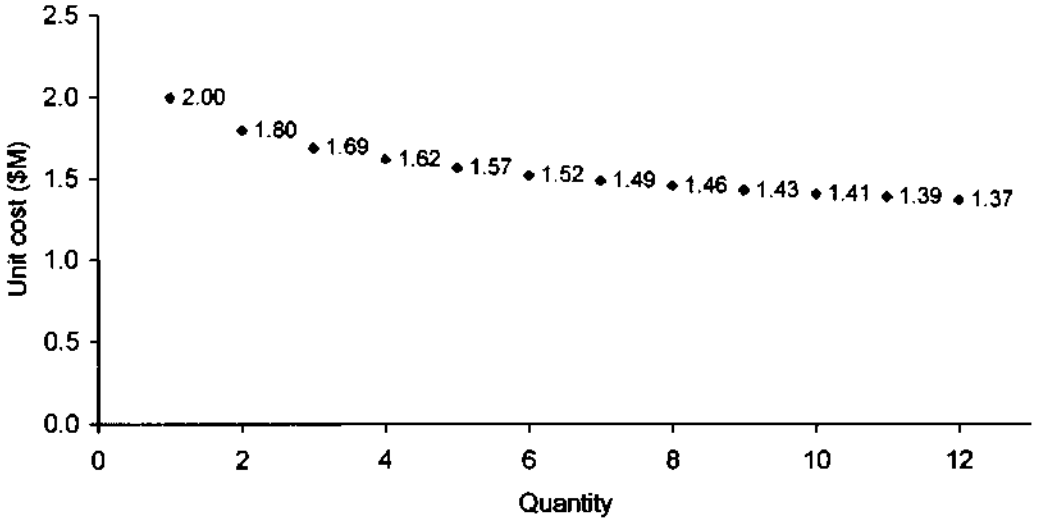


Figure 1.1. Unit Costs within the First Lot

This approach is generally unsatisfactory because the historical data do not always contain price quotes corresponding to alternative lot sizes. Instead, the buyer may ask the seller for a single price quote corresponding to a single delivery schedule. In effect, there would be only one data point, precluding estimation of the learning curve. In addition, even when multiple price quotes are available, they are only hypothetical and do not represent the actual or historical costs of production. Finally, multiple price quotes would confound pure learning with the so-called rate effect that arises from bunching various numbers of units in a single year; we discuss the rate effect in a later section.

A better approach uses only the actual data from production programs, as opposed to hypothetical price quotes. The key is simply to compare the average costs of successive lots from the same production program. For the remainder of this monograph, we use as our primary example the time-series data originally reported by Lee (1997) on lot sizes and lot costs for a tactical missile program. We reproduce those data here as Table 1.1.

The second and third columns of Table 1.1 give the unit numbers for each lot in the production sequence. For example, the initial lot runs from unit #1 to unit #218, and contains 218 units; the second lot runs from unit #219 to unit #1,158, and contains 940 units; and so on. The *incremental lot cost* is the cost of a particular lot, *not* the cumulative

cost of the entire production program. Finally, the lot average cost is computed as the ratio of the incremental lot cost and the lot size. Interestingly, unlike in the notional data shown previously in Figure 1.1, most of the learning in the actual data for the tactical missile program occurs between the first and third lots. Moreover, lot average cost actually increases slightly from lot #4 to lot #5 and again from lot #6 to lot #7. Later in this chapter, and again in Chapter 4, we discuss the fit of the smooth learning curve to these data.

Table 1.1. Data for a Tactical Missile Program

Lot number	Lot start	Lot end	Lot size	Incremental lot cost ($M)	Lot average cost ($M)
1	1	218	218	102.765	0.471
2	219	1,158	940	212.158	0.226
3	1,159	3,200	2,042	321.819	0.158
4	3,201	5,900	2,700	333.720	0.124
5	5,901	7,591	1,691	212.558	0.126
6	7,592	10,011	2,420	227.238	0.094
7	10,012	11,668	1,657	157.912	0.095
8	11,669	14,436	2,768	171.339	0.062

Source: Lee (1997), p. 50. Although he leaves the matter ambiguous, we presume that the final two columns are measured in millions of dollars (e.g., by the eighth lot, the average cost of a missile has fallen to $62,000 in some base year's dollars).

Unlike tactical missiles, military aircraft typically follow a 3-year production cycle. As illustrated in Figure 1.2, a contract that delivers aircraft within a particular fiscal year may involve costs during the two previous fiscal years as well. Conversely, the costs incurred in a particular fiscal year may be attributable to as many as three distinct aircraft lots. When dealing with multi-year production cycles, we interpret the incremental lot cost as the sum across fiscal years of all the costs attributable to a particular lot. This interpretation necessarily involves an allocation of plant-wide overhead costs among the various lots in progress during a particular fiscal year (as well as overhead allocations to other systems — presumably other aircraft models — being produced concurrently in the same plant). In Figure 1.2, we would horizontally (rather than vertically) sum the costs attributable to a particular lot.[2]

---

[2] Balut, Gulledge, and Womer (1989) thoroughly discuss the costs associated with multi-year production, including the allocation of overhead costs across concurrent programs. Womer (1984) describes the biases from ignoring multi-period production (actually, using monthly rather than annual data)

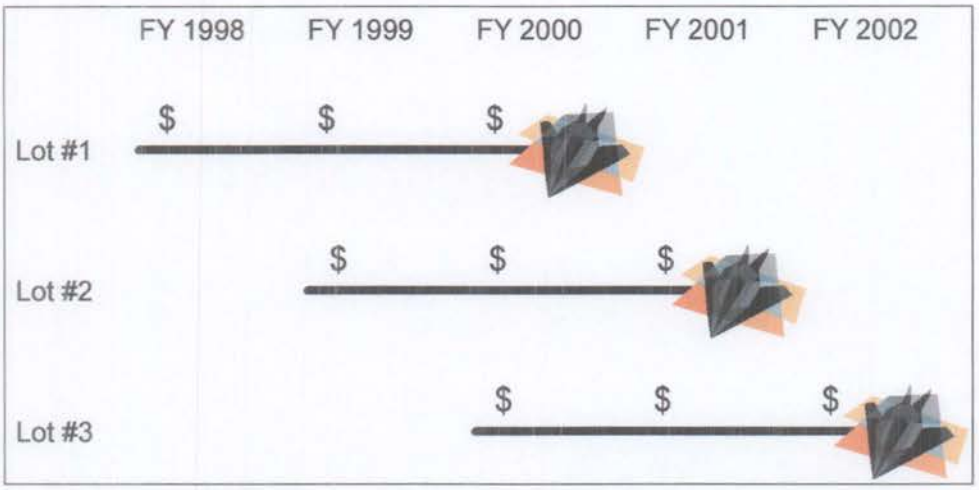


Figure 1.2. Distinction between Fiscal-Year Costs and Lot Costs

Much of this monograph is devoted to estimating the trend in learning from data on lot sizes and lot costs. One of the statistical methods we develop, when applied to Lee's data, results in the learning curve previewed here as Figure 1.3. The height of each data point represents the average cost of the lot. The horizontal coordinates are the "lot midpoints," a concept we discuss in the next section. The figure shows the fitted learning curve, as well as the ± 2 standard deviation ("sigma") confidence band around the learning curve. The formula for the confidence band is not widely known and is seldom used by cost analysts. We develop this formula in Chapter 2.

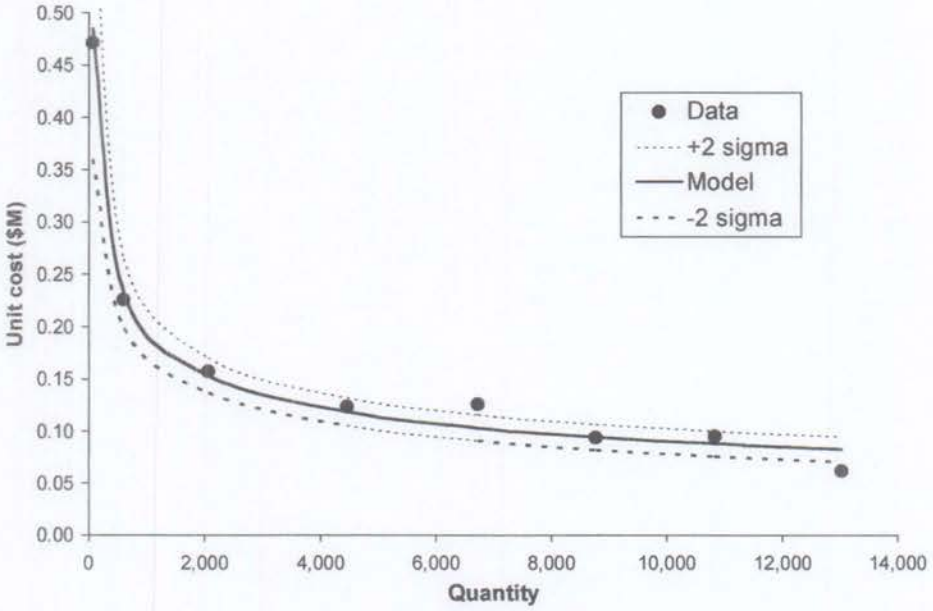


Figure 1.3. Learning Curve Fit to Tactical Missile Data

A few cautions are in order before proceeding further. In defense procurement, the actual execution of a production program almost always deviates from the initial delivery schedule. Increases in total quantity (or accelerated delivery of a fixed total quantity) could result in retooling, capacity expansion, and overtime labor costs, possibly offset by reduced overhead burdening. Decreases in total quantity could result in penalty clauses, severance and shutdown costs, and increased overhead burdening. In addition, technical upgrades (e.g., enhanced aircraft radars) during the course of a production program may make it difficult to compare the later units with the earlier units, unless some adjustment is made. We ignore these complications and assume that, notwithstanding any trend in lot average costs, the items produced are all observationally equivalent from the final customer's (e.g., the aircraft squadron's) perspective. Stated more directly, our notion of learning is the manufacturer's ability to produce successive, observationally equivalent units at declining unit cost. We refer the reader to the published literature for a discussion of adjustments for quantity deviations, technical upgrades, and so on.[3]

1.2 A learning-curve model

Let Q denote the sequence number of a particular unit in the production run. The learning curve is most often specified so that the cost of unit Q — the marginal cost — is a power function of Q:

$$MC(Q) = T_1 Q^b \tag{1.1}$$

for $Q > 0$, where $T_1 > 0$ and b are parameters to be estimated.

The "learning slope" is defined as the ratio of marginal costs between unit $2Q$ and unit Q:

$$\rho = MC(2Q)/MC(Q) = 2^b. \tag{1.2}$$

Marginal cost is presumed to decline with increasing quantity. However, as we argue in Chapter 2, it is implausible that marginal cost would decline by as much as 50% when quantity doubles. The plausible range of $\frac{1}{2} < \rho \leq 1$ for the learning slope translates into a corresponding range $-1 < b \leq 0$. For example, with $\rho = 0.9$ or a 90% learning slope

---

[3] A good recent example of this literature is Harmon, Touw, and Woolsey (2000).

(which equates to $b = -0.152$), the second unit costs only 90% as much as the first unit; the fourth unit costs 90% as much as the second unit, or 81% as much as the first unit; and so on.

If we take the logarithms on both sides of equation (1.1), we appear to have a model that can be estimated using ordinary least squares (OLS):

$$\ln\left[MC(Q)\right] = \ln T_1 + b \times \ln Q, \tag{1.3}$$

where "ln" denotes the natural logarithm. The difficulty, however, is that we are given only data on lot size and total lot cost, *not* the cost of individual units. Thus, the left-hand side of equation (1.3) cannot be computed for the individual units $Q = 1, 2, 3, \ldots$.

Instead, the most common solution is to find a "typical unit" within each lot, use the sequence number of that unit in place of Q on the right-hand side of equation (1.3), and use the *average cost* of the entire lot in place of $MC(Q)$ on the left-hand side. The lot average cost is computed simply as the ratio of total lot cost and lot size, both of which are observable. The "typical unit" is traditionally called the "lot midpoint." The regression analysis is then conducted on the lot midpoints (one per lot) rather than on the individual units. Letting $\bar{Q}_i$ denote the midpoint of the i^{th} lot and LAC_i the lot average cost, OLS is actually applied to the following model:

$$\ln(LAC_i) = \ln T_1 + b \times \ln \bar{Q}_i(b), \tag{1.4}$$

for lots $i = 1, \ldots, n$. We explicitly write the lot midpoint as $\bar{Q}_i(b)$, a function of the exponent b. We do so because, as we will see in a moment, the lot midpoint cannot be computed without knowledge (or at least an estimate) of the exponent b (or the corresponding learning slope).

The following simple example illustrates the calculation of lot midpoints.[4] Consider a production process with $b = -0.152$ (or $\rho = 0.9$) and $T_1 = 2.0$. Suppose the initial lot consists of two units. The first unit costs $T_1 = \$2.00$, and the second unit costs $2.0 \times 2^{-0.152} = \$1.80$ (with a 90% learning slope, the second unit costs only 90% as much as the first unit). The average cost of the entire lot is \$1.90, the average of \$2.00 and

---

[4] This example is adapted from Eskew (2000). This is not exactly the conventional lot-midpoint calculation, but it serves to illustrate the concept using a minimum of mathematics. We report the conventional lot-midpoint calculation later in this chapter, and more fully derive and critique it in Chapter 2.

$1.80. The lot midpoint is defined as the (generally non-integer) quantity whose marginal cost (left-hand side of the following equation) is equal to the lot average cost (right-hand side):

$$2.0 \times \bar{Q}_1^{-0.152} = \$1.90. \tag{1.5}$$

The solution to this equation is $\bar{Q}_1 = 1.40$. Thus, in the regression analysis on the lot midpoints, the midpoint for the first lot would be "unit" 1.40. The midpoints of the later lots would be computed in a similar fashion.[5]

Note, however, the circularity in this procedure. We had to *assume* a 90% learning slope in order to compute the lot midpoint. But if we already knew the learning slope, we would not have to proceed with the regression analysis. On the other hand, if we did *not* already know the learning slope (or the corresponding exponent; in this case, $b = -0.152$), how would we apply equation (1.5) to compute the lot midpoints?

One solution is to iterate: start with an initial guess of the learning slope, and then alternate between the two steps of estimating the regression exponent in equation (1.4) and updating the lot midpoints in equation (1.5). The iteration ends when (and if) two successive iterations yield the same value for the regression exponent, within a pre-specified numerical tolerance. We refer to this procedure as *lot-midpoint iteration.*

1.3 Estimation of the continuous learning-curve model

In Chapter 2, we review the theory of continuous learning curves, which leads to the following expression for the lot average cost:

$$LAC_i = \frac{TC_i - TC_{i-1}}{Q_i - Q_{i-1}} \approx \frac{T_1}{(1+b) \times (Q_i - Q_{i-1})} \times \left[(Q_i + 0.5)^{1+b} - (Q_{i-1} + 0.5)^{1+b} \right]. \tag{1.6}$$

In this notation, the i^{th} lot runs through unit Q_i. Similarly, the preceding lot $(i-1)$ ran through unit Q_{i-1}. Thus, the i^{th} lot begins with unit $Q_{i-1}+1$ (the unit *after* the one that completed the preceding lot) and runs through unit Q_i.[6] The variable TC_i is the

---

5 The preceding calculations are based on a discrete learning curve. In Chapter 2, we develop the more common, continuous approximation to the learning curve. In contrast to the midpoint of $\bar{Q}_1 = 1.40$ we just computed for the initial lot, the continuous approximation yields a slightly smaller midpoint of $\bar{Q}_1 = 1.36$.

6 For example, in the data of Table 1.1, the second lot contains units $Q_1 + 1 = 219$ through $Q_2 = 1{,}158$.

cumulative total cost of the production program through the i^{th} lot, so that $TC_i - TC_{i-1}$ represents the incremental cost of the i^{th} lot. Finally, $Q_i - Q_{i-1}$ represents the lot size, and lot average cost is computed as the ratio of incremental lot cost and lot size.

The midpoint of the i^{th} lot, $\overline{Q}_i(b)$, is defined as follows:

$$\overline{Q}_i(b) = \left(\frac{\left[(Q_i + 0.5)^{1+b} - (Q_{i-1} + 0.5)^{1+b} \right]}{(1+b) \times (Q_i - Q_{i-1})} \right)^{1/b}, \tag{1.7}$$

for $-1 < b < 0$. A comparison of equations (1.6) and (1.7) shows that the marginal cost of the lot midpoint is equal to the lot average cost, $LAC_i = T_1 \times [\overline{Q}_i(b)]^b$. Taking logarithms we recover equation (1.4), $\ln(LAC_i) = \ln T_1 + b \times \ln \overline{Q}_i(b)$.

Iterative estimation of equation (1.4), or lot-midpoint iteration, has been the norm in cost analysis for nearly half a century, since the days of Asher (1956). This practice was necessitated by the lack of either computer hardware or software capable of estimating non-linear least squares (NLS), as opposed to OLS regression. The definition of lot midpoint, along with the logarithmic transformation, resulted in equation (1.4) which looks tantalizingly close to OLS regression. In fact, given the technology of the day, equation (1.4) could be estimated only by alternating between the two steps of OLS regression and updating the lot midpoints.

We consider it extremely unlikely that a modern statistician, confronted with this problem, would advocate lot-midpoint iteration. If one insisted on retaining the artifice of lot midpoints, then equation (1.4) could be estimated in a single step using NLS. The right-hand side of equation (1.4) is a non-linear function of the exponent b, which pre-multiplies the lot midpoint and, from equation (1.7), is also embedded within the definition of the lot midpoint. Despite the two roles that b plays on the right-hand side, an estimate of b is still readily available. Simply choose b to minimize the sum-of-squared errors between the (non-linear) right-hand predictor and the actual values of the logarithmic lot average cost:

$$\sum_{i=1}^{n} \left(\ln(LAC_i) - \ln(T_1) - b \times \ln[\overline{Q}_i(b)] \right)^2, \tag{1.8}$$

where n is the number of lots in the data sample and $\overline{Q}_i(b)$ is the lot midpoint as given previously in equation (1.7).[7]

Statistical software to minimize expression (1.8) is widely available. The statistical properties of this problem, such as regression standard errors, confidence intervals, and significance tests, are well known. The convergence properties of various algorithms (such as Gauss-Newton) for locating the minimum are equally well known.[8] To clarify our earlier statement, the minimization algorithms require only a "single step" in the sense that the user need only specify the right-hand predictor (the right-hand side of equation (1.4)) *once*, as a parametric function of the unknown values T_1 and b. This situation contrasts with lot-midpoint iteration, during which the user must manually update all n lot midpoints from equation (1.7) *at every iteration*.

As yet another alternative, one could jettison entirely the artifice of lot midpoints, and simply treat equation (1.6) as a non-linear predictor of the lot average cost (*not* its logarithm). An estimate of b is available by minimizing the sum-of-squared errors between the right-hand predictor and the actual values of the lot average cost:

$$\sum_{i=1}^{n}\left(LAC_i \; - \; \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})}\times\left[(Q_i+0.5)^{1+b} \; - \; (Q_{i-1}+0.5)^{1+b}\right]\right)^2 . \quad (1.9)$$

It turns out, from the definition of lot midpoints, that expression (1.9) is equivalent to:

$$\sum_{i=1}^{n}\left(LAC_i \; - \; T_1\times[\overline{Q}_i(b)]^b\right)^2 . \quad (1.10)$$

Thus, in one sense, expression (1.8) represents NLS applied to lot-midpoint data after a logarithmic transformation, whereas expression (1.10) merely omits the logarithmic transformation. However, we can equally well arrive at expression (1.9) without ever considering or even being aware of the notion of lot midpoints. We postulate that our

---

7 Lee (1997, p. 56, equation 79) contemplates exactly this minimization problem. However, rather than advocating direct (albeit non-linear) minimization via NLS, Lee veers into a discussion of lot-midpoint iteration.

8 Two large treatises, concentrating on the statistical properties of NLS, appeared in the late 1980s: Gallant (1987) and Seber and Wild (1989). The algorithmic convergence properties of NLS are discussed in Dennis and Schnabel (1996), a reprint of an earlier monograph first published in 1983. An even earlier book by Bard (1974) quite thoroughly addressed both the statistical and algorithmic convergence properties of NLS.

modern statistician, upon viewing the model for lot average cost in expression (1.6), would immediately jump to expression (1.9) and apply NLS. The right-hand predictor, as we show in Chapter 2, is just the area under the continuous approximation to the learning curve, divided by the size of the i^{th} lot. The notion of lot midpoints is completely superfluous to this development. Our statistician might gravitate toward expression (1.8) only if the error terms appeared ill-behaved, and a logarithmic transformation was applied in an attempt to restore a normal error distribution or to stabilize the variance. However, we demonstrate several other estimation methods in Chapter 3 that can be used to restore normality or stabilize the variance, again without the artifice of lot midpoints.

1.4 What's wrong with lot-midpoint iteration?

Although NLS estimation of expression (1.9) may seem compelling, what harm is done by continuing to apply lot-midpoint iteration, as remains the norm in cost analysis? The harm is that neither the mathematical nor the statistical properties of lot-midpoint estimation are known. Indeed, a major motivation of the current research was to ascertain these previously unexplored properties.

In Chapter 2 we attempt to answer the following seven questions regarding lot-midpoint estimation:

1. Is lot-midpoint iteration equivalent to (i.e., does it yield the same point estimates as) NLS?
2. Is there a distributional assumption under which lot-midpoint iteration is equivalent to maximum-likelihood estimation (MLE[9])?
3. Does lot-midpoint iteration maximize or minimize *any* continuously differentiable function of the parameters T_1 and b (if not a sum-of-squares or a likelihood function, perhaps some other function)?
4. Is lot-midpoint iteration guaranteed to converge, or might the iteration continue forever?
5. If lot-midpoint iteration does converge, is the solution unique; or might the iteration converge to two (or more) distinct solutions depending upon the starting values?

---

[9] MLE is probably the most widely used estimation technique in all of statistics. For example, under the appropriate distributional assumptions, the use of sample moments (means, variances, and so on) to estimate their population counterparts is equivalent to MLE. Similarly, least-squares regression methods are often equivalent to MLE. In Chapter 3, we discuss MLE in the context of learning curves and CERs.

6. If a particular lot-midpoint iteration has two distinct solutions, on what basis do we choose one over the other?
7. If lot-midpoint iteration does converge, how accurate are the standard errors from the final regression step?

We consider it quite remarkable that cost analysts have blithely applied lot-midpoint iteration for nearly half a century without the answers to or, to our knowledge, even having asked these questions. Moreover, both the mathematical and statistical properties of NLS have been well established and disseminated at least since the publication of Bard (1974). Computer hardware and software capable of estimating NLS may have been scarce back in 1974, but they have been widely available and reliable for easily the past 15 years and arguably the past 20 years.

We found it surprisingly difficult to answer the seven questions regarding lot-midpoint iteration. However, we were able to establish the following theoretical properties:

- Lot-midpoint iteration is *not* equivalent to either NLS or MLE.
- Lot-midpoint iteration does not maximize or minimize *any* continuously differentiable function of the parameters T_1 and b.
- There is no universal guarantee that a solution pair T_1 and b exists to balance equation (1.4); that a solution, if it exists, is unique; or that a solution can be approximated by a finite number of steps of lot-midpoint iteration. The standard sufficient conditions that guarantee existence, uniqueness, and convergence may or may not hold for the lot-midpoint problem.
- Lot-midpoint iteration may still converge, despite the failure of the standard sufficient conditions, because these conditions are not actually necessary.
- In a maximization problem, we can always compare the value of the objective function at two distinct local maxima, disposing of the smaller value because it cannot be the global maximum. But because lot-midpoint iteration does not maximize any continuously differentiable objective function, we have no basis to choose between two distinct solutions.

Our theoretical analysis of lot-midpoint iteration does not provide a compelling motivation to use that technique. Although we were unable to develop any theoretical guarantee that lot-midpoint iteration converges, our Monte Carlo analysis in Chapter 5 suggests that it does converge. Nor have we encountered multiple solutions in practice, at least when using reasonable starting values. However, we are still reluctant to endorse an estimation technique whose theoretical properties remain largely unknown.

1.5 Multiplicative regression models

We are generally accustomed to regression models in which the stochastic error term is *additive* to the model prediction:

$$y_i = f(x_i, \beta) + u_i, \tag{1.11}$$

where y_i is the observed response variable, x_i is an observed vector of k predictor variables, β is a vector of m coefficients to be estimated, and u_i is the unobserved error term for the i^{th} observation. The regression model is linear (as distinct from additive) in the special case where $f(x_i, \beta) = \sum_j x_{ij}\beta_j$. The error terms are often assumed to be statistically independent with zero mean and finite variance, and are often further assumed to be normally distributed. However, none of these properties (including linearity) are essential to the definition of an additive regression model.

Even assuming that $f(x_i, \beta)$ is the correct model, we face two types of errors in attempting to predict the value of y_i for a new observation outside the original estimation sample (e.g., the cost of a new weapon system). First, we have only an estimate of β and not its true value. Second, the actual value of y_i will deviate from the model prediction $f(x_i, \beta)$ in light of the error term, u_i. Because the error term is additive to the model prediction, we may state roughly that the predictions of the additive regression model are accurate within $\pm x$ units (e.g., dollars).

By contrast, a multiplicative regression model has the form:

$$y_i = f(x_i, \beta) \times u_i, \tag{1.12}$$

where now u_i has mean 1.0. Once again, the assumptions of linearity, finite variance, and normal distribution are common, but not essential to the definition of a multiplicative regression model.[10] Because the error term is multiplicative to the model prediction, we may state roughly that the predictions of multiplicative regression model are accurate

---

[10] Lee (1997, p. 55–56) assumes that the error term u_i is normally distributed. A more common distributional assumption for the multiplicative regression model replaces the factor u_i with $\exp(v_i)$, where v_i is normally distributed. We contrast these two assumptions in Chapter 2. There we demonstrate that the two assumptions are nearly equivalent when the variance of the random error term is small. However, Lee's assumption is, strictly speaking, incompatible with certain estimation methods that are available under the alternative assumption.

within $\pm x\%$. For example, a random draw of $u_i = 1.30$ implies that $y_i = 1.30 \times f(x_i, \beta)$. The actual response variable is 30% larger than the model prediction; equivalently, the model underpredicts the actual response by 23% $\left[f(x_i, \beta) = 0.77 \times y_i \right]$. When comparing the model prediction to the actual response, it may seem more natural to treat the actual response as "truth" and, therefore, the base of the percentage difference. However, we shall more often treat the model prediction as the base of the percentage difference, *viz.*,

$$\frac{y_i - f(x_i, \beta)}{f(x_i, \beta)} = \frac{\left[1.30 \times f(x_i, \beta)\right] - f(x_i, \beta)}{f(x_i, \beta)} = 0.30 \ . \tag{1.13}$$

We follow this approach because two of the estimation methods that we explore — maximum likelihood and minimum percentage error — involve (at least approximately) minimizing the sum over all the observations of the squares of the percentage differences as defined in expression (1.13).

We have described three possible regression models for estimating the trend in learning: lot-midpoint iteration in equation (1.4), NLS applied to lot-midpoint data after a logarithmic transformation in expression (1.8), and NLS applied to lot-midpoint data without the logarithmic transformation in expression (1.10). In none of these instances have we specified the form in which the stochastic error enters the model. In principle, one could append additive, multiplicative, or even some other type of error term to any of the three regression models, yielding a multitude of possibilities.

An appealing specification would be to append a multiplicative error term to the predictor of lot average cost:

$$LAC_i = T_1 \times [\bar{Q}_i(b)]^b \times u_i \ . \tag{1.14}$$

Taking logarithms, we can transform this multiplicative model into an additive model for the logarithm of lot average cost:

$$\ln(LAC_i) = \ln(T_1) + b \ln[\bar{Q}_i(b)] + \ln(u_i) \ , \tag{1.15}$$

where $\ln(u_i)$ represents the additive error term.

The logarithmic transformation is tempting because, holding $\bar{Q}_i(b)$ constant during the regression step of lot-midpoint iteration, equation (1.15) is linear-in-parameters and thereby amenable to OLS. In order that the usual confidence intervals and

significance tests for OLS be exact in small samples, we require the additional assumption that the transformed error term, $\ln(u_i)$, is normally distributed. This will be the case only if the original error term, u_i, is log-normally distributed. Data analysts are often too quick to transform equation (1.14) into equation (1.15), obtaining a linear-in-parameters model but not checking whether the transformed error term $\ln(u_i)$ is indeed normally distributed. Only through serendipity does a single transformation both linearize a model and restore a normal error distribution. We note, however, that even if the transformed error term is non-normal, the usual confidence intervals and significance tests for OLS may still be valid in large samples.[11]

1.6 Rate effects in learning-curve models

Learning curves are sometimes augmented to include the rate of production in the current period in addition to the cumulative number of units produced (as typically measured by the unit number of the lot midpoint). The theory is that, learning notwithstanding, increases in the current rate of production could entail overtime labor costs, might drive up the short-run price of materials, or might increase the failure rate of manufacturing equipment.

In the multiplicative representation of lot average cost, the augmented learning-curve model would appear as follows:

$$LAC_i = T_1 \times [\bar{Q}_i(b_1)]^{b_1} \times \text{Rate}_i^{b_2} \times u_i \ . \tag{1.16}$$

Difficulties arise in attempting to measure the production rate. When using annual data aggregated to the entire system level (as in the U.S. Department of Defense's Selected Acquisition Reports (DoD SARs)), practitioners often equate production rate with the current lot size, $\text{Rate}_i = Q_i - Q_{i-1}$. For example, in the tactical missile data of Table 1.1, the column labeled "Lot size" might be used as a proxy for production rate.

In practice, the introduction of production rate has met with mixed success; see the discussion in Chapter 3 of Lee (1997). Many have argued that attempts to include the

---

[11] Schmidt (1976, pp. 55–64) showed that the OLS confidence intervals and significance tests are valid asymptotically if the error terms are independently and identically distributed, with finite variance (constant across all of the observations), and if certain other technical conditions hold; he does not require a normal distribution. White (1980) extended this result by deriving adjusted standard errors that yield asymptotic confidence intervals and significance tests under non-constant variance or heteroscedasticity.

production rate are doomed to failure because the current lot size is mechanically positively correlated (collinear) with cumulative quantity. For example, Large et al. (1974) state:

> In general, however, we must conclude that for predicting the overall effect of production rate on aircraft cost, generalized estimating equations [*i.e., including current lot size*] based on statistical analyses of our sample of military aircraft would be too unreliable to be useful.

Although we are not necessarily advocates of including the current lot size in the model, and although it *may* fail due to collinearity in particular instances, it is *not* mechanically correlated with cumulative quantity. Those who claim mechanical correlation are confusing the level of a time series with its rate of change. While the two concepts are clearly mathematically related, they are not *linearly* related, and correlation is a measure of linear association.

This confusion is compounded by the common practice of equating the current lot size with the theoretical production rate. Recalling Figure 1.2, when dealing with multi-year production cycles, several lots may be in progress concurrently at the same plant. The question arises of what exactly we are attempting to measure with production rate. If we believe that costs are driven by *all* activity in a plant, then we would vertically sum the number of units across all lots in progress during each fiscal year. In Figure 1.2, we would measure production rate in fiscal year 2000 as the sum of the quantities ordered (in DoD parlance, "authorized") in fiscal years 1998 (these units would be in their third and final year of production by 2000), 1999 (units in their second year of production), and 2000 (units in their initial year of production).

On the other hand, during the notional 3-year production cycle for military aircraft, a large portion of the elapsed time involves manufacturing sub-systems at subcontractors' plants. Final assembly at the prime contractor's plant may all occur during the final year of the production cycle. Activities that precede final assembly may be incidental to the prime contractor's plant, and might not drive overtime labor costs or failure rates of manufacturing equipment (at least, not at the prime contractor's plant, though possibly at the subcontractors' plants). By this argument, the prime contractor's production rate is perhaps better measured by the number of units in final assembly. The current lot size provides a serviceable approximation to this concept, although it too is somewhat flawed due to time lags. For example, an aircraft that completes final assembly and is delivered in the first month of DoD's fiscal year (October) would certainly have begun final assembly during the previous fiscal year.

We have developed a series of six charts, with these two objectives:

- Illustrate the complications in measuring production rate using aggregate annual data, and
- Debunk the assertion that production rate (however measured) and cumulative quantity are mechanically correlated.

To address the correlation issue in the simplest possible context, consider a production situation in which final assembly of each unit takes place within a single month. (One example might be assembly of full-up artillery rounds from existing components already in the inventory.) Thus, we temporarily avoid the problems of multi-year production and time lags across fiscal years. We can then safely equate production rate with the current lot size, because plant activity during each month involves only units that will be delivered during that month; by extension, plant activity during any fiscal year corresponds to that year's lot size. With these simplifying assumptions, we can concentrate on the correlation between production rate and cumulative quantity.

Figure 1.4 illustrates a production program with an oscillating production rate. The oscillating production rate is essentially uncorrelated with the steady increase in cumulative quantity; the correlation equals only 0.038. Figure 1.5 illustrates a production program with a steadily declining production rate. In this case the correlation is strongly *negative*, –0.969, contrary to the presumed positive mechanical correlation.

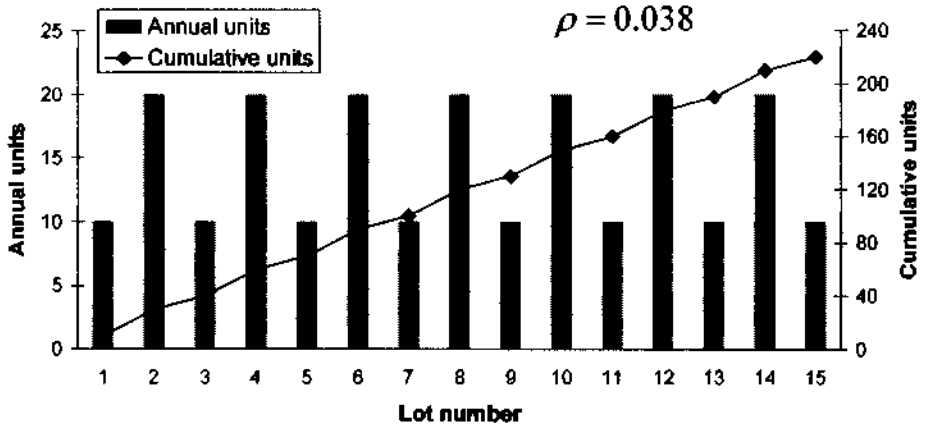


Figure 1.4. Production Program with Oscillating Production Rate

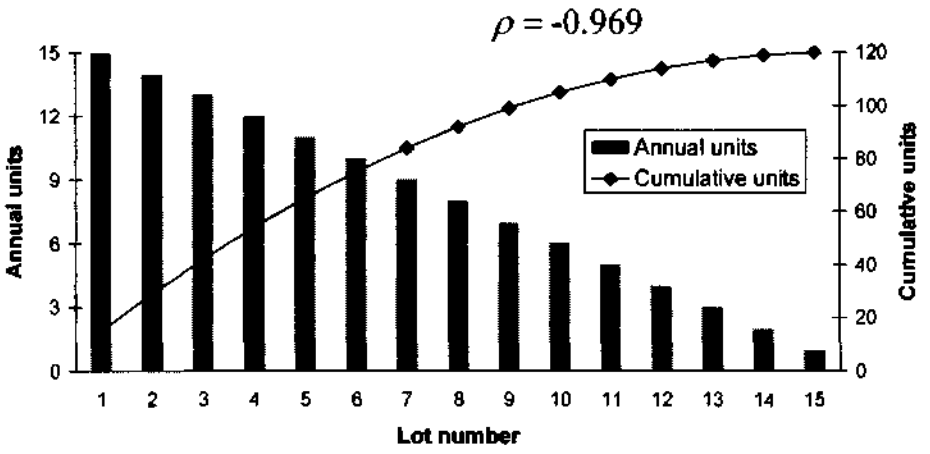


Figure 1.5. Production Program with Steadily Declining Production Rate

The situation is somewhat more complex if we reintroduce multi-year production, but the conclusion regarding the correlation remains essentially intact. We modify Figure 1.4 to reflect an assumed 3-year production cycle. The horizontal axis in Figure 1.6 now measures not the lot number, but rather the contract year. Thus, plant activity during the first contract year involves only the 10 units that are authorized and for which production begins that year. Plant activity during the second contract year involves those same 10 units, now in their second year of production, plus 20 new units. Plant activity during the third contract year involves all 40 units that were authorized during that year and the preceding two years. From that point forward, units in-process are measured over a three-year moving time window.

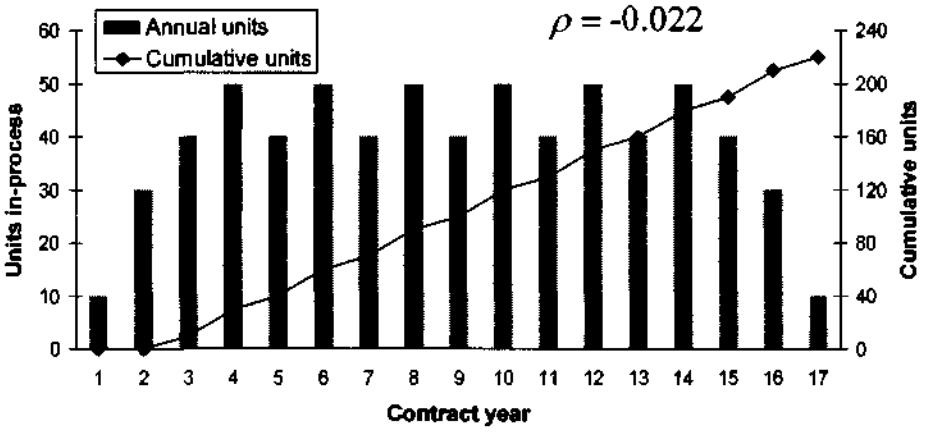


Figure 1.6. Multi-Year Production with Oscillating Production Rate

Apart from the ramping up and down at the two extremes, units in-process follow an oscillating pattern. Whereas the correlation in Figure 1.4 was 0.038, the correlation in Figure 1.6 is –0.022. In both instances, the alleged mechanical correlation between production rate and cumulative quantity is negligible.

We next modify Figure 1.5 to reflect a 3-year production cycle. In Figure 1.7 we again observe a ramping-up phenomenon, with 15 units in-process during the first contract year, followed by 29 units in-process during the second contract year, and 42 units during the third contract year. The ramping-up serves to dampen the negative correlation somewhat; the correlation is –0.742 in Figure 1.7 versus –0.969 in Figure 1.5. Nonetheless, even a correlation of –0.742 contradicts the assertion of a positive mechanical correlation.

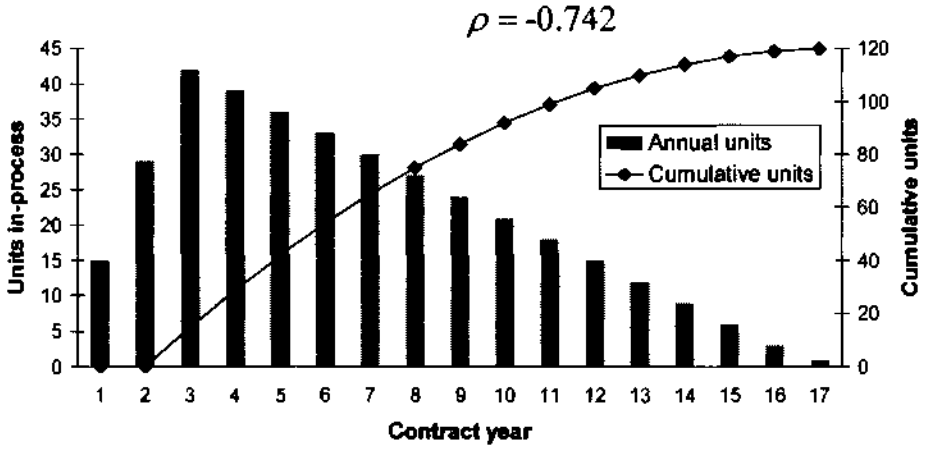


Figure 1.7. Multi-Year Production with Steadily Declining Production Rate

Finally, in case our examples appear contrived, Figure 1.8 displays the actual production program for the U.S. Air Force F–15E fighter. The figure covers the entire production program, for which production lots were authorized (with some breaks) between fiscal years 1986 through 2001. Figure 1.9 displays the units in-process, again assuming a 3-year production cycle. With this assumption, the final units will be delivered in fiscal year 2003. The correlation in Figure 1.9 equals –0.592, moderate in magnitude and opposite from the presumed positive direction.

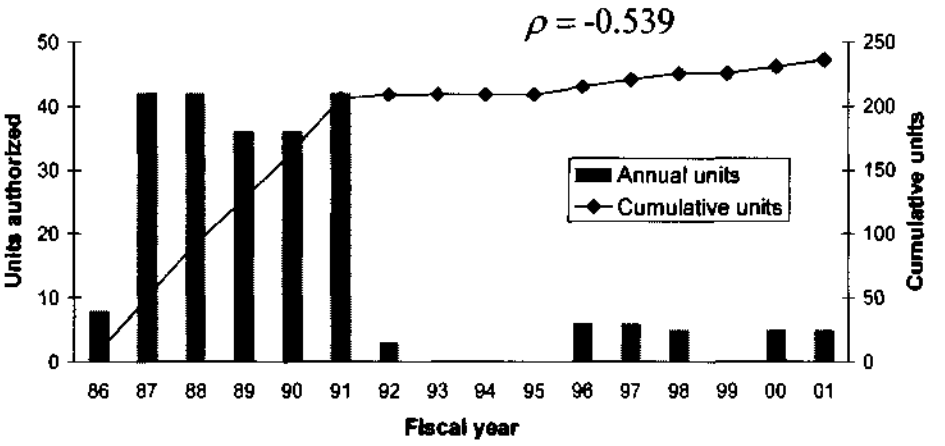


Figure 1.8. F–15E Production Program: Annual Units Authorized

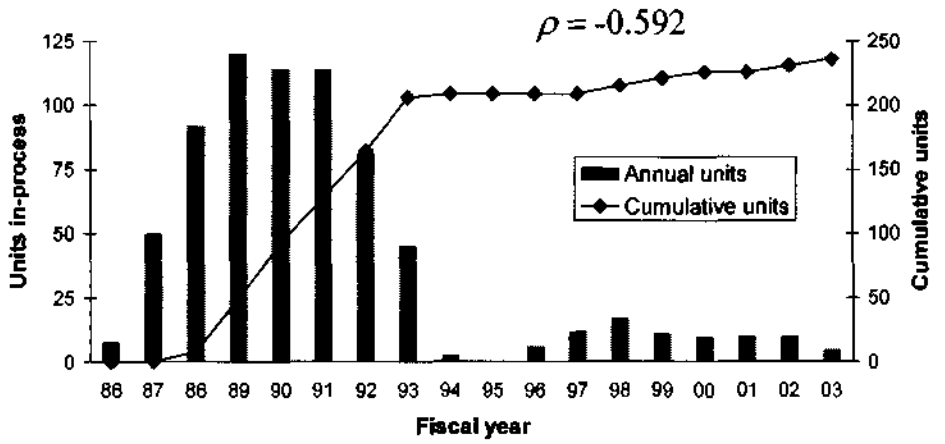


Figure 1.9. F–15E Production Program: Annual Units In-Process

We conclude that, although production rate and cumulative quantity *may* be correlated, precluding estimation of their separate effects, they need not be correlated. If production rate effects are thought to be important, it is worth the effort to attempt to include production rate in the learning-curve model. Moreover, even if collinearity proves to be a problem in a particular instance, there are statistical techniques that may overcome this problem and still allow estimation of the separate learning and rate effects.[12]

---

[12] See Judge, Griffiths, Hill, Lütkepohl, and Lee (1985), particularly their discussion of ridge regression.

1.7 Cost-estimating relationships

The learning curve is one of the two most pervasive models in cost analysis. The other is the cost-estimating relationship, a regression equation to predict the development or production cost of a system based on performance and technical characteristics such as weight, speed, and composite materials content.

A typical CER for production cost might take the form of the following non-linear, multiplicative regression model:

$$\text{Unit cost} = b_0 \times \text{Weight}^{b_1} \times \text{Speed}^{b_2} \times b_3^{\text{Remanufactured}} \times u_i \ , \qquad (1.17)$$

where the dummy variable "Remanufactured" equals 1.0 for remanufactured production items and 0.0 for those newly manufactured. Note that a dummy variable results in a proportional scale factor, rather than an additive factor as would be the case in a linear regression model.[13] If, for example, the coefficient b_3 in equation (1.17) were estimated as 0.9, we would infer that a remanufactured production item costs only 90% as much as a newly-manufactured item. Although the particular dummy variable for remanufacturing might not appear in most cost analyses, other dummy variables could reflect technical or programmatic characteristics such as multi-year contracting or follow-on systems (e.g., the U.S. Navy's F/A–18 C/D fighter/attack aircraft is a follow-on to the earlier F/A–18 A/B series).

The learning curve and the CER are two different "slices" of the same underlying data. Figure 1.10 shows hypothetical data from four different systems. The data from any one system indicate a trend in learning as we move horizontally from left to right. The data can also be compared vertically to study the differences in cost *between* systems. The latter comparison makes sense only if the systems under comparison are similar enough that the cost differences can reasonably be explained using regression variables such as weight, speed, and so on. For example, it is quite common and sensible to compare the eosts of various fighter aircraft models. However, it would be folly to use weight and speed in an attempt to understand why an aircraft carrier costs more than an F/A–18 C/D.

---

[13] The relevant property here is non-linearity (the manner is which the dummy variable enters the regression prediction), *not* additive versus multiplicative regression models (the manner in which the error term is appended to the regression prediction).

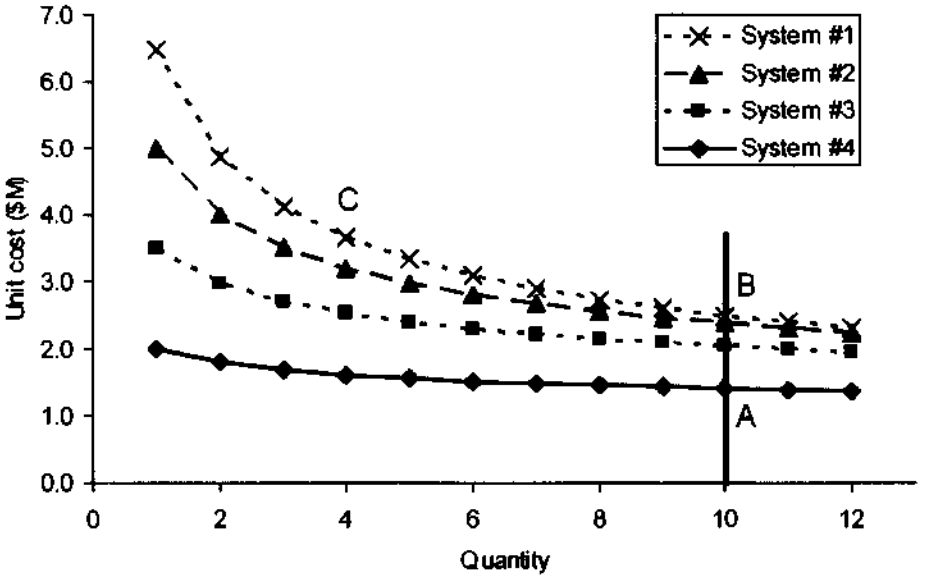


Figure 1.10. Distinction between Learning Curve and Cost-Estimating Relationship

Even when comparing like systems to estimate a CER, it is important to normalize the cumulative quantity in order to separate learning effects from true cost differences between systems. In Figure 1.10, System #1 has higher unit cost than System #4 at any common value of cumulative quantity. However, a crude comparison that did not normalize for quantity might grossly exaggerate the cost difference. At the 10^{th} unit, System #1 costs 77 percent more than System #4 (point B versus point A). But if we compared the 4^{th} unit of System #1 to the 10^{th} unit of System #4 (point C versus point A), we would report a 159-percent cost difference. The latter difference is misleading because System #4 has benefited from much more learning before reaching the 10^{th} unit.

In practice, CERs are estimated at a common quantity that lies well within the range of data for all the systems under comparison. Moreover, the cost of the initial production lot is often contaminated by non-recurring costs for items that the customer purchases in addition to completed production units: specialized tooling, test equipment, ground support equipment, and so on. Therefore, it is generally preferable to choose a common quantity that lies beyond the initial production lot. A typical point of comparison might be the 100^{th} unit for aircraft systems, but the $1{,}000^{th}$ unit for missile systems.

1.8 Estimation of multiplicative regression models

The CER in equation (1.17) is a multiplicative regression model, as is the learning-curve model for lot average cost in equation (1.16). Lot-midpoint iteration is a specialized technique for estimating power-function learning curves, with no counterpart for estimating CERs. However, there are several general-purpose estimation techniques that apply to all multiplicative regression models, including CERs as well as learning curves.

Both lot-midpoint iteration and lot-midpoint NLS (i.e., explicit minimization of expression (1.8)) are attempts to minimize the sum-of-squared errors in predicting the logarithm of lot average cost. Recently, Book and Young (1995, 1997) and Lee (1997) have proposed an alternative estimation method for multiplicative regression models. Their method minimizes the sum-of-squared *percentage* errors in predicting the *level* (not logarithm) of lot average cost. Accordingly, their method is known as Minimum Percentage Error (MPE). We show in Chapter 4 that the logarithmic and percentage fitting criteria are equivalent up to a first-order Taylor series approximation, but differ in the higher-order terms. Thus, the two fitting criteria generally lead to distinct estimates of the regression parameters.[14]

The choice of estimation method should be guided by the statistical properties of the resulting estimators, not the intuitive appeal of the fitting criterion being optimized. First, we seek an estimation method that requires minimal distributional assumptions. For example, we would almost certainly be willing to assume that the error term u_i in equation (1.12) has finite variance. However, we might not be nearly as willing to assume that the error term is normally distributed.

Another desirable property is that the estimator be *unbiased*. To understand this concept, suppose we repeated the estimation process on many different random samples (of the same, finite size) drawn from the same underlying population. We would want the average of the parameter estimates from these samples to equal the true (unknown) parameter value. We could tolerate (indeed, we would expect) an estimation error in any single sample, hut we would want this error to equal zero *on average*. The difference between the average of the parameter estimates and the true parameter value is known as the *bias*. An unbiased estimator has zero bias.

---

[14] Young (1999) also investigated the distinction between these two fitting criteria. As we argue in Chapter 4, however, his analysis was somewhat incomplete.

Some estimators are known to be biased, but the bias vanishes in large samples. This leads to the concept of *consistent* estimators. Suppose we select a small interval around the true parameter value, and we also specify a probability just short of 1.0. If an estimator is consistent, then we can find a sample size large enough so that the parameter estimate from the sample falls within the small interval with probability at least as large as the probability we pre-specified. Intuitively, a consistent estimator "approaches" the true parameter value in large samples. It can be shown that a biased estimator is consistent if both the bias and the standard error of the estimator approach zero in large samples.[15]

Two other desirable properties of an estimator concern its sampling distribution. First, there should be a formula available to compute the standard errors of the estimates and, more generally, their entire covariance matrix.[16] It is preferable to have an "exact" formula (i.e., one that is accurate even in small samples). If an exact formula is not available, we must sometimes settle for an asymptotic formula whose accuracy is, strictly speaking, guaranteed only in large samples. The use of asymptotic standard errors is somewhat problematic in cost analysis, because the sample sizes are often so small as to diminish the applicability of asymptotic properties. For certain estimation methods, however, there is no alternative because the exact standard errors are not known (e.g., this is the case for NLS).

Finally, in addition to having their standard errors, we require the sampling distribution of the estimates. It is convenient to divide a single coefficient by its standard error and label the result a "*t*–ratio" or "*t*–statistic." However, the mere computation of the "*t*–ratio" does *not* guarantee that its percentile points can be read off a published table of the *t*–distribution. Thus, to conduct statistical inference (e.g., to compute confidence intervals or significance tests), we need to know the sampling distribution of the estimates. Again, it is preferable to know the exact sampling distribution, but we must sometimes settle for the asymptotic sampling distribution.

---

15 Conversely, however, it is possible to construct a consistent estimator that has neither finite mean nor finite variance in large samples. The archetypical example was provided by Sewell (1969), and reproduced in the econometrics textbooks of Dhrymes (1974, pp. 87–89) and Johnston (1972, pp. 270–273).

16 The diagonal terms in the covariance matrix are the variances of the estimates (i.e., the squares of their respective standard errors). The off-diagonal terms are the covariances among the estimates.

Several other estimation methods, common in the statistical literature but unknown to most cost analysts, possess desirable statistical properties. For example, we show in Chapter 3 that the quasi-likelihood function for a multiplicative regression model is defined as follows:

$$q(\beta,\lambda) = -\frac{1}{\lambda}\sum_{i=1}^{n}\left[\frac{y_i}{f(x_i,\beta)} + \ln(f(x_i,\beta))\right], \tag{1.18}$$

where λ denotes the variance of u_i in equation (1.12). It turns out that by maximizing the quasi-likelihood function with respect to β, the resulting estimator of β is consistent. Moreover, the covariance matrix of this estimator follows a known formula, and the estimator is asymptotically normally distributed even though the regression error itself (u_i) need not be normally distributed.

When advocating quasi-likelihood estimation at professional conferences, we have been asked the question, "Why would you want to maximize such a non-intuitive function as $q(\beta,\lambda)$?" First, to reiterate our opinion, the choice of estimation method should be guided by the statistical properties of the resulting estimators, not the intuitive appeal of the fitting criterion being optimized. Second, we show in Chapter 3 that quasi-likelihood estimation of multiplicative regression models is equivalent to the better-known technique of iteratively reweighted least squares (IRLS). Indeed, the quasi-likelihood expression (1.18) provides tbe function that is implicitly being maximized when IRLS is performed.

1.9 Summary of comparisons among estimation methods

In the remainder of this monograph, we compare a total of six estimation methods:

- Lot-midpoint NLS,
- Lot-midpoint iteration,
- Minimum percentage error (MPE),
- Maximum likelihood,
- Iteratively reweighted least squares (IRLS), and
- Maximum quasi-likelihood.

The first two methods involve lot midpoints, thus these methods have little application outside the narrow realm of learning-curve models. However, the remaining four methods apply to the much broader class of multiplicative regression models, including multiplicative CERs as well as learning-curve models. We compare all six estimation methods with respect to all of the statistical properties described above.

Although the statistical properties of some of these methods can be derived theoretically, little could be proved theoretically about the others. To compare the statistical properties of *all* the methods, we conducted a series of Monte Carlo experiments. We generated data on lot average cost using known error structures and parameters values. Because the parameters values were known, we could directly compare the estimates produced by the different methods to the "truth." When assessing possibly biased estimators, we considered not only the variance of the estimates around the average estimate at any sample size, but also the bias in the estimate (i.e., the difference between the average estimate at any sample size and the true parameter value). We could also assess the rate at which the various estimates approach the true parameter values (i.e., the required sample size). In addition, in most cases, even if a formula for the covariance matrix is available, the matrix produced is only an asymptotic covariance matrix. The Monte Carlo experiments allowed us to compare the variances over a spectrum of sample sizes, ranging from very small (unfortunately, the typical situation in cost analysis) up to asymptotically large.

Most estimation methods are developed under a particular set of assumptions. Estimation methods are called *robust* if they continue to produce good estimates even when those assumptions are violated. None of the methods we compared rely on any particular assumption about the true learning slope, the number of units in a lot, or the standard deviation of the error term. However, it is still of interest to inquire whether the methods perform as well under a range of values for these parameters. Some of the methods rely on a particular distributional assumption, such as normally distributed errors. Thus, it is also of interest to inquire about the performance of the methods under alternative (non-normal) error distributions.

IRLS and lot-midpoint NLS produced unbiased estimates under all of the simulation excursions. The performance of these two methods was essentially unaffected by the substitution of either uniform or t–distributed errors for the normal errors found in the baseline experiment. Naturally, however, the parameter estimates became less precise during the excursion for which we doubled the standard deviation of the error terms.

The estimates produced by lot-midpoint iteration and lot-midpoint NLS are numerically distinct. However, with just one exception, the numerical differences between the two sets of parameter estimates (e.g., between the estimated learning slopes) were essentially negligible. Consequently, both of these methods produced unbiased estimates even for small numbers of lots. The one exception is that the parameter estimates from lot-midpoint iteration (though not lot-midpoint NLS) became much less precise under first-order serial correlation. The introduction of serial correlation led to a drop in precision nearly equal to that engendered by doubling the standard deviation of the error terms (but without serial correlation). None of the other estimation methods exhibited any sensitivity to serial correlation.

Notwithstanding this case, the performance of lot-midpoint iteration was much better than we had expected. Prior to the simulation experiments, there was no theoretical basis for lot-midpoint iteration and little was known about the behavior of its estimates. We show in Chapter 2 that lot-midpoint iteration does not minimize any continuously differentiable function. In a sense, that finding further undermines the theoretical basis for the method. Its apparently satisfactory performance characteristics, at least in the absence of serial correlation, remain a theoretical mystery.

The MPE estimates of T_1 were biased high, even in large samples, under every one of the simulation excursions. Similarly, the MPE predictions of lot average cost were also biased high. Moreover, the biases increased both when we doubled the standard deviation of the normal errors, and (unique to this method) when we substituted t–distributed errors for the normal errors. The latter result illustrates that the performance of MPE degrades when there are more outlier observations (in statistical parlance, the error distribution has "thicker tails") than would be expected under a normal error distribution. Because of these biases and sensitivities, we recommend against the use of MPE.

In light of the latter result, as well as the sensitivity of lot-midpoint iteration to serial correlation, we recommend either IRLS or lot-midpoint NLS as the estimation methods of choice. We sketched the concept of lot-midpoint NLS previously in this chapter (expression (1.8)); we give more details, including formulas for the standard errors of the parameter estimates, in Chapter 2. We give a full exposition of IRLS, including formulas for the standard errors, in Chapter 3. NLS is already available as an option in most statistical software packages. IRLS is becoming increasingly available as a built-in feature in many statistical packages, and the equivalent method of quasi-likelihood can be programmed quite easily using any computational software or even a

simple spreadsheet. There is no longer any excuse for cost analysts to use methods that produce inconsistent parameter estimates.

2. LEARNING CURVE MODELS

In this chapter, we first demonstrate the equivalence, under reasonable conditions, of two learning-curve models that are widely thought to be distinct. We then develop the concept of lot midpoint, which is often used as a single measure of cumulative quantity for production lots that span a range of units. We compare two methods for estimating learning-curve models using lot midpoints: non-linear least squares and lot-midpoint iteration. Among the issues that arise in this comparison are cumulative data versus data on individual production lots, admissible error distributions, computation of standard errors, and retransformation bias.

2.1 Two learning-curve models

Lee (1997, p. 11) distinguishes two learning-curve models: the Crawford model and the Wright model. The Crawford model expresses the marginal cost of unit Q as a power function:

$$MC(Q) = T_1 Q^b \tag{2.1}$$

for $Q > 0$, where $T_1 > 0$ and b are parameters to be estimated. Under this model, the ratio of marginal costs for any two units depends only on their relative (not absolute) position in the production sequence:

$$MC(\phi \times Q)/MC(Q) = \phi^b, \tag{2.2}$$

which is independent of Q.

In particular, the "learning slope" is defined as the ratio of marginal costs when $\phi = 2$:

$$\rho = MC(2Q)/MC(Q) = 2^b. \tag{2.3}$$

Lee (1997, p. 41) argues that the plausible range for the learning slope is $½ < \rho \leq 1$ or, correspondingly, $-1 < b \leq 0$. (We confirm Lee's argument in due course.)

By contrast, the Wright model expresses the cumulative average cost of the first Q units as a power function:

$$AC(Q) = A_1 Q^{\beta} \tag{2.4}$$

for $Q > 0$, where $A_1 > 0$ and β are parameters to be estimated. Note that Lee actually uses the same symbol for the exponents in equations (2.1) and (2.4). However, we use two different symbols to maintain, temporarily, Lee's apparent distinction between the two learning-curve models.

Lee treats the production quantities as discrete units, and uses arithmetic summation to compute the incremental cost of a lot or the cumulative cost of an entire production run. On the other hand, most cost analysts treat the production quantities as a continuum, and use integral calculus to approximate the incremental or cumulative cost. We mostly follow the continuous approach, while recognizing the ramifications of choosing one approach or the other.

2.2 Recurring, fixed, and variable costs

Whether using the discrete approach or the continuous approach, it is imperative to first define the universe of costs being modeled. One important distinction is between non-recurring costs and recurring costs. Non-recurring costs are paid only once, usually at the beginning of the production run. These costs are associated with such activities as designing the production process, recruiting the initial work crew, and purchasing or building specialized facilities and tooling. Recurring costs are paid in connection with each successive lot and in varying amounts, depending on the lot size and the cumulative amount of learning.

In studies of the learning curve, the response variable is often taken to be *direct labor hours*. One rationale behind this choice is an attempt to remove one-time activities that are not subject to learning. However, the focus on direct labor hours assumes that the industrial engineers who design the production process, and the personnel specialists who recruit the initial work crew, charge their time *indirectly* (i.e., charge to a corporate or plant-wide overhead account, rather than to a particular production program). In practice, the cost of the initial production lot is often contaminated because some of these non-recurring labor costs are charged *directly* to the production program. The large decline in average cost from the initial lot to the next few lots reflects, in part, the

payment of non-recurring costs in the initial lot. The degree of learning would be overstated if the entire decline were attributed to learning.

An extreme example of this effect is the case of naval ship construction. There, the "lead ship" (first unit) of a class is burdened with the full design costs of the class plus certain other non-recurring costs. For that reason, analysts virtually never include the lead ship in the database from which learning curves are estimated. Similar reasoning might motivate either exclusion of the first production lot, or use of a dummy variable to identify that lot, in the analysis of systems other than ships.

The cost analyst's distinction between non-recurring and recurring costs is somewhat different from the micro-economist's distinction between fixed and variable costs. The micro-economist defines fixed costs as costs that are independent of the number of units produced during a given time period (typically one year). More emphatically, fixed costs are paid even if *zero* units are actually produced during the time period.[17] Fixed costs might include rental or mortgage payments on land, buildings, and equipment that are not easily disposed of during a single time period.

The fundamental distinction is that the micro-economist's fixed costs are paid *repeatedly in every time period*, independent of the number of units produced during that period (including possibly zero units), as long as the firm maintains the product line. Thus, the cost analyst's *recurring* costs might well include some costs that the micro-economist would consider as *fixed* (e.g., the annual rental or mortgage payments), as well as other costs that the micro-economist would consider as variable. These cost categories are illustrated in Figure 2.1.

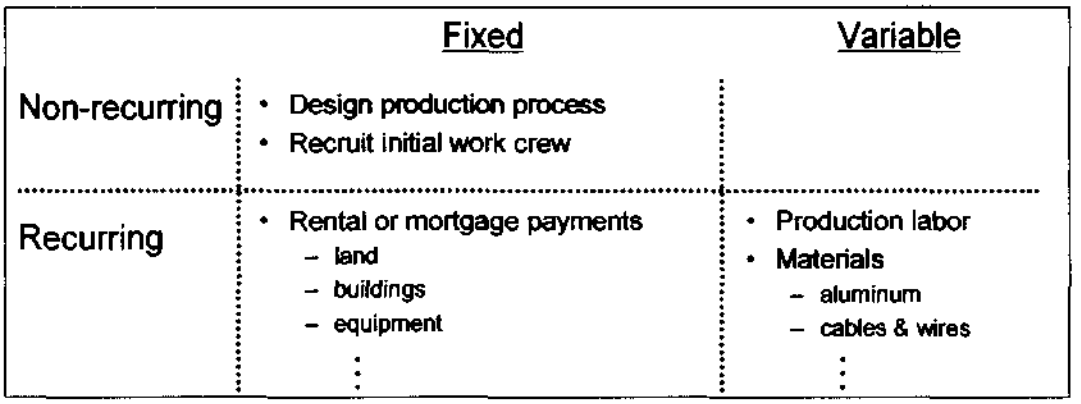


Figure 2.1. Illustration of Various Cost Categories

[17] See, for example, Henderson and Quandt (1980, chapter 4) or Varian (1992, chapter 5).

2.3 Equivalence between the two learning-curve models

We now return to the Crawford and Wright learning-curve models. We demonstrate that the two models are equivalent if:

- We use integral calculus to continuously approximate the incremental and cumulative costs; and
- Either non-recurring costs are equal to zero, or we are modeling only the recurring costs.

We begin with the Crawford model. The cumulative average cost of the first Q units is obtained by dividing the cumulative total cost by the cumulative number of units:

$$AC(Q) = \frac{1}{Q} \times \int_0^Q T_1 z^b \, dz = \frac{1}{Q} \times \left[NRC + \left(\frac{T_1}{1+b} \right) \times Q^{1+b} \right] = \frac{NRC}{Q} + \frac{T_1 \times Q^b}{(1+b)}, \quad (2.5)$$

where the constant of integration, NRC, may be interpreted as the non-recurring cost paid prior to the "zeroth" cumulative unit. If $NRC = 0$, the cumulative average cost reduces to:

$$AC(Q) = \left[T_1/(1+b) \right] \times Q^b \, . \quad (2.6)$$

Conversely, starting with the Wright model, the cumulative total cost is obtained by multiplying the cumulative average cost and the cumulative quantity:

$$TC(Q) = AC(Q) \times Q = A_1 Q^{1+\beta}, \quad (2.7)$$

and the marginal cost is the derivative of total cost with respect to cumulative quantity:

$$MC(Q) = \frac{d\,TC(Q)}{dQ} = A_1 \times (1+\beta) \times Q^{\beta} \, . \quad (2.8)$$

Now compare equations (2.1) and (2.8) for marginal cost, and equations (2.4) and (2.6) for cumulative average cost. The two learning-curve models are rendered equivalent by setting $b = \beta$ and $T_1 = A_1 \times (1+b)$. Thus, Lee was correct to use the same symbol for the exponents in equations (2.1) and (2.4). However, his use of two different symbols for the intercepts (T_1 and A_1) gives the impression that the two learning-curve models are distinct. Under the assumption of zero non-recurring costs, and using the continuous

approximation, the two models are actually equivalent.[18] If non-recurring costs are positive and are included in the model, it is more appropriate to use equation (2.5) from the Crawford model. The corresponding expression for cumulative average cost from the Wright model in equation (2.4) is incomplete because the term *NRC* (non-recurring cost) is absent.

To be precise, although the two learning-curve models are *mathematically* equivalent under the stated assumptions, their *statistical* properties may be different in small samples. For example, depending on the precise method of estimation employed, there is no guarantee that the ratio of the estimates of T_1 and $(1+b)$ from equation (2.6) will exactly equal the estimate of A_1 from equation (2.4). However, this equality will hold if the sample sizes are large enough and if consistent estimators are employed.

We can also see why Lee argues for the restriction $-1 < b \leq 0$ or $\frac{1}{2} < \rho \leq 1$. The cumulative average cost in equation (2.5) involves the following definite integral: $\int_0^Q T_1 z^b \, dz$. In the borderline case of $b = -1$, the anti-derivative of z^{-1} is the natural logarithm. The definite integral requires evaluation of the anti-derivative at the lower limit, $T_1 \times \lim_{z \to 0^+} [\ln(z)]$, which diverges. When $b < -1$, we encounter instead the expression $[T_1/(1+b)] \times \lim_{z \to 0^+} z^{b+1}$, which diverges because the exponent is negative, $b+1 < 0$. The power-function model simply makes no sense absent the constraint $-1 < b \leq 0$. Learning slopes smaller than 0.5, although theoretically possible, cannot be accommodated by this particular functional form.

An apparent discrepancy arises in evaluating the cost of the first unit produced (colloquially called the "T_1–cost"). Ignoring any non-recurring costs, the cumulative average cost and the marginal cost should be equal at the first unit. However, our equation (2.1) evaluates as $MC(1) = T_1$, but equation (2.6) evaluates instead as $AC(1) = T_1/(1+b)$. To resolve this discrepancy, recall that we are applying a continuous approximation to the learning-curve model. The incremental cost of a "lot" consisting of 1.0 units is given by the integral under the (Crawford) marginal cost curve,

---

[18] Lee (1997, pp. 41–42) did not use continuous approximation. He correctly demonstrated that, when output is measured in discrete units, the two learning-curve models are equivalent only asymptotically. However, contrary to our analysis, other authors such as Loerch (1999) have treated the two models as distinct even when using the continuous approximation. Importantly, these and other authors who have argued for a distinction between the two learning-curve models did *not* do so on the basis of non-recurring costs; they implicitly assumed that non-recurring costs were equal to zero.

$\int_0^1 T_1 z^b \, dz = T_1 / (1+b)$, which indeed equals $AC(1)$. Simply evaluating the marginal cost curve at the argument 1.0 is inexact because only the integral under the marginal cost curve is meaningful, not the curve's height.

Figure 2.2 illustrates this principle for a production process with $b = -0.152$ (implying, from equation (2.3), a 90% learning slope) and $T_1 = 2.0$. The height of the average cost curve at the argument 1.0 is 2.359. The height of the marginal cost curve is 2.000, but this value is not meaningful. Instead, the *integral* under the marginal cost curve (the shaded area in Figure 2.2) is meaningful and is equal to the earlier value 2.359. Strictly speaking, it is incorrect to interpret the parameter T_1 as the cost of the first unit.

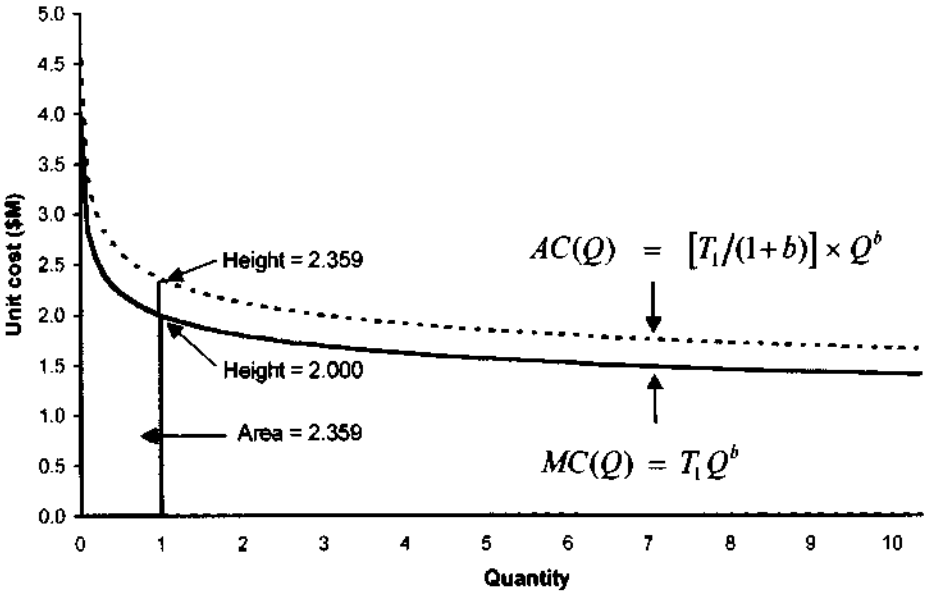


Figure 2.2. Proper Interpretation of First-Unit Cost

To better understand this principle, it may help to contemplate the analogous distinction between discrete and continuous probability density. When asked to interpret the height of a continuous probability density, even analysts with moderate amounts of statistical training might reply, "The height is just the probability [of occurrence for the event in question]." However, consider a continuous uniform density defined over the interval [0.0, 0.5]. The height of this density function must be 2.0 over the interval, to ensure that the entire probability (i.e., the area of the entire rectangle, both shaded and unshaded) in Figure 2.3 equals 1.0. Because probability is bounded above by 1.0, clearly the height of the density function here is *not* interpretable as a probability. Instead,

probability can be computed only as the *area* under a continuous density function. For example, the probability in the sub-interval [0.0, 0.2] is given by the shaded area, or 0.4.

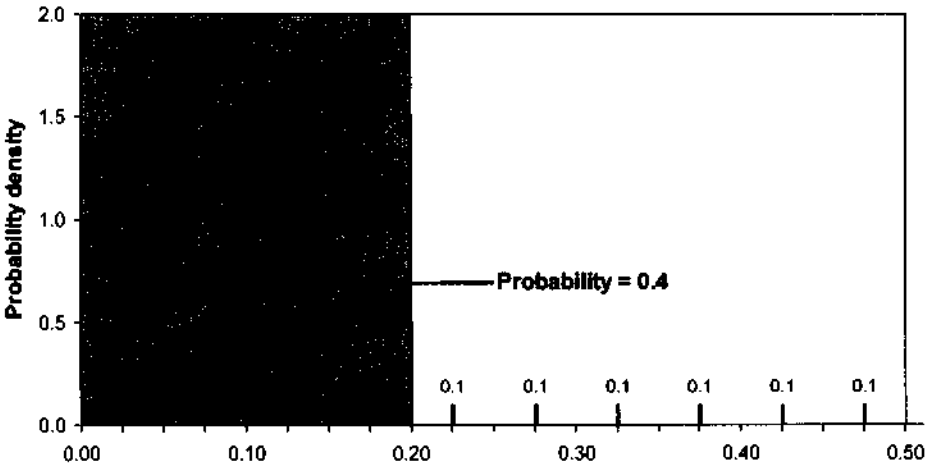


Figure 2.3. Discrete and Continuous Versions of Uniform Probability Density

The hypothetical analyst's reply, "The height is just the probability ...," is based on a discrete (*vs.* continuous) approach to the problem. For example, a *discrete* uniform density could be defined over the 10 points $0.025, 0.075, 0.125, \ldots, 0.475$ lying within the interval $[0.0, 0.5]$. In that case, the heights of 0.1 at each picket are indeed interpretable as probabilities; further, the probability of occurrence in a subinterval may be computed by arithmetically summing those heights. But when using the continuous approach, only the areas (i.e., the integrals) are meaningful, not the heights. Similarly, in our learning-curve model, only the area under the marginal cost curve is meaningful, not its height.

2.4 Cumulative data versus data on individual production lots

Given that the two learning-curve models are equivalent, which one should be used in estimation? One argument is that it makes no difference; use the Crawford model when the data are presented in terms of unit cost, and use the Wright model when the data are presented in terms of cumulative average cost. However, statistical estimation of a regression model, and estimation of an exact functional transformation of that model, do not necessarily yield identical parameter estimates (e.g., learning slopes), because the error terms have different properties after transformation.

First, we establish that it is always possible to transform one type of cost data to the other. It is obvious that a series of lot quantities and lot average costs can be transformed into a series of cumulative average costs. Conversely, suppose the analyst is presented with the cumulative quantities of k production lots, $Q_1,\ldots,Q_k$, where $0<Q_1<Q_2<\ldots<Q_k$, and the corresponding series of cumulative average costs, $AC(Q_1),\ldots,AC(Q_k)$. We can recover both the incremental cost of the i^{th} lot:

$$TC_i - TC_{i-1} = AC_i \times Q_i - AC_{i-1} \times Q_{i-1}, \tag{2.9}$$

and the lot average cost (LAC) of the i^{th} lot:

$$LAC_i \equiv \frac{TC_i - TC_{i-1}}{Q_i - Q_{i-1}} = \frac{AC_i \times Q_i - AC_{i-1} \times Q_{i-1}}{Q_i - Q_{i-1}}, \tag{2.10}$$

for $i=1,\ldots,k$ (with the convention that $Q_0=0$).

The data analyst might be tempted to work with the cumulative average costs because they are smoother than the lot average costs. In Figure 1.3 we plotted Lee's tactical missile data from Table 1.1. The height of each point in that figure represents the observed lot average cost. The horizontal coordinate represents the lot midpoint at convergence of the lot-midpoint NLS method. In Figure 2.4 we plot both the lot average costs and the cumulative average costs. The latter are plotted not at the lot midpoints, but rather at the lot endpoints, $AC(Q_i) = A_1 Q_i^{\beta}$. Comparing the two series, the cumulative average costs are much smoother because the effect of an apparent outlying lot (e.g., the fifth or eight lot) is averaged with all of the preceding lots.

The difference in fit is also reflected in the R-squared statistics. The logarithmic regression of lot average costs (equation 1.4) has a respectable R-squared of 0.951. Alternatively, we may take logarithms in equation 2.4 to obtain a regression of cumulative average costs, $\ln\left[AC(Q_i)\right] = \ln A_1 + \beta \times \ln Q_i$. The latter regression has a nearly perfect R-squared of 0.9986.

Although the R-squared statistics seem to favor using cumulative average costs, a deeper analysis of the statistical issues actually implies a preference for using lot average costs (i.e., the Crawford model rather than the Wright model). A series of cumulative average costs is almost certain to be serially correlated. For example, if the *4th* lot is particularly expensive, the cumulative average cost of the first 4 lots will tend to lie above the regression curve. Unless the *5th* lot is sufficiently cheap and contains

sufficiently many units, the cumulative average cost of the first 5 lots will also lie above the regression curve. Indeed, the anomaly in the cost of the *4th* lot will likely persist in the cumulative average cost of several subsequent lots. OLS regression estimation is inefficient (i.e., yields larger than the minimum possible standard errors) when applied to serially correlated data. A common remedy for serial correlation is to difference the data — essentially the procedure indicated in equation (2.9), which returns us to the Crawford model.[19]

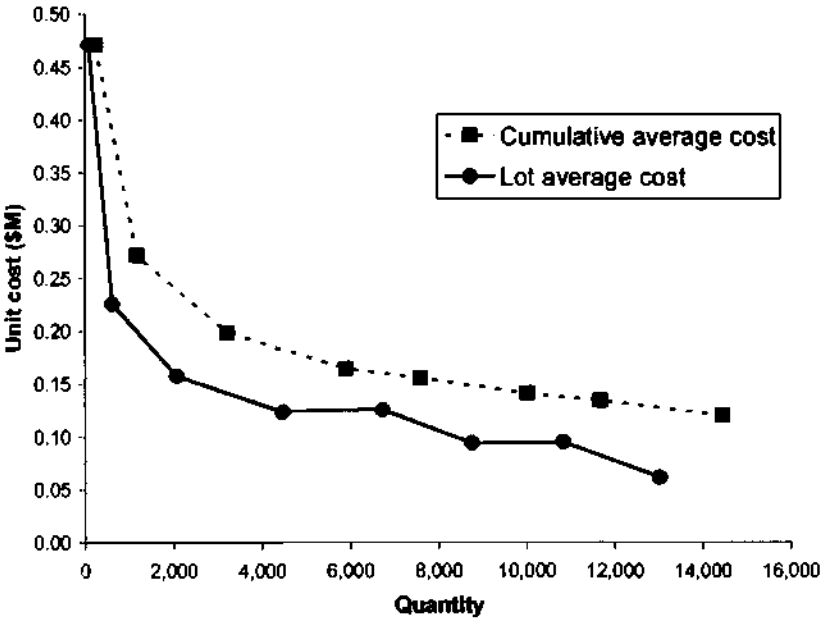


Figure 2.4. Cumulative and Lot Average Costs for Tactical Missile Data

Estimation using cumulative average costs may also lead to problems of non-constant variance or heteroscedasticity, again causing inefficiency in OLS estimation. The series on lot average costs and the series on cumulative average costs cannot both have constant variance — if one has constant variance, the other cannot. The lot average costs are more likely to have constant variance, in which case the cumulative average costs will tend to have decreasing variance as more lots are included in the cumulative average. To see this point, write the cumulative average cost as follows:

---

[19] See Womer and Patterson (1983) for a more thorough discussion of serial correlation in estimating models of incremental lot cost. Well aware that the series on lot average costs and the series on cumulative average costs cannot *both* be serially uncorrelated, they stated on p. 266, "Serial correlation of the residuals from one of the specifications is therefore expected."

$$AC_j = TC_j/Q_j = \sum_{i=1}^{j} LAC_i \times (Q_i - Q_{i-1})/Q_j \,. \tag{2.11}$$

It is immediately obvious that even if the series $LAC(Q_1),\ldots, LAC(Q_k)$ is serially uncorrelated, the series $AC(Q_1),\ldots, AC(Q_k)$ will be serially correlated (e.g., the expansions for AC_j and AC_{j+1} share the first j terms in common.) Turning to heteroscedasticity, if the series $LAC(Q_1),\ldots, LAC(Q_k)$ is serially uncorrelated with constant variance σ^2, the variance of AC_j turns out to be:

$$Var(AC_j) = \sigma^2 - (2\sigma^2/Q_j^2) \times \sum_{i=1}^{j-1} Q_i \times (Q_{i+1} - Q_i)\,, \tag{2.12}$$

which tends to decrease as more lots are cumulated. For example, if every lot contains the same number of units, $Q_i - Q_{i-1} = q$, the variance of cumulative average cost reduces to the familiar formula for the variance of a (non-weighted) average:

$$Var(AC_j) = \sigma^2/j = Var(LAC)/j \,. \tag{2.13}$$

Again, statistical considerations tend to favor estimation using lot average costs, not cumulative average costs.[20]

2.5 Lot midpoints

Some authors represent the incremental cost of the i^{th} lot as the sum of the discrete marginal costs:

$$TC_i - TC_{i-1} = \sum_{j=Q_{i-1}+1}^{Q_i} MC(j) = T_1 \times \sum_{j=Q_{i-1}+1}^{Q_i} j^b \,, \tag{2.14}$$

where the i^{th} lot begins at unit $Q_{i-1}+1$ (the unit after the one that completed the preceding lot) and ends at unit Q_i. However, this representation is inconvenient because it is not differentiable in the number of units, Q_i. Instead, the incremental lot cost is generally approximated by the integral under the marginal cost curve. Moreover, a continuity correction is generally applied that extends the range of integration by 0.5 units to the *left*

---

[20] This result is probably what Loerch (1999, p. 259) had in mind when he stated, "The cumulative average theory is used when the production environment is unstable, or when there is substantial variation in the costs of consecutive units. In a more stable environment, the unit [Crawford] theory variant is used."

of $Q_{i-1}+1$, or the point $[(Q_{i-1}+1)-0.5] = Q_{i-1}+0.5$; and by 0.5 units to the *right* of Q_i, or the point $Q_i+0.5$. The continuity correction of ±0.5 ensures that the range of integration equals the lot size, $(Q_i+0.5)-(Q_{i-1}+0.5) = Q_i-Q_{i-1}$; absent the correction, the range of integration would fall short of the lot size by 1.0.

Performing the integration, the incremental lot cost is approximated by:

$$TC_i - TC_{i-1} \approx \int_{Q_{i-1}+0.5}^{Q_i+0.5} T_1 z^b \, dz = \frac{T_1}{1+b} \times \left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b} \right], \qquad (2.15)$$

with corresponding lot average cost:[21]

$$LAC_i = \frac{TC_i - TC_{i-1}}{Q_i - Q_{i-1}} \approx \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})} \times \left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b} \right]. \qquad (2.16)$$

The midpoint of the i^{th} lot, $\overline{Q}_i(b)$, is defined as the quantity whose marginal cost is equal to the lot average cost. Setting the marginal cost $T_1 \times [\overline{Q}_i(b)]^b$ equal to LAC_i and solving yields the lot midpoint:

$$\overline{Q}_i(b) = \left(\frac{\left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b} \right]}{(1+b)\times(Q_i - Q_{i-1})} \right)^{1/b}, \qquad (2.17)$$

for $-1 < b < 0$. Note the functional dependence of the lot midpoint on the unknown coefficient, b.

The lot midpoint is illustrated in Figure 2.5 for an initial lot consisting of 20 units with learning slope $\rho = 0.9$. The existence of $(Q_{i-1}+0.5) \le \overline{Q}_i(b) \le (Q_i+0.5)$ is guaranteed because the integrand in equation (2.14) is continuous.[22] Thus, there always exists $\overline{Q}_i(b)$ such that $TC_i - TC_{i-1}$ may be written as the integrand at $\overline{Q}_i(b)$ multiplied by the range of integration, $(Q_i+0.5)-(Q_{i-1}+0.5) = Q_i-Q_{i-1}$. That is, $TC_i - TC_{i-1} = T_1 \times [\overline{Q}_i(b)]^b \times (Q_i - Q_{i-1})$, or LAC_i equals the marginal cost at unit $\overline{Q}_i(b)$.

---

21 The continuity correction of ±0.5 is explored by Camm, Evans and Womer (1987). They conclude that the correction, while not exactly reproducing the discrete sum, provides a close approximation. The exact correction always differs from ±0.5, but cannot be determined in advance without knowledge of the learning coefficient, b. Lee (1997, pp. 35–41) also investigates the accuracy of the continuity correction, but the additional terms that he suggests (based on the Euler-Maclaurin summation formula) are cumbersome in practice.

22 This result is the mean value theorem for integrals; see Taylor and Mann (1972, p. 47).

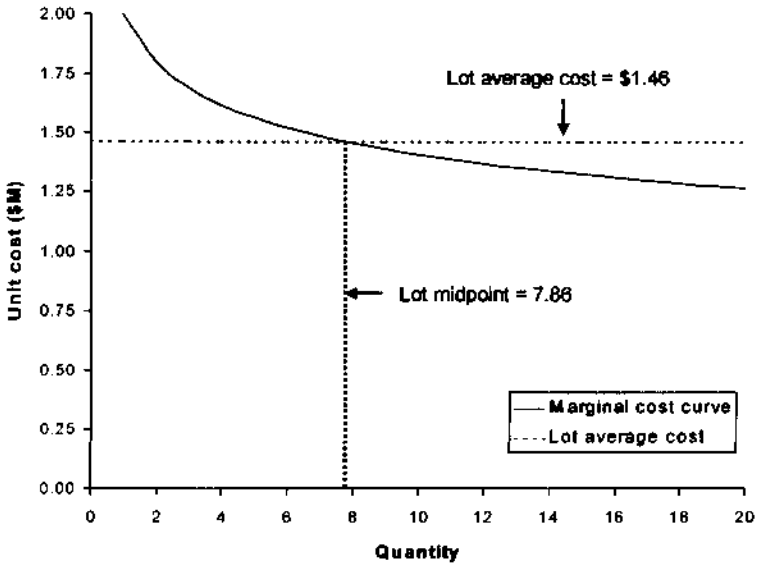


Figure 2.5. Illustration of Lot-Midpoint Calculation

At $b = 0$, there is no learning; thus, *any* point in the interval serves as a lot midpoint. As $b \to -1$ the anti-derivative in equation (2.14) approaches a logarithmic function, and $\overline{Q}_i(b)$ approaches the polar form $(Q_i - Q_{i-1}) / \ln[(Q_i + 0.5)/(Q_{i-1} + 0.5)]$, which also can be shown to lie in the interval $[(Q_{i-1} + 0.5),\ (Q_i + 0.5)]$.

2.6 Error distributions for learning curves and CERs

The error distributions for both learning curves and CERs may take a variety of forms. Figures 2.6 through 2.8 illustrate three possibilities. Figure 2.6 depicts a CER in which the error terms are:

- Symmetric (in fact, normally distributed), and
- Constant variance for all values of the cost driver (in this case, weight).

A mathematical expression of this CER might be:

$$\text{Unit cost} = b_0 + b_1 \times \text{Weight} + u_i\ , \tag{2.18}$$

where u_i is normally distributed with mean zero.

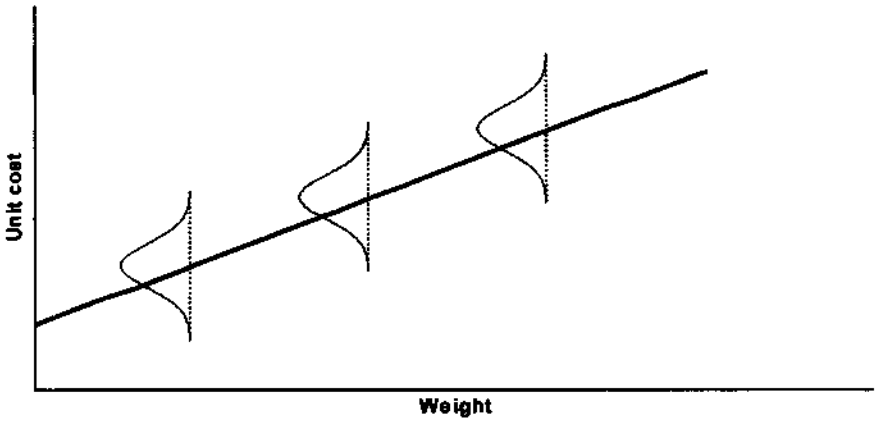


Figure 2.6. CER with Additive, Normal Errors

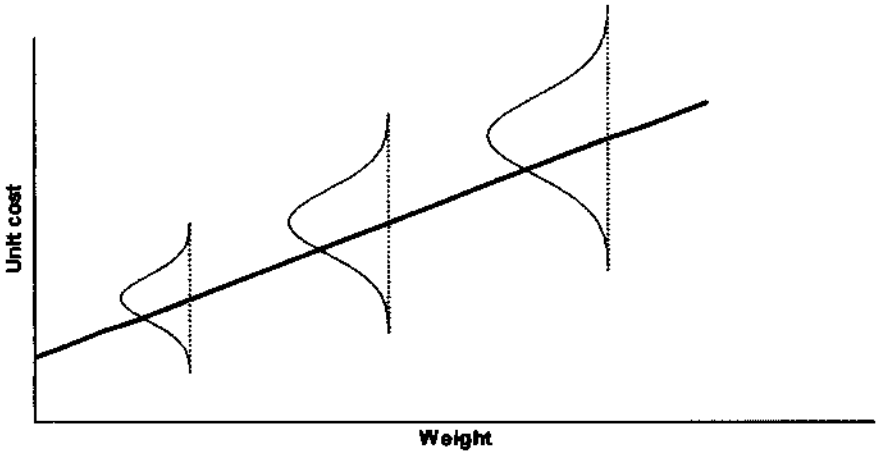


Figure 2.7. CER with Multiplicative, Normal Errors

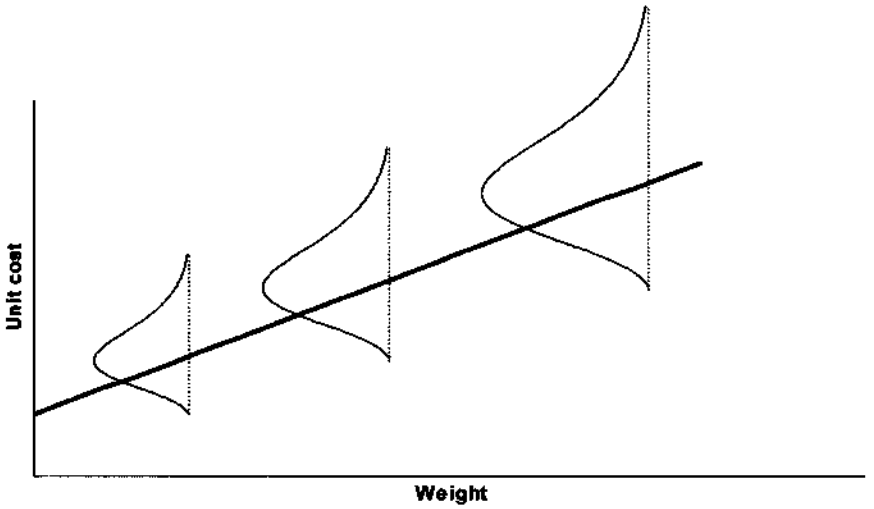


Figure 2.8. CER with Multiplicative, Log-Normal Errors

Figures 2.7 and 2.8 illustrate two different multiplicative regression models. Figure 2.7 depicts a CER in which the error terms are:

- Symmetric (again, normally distributed), but
- Standard deviation proportional to the value of the cost driver.

A mathematical expression of this CER might be:

$$\text{Unit cost} = \left(b_0 + b_1 \times \text{Weight}\right) \times u_i \ , \tag{2.19}$$

where u_i is now normally distributed with mean 1.0. We will refer to this assumption hereafter as the "multiplicative normal assumption."

Finally, Figure 2.8 depicts a CER in which the error terms are:

- Skewed to the right (in fact, log-normally distributed), and
- Standard deviation proportional to the value of the cost driver.

A mathematical expression of this CER might be:

$$\text{Unit cost} = \left(b_0 + b_1 \times \text{Weight}\right) \times e^{v_i} \ , \tag{2.20}$$

where v_i is zero-mean normally distributed and e^{v_i} is, therefore, log-normally distributed. This assumption (hereafter, the "log-normal assumption") was made in the seminal papers on log-linear regression by Goldberger (1968), Heien (1968), and Bradu and Mundlak (1970), among others.

The three candidate distributions differ in two major respects:

- Additive versus multiplicative errors, and
- Symmetric versus right-skewed errors.

Figures 2.7 and 2.8, representing multiplicative regression models, allow for non-constant variance or heteroscedasticity. This property is probably more compelling for CERs than for learning curves. A single CER might be estimated over a wide range of systems that vary greatly in weight, speed, and most importantly, unit cost. The error variance is often larger for the heavier, faster, and more expensive systems, so that heteroscedasticity becomes an important property to accommodate. By contrast, the sequential unit costs for a single system, modeled with a learning curve, seldom vary by an order of magnitude. Thus, heteroscedasticity is a less important property for learning

curves than for CERs. In the case of CERs, the remaining issue is whether the errors are symmetric (Figure 2.7) or skewed (Figure 2.8). Skewness reflects the common observation that, at least for military weapon systems, large over-runs are more common than large under-runs of the same absolute (i.e., dollar) magnitude.[23]

Although both heteroscedasticity and skewness have been observed in cost data for weapon systems, it is also tempting to argue, in light of the Central Limit Theorem (CLT), that the error distribution must be additive and symmetric normal. The cost of a weapon system may always be expressed additively as the sum of the costs of its sub-systems; in turn, as the grand sum of the costs of their sub-sub-systems, etc. This hierarchical linear structure, known as a work breakdown structure (WBS), is illustrated in Table 2.1 for an unmanned space vehicle. The entry in the "Index" column indicates the position of each element down to the third level of indenture in the WBS.

The elementary textbook version of CLT states that a sum of independent, identically-distributed (*iid*) random variables approaches a normal distribution as the number of terms tends to infinity. A more sophisticated version of CLT allows for non-identical distributions, as long as each term in the sum has finite variance, and each term contributes at most a negligible fraction to the overall variance of the sum (the latter known as the non-domination condition; see Feller (1971, p. 262) or Rao (1973, p. 128)). Note, however, that even the more sophisticated version of CLT apparently requires independence among all the random variables.

---

[23] Under the log-normal assumption, errors of (e.g.) +0.4 (or greater) and –0.4 (or greater) are equally likely in predicting the *natural logarithm* of cost. A logarithmic error of +0.4 implies that cost exceeds the model prediction by about 50 percent [{exp(+0.4)}-1.0 = 0.492]. However, an equally likely logarithmic error of –0.4 implies a cost under-run of only about 33 percent. Thus, the errors in predicting dollar costs are skewed to the right.

Table 2.1 Work Breakdown Structure for an Unmanned Space Vehicle

Index	Level 1	Level 2	Level 3
1.	Spacecraft		
1.1		Structure, Interstage / Adapter	
1.2		Thermal Control	
1.3		Attitude Determination Control System	
1.3.1			Attitude Determination
1.3.2			Reaction Control System
1.4		Electrical Power Supply	
1.4.1			Power Generation
1.4.2			Power Storage
1.4.3			Power Conditioning & Distribution
1.5		Telemetry, Tracking & Command	
1.5.1			Transmitter
1.5.2			Receiver / Exciter
1.5.3			Transponder
1.5.4			Digital Electronics
1.5.5			Analog Electronics
1.5.6			Antennas
1.5.7			RF Distribution
1.6		Propulsion – Apogee Kick Motor	
2.	Communications Payload		
2.1		Transmitter	
2.2		Receiver / Exciter	
2.3		Transponder	
2.4		Digital Electronics	
2.5		Analog Electronics	
2.6		Antennas	
2.7		RF Distribution	
3.	Integration, Assembly & System Test (IA&T)		
4.	Program Level		
4.1		Program Management	
4.2		Systems Engineering	
4.3		Data	

If correlations are present, it may be possible to circumvent the independence condition by combining the random variables into aggregates that absorb the correlations, such that there are no correlations among the aggregates. Then we would attempt to apply CLT to the aggregates. In the example of the unmanned space vehicle, we might have the following correlation sub-matrix for the IA&T sub-system (first column) and the three sub-sub-systems under Program Level (second through fourth columns):

$$\begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.871 & 0.517 \\ 0.0 & 0.871 & 1.0 & 0.871 \\ 0.0 & 0.517 & 0.871 & 1.0 \end{pmatrix} \tag{2.21}$$

The three sub-sub-systems must be combined in order to absorb the correlations among them. The resulting aggregate, Program Level costs, is uncorrelated with IA&T costs (because each of the sub-sub-systems under Program Level is uncorrelated with IA&T). However, we have reduced the number of distinct terms from 4 down to 2 in the (partial) sum defining total system cost.

When working with uncorrelated aggregates, two questions remain:

- (i) Are there enough uncorrelated aggregates to approximate the infinite number of terms that CLT requires?
- (ii) Do the aggregates satisfy the non-domination condition?

Proceeding in the other direction, simply subdividing the system under study into sub-systems, sub-sub-systems, etc. will increase the number of terms in the sum defining total system cost. Unfortunately, however, it will also almost inevitably increase the correlations among the terms. For example, consider subdividing an aircraft's wings (typically a single cost term) into distinct left and right wings. Because the left and right wings are manufactured (often by a subcontractor) as a single package, they will have identical costs and thus a correlation of 1.0. Continuing, one could further subdivide the surface of a single wing into a large number of square-inch sub-surfaces. However, the costs of adjacent sub-surfaces will again be highly (though perhaps not perfectly) correlated. We see that attempts to increase the number of terms through subdivision will violate the assumption of independence among the terms. The total cost of the aircraft's wings may not be amenable to subdivision, but could still be uncorrelated with the costs of the aircraft's other major sub-systems. But then, we again face the two issues of number of terms and non-domination.

The theoretical hypotheses of the CLT may not hold for military weapon systems and, as we have already pointed out, both heteroscedasticity and skewness have been observed in cost data for many such systems. Thus, the CLT's conclusion of additive, normal errors is far from inevitable.

2.7 Error distributions for multiplicative learning curves

Although additive normal errors are not inevitable, they may nonetheless be the correct specification for many learning curves. As we argued in Chapter 1, a modern statistician might construct an additive regression model from the learning curve's prediction of lot average cost:

$$LAC_i = \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})}\times\left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b}\right] + u_i \ , \quad (2.22)$$

and apply NLS to minimize the regression sum-of-squares:

$$\sum_{i=1}^{n}\left(LAC_i - \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})}\times\left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b}\right]\right)^2 . \quad (2.23)$$

In the remainder of this section, however, we will investigate instead the two multiplicative regression models (depicted in Figures 2.7 and 2.8). We do so because the two estimation techniques that we wish to discuss for the duration of this chapter — lot-midpoint iteration and lot-midpoint NLS — both (at least implicitly) assume that the error terms are additive on the *logarithmic* scale, thus multiplicative on the original (dollar) scale. In addition, multiplicative regression models accommodate heteroscedasticity which, although less compelling for learning curves than for CERs, is still sometimes observed.

The two multiplicative CERs from the previous section can be adapted as learning curves. Under the multiplicative normal assumption we have the following model for lot average cost:

$$LAC_i = T_1\times[\bar{Q}_i(b)]^b \times u_i \ , \quad (2.24)$$

where u_i is normally distributed with mean 1.0 and constant variance for all lots $i = 1,\ldots,n$. This model was proposed by Lee (1997, pp. 55–56). A logarithmic transformation yields:

$$\ln(LAC_i) = \ln(T_1) + b\ln[\bar{Q}_i(b)] + \ln(u_i)\ . \tag{2.25}$$

Alternatively, under the log-normal normal assumption we have instead the following model:

$$LAC_i = T_1 \times [\bar{Q}_i(b)]^b \times e^{v_i}, \tag{2.26}$$

where v_i is normally distributed with mean 0.0 and constant variance for all lots $i = 1,\ldots,n$. In this case, e^{v_i} is log-normally distributed and a logarithmic transformation yields:

$$\ln(LAC_i) = \ln(T_1) + b\ln[\bar{Q}_i(b)] + v_i\ . \tag{2.27}$$

Both equation (2.25) and equation (2.27) suggest the use of regression analysis to estimate the parameters T_1 and β.

Lee appeals to a first-order Taylor series approximation, specifically $\ln(u_i) = \ln[1 + (u_i - 1)] \approx u_i - 1$, to argue that the error term in the logarithmic equation (2.25) is approximately zero-mean normally distributed. However, if u_i is normally distributed with mean 1.0, then the event $u_i < 0$ occurs with positive probability, yet leaves $\ln(u_i)$ undefined. Strictly speaking, although equation (2.24) is certainly an admissible representation of the lot-midpoint model, one should avoid the logarithmic transformation to equation (2.25) because the error term is not well-defined. This argument appears to further suggest that, under Lee's multiplicative normal assumption, one should avoid estimation methods that operate on the logarithmic data (e.g., lot-midpoint iteration or non-linear least squares applied to equation (2.25)).

This advice is a bit too severe. One should, of course, avoid any attempts to take logarithms of *measured* predictor variables (x_i) that can range over non-positive values. However, statistical software can certainly execute lot-midpoint iteration or NLS independent of the analyst's technical assumptions on the error term. Speaking anthropomorphically, when executing one of these estimation methods, the computer does not "know" that there is a minor technical problem with the definition of the error term; the computer cannot distinguish between the representations (2.25) and (2.27). We conclude that all of the same estimation methods may, at least in a mechanical sense, be applied under either representation of the lot-midpoint model — even though methods that require the logarithmic transformation are, strictly speaking, incompatible with Lee's multiplicative normal assumption. Of course, there is no such difficulty with estimation

methods that avoid the logarithmic transformation and operate on equation (2.24) directly (e.g., NLS applied to equation (2.24)).

Further, when the variance of the error term is small, the multiplicative-normal and log-normal distributional assumptions are nearly equivalent in a numerical sense. The solid curve in Figure 2.9 is a normal distribution with mean 1.0 and standard deviation 0.15. As we show in Chapter 4, the latter value is the standard error (at convergence) of the lot-midpoint iteration applied to the data of Table 1.1; moreover, a 15-percent relative error is fairly typical for learning curves. The dashed curve in Figure 2.9 is a log-normal distribution calibrated to have the same mean of 1.0 and standard deviation of 0.15. Although slightly skewed, it seems appropriate to describe the latter distribution as approximately normal. In particular, the skewness and kurtosis of 0.45 and 3.37 are close to the theoretical normal values of 0.0 and 3.0.

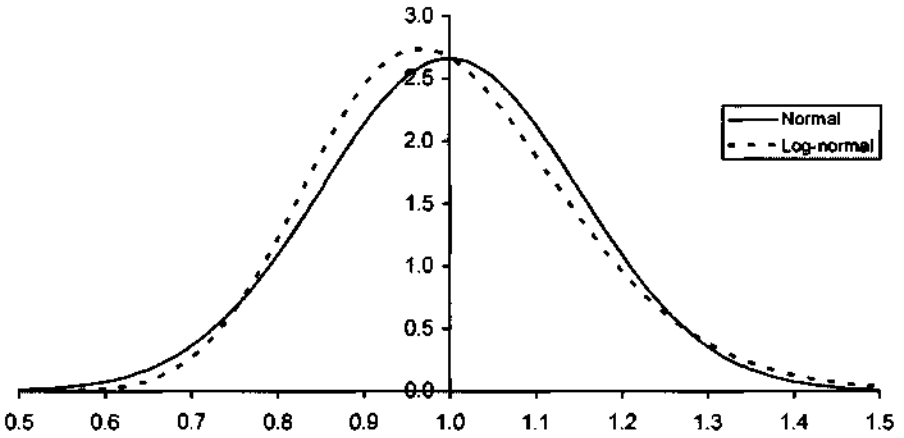


Figure 2.9. Normal Approximation to the Error Distribution, Standard Deviation = 0.15

In addition, as Figure 2.10 indicates, the skewness and kurtosis of the log-normal approach the normal values even more closely as the standard deviation shrinks (the normal kurtosis of 3.0 is shown as a benchmark).[24] We conclude that the multiplicative normal assumption and the log-normal assumption are nearly equivalent when the variance of the error term is small.

---

[24] The approximation of a log-normal distribution by a normal distribution, when the variance is small, is sketched in Johnson and Kotz (1970, Volume 1, Chapter 14, pp. 117–118).

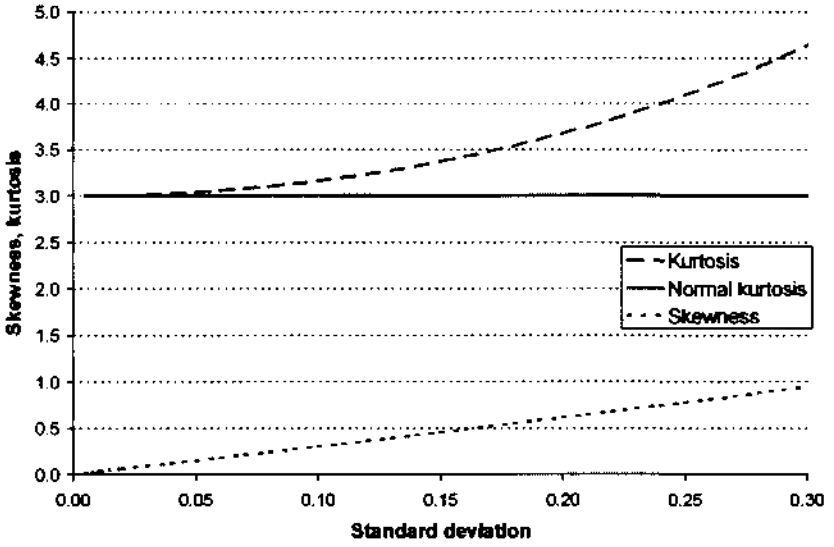


Figure 2.10. Skewness and Kurtosis of Log-Normal Distribution

2.8 Non-linear least squares (NLS)

Equations (2.25) and (2.27) are not immediately amenable to linear least squares, because $\overline{Q}_i(b)$ is functionally dependent on b. At least two approaches are available to resolve this non-linearity. First, because both equations have an additive, homoscedastic (constant variance) error structure, they are amenable to non-linear least squares. That is, we may choose the parameters T_1 and b to minimize the regression sum-of-squares:

$$\sum_{i=1}^{n}\left(\ln(LAC_i) - \ln(T_1) - b\ln[\overline{Q}_i(b)]\right)^2. \tag{2.28}$$

The log-normal assumption exactly describes equation (2.27) and, as we have argued, approximately describes equation (2.25). Under this assumption, the NLS estimator is consistent and asymptotically normally distributed, and its asymptotic covariance matrix may be developed as follows.[25] For the general non-linear regression model, $y_i = f(x_i, \beta)$, let J denote the $n \times m$ Jacobian matrix of the predictor function $f(x_i, \beta)$ with respect to the m parameters β:

---

[25] See Bard (1974, pp. 176–179) or Seber and Wild (1989, pp. 21–25). Donaldson and Schnabel (1987) demonstrate the superiority of this form of the covariance matrix over two asymptotically equivalent alternatives.

$$J_{n\times m} \equiv \left[\partial f(x_i,\beta)/\partial \beta_j\right]_{n\times m}, \tag{2.29}$$

with rows $J_1,\ldots,J_n$. The $m \times m$ asymptotic covariance matrix is given by:

$$Var(\hat{\beta}) = \sigma^2 \times \left(J^T J\right)^{-1}, \tag{2.30}$$

where the dispersion σ^2 is estimated by the minimized regression sum-of-squares, expression (2.28), divided by the degrees-of-freedom, $n-m$. Equivalently, the asymptotic covariance matrix may be written as:

$$Var(\hat{\beta}) = \sigma^2 \times \left(\sum_{i=1}^{n}[J_i^T J_i]\right)^{-1}, \tag{2.31}$$

where each term $[J_i^T J_i]$ is itself an $m \times m$ outer product matrix.

Note that, in the case of lot-midpoint estimation, the predictor function $f(x_i,\beta) = \ln(T_1) + b\ln[\bar{Q}_i(b)]$ is highly non-linear in light of the definition of the lot midpoint (equation (2.17)). The Jacobian matrix of this function is particularly difficult to compute analytically. However, software packages that compute NLS estimates also provide the asymptotic covariance matrix. They generally approximate the Jacobian matrix by numerical differentiation.

As a special case, a Wald test may be used to test a single coefficient against zero, using the result $\hat{\beta}_j / [Var_{jj}(\hat{\beta})]^{1/2} \rightarrow N(0,1)$. It may seem tempting to use the t–distribution for testing in finite samples, because the t–test is exact in linear models (i.e., those estimated using OLS) and because the t–distribution tends toward normality in large samples. However, the properties required to construct the exact t–test (i.e., $\hat{\beta}$ normally distributed, $\hat{\sigma}^2$ proportional to a χ^2 random variable, and $\hat{\sigma}^2$ independent of $\hat{\beta}$) are guaranteed to hold only asymptotically in non-linear regression models. Although the true, finite-sample distribution of the "t–statistic" tends toward normality, as does the t–distribution, the finite-sample distribution is not necessarily a t–distribution. Some have argued that the t–distribution is no more accurate than simply applying the asymptotic normal distribution in finite samples.[26] However, Gallant (1987, pp. 24-25) offers limited Monte Carlo evidence in favor of using the t–distribution.

---

[26] See Dhrymes (1974, pp. 166–167) or Schmidt (1976, pp. 60–61).

Alternatively, under the log-normal assumption, we can also apply likelihood ratio tests. Let SSE^r denote the sum-of-squares in expression (2.28) under r independent linear restrictions, and let SSE^u denote the unrestricted sum-of-squares. Then the likelihood-ratio statistic $-2\times\ln(L^r/L^u) = n\times\ln(SSE^r/SSE^u)$ has an asymptotic χ^2_r distribution.[27]

Hypothesis tests based on the asymptotic covariance matrix (i.e., Wald tests) and those based on the likelihood-ratio statistic are asymptotically equivalent. The two methods differ only in the required computation. The asymptotic covariance matrix requires the matrix calculation indicated in equation (2.30) or (2.31). However, the regression model need only be estimated once. By contrast, likelihood ratio testing avoids the matrix calculation, but requires estimation of the unrestricted model as well as separate estimation of each restricted model under test.

2.9 Lot-midpoint iteration

Lot-midpoint iteration is an alternative approach to resolving the non-linearity in $\overline{Q}_i(b)$. Begin with an initial estimate of b, denoted $b^{(0)}$. Fix $b=b^{(0)}$ in the definition of the lot midpoint, $\overline{Q}_i(b^{(0)})$, and minimize the regression sum-of-squares (expression (2.28)) with respect to b as the *regression coefficient only*. The minimum occurs at a new estimate, $b^{(1)}$. Now fix $b=b^{(1)}$ in the definition of the lot midpoint $\overline{Q}_i(b^{(1)})$ and again minimize with respect to b as the regression coefficient only. In general, estimate the following sequence of regressions:

$$\ln(LAC_i) = \ln(T_1) + b^{(p+1)}\ln[\overline{Q}_i(b^{(p)})] + v_i \ , \tag{2.32}$$

for $p=0,1,2,\ldots$. Finally, the lot-midpoint estimator is defined as the limit of the sequence:

$$b_1 = \lim_{p\to\infty} b^{(p)} \ , \tag{2.33}$$

when the limit exists. In practice, the lot-midpoint estimator is taken where the sequence converges within a pre-specified numerical tolerance.[28]

---

[27] See Goldfeld and Quandt (1972, p. 74) or Seber and Wild (1989, p. 230).

[28] Although lot-midpoint iteration is ubiquitous in cost analysis, we do not know its exact origins. However, Womer and Patterson (1983) attribute it to RAND Corporation (1971).

Cost analysts have been using lot-midpoint iteration for nearly half a century without questioning the theoretical basis for this method. We asked the following seven questions regarding lot-midpoint estimation:

1. Is lot-midpoint iteration equivalent to (i.e., does it yield the same point estimates as) non-linear least squares?
2. Is there a distributional assumption under which lot-midpoint iteration is equivalent to maximum-likelihood estimation?
3. Does lot-midpoint iteration maximize or minimize *any* continuously differentiable function of the parameters T_1 and b (if not a sum-of-squares or a likelihood function, perhaps some other function)?
4. Is lot-midpoint iteration guaranteed to converge, or might the iteration continue forever?
5. If lot-midpoint iteration does converge, is the solution unique; or might the iteration converge to two (or more) distinct solutions depending upon the starting values?
6. If a particular lot-midpoint iteration has two distinct solutions, on what basis do we choose one over the other?
7. If lot-midpoint iteration does converge, how accurate are the standard errors from the final regression step?

We were able to answer some, but not all of these questions, using theoretical analysis. That analysis, involving rather advanced mathematics, is presented in its entirety in the Appendix and merely summarized here. We also learned more about lot-midpoint iteration from the Monte Carlo analysis reported in Chapter 5.

We demonstrate in the Appendix that lot-midpoint iteration is *not* equivalent to either NLS or MLE. In fact, lot-midpoint iteration does not maximize or minimize *any* continuously differentiable function of the parameters T_1 and b.

The issues of existence, uniqueness, and convergence to a solution would typically be addressed by the theory of contraction mappings. To understand that theory, consider the elementary case of the geometric sequence $b, b^2, b^3, \ldots$. That sequence converges to zero if $|b| < 1$. For example, if $b = 1/2$ we have the convergent sequence $1/2, (1/2)^2, (1/2)^3, \ldots$; or if $b = -1/2$ we again have a convergent sequence $-1/2, (-1/2)^2, (-1/2)^3, \ldots = -1/2, 1/4, -1/8, \ldots$. Conversely, the geometric sequence diverges to infinity if $|b| > 1$. For example, if $b = 2$ we have the divergent sequence $2, 2^2, 2^3, \ldots$.

This elementary theory can be generalized to a non-linear, multivariate situation such as lot-midpoint iteration. As we develop in the Appendix, there is a Jacobian matrix associated with lot-midpoint iteration, and we may compute the eigenvalues of that matrix. By Ostrowski's theorem, if the eigenvalues are all less than 1.0 in absolute value throughout a region of parameter space (or, equivalently, if the maximum absolute eigenvalue is less than 1.0 throughout the region), then quite remarkably:

- there exists a pair of values T_1 and b in the region that balance equation (2.32);
- the pair T_1 and b is unique in the region; and
- iteration, starting from any point in the region, generates a sequence that converges to the unique root.

On the other hand, if the maximum absolute eigenvalue exceeds 1.0, there is no universal guarantee that a solution pair T_1 and b exists to balance equation (2.32); that a solution, if it exists, is unique; or that a solution can be approximated by a finite number of steps of lot-midpoint iteration.

In the lot-midpoint problem, the maximum absolute eigenvalue may lie on either side of 1.0. As we will see in Chapter 4, the maximum absolute eigenvalue exceeds 1.0 for Lee's tactical missile data. Lot-midpoint iteration may still converge (and, indeed, it does converge when applied to Lee's data), because the eigenvalue condition is *sufficient but not necessary*. That is, an iterative scheme *must* converge if the eigenvalue condition is satisfied; it *may* still converge even if the eigenvalue condition is violated. Again, there is no universal guarantee — existence, uniqueness, and convergence of lot-midpoint iteration may vary from one data set to another.

The theoretical possibility of multiple solutions is particularly disquieting in light of the failure of lot-midpoint iteration to maximize any continuously differentiable objective function. In a maximization problem, we can always compare the value of the objective function at two distinct local maxima, disposing of the smaller value because it cannot be the global maximum. But because lot-midpoint iteration does not maximize any such objective function, if two distinct solutions are located we have no basis to choose between them.

Finally, we briefly turn to the statistical (as opposed to mathematical) properties of lot-midpoint iteration. Under the log-normal assumption, the regression standard errors, confidence intervals, and significance tests are apparently available from

conventional OLS regression theory applied to equation (2.32) at convergence. However, the lot-midpoint variable $\overline{Q}_i(b)$ is unknown even at convergence, and is replaced with the estimate $\overline{Q}_i(b_1)$. The conventional approach does not recognize this additional uncertainty, thereby leading to an underestimate of the standard error of b.

More ominously, Schmidt (1976, pp. 93-96) reports that even if the correct standard errors were known, the "t–statistics" would not necessarily follow a t–distribution. This difficulty arises because the errors in the final lot midpoints propagate into the estimated slope parameter (b_1) such that the latter is no longer normally distributed. In addition, the usual theoretical guarantees of consistent OLS estimates no longer apply. The lot-midpoint estimates may still be consistent, but that determination would require either a special theoretical investigation or an exhaustive Monte Carlo analysis.

In fact, our Monte Carlo results presented in Chapter 5 suggest that lot-midpoint iteration is consistent. However, those results also show that among all the estimation methods we consider, only lot-midpoint iteration is sensitive to serial correlation in the error terms. Serial correlation is ubiquitous in cost analysis although, as we argued in Section 2.4, serial correlation can often be reduced by transforming the data series from cumulative average cost to lot average cost prior to estimation. Nonetheless, both the lack of a theoretical foundation and the sensitivity to serial correlation conspire to render lot-midpoint iteration an unattractive statistical procedure.

2.10 Retransformation bias

The issue of retransformation bias arises regardless of whether the lot model is estimated by NLS or by lot-midpoint iteration. Either approach yields estimates of the parameters in equation (2.26). But with v_i normally distributed in equation (2.27), e^{v_i} is *log-normally* distributed in equation (2.26). Letting $\hat{\sigma}$ denote the standard error of the logarithmic lot-midpoint regression, the mean of e^{v_i} is consistently estimated by $\exp(\hat{\sigma}^2/2) > 1$. Unless this factor is accommodated, the predictions of lot average cost from equation (2.26) will be systematically too low. One way to accommodate the log-normal mean is to replace the estimated intercept $\hat{T}_1$ for that equation with $\hat{T}_1 \times \exp(\hat{\sigma}^2/2)$.

This correction factor is commonly used, and is advocated in a well-known paper by Miller (1984) among others. However, alternative retransformation factors are available that do not rely as heavily on distributional assumptions. The following method

relies only on the logarithmic nature of the transformation, without reference to the distributional assumption at all.[29] Suppose a random variable X has known mean μ and standard deviation σ. Consider an exact transformation $Y = f(X)$. We can expand Y around μ in a second-order Taylor series approximation:

$$Y = f(X) \approx f(\mu) + (X-\mu)\times f'(\mu) + \frac{(X-\mu)^2 f''(\mu)}{2} . \qquad (2.34)$$

Taking the expectation of both sides of equation (2.34) yields:

$$E(Y) \approx f(\mu) + \frac{\sigma^2 f''(\mu)}{2} . \qquad (2.35)$$

In our situation, we have estimated the mean ($\mu = 0$) and the standard deviation (σ) of v_i in equation (2.27). Our objective is to estimate the mean of e^{v_i} in equation (2.26). The transformation $Y = f(X)$ is now the exponential function. Specializing equation (2.35) to the exponential function (and recalling that $\mu = 0$) yields:

$$E(e^{v_i}) = 1 + (\sigma^2/2) , \qquad (2.36)$$

which is itself a first-order approximation to the log-normal correction factor, $\exp(\sigma^2/2)$. Along similar lines, all of the other correction factors in Miller (1984) (e.g., roots and powers) can be reproduced using only the form of the transformation, without reference to the distributional assumption.

The use of correction factors may affect not only the intercept T_1, but also its standard error. Two different cases must be distinguished. Using lot-midpoint NLS, we can parameterize the model to estimate T_1 directly rather than its logarithm. Using lot-midpoint iteration, however, we estimate the intercept as $\ln(T_1)$. The statistical software generally provides a standard error for this quantity, but not for T_1 itself. Moreover, we are ultimately interested in the standard error of $\hat{T}_1 \times \exp(\hat{\sigma}^2/2)$. The composite effect of the anti-logarithmic transformation and the log-normal correction factor will now be calculated.

---

[29] See Seiler (1987) or Lurie, Goldberg, and Robinson (1993, p. 6).

Before proceeding, we must collect a few results on non-linear regression.[30] We have an estimate of $a = \ln(T_1)$ from lot-midpoint iteration, its sampling variance $V_{11} = Var(\hat{a})$, and the residual variance $\hat{\sigma}^2$. The residual variance itself has asymptotic sampling variance $Var(\hat{\sigma}^2) \to 2\sigma^4/n$, where n is the sample size; moreover, $\hat{a}$ and $\hat{\sigma}^2$ are asymptotically independent. Thus, the asymptotic covariance matrix of the two parameters is given by:

$$Cov(\hat{a}, \hat{\sigma}^2) \to \begin{pmatrix} V_{11} & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}. \tag{2.37}$$

We are interested in computing the asymptotic variance of $T^* = \hat{T}_1 \times \exp(\hat{\sigma}^2/2) = \exp[\hat{a} + (\hat{\sigma}^2/2)]$. The gradient of T^* with respect to the two parameters is:

$$\nabla T^* = \frac{\partial T^*}{\partial(\hat{a}, \hat{\sigma}^2)} = \begin{pmatrix} T^* \\ T^*/2 \end{pmatrix}. \tag{2.38}$$

The asymptotic variance of the transformed intercept follows from a first-order Taylor series approximation:

$$Var(T^*) \to (\nabla T^*)' \, Cov(\hat{a}, \hat{\sigma}^2) \, \nabla T^* = \left[V_{11} + \frac{\sigma^4}{2n} \right] \times (T^*)^2 . \tag{2.39}$$

Of course, remember that the entire formula may represent an underestimate because the sampling variance from lot-midpoint iteration V_{11} is itself underestimated.

Duan (1983, p. 608) extends this result to compute the asymptotic variance of the prediction of lot average cost for any observation, again assuming estimation by (in our terminology) lot-midpoint iteration and use of the log-normal correction factor.[31] Our prediction is $\hat{LAC}_i = \exp[\hat{a} + \hat{b} \ln \bar{Q}_i + (\hat{\sigma}^2/2)]$, and its asymptotic variance is given by:

---

[30] These results are found in Seber and Wild (1989), sections 2.1.2, 2.2.1, and 5.1; or Gallant (1987), pp. 47 and 260–261.

[31] The details of Duan's derivation are found in Appendix B to Duan, Manning, Morris, and Newhouse (1982).

$$Var(\hat{LAC}_i) \rightarrow \left[w_i V w_i^T + \frac{\sigma^4}{2n} \right] \times (\hat{LAC}_i)^2, \qquad (2.40)$$

where $w_i = \left(1, \ln \bar{Q}_i\right)$ and V is the 2×2 covariance matrix of $(\hat{a}, \hat{b})$ at convergence.[32]

Using the same methods as Duan, we can compute the asymptotic variance of the prediction assuming estimation by lot-midpoint NLS (parameterized in terms of T_1 directly) and use of the log-normal correction factor. In this case our prediction is $\hat{LAC}_i = \exp[\ln \hat{T}_1 + \hat{b} \ln \bar{Q}_i + (\hat{\sigma}^2/2)]$, and its asymptotic variance is given by:

$$Var(\hat{LAC}_i) \rightarrow \left[\sigma^2 \times w_i \left(J^T J\right)^{-1} w_i^T + \frac{\sigma^4}{2n} \right] \times (\hat{LAC}_i)^2, \qquad (2.41)$$

where we redefine $w_i = \left(1/\hat{T}_1, \ \ln \bar{Q}_i\right)$.

We close this chapter with a reminder that the sample sizes in cost analysis are often quite small, making the usefulness of asymptotic properties somewhat problematic. Unfortunately, when working with highly non-linear combinations of random variables, asymptotic properties are often the only analytical tools we have available for statistical inference. More empirically based methods such as bootstrapping, though not traditionally applied in cost analysis, are also worthy of consideration.[33]

---

32 Our prediction is consistent, but may be biased in small samples. Eskew and Lawler (1993, 1994) propose an alternative prediction, which they argue has a smaller bias: $\hat{LAC}_i = \exp[\hat{a} + \hat{b} \ln \bar{Q}_i + (\hat{\sigma}^2/2) - (w_i V w_i^T/2)]$. They also cite Bradu and Mundlak (1970) for an unbiased prediction. However, the latter involves a cumbersome, infinite series expansion that must be truncated for practical application.

33 See, for example, Efron and Tibshirani (1993) or Davison and Hinkley (1997).

This page intentianally left blank.

3. ALTERNATIVE ESTIMATION METHODS

We discuss four estimation methods in this chapter. We frame much of the discussion in terms of a particular distributional assumption, namely multiplicative normal errors. *Some* distributional assumption (though not necessarily this one) is required for one of the estimation methods, maximum likelihood. The other three methods are minimum percentage error, iteratively reweighted least squares, and maximum quasi-likelihood. We conclude the chapter with a comparison of the four estimation methods, extending the comparison to include the two estimation methods (lot-midpoint NLS and lot-midpoint iteration) that apply to learning curves but not to CERs.

3.1 Definitions and assumptions

A multiplicative regression model has the form:

$$y_i = f(x_i, \beta) \times u_i, \tag{3.1}$$

where y_i is the observed response variable, x_i is an observed vector of k predictor variables, β is a vector of m coefficients to be estimated, and u_i is the unobserved error term for the i^{th} observation. At this juncture, we assume only that the error terms $\{u_i\}$ have finite variance, are statistically independent of each other, and statistically independent of the predictor variables $\{x_i\}$. However, we make no particular distributional assumption on the $\{u_i\}$.

In this model, we note that y_i has mean:

$$E(y_i) = f(x_i, \beta) \times E(u_i), \tag{3.2}$$

which equals $f(x_i, \beta)$ if u_i has mean 1.0. Also, y_i has variance:

$$Var(y_i) = [f(x_i, \beta)]^2 \times V(u_i). \tag{3.3}$$

This chapter differs from the previous one in that the predictor function $f(x_i,\beta)$ may take on a wider variety of forms. Recall that, in the discussion of lot-midpoint estimation, the predictor function was $f(x_i,\beta) = \ln(T_1) + b\ln[\bar{Q}_i(b)]$. As we noted in Section 2.9, that formulation is somewhat problematic because the lot-midpoint variable $\bar{Q}_i(b)$ is unknown — and has an unknown random distribution — even at convergence.

In the current chapter, we return to the multiplicative regression model of equation (3.1). In the learning-curve context, $f(x_i,\beta)$ would equal expression (1.6) for lot-average cost, repeated here for convenience:

$$f(x_i,\beta) = LAC_i = \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})}\times\left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b}\right]. \qquad (3.4)$$

Thus, we are directly estimating the parameters T_1 and β in the non-linear model for lot-average cost obtained by integrating under the (Crawford) marginal cost curve. We are eschewing the device of lot midpoints and the logarithmic transformation (i.e., equations (2.24) and (2.25)). Moreover, unlike the lot midpoints, the predictor variables in equation (3.4) (i.e., the lot endpoints Q_i and Q_{i-1}) are known and non-random.

In the CER context, $f(x_i,\beta)$ would simply be the CER itself, e.g., equation (1.17), repeated here for convenience:

$$f(x_i,\beta) = \text{Unit cost} = b_0\times\text{Weight}^{b_1}\times\text{Speed}^{b_2}\ . \qquad (3.5)$$

Hence, the methods of this chapter apply equally to both of the primary models used in cost analysis.

3.2 Minimum percentage error

Lee (1997, pp. 47–49) investigated estimation of equation (3.1) when the error terms u_i are statistically independent, and normally (*not* log-normally) distributed with mean 1.0 and variance θ. The likelihood function for this model may be written as:

$$L(\beta,\theta) = \frac{\exp\left[\dfrac{-1}{2\theta}\displaystyle\sum_{i=1}^{n}\left(\dfrac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)^2\right]}{(2\pi\theta)^{n/2}\times\displaystyle\prod_{i=1}^{n} f(x_i,\beta)}. \qquad (3.6)$$

The log-likelihood function is equal to:

$$\begin{aligned} l(\beta,\theta) &= \ln[L(\beta,\theta)] \\ &= \frac{-1}{2\theta}\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)^2 - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta) - \sum_{i=1}^{n}\ln f(x_i,\beta). \end{aligned} \tag{3.7}$$

Ignoring θ for the moment, one is tempted to estimate the parameter vector β by minimizing the sum-of-squares that appears in the numerator of equation (3.6), or equivalently, by minimizing the same sum-of-squares that appears as the first term on the final line of equation (3.7). Lee (1997, p. 48) explicitly advocates this approach, asserting that, "the exponential in the numerator of [equation (3.6)] is much more sensitive to variations in [β] than is the denominator." This approach is also recommended by Book and Young (1995, 1997), who label it the Minimum Percentage Error (MPE) procedure. The resulting estimator, which we denote β_2, is characterized by:

$$\beta_2 = \arg\min_{\beta}\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)^2 . \tag{3.8}$$

The MPE estimator is different from the maximum-likelihood estimator because, as noted by Lee, the former ignores the variation of the denominator of equation (3.6) (or, equivalently, the final term in equation (3.7)) with respect to β. This general approach, maximizing an approximate or truncated form of the likelihood function, is known as pseudo-likelihood estimation. This approach might be rigorously justified if the approximation error or the truncated terms were shown to vanish as the sample size increased. Absent such justification, there is no general guarantee that a pseudo-likelihood estimator behaves like the MLE. Instead, the properties of a pseudo-likelihood estimator must be established on a case-by-case basis.

Another way to view this problem is to examine the *concentrated* log-likelihood function.[34] First, maximize the log-likelihood function with respect to θ by setting to zero the derivative with respect to that parameter; then substitute the resulting estimate of θ back into the log-likelihood function to obtain a function of β alone. The first step yields the MLE of θ conditional on β (which we denote θ_2):

---

[34] See Seber and Wild (1989, pp. 37–42).

$$\theta_2 = \frac{1}{n} \times \sum_{i=1}^{n} \left(\frac{y_i - f(x_i, \beta)}{f(x_i, \beta)} \right)^2 ; \tag{3.9}$$

and the second step yields the concentrated log-likelihood function:

$$\begin{aligned} l^*(\beta) &= l(\beta, \theta_2) \\ &= -\frac{n}{2} \times \ln(2\pi e / n) - \frac{n}{2} \times \ln\left[\sum_{i=1}^{n} \left(\frac{y_i - f(x_i, \beta)}{f(x_i, \beta)} \right)^2 \right] - \sum_{i=1}^{n} \ln f(x_i, \beta), \end{aligned} \tag{3.10}$$

where e is the base of the natural logarithms. Once again, the MPE estimator considers only the variation in the middle term with respect to β, but ignores the variation in the final term. By contrast, the full MLE of β maximizes the entire expression (3.10) with respect to β.

The choice between the MPE estimator β_2 and the full MLE of β should hinge on the relative statistical properties of the two estimators. The MPE estimator is intuitively appealing because the regression model (3.1) is heteroscedastic: from equation (3.3), the standard deviation of y_i is proportional to $f(x_i, \beta)$. The MPE estimator minimizes the *weighted* sum-of-squares, correcting for heteroscedasticity by giving relatively more weight to the less variable observations. Put differently, the MPE estimator minimizes the sum-of-squares of *relative* (i.e., percentage) prediction errors.

Intuition notwithstanding, the MPE estimator is unsatisfactory because it is inconsistent: the estimator is biased, and the bias remains even as the sample size grows infinitely large.[35] To understand the bias, consider again equation (3.8). The optimization that defines MPE has two avenues for minimizing the sum-of-squares. First, accurate and unbiased predictions will bring the $f(x_i, \beta)$ in line with the y_i, thereby minimizing the numerator of equation (3.8). However, simply inflating the predictions $f(x_i, \beta)$ in the denominator will tend to deflate the percentage errors, albeit at the expense of worsening the fit in the numerator. The net result of these two effects is that the predictions $f(x_i, \beta)$ tend to be somewhat inflated, leading to biased parameter estimates. In particular, when modeling lot-average cost as $f(x_i, \beta) = T_1 \times [\bar{Q}_i(b)]^b$, the T_1 parameter tends to be biased upward. As we demonstrate in the Monte Carlo results in Chapter 5, the bias in T_1 tends

---

[35] See Seber and Wild (1989, pp. 88–89), especially the discussion immediately following their equation 2.183.

to increase with the variance θ. The logic behind this result is explored later in this chapter, during the discussion of iteratively reweighted least squares.[36]

The covariance matrix of the MLE is estimated by the negative inverse Hessian of the log-likelihood function. This matrix is not always available in closed-form, but numerical approximation is geuerally feasible. However, the covariance matrix of the MPE estimator has never been derived.

The MLE may be computed by numerically maximizing equation (3.7) with respect to β and θ. An alternative is to numerically maximize the concentrated log-likelihood function, equation (3.10), with respect to β; then compute the MLE of θ from equation (3.9). The latter method presents a slightly lower-dimensional maximization problem, because the estimate of θ conditional on β (i.e., θ_2) is known in closed form.

Under the current assumption of multiplicative normal errors, the likelihood-ratio statistic differs from the expression given earlier in the case of log-normal errors, which was $n \times \ln(SSE^r / SSE^u)$. Let θ_2^r and θ_2^u denote the restricted and unrestricted variance estimates, and let $\hat{f}_i^r$ and $\hat{f}_i^u$ denote the corresponding model predictions for observations $i = 1, \ldots, n$. Then under the multiplicative normal assumption, the likelihood-ratio statistic is given by: $-2 \times \ln(L^r / L^u) = n \times \ln(\theta_2^r / \theta_2^u) + 2 \times \sum_{i=1}^{n} \ln(\hat{f}_i^r / \hat{f}_i^u)$. This statistic has an asymptotic χ_r^2 distribution.[37]

3.3 Iteratively reweighted least squares

IRLS differs from the MPE estimator in a subtle but important way. Begin with an initial estimate of β, denoted $\beta^{(0)}$. In the minimand in expression (3.8), fix $\beta = \beta^{(0)}$ in the denominator, and minimize with respect to β in the *numerator only*. The minimum occurs at a new estimate, $\beta^{(1)}$. Now fix $\beta = \beta^{(1)}$ in the denominator, and again minimize

---

36 One might speculate whether the bias in MPE would vanish if $f(x_i, \beta)$ were replaced by y_i in the denominator of equation (3.8); i.e., if the objective function were $\sum \left[\left(y_i - f(x_i, \beta) \right) / y_i \right]^2$. We leave this question open for future researchers.

37 The likelihood-ratio statistic must be non-negative. Suppose we apply MPE first, treat the resulting estimates of β as fixed values, and test the MLEs against these fixed values. The first term in the likelihood-ratio statistic will be negative, because MPE explicitly minimizes θ. However, the entire statistic will still be non-negative, because the second term (which measures the superior fit of the MLE under the model assumptions) always dominates.

with respect to β in the numerator only. In general, compute the following sequence of estimators:

$$\beta^{(p+1)} = \arg\min_{\beta} \sum_{i=1}^{n} \left(\frac{y_i - f(x_i, \beta)}{f(x_i, \beta^{(p)})} \right)^2 , \tag{3.11}$$

for $p = 0, 1, 2, \ldots$. Finally, the IRLS estimator is defined as the limit of the sequence:

$$\beta_3 = \lim_{p \to \infty} \beta^{(p)} , \tag{3.12}$$

when the limit exists. In practice, the IRLS estimator is taken where the sequence converges within a pre-specified numerical tolerance.

It is best not to regard IRLS as yet another pseudo-likelihood estimator. Rather, IRLS is a classical technique that can be motivated in many different ways without reference to any likelihood function. In contrast to likelihood methods, IRLS does not require any parametric distributional assumption (e.g., normality or log-normality).

IRLS is numerically distinct from MPE estimation. Although IRLS may appear to minimize expression (3.8), the gradient of expression (3.8) with respect to β is generally non-zero at the IRLS solution. We will demonstrate this point later by numerical example.

IRLS yields consistent estimates of regression model (3.1) under quite general conditions: the only essential distributional assumption is finite variance. Moreover, the covariance matrix of the estimator follows a known formula, and the estimator is asymptotically normally distributed even though the regression error itself (u_i or v_i) need not be normal.[38]

IRLS has periodically been rediscovered; for example, Book and Lao (1996) and Book and Young (1997) label it the Minimum Unbiased Percentage Error (MUPE) procedure.[39] We saw that the MPE estimator is inconsistent, and the MLE is consistent if

---

[38] Specifically, Seber and Wild (1989, pp. 88–89) report that the asymptotic sampling distribution of the IRLS estimator is normal with mean equal to the true (unknown) parameter vector.

[39] Some references from the 1970s are: Bradley (1973), Jennrich and Moore (1975), and Charnes, Frome and Yu (1976). Another resurgence of interest occurred during the 1980s: Jorgensen (1983, 1984) and Green (1984). Incidentally, Seber and Wild's (1989) analysis serves to definitize Book and Young's (1997, p. 13) empirical observation that the bias in IRLS / MUPE is "apparently asymptotically zero" for non-linear regression functions.

the distributional assumption is valid. Remarkably, the IRLS estimator is consistent without any distributional assumption except finite variance (though it may be biased in small samples, so the "U" in the MUPE terminology is misleading). Despite the desirable property of consistency, Book and Lao (1996, p. 10) criticize the IRLS estimator because it is, "not clear that MUPE [i.e., IRLS] is optimal with respect to any relevant criterion." Again, Book and Young (1997, p. I3) describe IRLS as, "converging to a parameter vector...that may or may not be optimal with respect to some appropriate criterion." In the next section, we exhibit the criterion under which IRLS is optimal, and argue for its relevance.

3.4 Quasi-likelihood estimation

Quasi-likelihood is a remarkable statistical concept that yields estimators sharing many of the desirable properties of MLEs, but without the need for precise distributional assumptions.[40] In the case of independent observations (our maintained assumption throughout this entire work), quasi-likelihood estimation requires only:

- a mapping from the predictor variables to the mean of the response variable; and
- a functional relationship between the variance (assumed finite) of the response variable and the mean, up to a scaling constant (i.e., $V(y_i) = \lambda \times g[E(y_i)] < \infty$). The function $g[\ ]$ must be continuous, but not necessarily monotonic.

Returning to equation (3.2), and assuming that $E(u_i) = 1.0$, the first requirement is satisfied by the equation $E(y_i) = f(x_i, \beta)$. Assuming that $V(u_i)$ is constant for all observations ($i = 1, \ldots, n$), the second requirement is satisfied as well; setting $\lambda = V(u_i)$, equation (3.3) becomes $V(y_i) = \lambda \times [E(y_i)]^2$.

Letting $\mu_i = E(y_i) = f(x_i, \beta)$, the contribution of the i^{th} observation to the log-quasi-likelihood function is the solution to the differential equation:

$$\frac{\partial q_i(\mu, \lambda)}{\partial \mu_i} = \frac{y_i - \mu_i}{\lambda \times g(\mu_i)}. \tag{3.13}$$

In our example of a multiplicative regression model, we have $g(\mu_i) = \mu_i^2$, so that the differential equation becomes:

---

40 See McCullagh and Nelder (1989, chapter 9), or Seber and Wild (1989, pp. 42–48).

$$\frac{\partial q_i(\mu,\lambda)}{\partial \mu_i} = \frac{y_i - \mu_i}{\lambda \times \mu_i^2} = \frac{y_i}{\lambda \times \mu_i^2} - \frac{1}{\lambda \times \mu_i}, \tag{3.14}$$

with solution:

$$q_i(\mu,\lambda) = -\frac{y_i}{\lambda \times \mu_i} - \frac{\ln(\mu_i)}{\lambda}. \tag{3.15}$$

The sample log-quasi-likelihood function is given by the sum over all the observations:

$$\begin{aligned} q(\beta,\lambda) &= -\frac{1}{\lambda}\sum_{i=1}^{n}[(y_i/\mu_i) + \ln(\mu_i)] \\ &= -\frac{1}{\lambda}\sum_{i=1}^{n}\left[\frac{y_i}{f(x_i,\beta)} + \ln(f(x_i,\beta))\right]. \end{aligned} \tag{3.16}$$

The maximum quasi-likelihood estimator of β is the value that minimizes the summation in equation (3.16). The resulting estimator, which we denote β_4, is characterized by:

$$\beta_4 = \arg\min_{\beta} \sum_{i=1}^{n}\left[\frac{y_i}{f(x_i,\beta)} + \ln(f(x_i,\beta))\right]. \tag{3.17}$$

Finally, quasi-likelihood estimation of β_4 is separable from estimation of λ. The latter parameter is conventionally estimated by the generalized Pearson statistic:

$$\lambda_4 = \frac{1}{n-m} \times \sum_{i=1}^{n} \frac{[y_i - f(x_i,\beta_4)]^2}{g[f(x_i,\beta_4)]}, \tag{3.18}$$

or in our multiplicative example:

$$\lambda_4 = \frac{1}{n-m} \times \sum_{i=1}^{n}\left[\frac{y_i - f(x_i,\beta_4)}{f(x_i,\beta_4)}\right]^2. \tag{3.19}$$

We now respond to Book and Lao's criticism and demonstrate that the IRLS estimator maximizes the quasi-likelihood. To characterize the maximum quasi-likelihood, we set to zero the gradient of the right-hand side of equation (3.17) with respect to the parameter vector β:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \left[\frac{y_i}{f(x_i,\beta)} + \ln(f(x_i,\beta)) \right] \\
&= \sum_{i=1}^{n} \left[\frac{(f(x_i,\beta) - y_i)}{(f(x_i,\beta))^2} \times \frac{\partial f(x_i,\beta)}{\partial \beta_j} \right],
\end{aligned}
\tag{3.20}
$$

for $j = 1, \ldots, m$.

Alternatively, consider the sequence of estimators generated by IRLS. In particular, $\beta^{(p+1)}$ sets to zero the gradient of the right-hand side of equation (3.11) with respect to β. Again, we minimize with respect to β in the *numerator only*, fixing $\beta = \beta^{(p)}$ in the denominator:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta^{(p)})} \right)^2 \\
&= 2 \times \sum_{i=1}^{n} \left[\frac{\left(f(x_i,\beta^{(p+1)}) - y_i\right)}{\left(f(x_i,\beta^{(p)})\right)^2} \times \frac{\partial f(x_i,\beta^{(p+1)})}{\partial \beta_j} \right],
\end{aligned}
\tag{3.21}
$$

for $j = 1, \ldots, m$. Note that we replace β by $\beta^{(p+1)}$ in the numerator, to indicate that the gradient vanishes at the new estimate, $\beta^{(p+1)}$. At convergence, however, $\beta^{(p)} = \beta^{(p+1)}$, so equation (3.21) reduces to:

$$
0 = 2 \times \sum_{i=1}^{n} \left[\frac{\left(f(x_i,\beta^{(p)}) - y_i\right)}{\left(f(x_i,\beta^{(p)})\right)^2} \times \frac{\partial f(x_i,\beta^{(p)})}{\partial \beta_j} \right],
\tag{3.22}
$$

which is identical to the condition for maximum quasi-likelihood.[41]

We may also use this analysis to gain some understanding into the bias in the MPE estimator and, in particular, its sensitivity to the variance. Referring back to equation (3.8), the MPE estimator minimizes the sum-of-squares with respect to β as it appears in both the numerator and denominator, whereas IRLS fixes β in the denominator

---

[41] Under certain conditions, IRLS can be used to maximize the likelihood (not quasi-likelihood) function as well, rendering all three estimators (IRLS, maximum quasi-likelihood, MLE) identical. The main regularity condition is that the density function belongs to the exponential family; see Bradley (1973), Jennrich and Moore (1975), and Charnes, Frome and Yu (1976). Lee's multiplicative normal density (our equation (3.6)) does not belong to this family and, as we will see in the numerical examples, the MLE is quite distinct from the IRLS/maximum quasi-likelihood estimates.

and minimizes with respect to β in the numerator only. The gradient equation that defines MPE may be decomposed into two terms: the first term represents the gradient with respect to β in the numerator only, as in equation (3.22) (i.e., IRLS); the second term represents the bias away from the IRLS solution. Thus, the gradient of the minimand in equation (3.8) may be decomposed as:[42]

$$2\times\sum_{i=1}^{n}\left[\frac{(f(x_i,\beta)-y_i)}{(f(x_i,\beta))^2}\times\frac{\partial f(x_i,\beta)}{\partial\beta_j}\right] \;-\; 2n\times\sum_{i=1}^{n}\left[\frac{\partial \ln f(x_i,\beta)}{\partial\beta_j}\times\frac{1}{n}\left(\frac{y_i-f(x_i,\beta)}{f(x_i,\beta)}\right)^2\right]. \qquad (3.23)$$

Comparing the variance estimator in equation (3.9), the second term here is more important, roughly speaking, when the variance is larger. Thus, the bias in MPE is more severe for large-variance problems. We confirm this finding in the Monte Carlo results in Chapter 5.

Because IRLS and quasi-likelihood estimation are identical, they share the properties of consistency and asymptotic normality under the minimal assumptions of finite variance and continuity of the variance-to-mean function. To quote Seber and Wild (1989, p. 46):

> An attractive aspect of quasi-likelihood theory...is the following. When the data analyst is fairly confident that the mean function and the relationship between the mean and variance has been modeled fairly well, but is unsure of the other aspects of the parametric distribution used, quasi-likelihood theory assure him or her of the asymptotic correctness of the resulting inferences. In this way it is a generalization of the asymptotic applicability of least-squares theory beyond the restrictive assumption of normally distributed errors.

The asymptotic covariance matrix of the estimator may be developed as follows. Again let J denote the $n\times m$ Jacobian matrix of the mean function with respect to the m parameters β. Also, let G denote the diagonal matrix of relative variances of the observations:

$$G \;\equiv\; diag\{g(\mu_1),\ldots,g(\mu_n)\}\,. \qquad (3.24)$$

The $m\times m$ asymptotic covariance matrix is given by:

$$Var(\beta_4) \;=\; \lambda_4\times\left(J^TG^{-1}J\right)^{-1}, \qquad (3.25)$$

---

[42] This equation is essentially the same as equation (2.183) on p. 89 of Seber and Wild (1989).

where the dispersion estimate λ_4 was given previously in equation (3.18). Specializing to our situation with $g(\mu_i) = \mu_i^2$, the asymptotic covariance matrix may be written as:

$$\begin{aligned} Var(\beta_4) &= \lambda_4 \times \left(\sum_{i=1}^{n} [J_i^T J_i] / [f(x_i, \beta_4)]^2 \right)^{-1} \\ &= \frac{1}{n-m} \times \sum_{i=1}^{n} \left(\frac{y_i - f(x_i, \beta_4)}{f(x_i, \beta_4)} \right)^2 \times \left(\sum_{i=1}^{n} [J_i^T J_i] / [f(x_i, \beta_4)]^2 \right)^{-1}, \end{aligned} \tag{3.26}$$

where each term $[J_i^T J_i]$ is again an $m \times m$ outer product matrix.

Although the calculations here appear formidable, they are in one sense simpler than the corresponding calculations under lot-midpoint NLS. In the latter situation, the Jacobian matrix had to be computed for the predictor function $f(x_i, \beta) = \ln(T_1) + b \ln[\bar{Q}_i(b)]$, which is highly non-linear in light of the definition of the lot midpoint. Under IRLS, the predictor function generally takes a much simpler form. In the learning-curve context, the predictor function is the non-linear model for lot-average cost obtained by integrating under the (Crawford) marginal cost curve, equation (3.4). The calculation is even simpler in the CER context, in which case the predictor function is simply the CER itself, equation (3.5).

3.5 Comparison of the six estimation methods

Table 3.1 summarizes our comparison of the six estimation methods considered in this monograph. Lot-midpoint regression assumes a log-normal error distribution. The NLS estimator is consistent and asymptotically normally distributed. The covariance matrix is available as equation (2.30) above. A closely related method is lot-midpoint iteration. Although widely used in cost analysis, the asymptotic properties of this estimator are not currently known. In particular, the conventional formula is an underestimate of the standard error of b. Moreover, there are no theoretical guarantees of existence or uniqueness of the solution, or of convergence even when a solution does exist.

We cannot endorse MPE because it is biased and inconsistent, and its covariance matrix has not been derived in the literature. Maximum likelihood is probably the most ubiquitous estimation method in statistics, but it requires a particular distributional form. In addition, although the MLE covariance matrix follows a well-known formula, evaluation of that formula may require numerical approximation.

Table 3.1. Comparison of Six Estimation Methods for Learning-Curve Models

Estimation method	Distributional assumptions	Asymptotic properties	Covariance matrix
Lot-midpoint non-linear least squares (NLS)	log-normal	consistent and asymptotically normal	formula available
Lot-midpoint iteration	log-normal	unknown	conventional formula, an underestimate
Minimum percentage error (MPE)	multiplicative model, finite variance	biased and inconsistent	unknown
Maximum likelihood estimation (MLE)	particular distributional form	consistent if correct distributional form	well-known formula, may require numerical approximation
Iteratively reweighted least squares (IRLS) / Minimum unbiased percentage error (MUPE)	multiplicative model, finite variance, continuous variance-to-mean function	consistent and asymptotically normal	formula available
Maximum quasi-likelihood	multiplicative model, finite variance, continuous variance-to-mean function	consistent and asymptotically normal	formula available

Finally, as we (and others) have shown, the IRLS (recently renamed MUPE) and maximum quasi-likelihood estimators are identical, thus they share all of the same properties. In particular, these estimators are consistent and asymptotically normally distributed. The covariance matrix is available in closed-form as equation (3.26) above.

Reviewing Table 3.1, IRLS appears to produce the best estimates for the fewest assumptions. It does not require a log-normal or any other particular distributional form, only finite variance. Moreover, its asymptotic properties are the best that can be hoped for in a non-linear model, and its covariance matrix is easily computed.

Some would question our harsh assessment of the MPE method. Book and Young (1997, especially pp. 6–7) observe that the minimized sum-of-squares is generally lower for MPE (our expression (3.8)) than for IRLS / MUPE (our expression (3.11) evaluated at convergence). They engage in a rather lengthy discussion of the tradeoff between MPE, which is biased and inconsistent but has a smaller sum-of-squares, and IRLS/MUPE, which is consistent (though possibly biased in small samples) but may have a considerably larger sum-of-squares. In our view, reducing bias should always be a higher priority than reducing the sum-of-squares. First, the sum-of-squares can always be artificially reduced to zero by regressing a times series of lot costs on a sufficiently high-order polynomial in any single predictor such as calendar time or lot size. However, such

a polynomial equation, with multiple points of inflection or even non-monotonicities, is of virtually no value in forecasting the cost of future lots.[43]

Further, a common use of regression models in cost analysis is to forecast the growth in unit cost due to a deviation from baseline assumptions (e.g., an increase in system weight or a smaller production run). These exercises require "good" estimates of the weight coefficient in a CER or the learning slope. A "good" estimate is one possessing both low bias and low variability (i.e., a small standard error). Although Book and Young make claims for MPE based on its smaller regression sum-of-squares, they have not demonstrated that the parameter vector estimated via MPE has smaller standard errors than the one estimated via IRLS/MUPE. As we have pointed out, the covariance matrix of the MPE estimator is unknown. Hence, there is currently no basis for claiming that the MPE estimator has lower variability—it may well have higher variability.

3.6 Correction for serial correlation

Our maintained assumption throughout this entire work has been that the costs of successive lots are statistically independent. In particular, the derivations of the various estimation methods have all assumed the absence of serially correlated errors. In the Monte Carlo analysis of Chapter 5, we measure the loss of precision that occurs when serial correlation is present, despite the modeling assumption to the contrary. We show there that all of the methods considered, except for lot-midpoint iteration, are robust to serial correlation. Moreover, as we argued in Chapter 2, serial correlation can often be reduced by transforming the data series from cumulative average cost to lot average cost prior to estimation (equations (2.9) and (2.10)).

Nonetheless, we want to give the reader some indication of the estimation technique when serial correlation is present and perceived as a serious problem. We do not pursue this extension for either lot-midpoint iteration or MPE, because we do not recommend these methods even under the best of circumstances. The extension of MLE to serially correlated errors is covered in many sources; Womer and Patterson (1983) apply this method, and Seber and Wild (1989, Chapter 6) explicitly give the estimating equations. Therefore, we restrict our discussion to NLS estimation in the presence of serial correlation.

---

[43] Lee (1997, pp. 79–81) gives an excellent example of the pitfalls that arise when attempting to forecast using models that were selected solely on the basis of in-sample goodness-of-fit.

We consider the particular case of first-order serial correlation, or so-called AR(1) errors. For this analysis it is convenient to zero-out the mean of the error term, thus we write the model as:

$$y_i = f(x_i,\beta)\times(1+u_i)\ , \tag{3.27}$$

where now $E(u_i)=0$. However, the errors now have the AR(1) structure:

$$u_i = \rho\times u_{i-1}+\varepsilon_i\ , \tag{3.28}$$

where $|\rho|<1$ and the error terms $\{\varepsilon_i\}$ independent normal, $\varepsilon_i \sim N\left(0,\sigma^2\right)$. The correlations between successive values of $\{u_i\}$ decline geometrically with their distance on the time scale, $Corr(u_i,u_j) = \rho^{|i-j|}$.

It can be shown that $Var(u_i) = Var(\varepsilon_i)/(1-\rho^2) = \sigma^2/(1-\rho^2)$. Thus, we have the variance of each observation:

$$Var(y_i) = [f(x_i,\beta)]^2\times V(u_i) = f_i^2\sigma^2/(1-\rho^2)\ , \tag{3.29}$$

where we have used the shorthand notation $f_i = f(x_i,\beta)$. We can also derive the covariance between any two observations:

$$Cov(y_i,y_j) = \rho^{|i-j|}f_i f_j\sigma^2/(1-\rho^2)\ . \tag{3.30}$$

We can array all of the variances and covariances into a matrix:

$$V = Cov(y_1,\ldots,y_n) = \sigma^2/(1-\rho^2)\times\begin{pmatrix} f_1^2 & \rho f_1 f_2 & \cdots & \rho^{n-1}f_1 f_n \\ \rho f_2 f_1 & f_2^2 & \cdots & \rho^{n-2}f_2 f_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1}f_n f_1 & \cdots & \rho f_n f_{n-1} & f_n^2 \end{pmatrix}. \tag{3.31}$$

We can also find the lower-triangular $n\times n$ matrix L such that $V = L^{-1}(L^T)^{-1}$ or $V^{-1} = L^T L$, where the superscript "T" indicates the matrix transpose:[44]

---

[44] The matrix L generalizes the standard factorization of a serially-correlated covariance matrix that was first derived by Kadiyala (1968) and reproduced in many places including Seber and Wild (1989, p. 276). Our generalization introduces the weights $\{f_i\}$ that account for heteroscedasticity.

$$L = \sigma^{-1} \times \begin{pmatrix} \frac{\sqrt{1-\rho^2}}{f_1} & 0 & \cdots & 0 \\ \frac{-\rho}{f_1} & \frac{1}{f_2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \frac{-\rho}{f_{n-1}} & \frac{1}{f_n} \end{pmatrix}. \tag{3.32}$$

Conventional NLS estimation of equation (3.27) would minimize the untransformed sum-of-squares, $\sum(y_i - f(x_i, \beta))^2$. However, the covariance matrix of the observations, V, exhibits both heteroscedasticity (unequal diagonal elements) and serial correlation (non-zero off-diagonal elements). Thus, conventional NLS estimates are consistent but not efficient (i.e., do not have the minimum sampling variance among all consistent estimates).

Efficient estimates may be obtained by applying non-linear generalized least squares (NGLS). We minimize instead the weighted sum-of-squares, which is represented in matrix form as:

$$\begin{aligned} [\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta)]^T V^{-1} [\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta)] &= [\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta)]^T (L^T L)[\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta)] \\ &= [L \times (\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta))]^T [L \times (\mathbf{y} - \mathbf{f}(\mathbf{x}, \beta))] \\ &= [L \times \mathbf{y} - L \times \mathbf{f}(\mathbf{x}, \beta)]^T [L \times \mathbf{y} - L \times \mathbf{f}(\mathbf{x}, \beta)], \end{aligned} \tag{3.33}$$

where $\mathbf{y}$ is the $n \times 1$ vector of response variables and $\mathbf{f}(\mathbf{x}, \beta)$ is the $n \times 1$ vector of model predictions. Expression (3.33) differs from the untransformed sum-of-squares by the insertion of the weighting matrix, V^{-1}. However, we also see from the final line that NGLS is achieved by transforming both the response variables and the model predictions by the matrix L prior to estimation. That is, expression (3.33) reduces to the sum-of-squared differences between the transformed response variables $L \times \mathbf{y}$ and the transformed model predictions $L \times \mathbf{f}(\mathbf{x}, \beta)$ (both $n \times 1$ vectors).

The matrix L contains the unknown parameters ρ and σ, as well as the parameters β that are embedded in the model predictions $\{f_i\}$. This situation suggests an iterative procedure in which we first estimate all of the parameters (most likely by conventional NLS), use those estimated parameters to form the L matrix, estimate the transformed model, then possibly continue iterating until convergence (i.e., until the parameters β and perhaps also ρ and σ stabilize). However, several points must be noted here. First, Gallant and Goebel (1976) reported that the NGLS estimates obtained after a single

round of transformation have the same asymptotic distribution as the estimates obtained by iterating until convergence. Except for some very special cases that have been studied in the literature, the justification for NLS is based solely on its asymptotic properties. Thus, there appears to be little gain from continuing beyond the first round of transformation.

Second, when estimating the transformed model, we must hold the L matrix fixed and minimize the sum-of-squares with respect to the parameters β only as they enter the model predictions $\mathbf{f}(\mathbf{x}, \beta)$, *not* as they feed back through the L matrix. This distinction is quite analogous to the one we emphasized in Chapter 3, where IRLS sought the minimum sum-of-squares in the prediction errors alone, expression (3.11), but MPE allowed the same parameters to vary in the weights as well as in the prediction errors, expression (3.8). We saw there that the latter approach leads to biased parameter estimates.[45]

To make it absolutely clear that the weights $\{f_i\}$ are fixed in the L matrix, we re-write that matrix with the notation $\{w_i\}$ replacing $\{f_i\}$; we regard the weights $\{w_i\}$, like ρ and σ, as fixed elements during the minimization that yields the estimate of β. Moreover, we can suppress σ because (as is easily demonstrated) doing so does not affect the estimate of β. Instead, σ^2 may be recovered in the usual way at the end of the process, as the minimized regression sum-of-squares divided by the degrees-of-freedom. Thus, without any loss of generality, we re-write the L matrix as:

$$L = \begin{pmatrix} \frac{\sqrt{1-\rho^2}}{w_1} & 0 & \cdots & 0 \\ \frac{-\rho}{w_1} & \frac{1}{w_2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \frac{-\rho}{w_{n-1}} & \frac{1}{w_n} \end{pmatrix}. \tag{3.34}$$

Returning to equation (3.33), we can now explicitly display the transformations applied to the response variables and the model predictions prior to estimation (both $n \times 1$ vectors):

---

[45] By contrast, MLE treats all of the parameters (including ρ and σ) on an equal footing, wherever they may appear in the likelihood function. For the case of AR(1) errors (though without the additional complication of heteroscedasticity), see Seber and Wild (1989, section 6.2.2).

$$L \times \mathbf{y} = \begin{pmatrix} \sqrt{1-\rho^2}\, y_1/w_1 \\ (y_2/w_2) - \rho(y_1/w_1) \\ \vdots \\ (y_n/w_n) - \rho(y_{n-1}/w_{n-1}) \end{pmatrix}, \tag{3.35}$$

and

$$L \times \mathbf{f}(\mathbf{x},\beta) = \begin{pmatrix} \sqrt{1-\rho^2}\, f(x_1,\beta)/w_1 \\ (f(x_2,\beta)/w_2) - \rho(f(x_1,\beta)/w_1) \\ \vdots \\ (f(x_n,\beta)/w_n) - \rho(f(x_{n-1},\beta)/w_{n-1}) \end{pmatrix}. \tag{3.36}$$

We have not yet indicated the procedure for estimating ρ. Inverting equation (3.27), we can estimate the individual error term u_i as the percentage prediction error, $\hat{u}_i = \left[y_i - f(x_i,\beta)\right]/f(x_i,\beta)$. Goldberger (1964, p. 243) gives the expression for the first-order serial correlation coefficient:

$$\frac{\sum_{i=2}^{n} \hat{u}_i \hat{u}_{i-1}}{\sqrt{\sum_{i=2}^{n} \hat{u}_i^2}\sqrt{\sum_{i=2}^{n} \hat{u}_{i-1}^2}}. \tag{3.37}$$

We give an example of the NGLS procedure in the next chapter. However, we must alert the reader to one oddity before ending this discussion. When first forming the vector $L \times \mathbf{f}(\mathbf{x},\beta)$, we set $w_i = f(x_i,\beta^{(0)})$ where $\beta^{(0)}$ is the estimate obtained by conventional NLS (i.e., by simply minimizing $\sum\left(y_i - f(x_i,\beta)\right)^2$). Thus, the terms $\{f_i/w_i\}$ reduce to unity and the vector $L \times \mathbf{f}(\mathbf{x},\beta)$ numerically computes as:

$$L \times \mathbf{f}(\mathbf{x},\beta) = \begin{pmatrix} \sqrt{1-\rho^2} \\ 1-\rho \\ \vdots \\ 1-\rho \end{pmatrix}. \tag{3.38}$$

However, when programming the NGLS algorithm, it is imperative to write the terms $\{f_i\}$ *functionally* (vs. *numerically*) in terms of the coefficient vector β, as we have done in equation (3.36) with the explicit notation $f_i = f(x_i,\beta)$. The NGLS algorithm

will adjust β, and thus adjust the numerators of the ratios $\{f_i/w_i\}$, in an attempt to minimize the weighted prediction errors. If the vector $L \times \mathbf{f}(\mathbf{x}, \beta)$ is revisited at the end of the process, it will be seen to differ numerically from expression (3.38). That difference reflects the improvement due to the single NGLS step.

4. APPLICATION OF THE ESTIMATION METHODS TO LEE'S DATA

In this chapter, we apply the various estimation methods to Lee's (1997) data on a tactical missile program, reproduced earlier as Table 1.1. This exercise reinforces several of our theoretical results and illustrates the magnitudes of the differences among the various estimation methods. In particular, we show that lot-midpoint iteration converges on Lee's data, even though the eigenvalue $\partial b^{(p+1)}/\partial b^{(p)}$ evaluates as -1.041 at the starting point. This example confirms our theoretical finding from Chapter 2 that an absolute bound of 1.0, while sufficient for convergence, is not necessary. We also directly compare two fitting criteria: the sum-of-squared errors in predicting the logarithm of lot average cost, and sum-of-squared *percentage* errors in predicting the *level* (not logarithm) of lot average cost. Contrary to Young's (1999) somewhat ambiguous assessment, we show that for a symmetric data set with no extreme outliers, the logarithmic sum-of-squares is generally larger than the percentage sum-of-squares.

4.1 Non-linear least squares

Lee presents only a single set of numerical estimates. He applied NLS to a regression of incremental lot cost, as in our equation (2.15). That is, Lee minimized the following quantity:[46]

$$\sum_{i=1}^{n}\left(\left[TC_i - TC_{i-1}\right] - \frac{T_1}{1+b}\times\left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b}\right]\right)^2 . \qquad (4.1)$$

There is some evidence of heteroscedasticity in the data (i.e., lots containing more units also exhibit greater variability in incremental lot cost). To restore variance homogeneity, we also applied NLS to a regression of lot average cost, as in our equation (2.16). That is, we minimized, instead, the following quantity:

---

[46] More precisely, Lee (1997, pp. 35–41) replaced the right-hand side of our expression (4.1) with a more exact expression for incremental lot cost, based on the Euler-Maclaurin summation formula. However, his procedure yields a learning coefficient (b) that differs from our estimate by only 10^{-4}.

$$\sum_{i=1}^{n}\left(LAC_i \; - \; \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})}\times\left[(Q_i+0.5)^{1+b} \; - \; (Q_{i-1}+0.5)^{1+b}\right]\right)^2 \quad . \quad (4.2)$$

These two sets of estimates appear in the first two rows of Table 4.1.

Using the definition of the lot midpoint, the latter estimate is exactly what one would obtain by applying NLS directly to equation (2.26). However, equation (2.26) has *multiplicative* error structure, whereas NLS implicitly assumes an *additive* error structure. Thus, NLS would more appropriately be applied to equation (2.27), which indeed has an additive error structure. We label this method "Lot-midpoint NLS" in Table 4.1; it is the non-linear estimator that minimizes expression (2.28). The resulting estimates appear in the third row of Table 4.1. We adjusted the intercept by the log-normal correction factor, $\exp(\sigma^2/2)$, to enable consistent predictions of lot average cost (rather than its natural logarithm). Note that the first three rows of Table 4.1 all use different response variables: incremental lot cost (expression (4.1)), lot average cost (expression (4.2)), and the natural logarithm of lot average cost (expression (2.28)), respectively. Thus, although the parameter estimates are comparable, the sums-of-squared errors are not.

Figure 4.1 compares the intercepts and learning slopes for these and all of the other estimation methods considered in this chapter. Methods that yield higher intercepts for this particular data set compensate with "steeper" (numerically smaller) learning slopes, else the fitted learning curve would bypass the centroid of the data. Within the small ranges of slopes in this example, the relationship between slope and intercept is remarkably linear. The NGLS estimates do not appear in Table 4.1, but are highlighted for discussion in a later section.

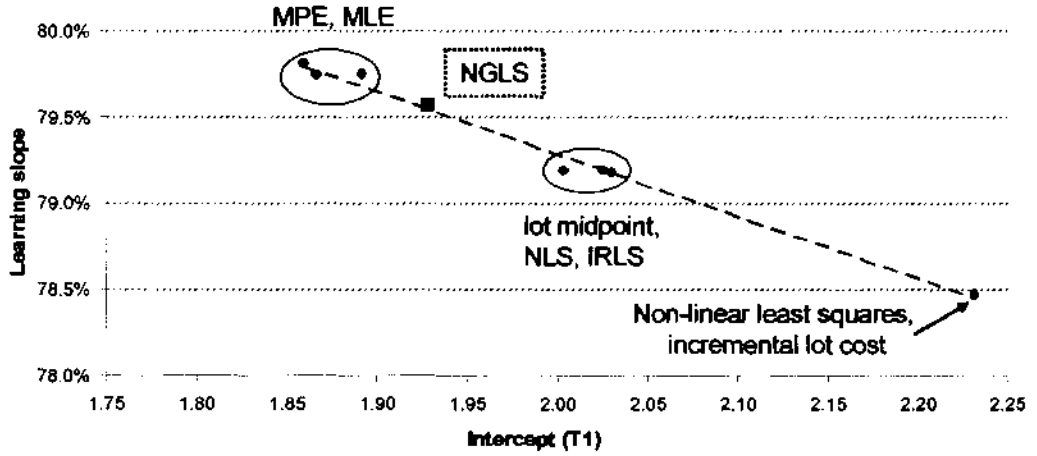


Figure 4.1. Comparison of Intercepts and Slopes for Various Estimation Methods

Table 4.1. Alternative Learning-Curve Estimates using Lee's Data for a Tactical Missile Program

Estimation method	Intercept	Quantity exponent	Standard error of exponent	Learning slope	Sum-of-squared errors			Standard error of estimate (pct.)	R-squared	Log-likelihood	Log-quasi-likelihood
					Arithmetic	Logarithmic	Percentage				
Non-linear least squares, incremental lot cost	2.2319	–0.3496	0.0381	78.48%	5,085.6	—	—	—		20.842	–50.6320
Non-linear least squares, lot average cost	1.8593	–0.3254	0.0118	79.81%	0.00102	—	—	—		21.226	–50.6296
Lot-midpoint NLS	2.0256	–0.3366	0.0300	79.19%	—	*0.1328*	0.1218	0.1425	0.951	21.155	–50.6289
Lot-midpoint iteration	2.0301	–0.3369	0.0312	79.18%	—	*0.1328*	0.1188	0.1407	0.951	21.150	–50.6289
Minimum percentage error (MPE)	1.8921	–0.3265	not available	79.75%	—	0.1394	*0.1153*	*0.1386*		21.178	–50.6301
Maximum likelihood estimation (MLE)	1.8663	–0.3266	0.0248	79.74%	—	0.1358	0.1170	0.1396		*21.236*	–50.6292
IRLS / MUPE / Maximum quasi-likelihood	1.9649	–0.3331	0.0283	79.38%	—	0.1336	0.1180	0.1402		21.202	*–50.6287*

Note: entries in italics represent minimum sum-of-squared errors, or maximum likelihood or quasi-likelihood.

We next compute the fitted learning curve that results from lot-midpoint NLS, as well as the $\pm 2\sigma$ confidence band around the fitted curve. Because the formula for the confidence band is not widely known by cost analysts, we present most of the details behind this calculation. Table 4.2 summarizes some of the key ingredients required for the calculation.

Table 4.2. Parameter Estimates from Lot-Midpoint NLS

Parameter	Estimate
Sum-of-squared errors	0.1328
Sample size (n)	8
Number of parameters (k)	2
Degrees of freedom ($n-k$)	6
Standard error of regression (σ)	0.1488
T_1, pre-adjustment	2.0033
Log-normal correction factor	1.0111
T_1, post-adjustment	2.0256
Exponent (b)	–0.3366
Learning slope	79.19%

The formula for the asymptotic variance of the prediction was given previously in equation (2.41). We repeat that formula here, except that we subsume the 2×2 covariance matrix of the NLS parameter estimates $V = \sigma^2 \left(J^T J\right)^{-1}$ into the single term V that is available directly from the regression output:

$$Var(\hat{LAC}_i) \;\rightarrow\; \left[w_i V w_i^T + \frac{\sigma^4}{2n} \right] \times (\hat{LAC}_i)^2 , \tag{4.3}$$

where $w_i = \left(1/\hat{T}_1, \;\; \ln \overline{Q}_i\right)$. Only two terms in this formula vary across the observations: $\ln \overline{Q}_i$ and $\hat{LAC}_i$. We present these terms in Table 4.3. In particular, the middle column gives the logarithmic lot midpoint ($\ln \overline{Q}_i$) for each lot, and the final column gives the predicted lot average cost ($\hat{LAC}_i$).

Table 4.3. Terms that Vary Across the Observations

Lot number	Lot midpoint	Logarithmic lot midpoint	Actual lot average cost ($M)	Predicted lot average cost ($M)
1	67.6	4.214	0.471	0.490
2	606.4	6.408	0.226	0.234
3	2,066.6	7.634	0.158	0.155
4	4,459.3	8.403	0.124	0.120
5	6,722.3	8.813	0.126	0.104
6	8,764.3	9.078	0.094	0.095
7	10,825.9	9.290	0.095	0.089
8	13,019.7	9.474	0.062	0.083

Considering, for example, the second lot, the asymptotic variance of the prediction $Var(\hat{LAC}_2)$ would be calculated as follows:

$$\left[(0.499 \quad 6.408)\times\begin{pmatrix} 0.2355 & -0.0142 \\ -0.0142 & 0.0009 \end{pmatrix}\times\begin{pmatrix} 0.499 \\ 6.408 \end{pmatrix} + \frac{(0.1488)^4}{2\times 8}\right]\times(0.234)^2 \quad . \qquad (4.4)$$

Note that we invert the *pre-adjustment* value of T_1, $1/\hat{T}_1 = 1/2.0033 = 0.499$. Expression (4.4) evaluates as $(0.0162)^2$, so the standard error of the prediction for the second lot equals 0.0162.

Figure 4.2 illustrates the close fit of the two-parameter learning-curve model to the tactical missile data. Each data point represents the computed lot midpoint (second column of Table 4.3) and the actual lot average cost (fourth column of Table 4.3). The solid curve represents the smooth model prediction of marginal cost (i.e., the Crawford model). The data points would ideally fall along the solid curve, because the lot midpoint is calculated such that its marginal cost (the height of the solid Crawford curve) equals the predicted lot average cost (the predicted height of the data point). Although the figure is drawn for the lot-midpoint NLS estimates, all of the alternative estimates are numerically close and the visual representations are indistinguishable.

The $\pm 2\sigma$ confidence band reveals two minor outliers at the fifth and eight lots. However, these outliers are departures from the two-parameter learning-curve function and cannot be resolved by mere recalibration of that function. The analyst's only choices are to:

- Review the data for possible errors,
- Expand the two-parameter functional form,
- Add more predictor variables (e.g., production rate), or

- Simply live with the two minor outliers.[47]

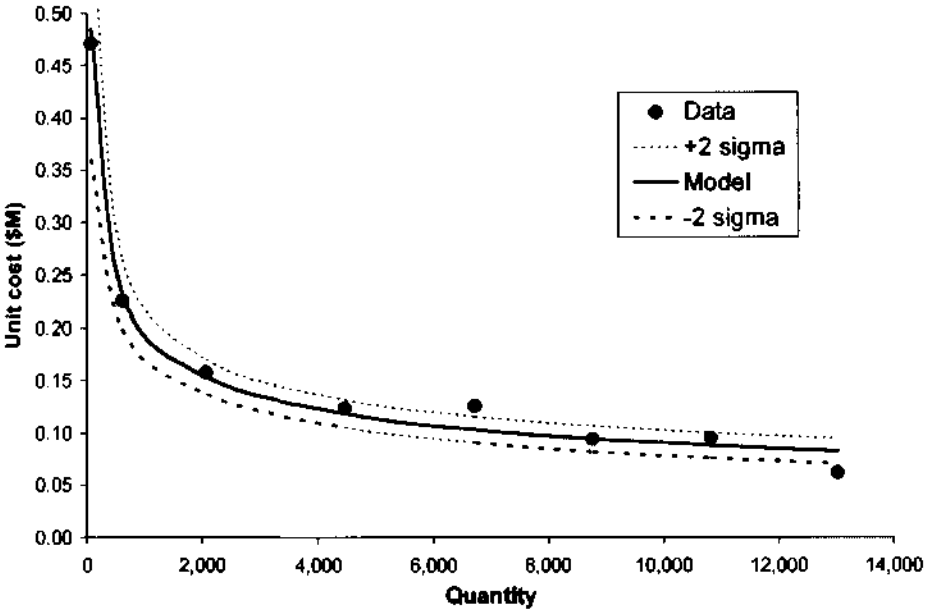


Figure 4.2. Learning Curve Fit to Tactical Missile Data

The prediction errors in Figure 4.2 do not show any indication of serial correlation. We again use Goldberger's (1964) expression for the first-order serial correlation coefficient:

$$\frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sqrt{\sum_{i=2}^{n} e_i^2}\sqrt{\sum_{i=2}^{n} e_{i-1}^2}}, \tag{4.5}$$

where e_i is the prediction error for the i^{th} lot. The serial correlation among the errors in predicting lot average cost (*not* its logarithm) is only 0.054.

Recall our previous assertion that serial correlation is more likely in a series of cumulative average costs (i.e., the Wright model) than in a series of lot average costs (i.e., the Crawford model). To test this assertion, we used the same lot-midpoint NLS estimates to predict the cumulative average cost for each lot, as in equation (2.5) (with

---

[47] We also computed the nearly unbiased predictions suggested by Eskew and Lawler (1993, 1994). The two sets of predictions differed by 0.3% on average, with a maximum difference of 0.8%. Thus, our consistent predictions appear to be essentially unbiased even in a sample containing only eight lots.

non-recurring cost set equal to zero). We then computed the prediction errors and, finally, the serial correlation among these errors. In this instance, the serial correlation coefficient evaluates much higher, 0.884.

The confidence band is useful not only for displaying the model fit within the estimation sample, hut also for predicting the next observation beyond the current sample. For example, suppose we had applied lot-midpoint NLS to Lee's data after observing only seven lots, not all eight lots, and we attempted to predict the average cost for the (as yet unobserved) eighth lot. Table 4.4 summarizes the model estimates from the sub-sample consisting of the first seven lots.

Table 4.4. Parameter Estimates from Lot-Midpoint NLS, Sub-Sample of Seven Lots

Parameter	Estimate
Sum-of-squared errors	0.0250
Sample size (n)	7
Number of parameters (k)	2
Degrees of freedom ($n-k$)	5
Standard error of regression (σ)	0.0707
T_1, pre-adjustment	1.7173
Log-normal correction factor	1.0025
T_1, post-adjustment	1.7216
Exponent (b)	–0.3112
Learning slope	80.60%

Now we are asked to predict the average cost for eighth lot, consisting of 2,768 units beginning with unit #11,669 and ending with unit #14,436. We estimate the midpoint of the eighth lot as unit #13,020, and we predict the lot average cost as $1.7216 \times 13{,}020^{-0.3112} = 0.0903$ (i.e., \$90,300). The asymptotic variance of the prediction $Var(\hat{LAC}_8)$ would be calculated as follows

$$\left[(0.5823 \quad 9.474)\times\begin{pmatrix} 0.0431 & -0.0031 \\ -0.0031 & 0.0002 \end{pmatrix}\times\begin{pmatrix} 0.5823 \\ 9.474 \end{pmatrix} + \frac{(0.0707)^4}{2\times 7}\right]\times(0.0903)^2. \qquad (4.6)$$

Expression (4.6) evaluates as $(0.0035)^2$, so the standard error of the prediction for the unobserved eighth lot equals 0.0035. Thus, the $\pm 2\sigma$ prediction interval for the average cost of the eighth lot is $0.0903 \pm 2\times 0.0035 = (0.0833, 0.0973)$, or \$83,300 to \$97,300. However, the actual average cost of the eighth lot is only \$62,000. This situation is illustrated in Figure 4.3, where that actual average cost (square data point) lies

not only below the predicted average cost (diamond), but actually below the entire prediction interval (range between the two dashed curves). The relatively low cost of the eighth lot would have come as a surprise to the cost analyst who had observed only the first seven lots.

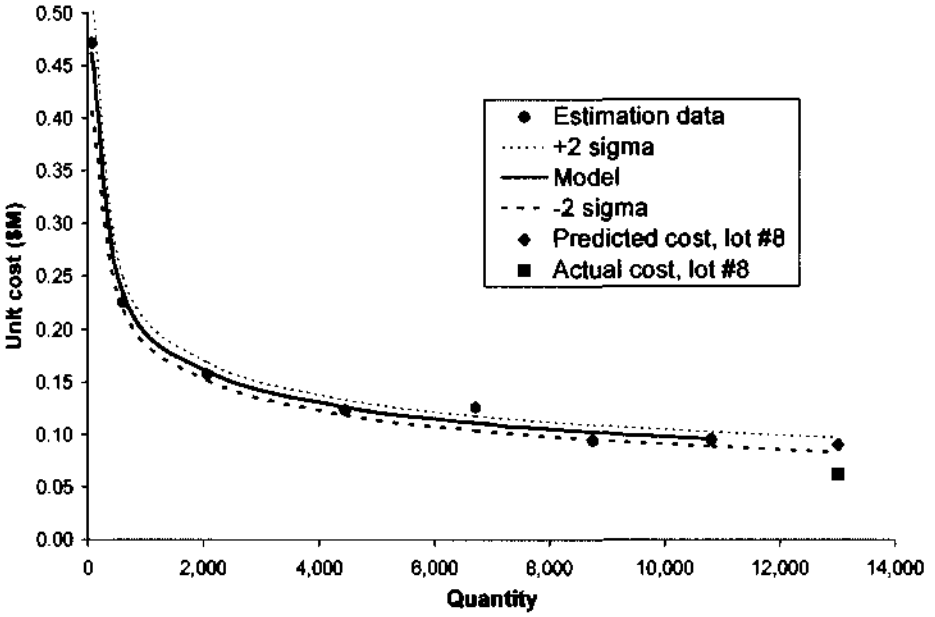


Figure 4.3. Prediction Interval for Eighth Lot from Sub-Sample of Seven Lots

4.2 Lot-midpoint iteration

We next applied lot-midpoint iteration to equation (2.27), following the algorithm outlined in equations (2.32) and (2.33). This procedure converged in four iterations, starting from the NLS estimates of incremental lot cost (i.e., starting from the first row in Table 4.1). The eigenvalue $\partial b^{(p+1)}/\partial b^{(p)}$ evaluates as –1.041 at the starting point, illustrating that an absolute bound of 1.0, while sufficient for convergence, is not necessary.

We again adjusted the intercept by the log-normal correction factor, $\exp(\sigma^2/2)$. Using equation (2.39), the adjusted intercept of 2.0301 has a standard error of 0.512. We also report the standard error of the learning coefficient b, recalling our earlier claim that the standard error is underestimated because the true lot-midpoint variable is unknown even at convergence. Even the underestimated standard error from lot-midpoint iteration is larger than the standard error from lot-midpoint NLS (0.0312 versus 0.0300); the true standard error from lot-midpoint iteration must be larger still. Thus, although the

numerical estimates are virtually identical, lot-midpoint iteration appears to be a less efficient estimation technique than lot-midpoint NLS. Moreover, there is no longer any computational advantage to using lot-midpoint iteration. Commercial spreadsheet programs now contain non-linear solvers that can easily perform lot-midpoint NLS (i.e., minimize expression (2.28)). Further, most statistical software packages automatically provide the NLS standard errors as well as the point estimates. Thus, manual calculation of the standard errors (i.e., evaluation of equation (2.31)) is no longer necessary.

We have argued rather vociferously against lot-midpoint iteration. In addition, as we show in Chapter 5, among all the estimation techniques that we compare, only lot-midpoint iteration is sensitive to serial correlation in the error terms (though that problem is not present in Lee's data). The lot midpoints themselves have the sole (and rather modest) virtue of providing plot points for each lot, as in Figure 4.2 above. Indeed, Lee (1997, p. 35) first introduces the lot midpoints merely as plot points. He then (pp. 55-56) goes on to describe lot-midpoint iteration, though he concludes rather pessimistically by stating that:

> While this procedure [lot-midpoint iteration] may appear to make the wealth of information that is known about linear regression available to the estimation of cost-progress curve parameters, the dependence of [the lot midpoint on the unknown exponent] is a complication whose consequences seem not easily seen. Today's practitioners almost always have more straightforward means of estimating cost-progress curve parameters.

4.3 Other estimation methods

The two lot-midpoint estimators attempt to fit equation (2.27) directly; that is, they attempt to minimize the differences between observed log-average cost, $\ln(LAC_i)$, and predicted log-average cost, $\ln(\hat{LAC}_i)$. Put differently, the two lot-midpoint estimators attempt to minimize the quantity:

$$\begin{aligned} \sum_{i=1}^{n}[\ln(LAC_i) - \ln(\hat{LAC}_i)]^2 &= \sum_{i=1}^{n}[\ln(LAC_i / \hat{LAC}_i)]^2 \\ &= \sum_{i=1}^{n}\left[\ln\left(1+\frac{(LAC_i - \hat{LAC}_i)}{\hat{LAC}_i}\right)\right]^2 . \end{aligned} \tag{4.7}$$

The logarithmic sums-of-squares reported in Table 4.1 are the minimized values of expression (4.7).

By contrast, the next three estimators operate on the percentage sum-of-squares:

$$\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)^2 , \tag{4.8}$$

where $y_i = LAC_i$ and $f(x_i,\beta)$ is the non-linear predictor for LAC_i given in equation (3.4). Put differently, these three estimators operate on the quantity:

$$\sum_{i=1}^{n}\left(\frac{LAC_i - \hat{LAC}_i}{\hat{LAC}_i}\right)^2 . \tag{4.9}$$

To compare the quality of estimators based on expression (4.7) with estimators based on expression (4.9), we must establish the mathematical relationship between these two measures of fit. We recall the second-order Taylor series approximation, $\ln(1+z) \approx z - z^2/2 < z$. Letting $z_i = (LAC_i - \hat{LAC}_i)/\hat{LAC}_i$, the two measures are first-order equivalent. However, their second-order relationship is theoretically indeterminate in sign. If $LAC_i > \hat{LAC}_i > 0$, then we have:

$$0 < \ln(LAC_i/\hat{LAC}_i) < \frac{(LAC_i - \hat{LAC}_i)}{\hat{LAC}_i} \tag{4.10}$$

and

$$[\ln(LAC_i / \hat{LAC}_i)]^2 < \left(\frac{LAC_i - \hat{LAC}_i}{\hat{LAC}_i}\right)^2 . \tag{4.11}$$

But if $0 < LAC_i < \hat{LAC}_i$, we have:

$$\ln(LAC_i/\hat{LAC}_i) < \frac{(LAC_i - \hat{LAC}_i)}{\hat{LAC}_i} < 0 \tag{4.12}$$

and

$$[\ln(LAC_i / \hat{LAC}_i)]^2 > \left(\frac{LAC_i - \hat{LAC}_i}{\hat{LAC}_i}\right)^2 . \qquad (4.13)$$

This indeterminacy cannot be resolved theoretically and, in fact, the ordering between expressions (4.7) and (4.9) depends on the particular data set. To make a fair comparison, consider a case in which $\sum_{i=1}^{n} z_i \approx 0$, as must be true for any reasonable estimation procedure. As indicated by inequalities (4.11) and (4.13), the squared percentage error exceeds the squared logarithmic error when $z_i > 0$ (i.e., for a data point lying above the model prediction — a positive outlier), but the relationship is reversed when $z_i < 0$ (a negative outlier).

The circles in Figure 4.4 represent a hypothetical data set containing a large positive outlier (i.e., a point lying far above its model prediction), yet balanced by four points lying slightly below their model predictions so that $\sum_{i=1}^{n} z_i = 0$. The positive outlier evaluates much higher on the function z^2 than on the function $[\ln(1+z)]^2$ in the right-hand side of the figure, dominating the other terms in the summation. Thus, for this data set, the percentage sum-of-squares is larger than the logarithmic variant (1.025 > 0.698). Conversely, the squares in Figure 4.4 represent a hypothetical data set containing a large negative outlier. For the latter data set, the negative outlier evaluates much higher on the function $[\ln(1+z)]^2$, dominating the other terms and causing the logarithmic sum-of-squares to be larger (5.473 > 1.025).

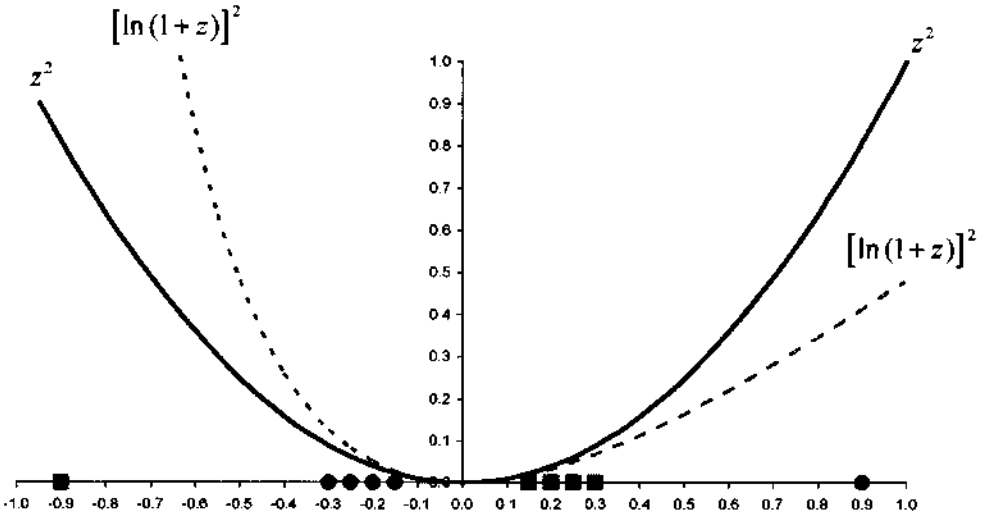


Figure 4.4. Two Measures of Fit, Hypothetical Data

For a symmetric data set with no extreme outliers, the logarithmic sum-of-squares will generally be larger than the percentage sum-of-squares. To see why, consider a data set containing $n/2$ under-predicted points $\{z_1,\ldots,z_{n/2}>0\}$ and, symmetrically, $n/2$ over-predicted points $\{-z_1,\ldots,-z_{n/2}<0\}$. Put differently, the percentage deviations in the data set are $\{\pm z_1,\ldots,\pm z_{n/2}\}$. The percentage sum-of-squares over the entire data set is simply:

$$\sum_{i=1}^{n/2} z_i^2 + \sum_{i=1}^{n/2}(-z_i)^2 = 2\times\sum_{i=1}^{n/2} z_i^2 . \tag{4.14}$$

Again using the second-order approximation $\ln(1+z) \approx z - z^2/2$, the logarithmic sum-of-squares over the entire data set is larger:

$$\begin{aligned}\sum_{i=1}^{n/2}[\ln(1+z_i)]^2 + \sum_{i=1}^{n/2}[\ln(1-z_i)]^2 &\approx \sum_{i=1}^{n/2}[z_i-(z_i^2/2)]^2 + \sum_{i=1}^{n/2}[-z_i-(z_i^2/2)]^2 \\ &\approx 2\times\sum_{i=1}^{n/2} z_i^2 + \sum_{i=1}^{n/2}(z_i^4/2) > 2\times\sum_{i=1}^{n/2} z_i^2 .\end{aligned} \tag{4.15}$$

Young (1999) also compared the logarithmic and percentage sums-of-squares. However, he did not notice our inequalities (4.11) and (4.13); more importantly, he did not impose the symmetry condition $\sum_{i=1}^{n} z_i \approx 0$. Figure 4.5 shows the values of z for Young's two examples.[48] His Example 1, depicted as circles, has $\sum_{i=1}^{3} z_i = 0.4$ in our notation, with two positive outliers. His Example 2, depicted as squares, has $\sum_{i=1}^{3} z_i = -0.99$ with two large (but symmetrical) outliers as well as one extremely large negative outlier. The value $z_3 = (y_3 - \hat{y}_3)/\hat{y}_3 = -0.99$ implies that $\hat{y}_3 = 100\times y_3$, so the prediction error is hundred-fold. The prediction errors for the two symmetrical outliers are ten-fold. In either case, the condition $\sum_{i=1}^{n} z_i \approx 0$ is clearly violated, thus Young's examples shed little insight on the general relationship between the logarithmic and percentage sum-of-squares.

---

[48] Note that Young's variable z corresponds to $z-1$ in our notation. We will use our notation throughout the discussion of his examples.

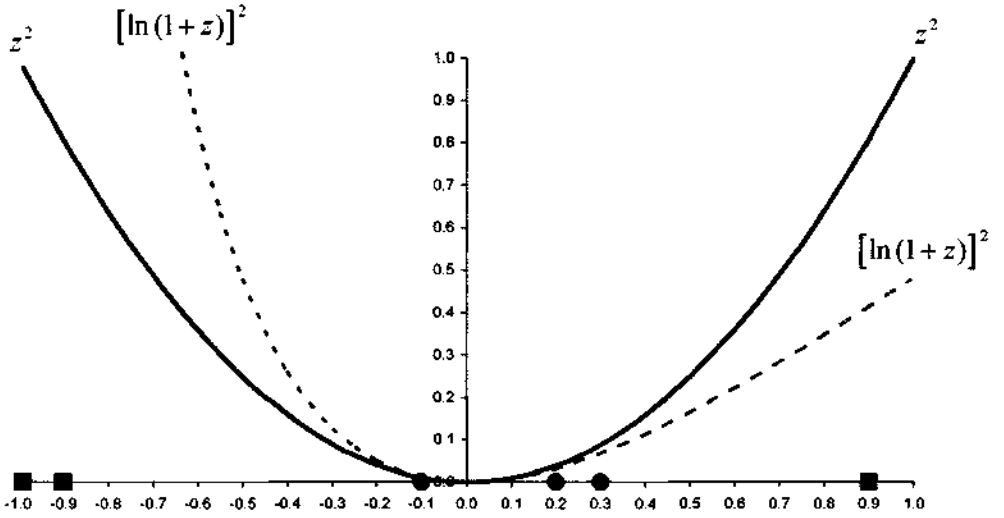


Figure 4.5 Two Measures of Fit, Young's Data

Finally, Figure 4.6 results when lot-midpoint NLS is applied to Lee's data set. The values of z are simply the percentage errors from Figure 4.2. We again observe two minor outliers at the fifth ($\hat{y}_5 = 0.821 \times y_5$) and eight ($\hat{y}_8 = 1.334 \times y_8$) lots. The logarithmic sum-of-squares is larger than the percentage sums-of-squares (0.1328 > 0.1218). Indeed, this ordering holds not only for the lot-midpoint NLS estimates, but for all of the estimates reported in Table 4.1. This pattern is consistent with inequality (4.15), because the data set is fairly symmetric, the errors sum to approximately zero ($\sum_{i=1}^{8} z_i = 0.064$), and the two outliers are modest.

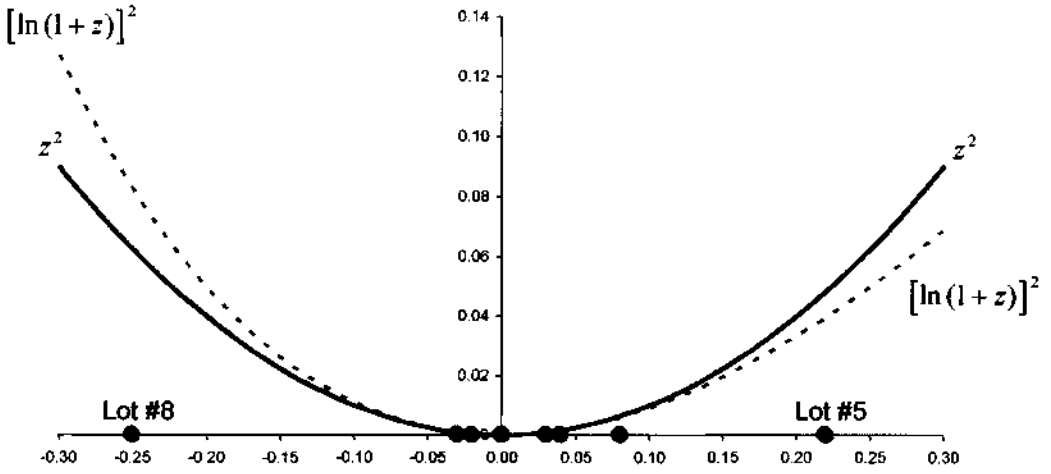


Figure 4.6 Two Measures of Fit, Lee's Data

We also computed the standard error of the estimate, dividing the percentage sum-of-squares in expression (4.8) by the degrees of freedom $(n-m)$, then taking the square root. This procedure is suggested by the quasi-likelihood dispersion estimate in equation (3.19). Because the standard error, so computed, is monotone in the percentage sum-of-squares, these two measures provide identical rankings of the various estimators; the standard error is simply a more familiar metric.

The MPE, MLE (assuming multiplicative normal errors, as in equation (3.7)), and IRLS estimators round out Table 4.1. All of these estimators avoid the logarithmic transformation and operate directly on the lot-average cost, equation (3.4). The MLE is suggested, but not actually applied to this data set, by Lee (1997, pp. 47–49). We computed the covariance matrix of the MLE as the negative inverse Hessian of the concentrated log-likelihood function; we approximated the Hessian via numerical second differencing in the neighborhood of the maximum.[49]

Finally, the IRLS (MUPE) estimates are identical to maximum quasi-likelihood. The IRLS estimates converged in three iterations, starting from the first row in Table 4.1. As previously remarked, IRLS does not minimize the percentage sum-of-squares (0.1180 versus 0.1153 at the MPE solution). The gradient of the percentage sum-of-squares (expression (4.8)), evaluated at the IRLS estimate, is: $\partial/\partial(T_1, b) = (-0.120, -2.163) \neq 0$.

Although all of the alternative estimates are numerically close, a few interesting differences emerge from Table 4.1. Because the two lot-midpoint estimators attempt to predict log-average cost directly (equation (4.7)), they score the best in terms of logarithmic sum-of-squares. By contrast, the MPE estimator explicitly minimizes the percentage sum-of-squares (equation (4.9)), thus MPE scores the best in terms of this metric as well as the monotonically-related standard error of estimate. Book and Young (1997) report a bias in their MPE estimates as high as 29%, though typically closer to 8%. While we do not know the true parameter values, the MPE estimates for Lee's data lie within the range of the other estimates that are known to be unbiased. Thus, we find no evidence of bias when MPE is applied to this particular data set. However, the Monte Carlo experiments reported in Chapter 5 reveal considerable bias in the MPE estimates.

We also report the log-likelihood and log-quasi-likelihood values not only at their respective maxima, but also evaluated at each of the other estimates in Table 4.1. The log-likelihood function, again assuming multiplicative normal errors, is relatively flat in

---

[49] Numerical differentiation is covered by Dennis and Schnabel (1996, p. 80 and pp. 103–106).

the neighborhood of the MLE, with most differences among the estimates appearing in the *4th* significant digit. The log quasi-likelihood function is even flatter, with most differences appearing in the *5th* significant digit. Thus, there is little basis for distinguishing the alternative estimates for this particular data set.

The results of this chapter, though illuminating, are not definitive because we do not know the true parameter values that generated Lee's data. Therefore, we supplement Lee's data with a series of Monte Carlo experiments, in which we know the true parameter values. We report the results of these Monte Carlo experiments in Chapter 5.

4.4 Correction for serial correlation

When using lot-midpoint NLS, we found serial correlation of only 0.054 among the errors in predicting lot average cost. However, the serial correlation coefficient will vary somewhat with the method of estimation. Moreover, for the sake of completeness, we want to illustrate the correction for serial correlation under NLS. To make the problem more interesting, we consider NLS applied directly to a regression of lot average cost, without the artifice of lot midpoints. Thus, we return to expression (4.2) and the corresponding estimates that appear in the second row of Table 4.1. For that estimation method, the serial correlation coefficient evaluates as –0.257.

Expression (4.2) is consistent with our conjecture, back in Chapter 1, that a modern statistician would simply apply NLS to the model for lot average cost based on the area under the continuous approximation to the learning curve. However, we must correct those estimates for serial correlation. In addition, we must also correct for heteroscedasticity if we make the now-familiar multiplicative error assumption:

$$LAC_i = \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})} \times \left[(Q_i + 0.5)^{1+b} - (Q_{i-1} + 0.5)^{1+b} \right] \times u_i \ . \qquad (4.16)$$

Table 4.5 shows the weights $w_i = f(x_i, \beta^{(0)}) = LAC(Q_i, Q_{i-1}; T_1^{(0)}, b^{(0)})$, where $(T_1^{(0)}, b^{(0)})$ are the starting estimates obtained by conventional NLS applied to the model for lot average cost. Table 4.6 shows the starting estimates (repeated from the second row of Table 4.1) and the final estimates after a single step of non-linear generalized least squares. In this instance, NGLS produces an almost imperceptibly steeper learning slope. The NGLS intercept is 3.7 percent higher than the NLS intercept, an apparently large difference. However, referring back to Figure 4.1, the NGLS parameters appear almost exactly on the line that interpolates between the various other sets of parameter estimates.

The higher intercept compensates for the steeper slope, ensuring that the NGLS learning curve passes through the centroid of the data.

Table 4.5. Weights for NGLS Estimation

Lot number	Actual lot average cost	Predicted lot average cost (w_i)
1	0.4714	0.4708
2	0.2257	0.2311
3	0.1576	0.1551
4	0.1236	0.1208
5	0.1257	0.1057
6	0.0939	0.0970
7	0.0953	0.0905
8	0.0619	0.0852

Table 4.6. Comparison of NLS and NGLS Estimates

	NLS	NGLS
Intercept	1.8593	1.9281
Quantity exponent	–0.3254	-0.3298
Learning slope	79.81%	79.57%

5. SIMULATION EXPERIMENTS

As shown in the previous chapter, there are limitations to comparing the effectiveness of the different estimation methods when using an actual data set, such as Lee's data on tactical missiles. The "true" values of the parameters are unknown, so it is impossible to say which of the methods' estimates are closest to "truth." The different methods minimize different functions, and comparing the relative merits of these functions is subjective. Nor is it possible to fully compare the estimation errors, because the theoretical form of the covariance matrix for some of the methods is unknown.

These difficulties motivated a series of Monte Carlo simulation experiments. Because the simulated data were generated using known parameters, the estimates produced by the different methods could be directly compared to "truth." The bias and random error in the parameter estimates could be separately measured, even for methods where theoretical values were unknown. In addition, in most cases, even if a formula for the covariance matrix is available, the matrix produced is only an asymptotic covariance matrix. Tbe simulation experiments allowed us to compare the variances over a spectrum of sample sizes, ranging from very small (unfortunately, the typical situation in cost analysis) up to asymptotically large.

Because actual data sets do not always have normally distributed random errors, we examined several alternative error structures. By varying the error structure, we could determine how robust the methods are even if their assumptions are incorrect. We also varied the underlying parameter values, because it was unclear from theory alone how the parameter values affected the covariance matrix of the lot-midpoint iteration and MPE estimates.

5.1 Basic methodology

We compared four of the estimation methods previously discussed. We did not include maximum likelihood because this method requires the most computation to converge, and because its properties (at least, asymptotically) are already well known. Each simulation is defined by the assumed values of the "true" parameters, the magnitude

of the random error, the structure of the random error, the number of lots, and the method of estimation.

We first conducted a baseline experiment under the following conditions:

- the true learning slope equals 80%,
- each lot contains 50 units,
- the error term u_i is normally distributed with standard deviation $\sigma = 0.15$, and
- the error terms $\{u_i\}$ are statistically independent.

We then conducted the following excursions, varying one assumption at a time relative to the baseline experiment:

- the true learning slope equals 90%;
- each lot contains 10 units;
- the error term is normally distributed with standard deviation $\sigma = 0.30$;
- the error term is uniformly distributed with standard deviation $\sigma = 0.15$;
- the error term is t–distributed with standard deviation $\sigma = 0.15$; and
- the error term is normally distributed with standard deviation $\sigma = 0.15$, but suffers from first-order serial correlation.

Once we selected parameter values and error structures, we calculated "true" lot average costs using the "true" parameter values and the theoretical formula from the Crawford model:

$$LAC_i = \frac{T_1}{(1+b)\times(Q_i - Q_{i-1})}\times\left[(Q_i+0.5)^{1+b} - (Q_{i-1}+0.5)^{1+b}\right]. \qquad (4.1)$$

We generated observed lot average costs by applying random error to the true lot average cost using a pre-determined error structure, discussed in each experiment below. The estimation method of interest was then applied to these observed costs to estimate the parameters T_1 and b. Finally, the estimated parameters were compared to the true parameters.

In each simulation experiment, and for each method of estimation, we varied the number of consecutive lots from 5 to 200. The number of lots represents the sample size, n, from the previous chapters. The range in sample size enables us to examine both the

small-sample and the asymptotic properties of each estimation method. For each sample size, we ran 3,000 repetitions of the simulation experiment. Thus, for each case, we produced 3,000 sets of estimated parameters for each method.[50]

We summarized the error in the estimate of each parameter using the following measures. First, we calculated the root mean squared error (RMSE) between "true" parameter and the estimates of that parameter:

$$RMSE_b = \sqrt{1/3000 \times \sum_{i=1}^{3000} (b - \hat{b}_i)^2}, \quad RMSE_{T_1} = \sqrt{1/3000 \times \sum_{i=1}^{3000} (T_1 - \hat{T}_{1i})^2}. \tag{4.2}$$

The RMSE includes both bias and random error. We used the following formulas to decompose the RMSE into its bias and random error components:

$$RMSE_b = \sqrt{bias_b^2 + ran\_err_b^2}, \tag{4.3}$$

where:

$$bias_b = 1/3000 \times \sum_{i=1}^{3000} (b - \hat{b}_i), \tag{4.4}$$

$$ran\_err_b = \sqrt{Var(b - \hat{b}_i)} = \sqrt{\left[1/3000 \times \sum_{i=1}^{3000} (b - \hat{b}_i)^2\right] - \left[1/3000 \times \sum_{i=1}^{3000} (b - \hat{b}_i)\right]^2}. \tag{4.5}$$

The corresponding formulas for T_1 have the same form.

Finally, we are interested in how errors in the parameter estimates propagate when attempting to predict unit cost (*not* lot average cost), $T_1 \times Q^b$, at a given cumulative quantity. Therefore, we report the bias and random errors for unit cost at the following cumulative quantities: 50, 100, and 1,000 units (i.e., 1, 2, and 20 lots).

---

[50] It appears that 3,000 repetitions are sufficient to capture the basic behavior of the estimation methods we compared. As we show in some of the summary plots, additional repetitions might have controlled the erratic behavior observed in a few of the simulation experiments. However, because of limitations on computer time, we did not perform additional repetitions in this study.

5.2 Simulation experiment 1: multiplicative, normal errors, learning slope = 80% (baseline)

For this simulation experiment, we chose the parameter values $b = -0.33$ and $T_1 = 1.8$ to resemble the estimates derived in Chapter 4 from Lee's tactical missile data. We calculated the true lot average cost using equation (4.1). We generated the observed lot average cost by applying a multiplicative normal error with standard deviation $\sigma = 0.15$:

$$Obs\_LAC_i = LAC_i \times u_i, \tag{4.6}$$

where $u_i \sim N\left(1.0, 0.15^2\right)$. We generated the error terms $\{u_i\}$ independently, without any serial correlation: $E(u_i u_j) = 0$ for all $i \neq j$.

In Chapter 3 we reviewed the MLE under a multiplicative normal error structure, and we showed that same error structure underlies the derivation of the MPE. The lot-midpoint NLS and lot-midpoint iteration methods assume, instead, a log-normal error structure. As Figure 2.9 revealed, these two error structures are quite similar for random errors of the magnitude $\sigma = 0.15$. Nonetheless, we are evaluating the two lot-midpoint methods under a slightly different error structure from the one assumed in their derivation.

Figures 5.1 through 5.5 compare the estimation errors for the different methods. Three of the methods — IRLS, lot-midpoint NLS, and lot-midpoint iteration — are consistent estimators and converge to the true parameter values at similar rates. Although it was known from theory that IRLS and lot-midpoint NLS would produce consistent estimates, we were surprised to find that lot-midpoint iteration performed about as well. While the latter method does not minimize any continuous function, it nonetheless produced consistent estimates. Moreover, the estimation errors from lot-midpoint iteration were quite close to those found in the two methods having a stronger theoretical basis.

MPE did not perform as well, producing biased estimates as predicted from the theory. The MPE estimates of b are biased for small numbers of lots, although the bias decreases as the number of lots increases. However, the bias in T_1 remained nearly constant even with large numbers of lots. The bias was small relative to the random error and is therefore not obvious when examining the RMSE for the parameter estimates. But when projections are made for unit cost, the bias is large enough to separate MPE from

the other methodologies. For example, when the cost of the 1,000th unit is estimated (Figure 5.5), the bias accounts for half of the total error. In contrast, none of the other methodologies showed substantial bias.

For all methods examined, the errors in estimating the cost at the 1,000th unit are smaller than the errors at the 50th unit. This is true for both the absolute RMSE errors (see Figures 5.3 and 5.5) and the percentage errors (see Figure 5.6). In general, the errors are smaller nearest to the observed data, and increase as larger extrapolations are made. In the percentage error plots, the errors for both the 50th unit (1 lot) and the 1,000th unit (20 lots) are close to 8% when only a few lots have been observed. However, the error for the cost of the 1,000th unit declines rapidly, while the error at the 50th unit remains nearly constant. This difference is due to the multiplicative nature of the model. Because the model is multiplicative, the predictor variable—cumulative quantity—should more properly be treated on a logarithmic scale. As more lots of equal size are accumulated, they become more tightly clustered on a logarithmic scale (see Figure 5.7). More data points are observed near the 1,000th unit than near the 50th unit, so the error at the 1,000th unit declines. Furthermore, because of the logarithmic scale, extrapolating forward to higher cumulative quantities will produce less error than extrapolating backward to lower cumulative quantities. Of course, if the lots were smaller at low quantities, predicting the cost at low quantities would involve interpolation rather than extrapolation and the errors might be smaller.

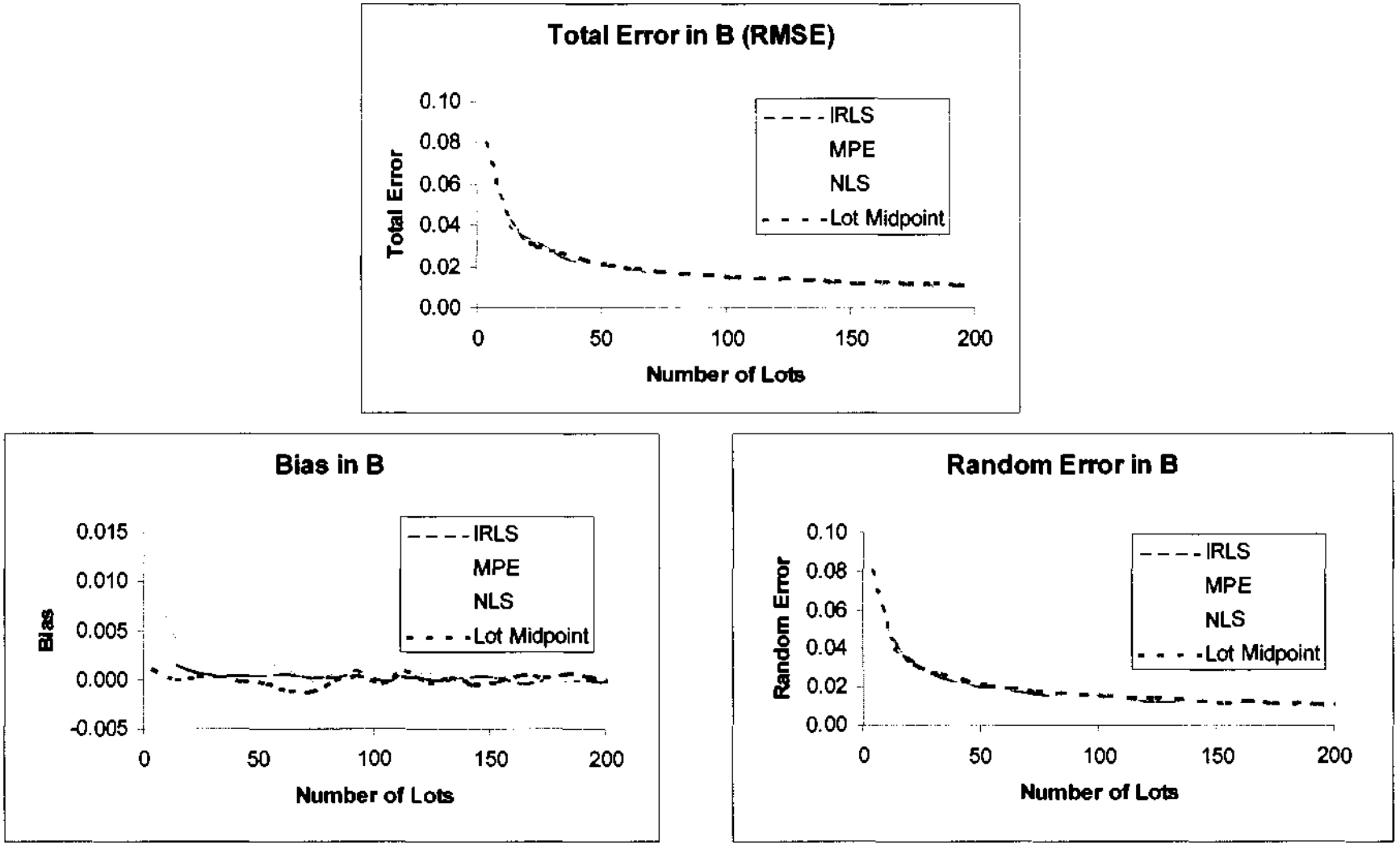
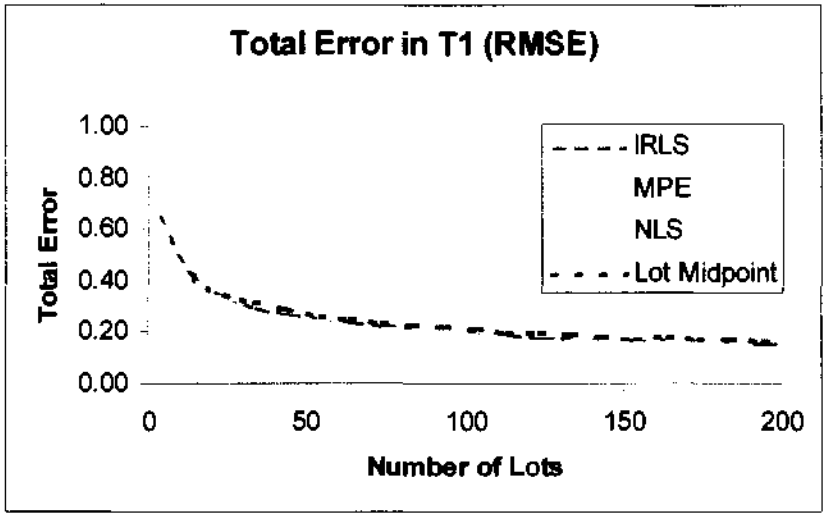
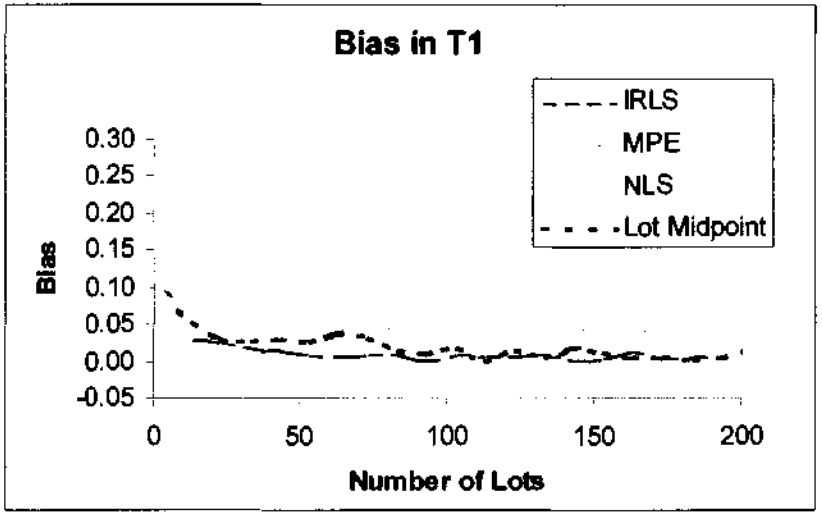


Figure 5.1. Simulation Experiment 1, Error in Slope Coefficient





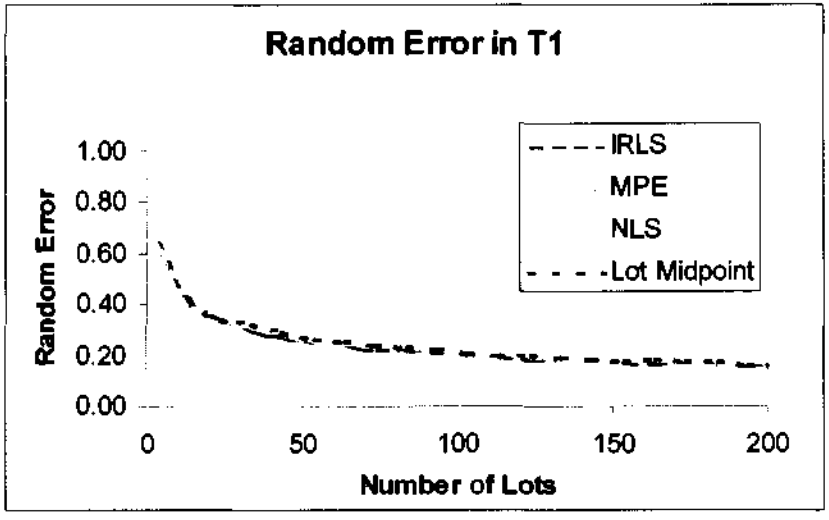
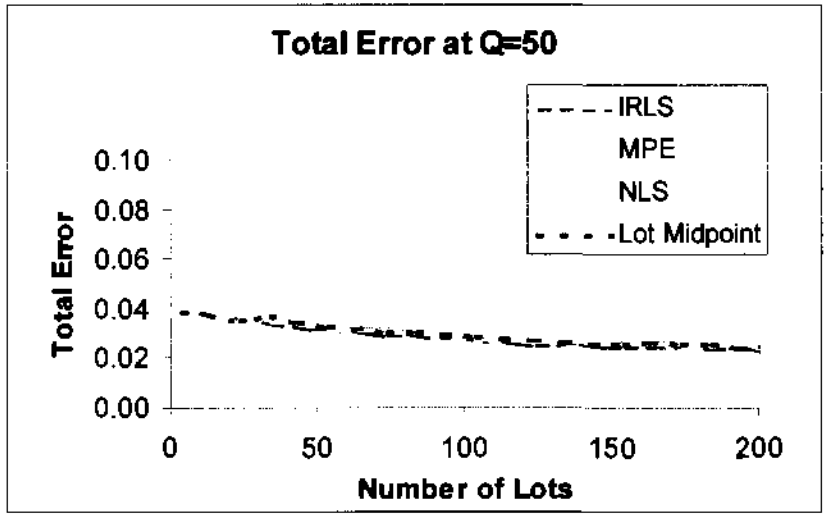
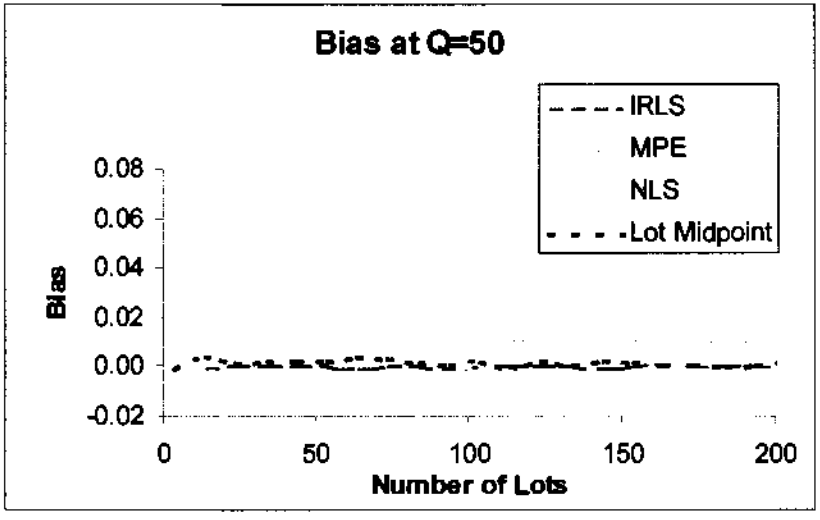


Figure 5.2. Simulation Experiment 1, Error in Intercept





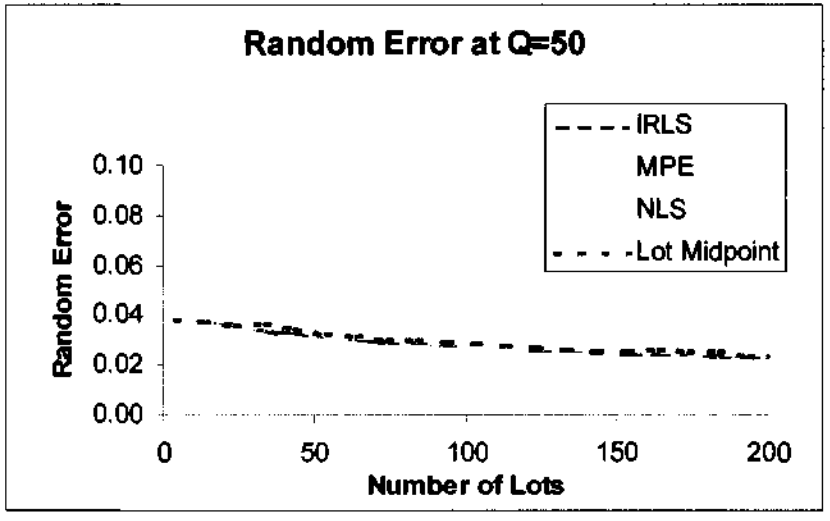
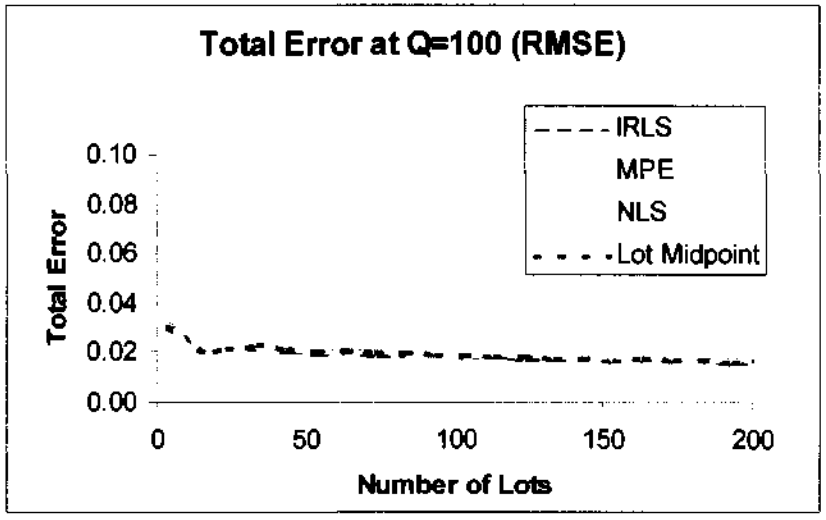
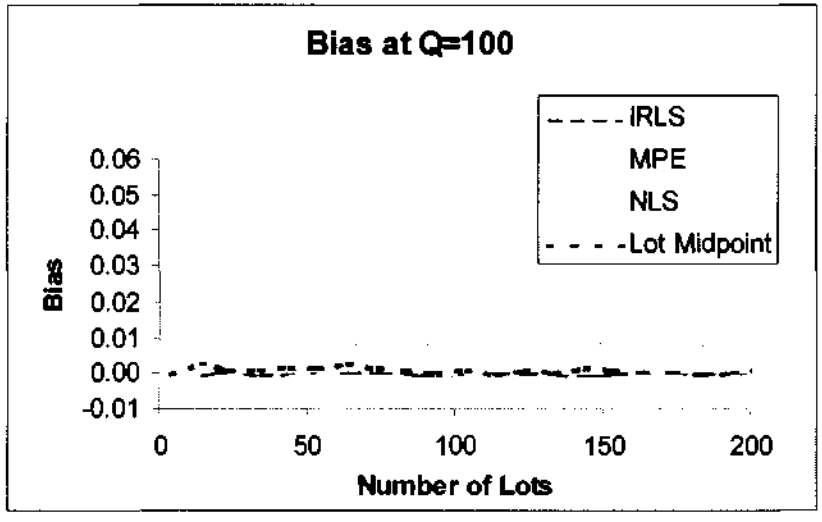


Figure 5.3. Simulation Experiment 1, Error in Predicting the Cost of Unit 50





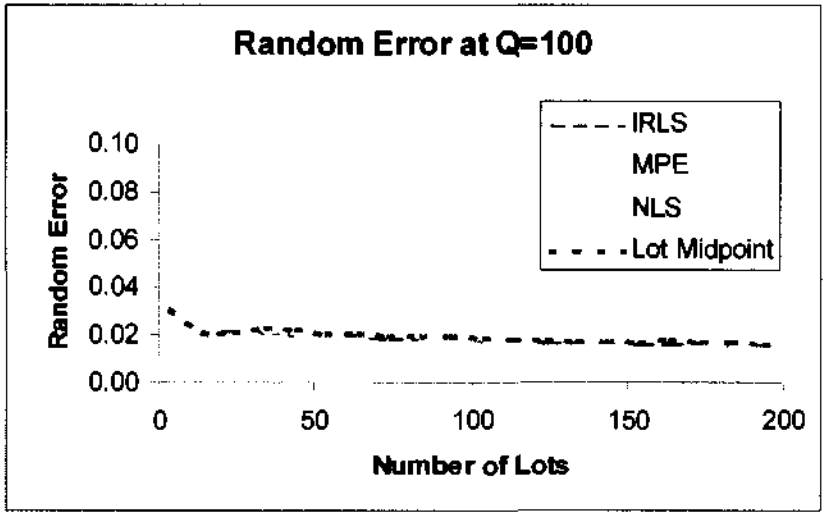
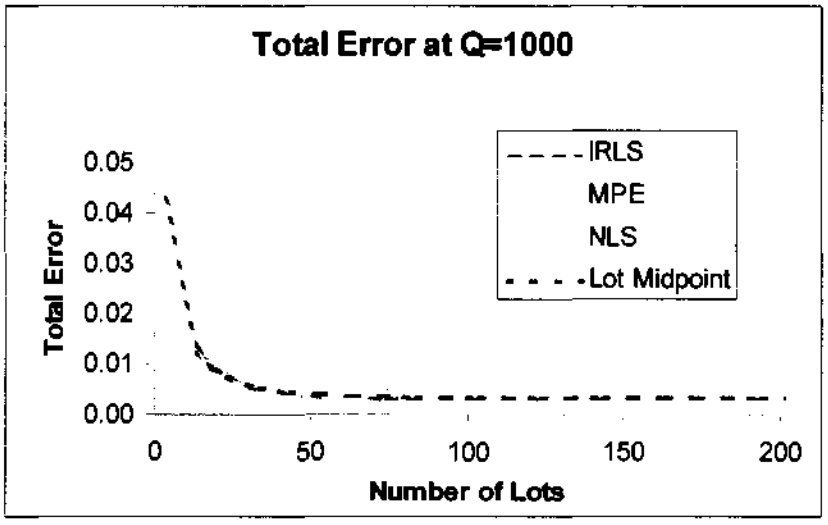
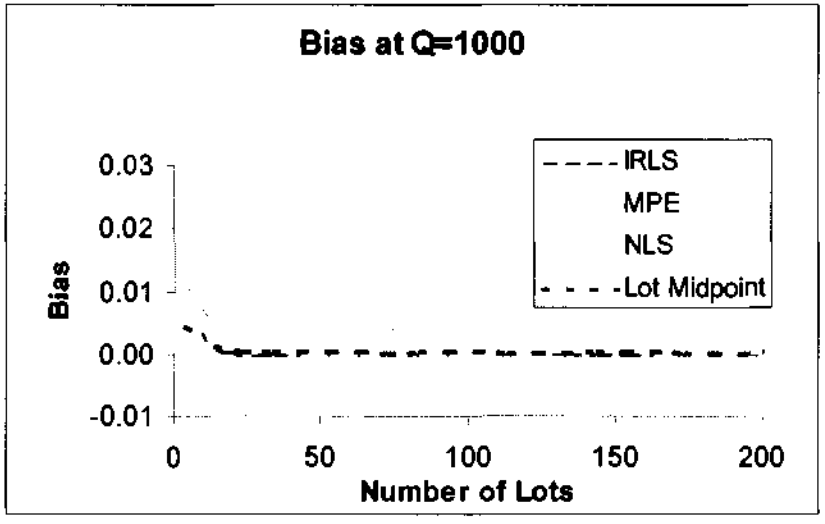


Figure 5.4. Simulation Experiment 1, Error in Predicting the Cost of Unit 100





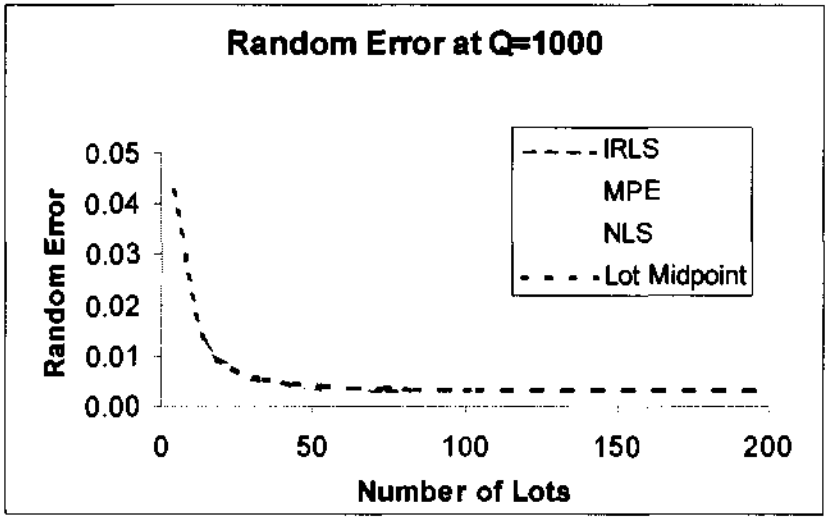


Figure 5.5. Simulation Experiment 1, Error in Predicting the Cost of Unit 1,000

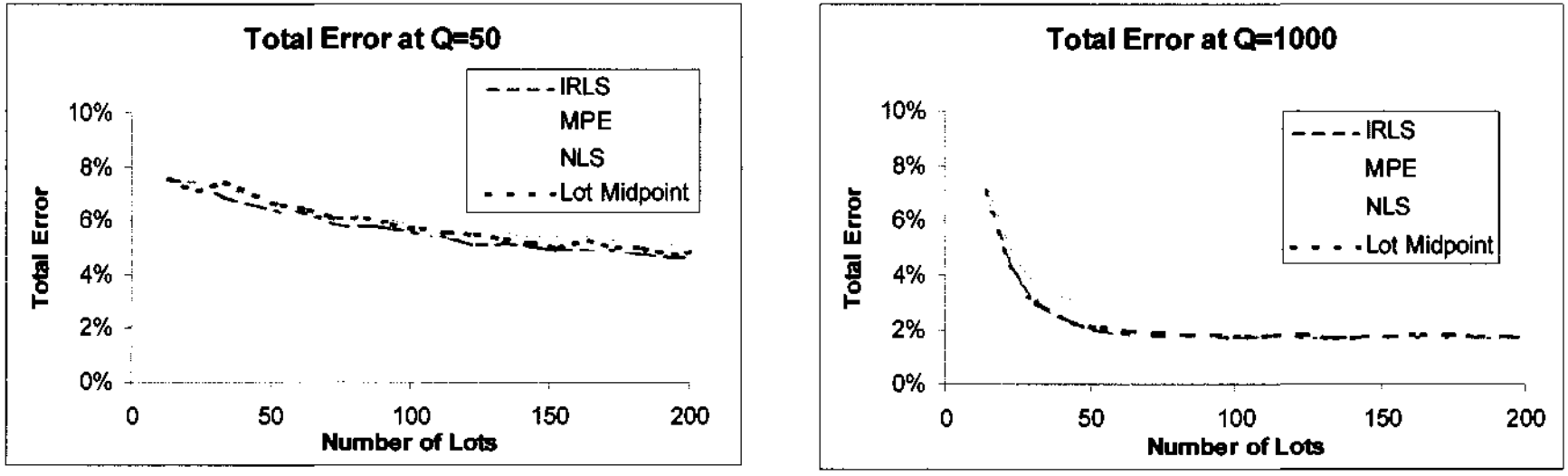


Figure 5.6. Simulation Experiment 1, Comparing Predictions at Different Cumulative Quantities

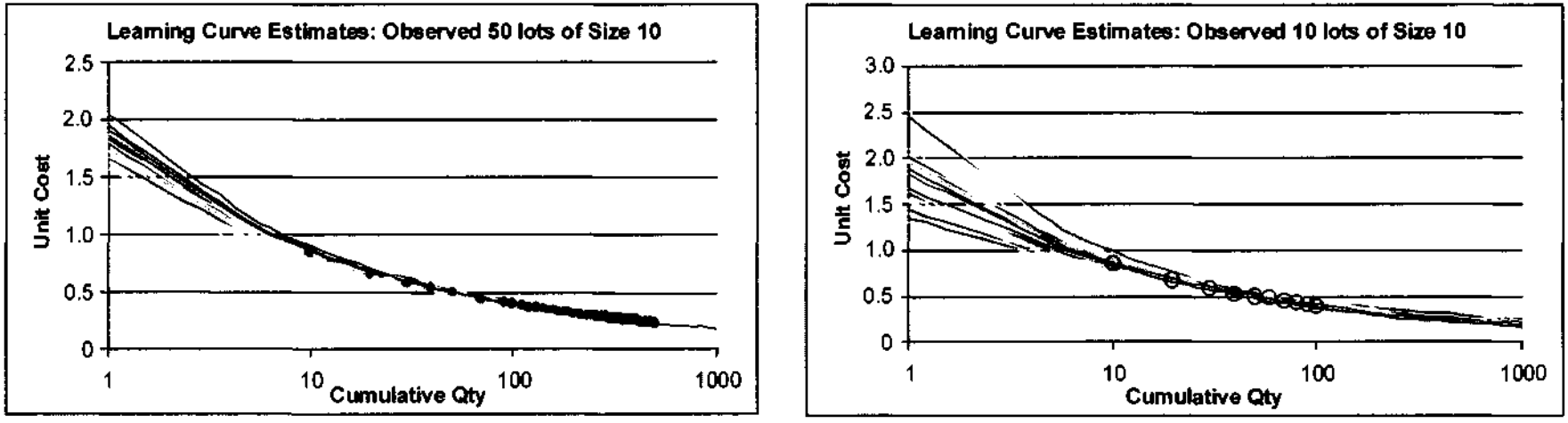


Figure 5.7. Effects of Number of Lots Observed on Extrapolated Errors

5.3 Simulation experiment 2: multiplicative, normal errors, learning slope = 90%

For this experiment, we retained $T_1 = 2.0$ but we changed the exponent to $b = -0.15$. Thus, we increased the learning slope from 80% to approximately 90%. Because of the way the learning slope is defined, 90% is a *shallower* slope and corresponds to less rapid learning than in the previous experiment. We also retained the multiplicative normal error structure from the previous experiment, again with $\sigma = 0.15$. Hence, we again generated the observed lot average cost as:

$$Obs\_LAC_i = LAC_i \times u_i, \tag{4.7}$$

where $u_i \sim N\left(1.0, 0.15^2\right)$.

Figures 5.8 through 5.12 present our results. Again, as expected, MPE produces biased estimates even with a large number of lots. The other methods are unbiased and all perform about equally well. As in the baseline experiment, the ability of lot-midpoint iteration to produce unbiased estimates is impressive though lacking in theoretical foundation.

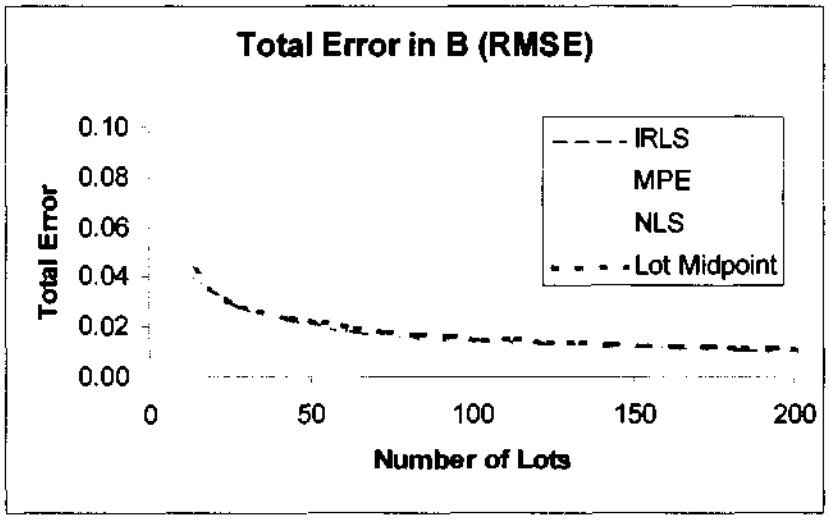
When compared to the steeper slope, the raw total error assuming a 90% slope is considerably higher (see Figure 5.13). To understand this finding, consider the following identity:

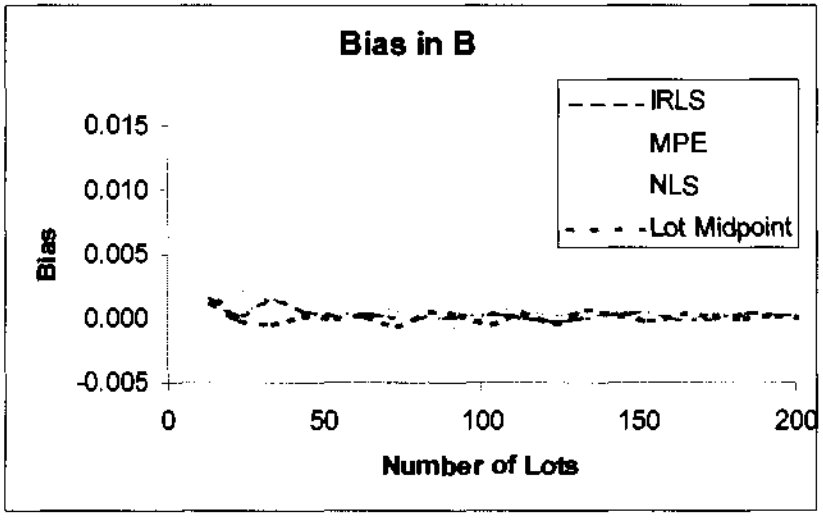
$$T_1 \times \bar{Q}^{-0.15} = T_1 \times \left(\bar{Q}^{0.5}\right)^{-0.3}. \tag{4.8}$$

Predicting the cost of a given cumulative quantity, $\bar{Q}$, for a 90% slope (left-hand side) is equivalent to predicting the cost of the *square root* of that quantity for an 80% slope (right-hand side). As already discussed, there is less error in predicting cost at higher quantities than at lower quantities, as long as some data have been observed near both. For a given quantity, $\bar{Q}$, more error can be expected in the estimate using a 90% slope than using an 80% slope, because the former is tantamount to predicting at a lower cumulative quantity.

Figure 5.13 shows that the absolute error assuming a 90% slope is higher than the absolute error assuming an 80% slope at the same cumulative quantity. However, the unit cost at that quantity is also higher for the assumed 90% slope. Figure 5.14 shows that the *percentage* prediction errors under the two assumptions are virtually identical. This

finding is plausible because the two simulations use the same percentage random error ($\sigma = 0.15$). Ultimately, it is the percentage random error, not the slope, that determines the accuracy of the predictions.





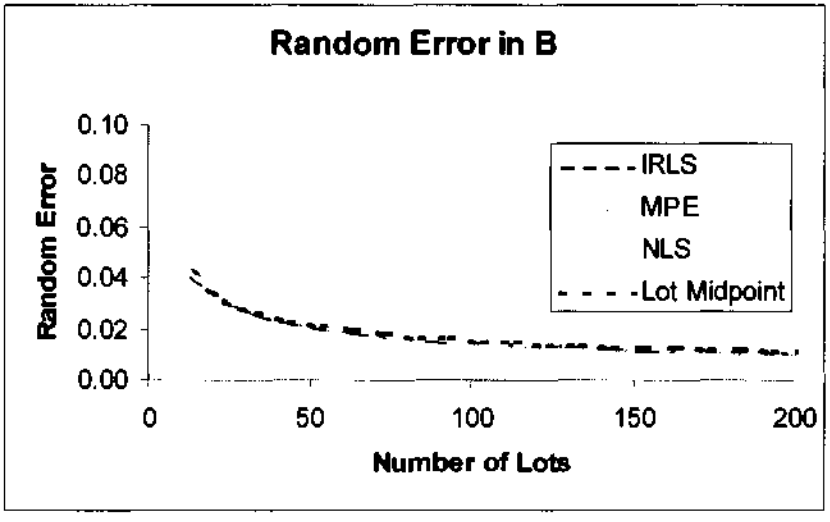
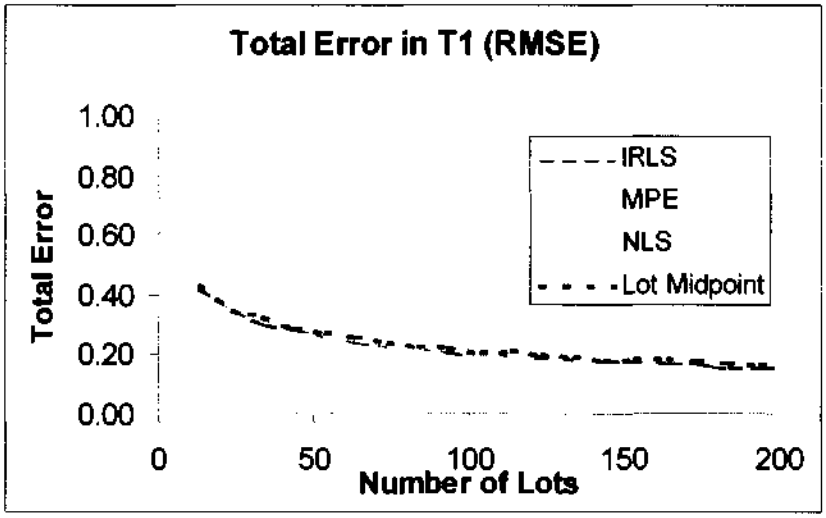
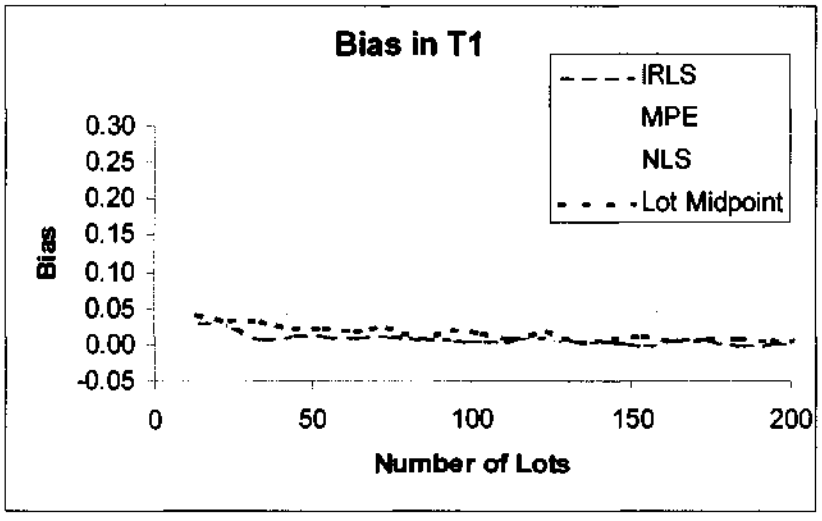


Figure 5.8. Simulation Experiment 2, Error in Slope Coefficient





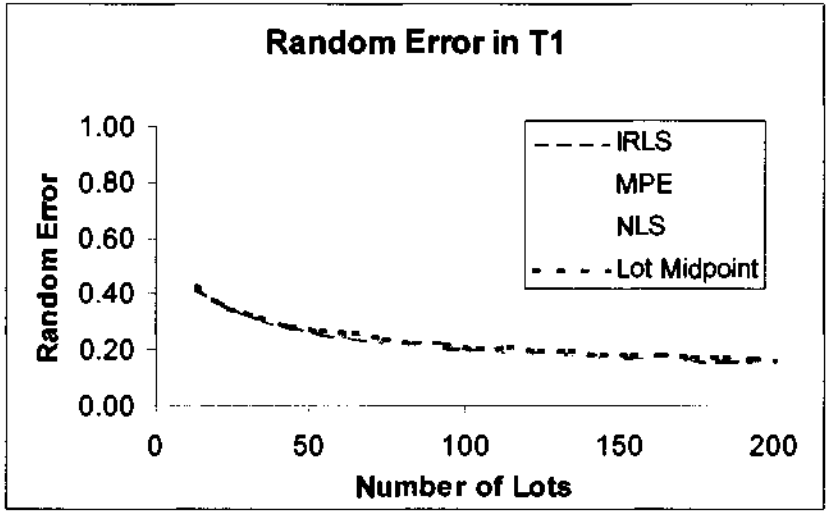
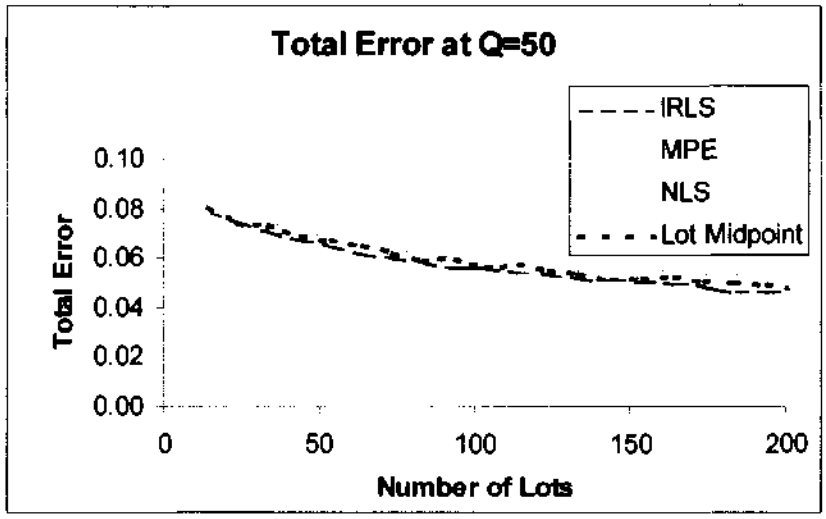
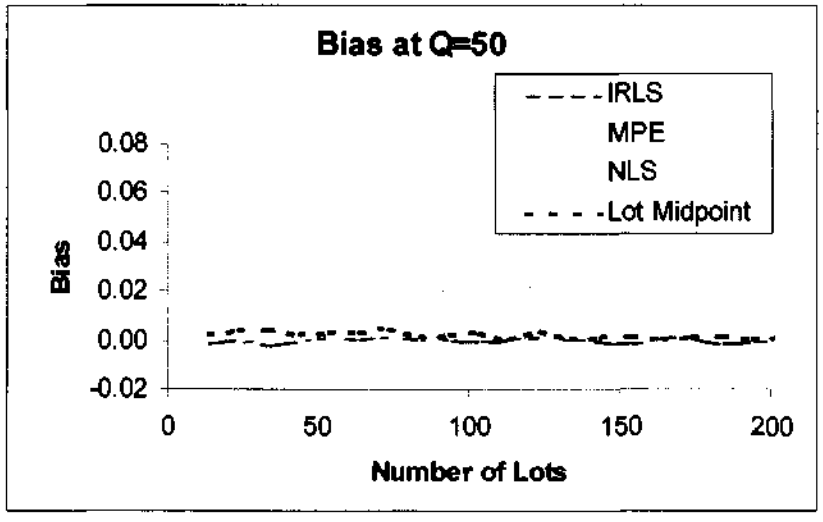


Figure 5.9. Simulation Experiment 2, Error in Intercept





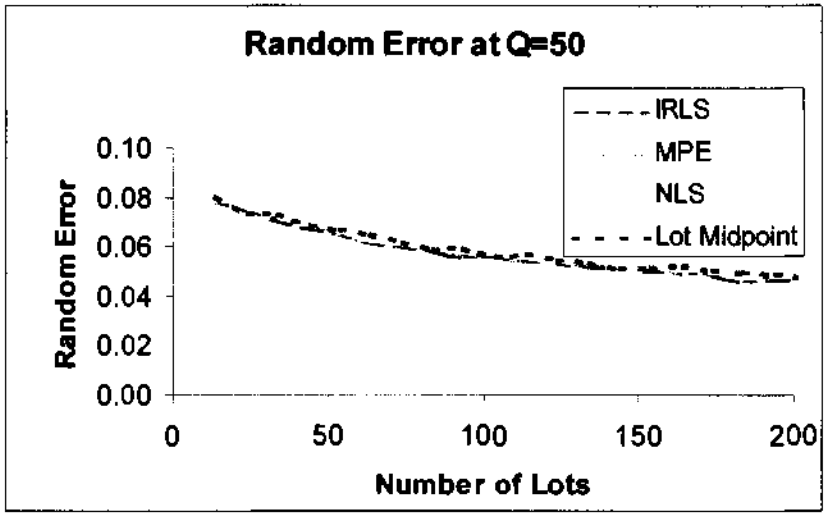
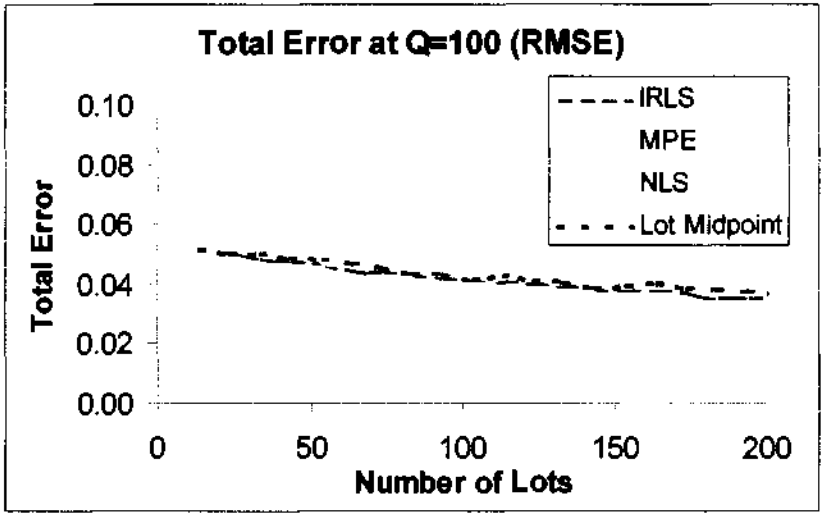
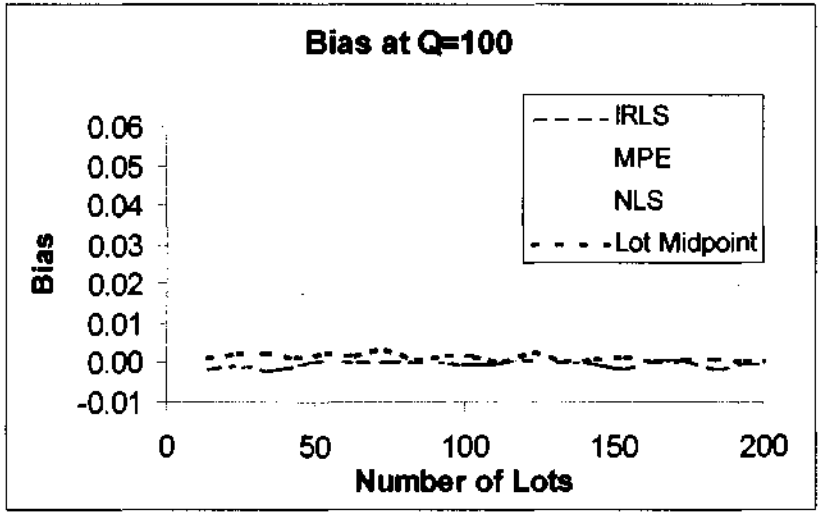


Figure 5.10. Simulation Experiment 2, Error in Predicting the Cost of Unit 50





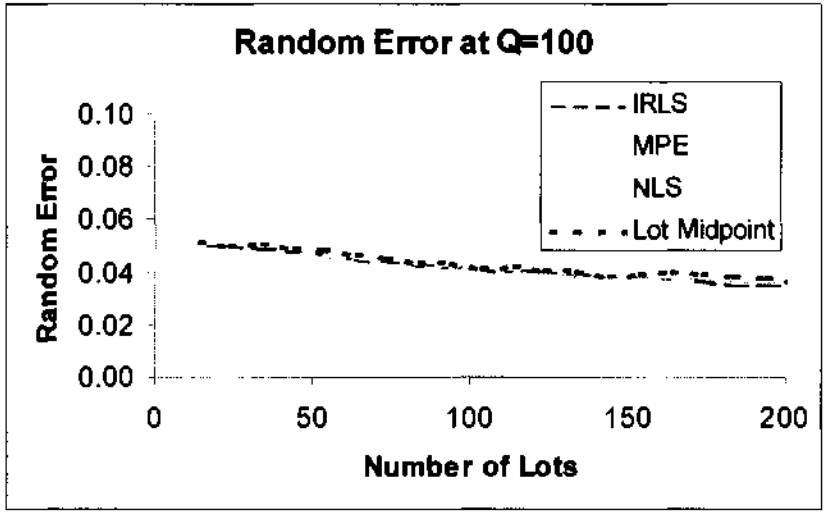
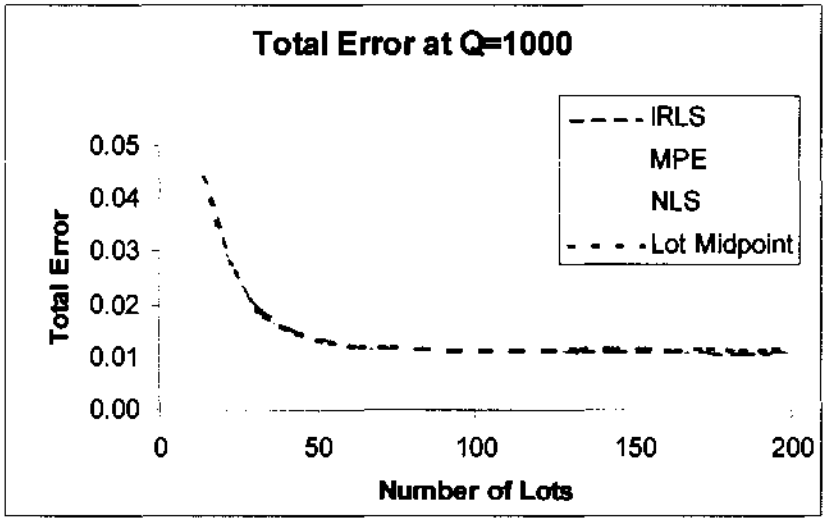
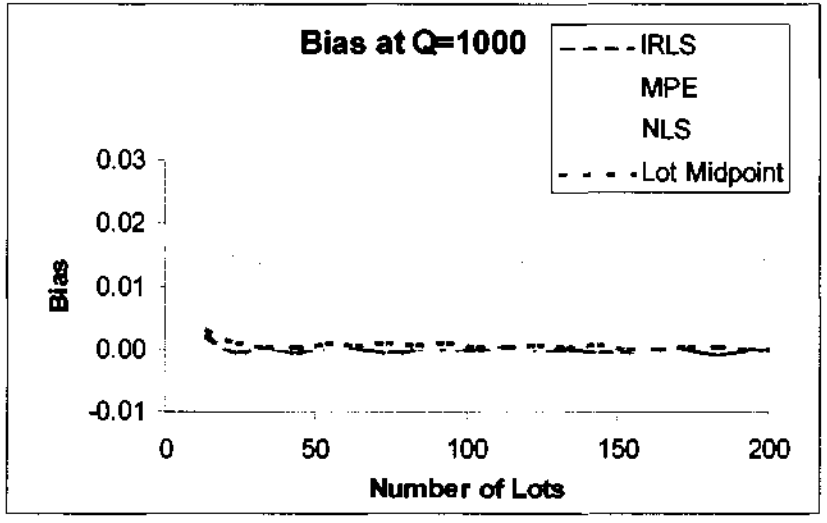


Figure 5.11. Simulation Experiment 2, Error in Predicting the Cost of Unit 100





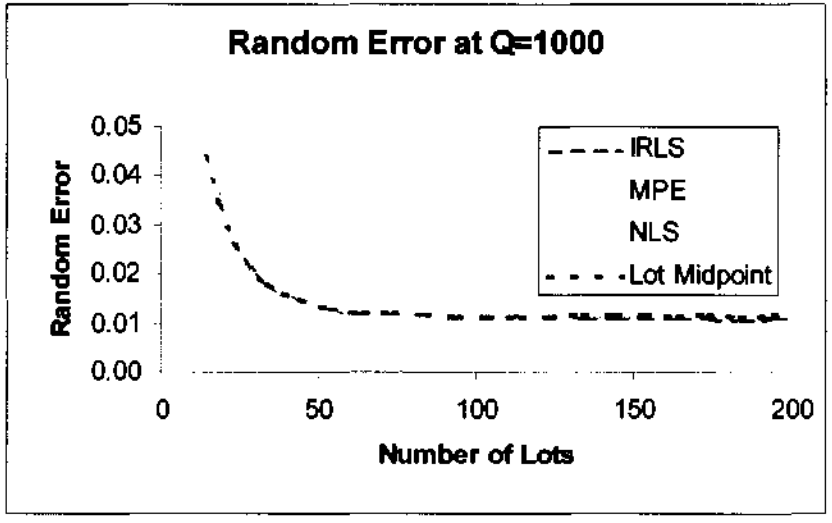


Figure 5.12. Simulation Experiment 2, Error in Predicting the Cost of Unit 1,000

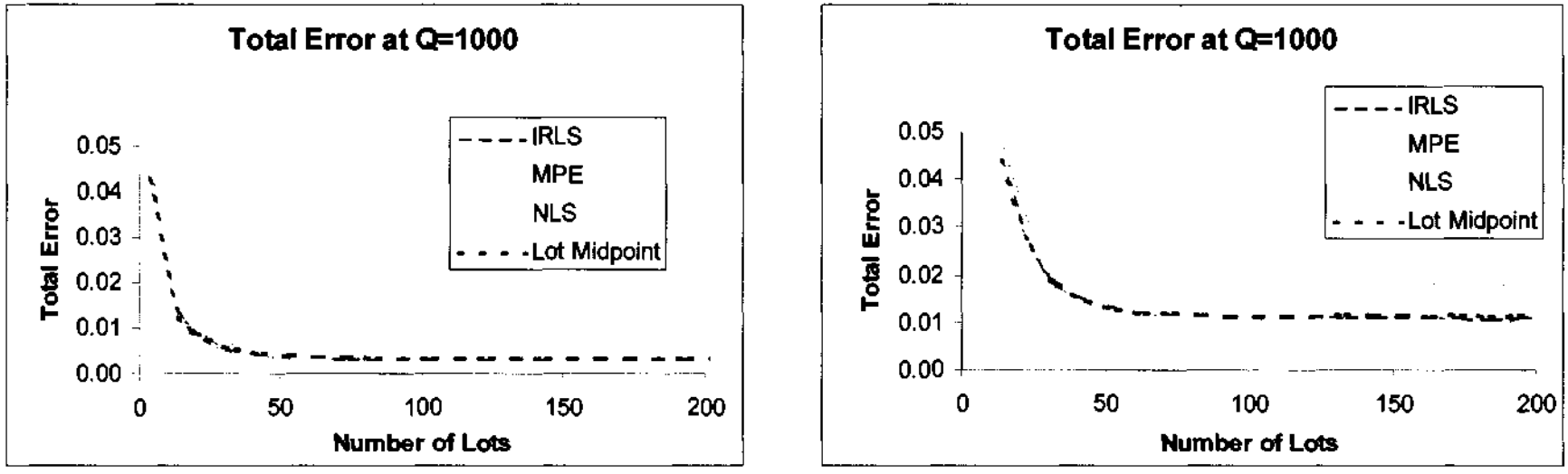


Figure 5.13. Effect of Slope on Estimated Cost of Unit 1,000
(Results from 80% slope are shown on left; results from 90% slope are shown on right)

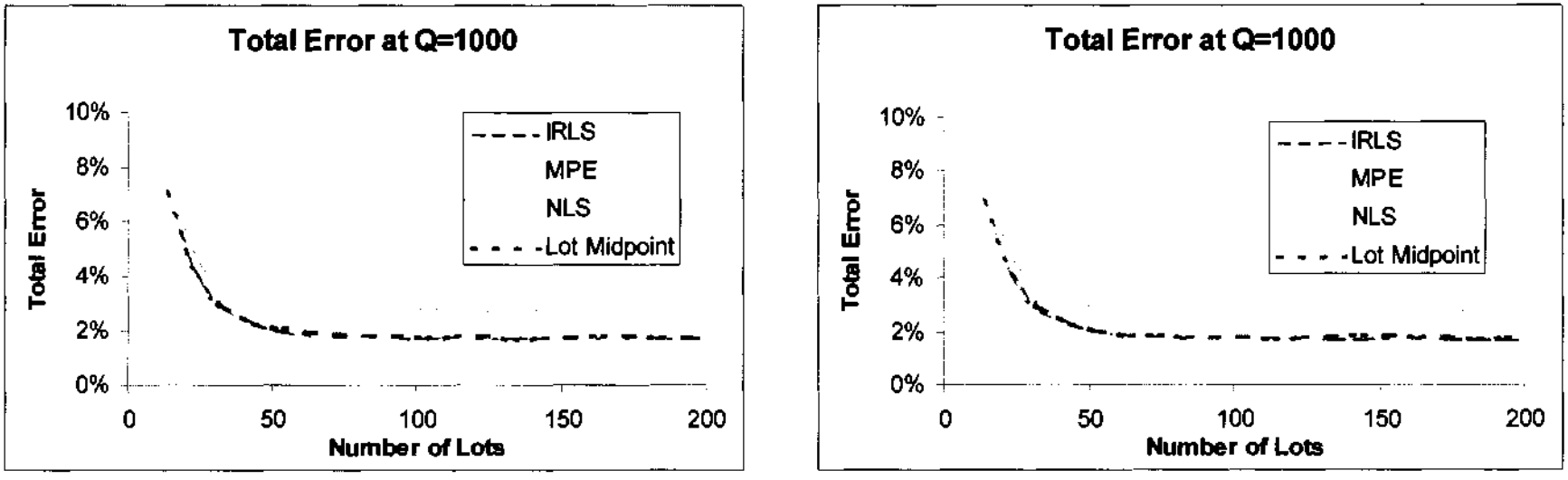
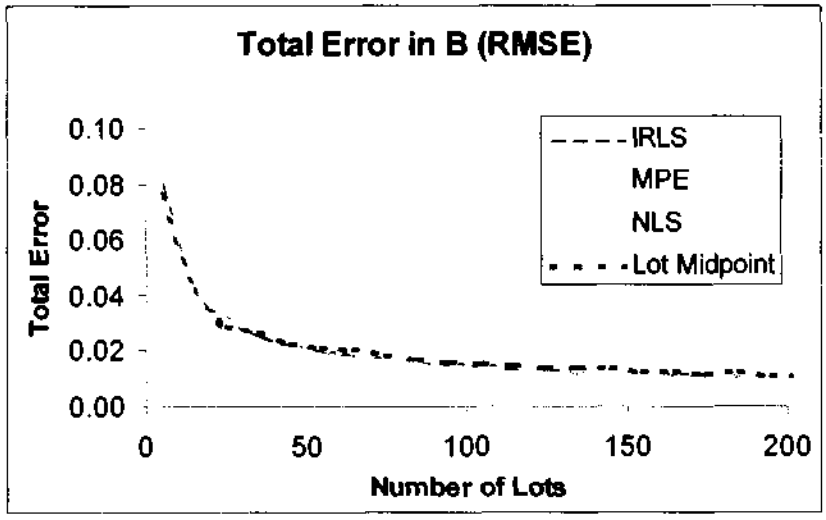


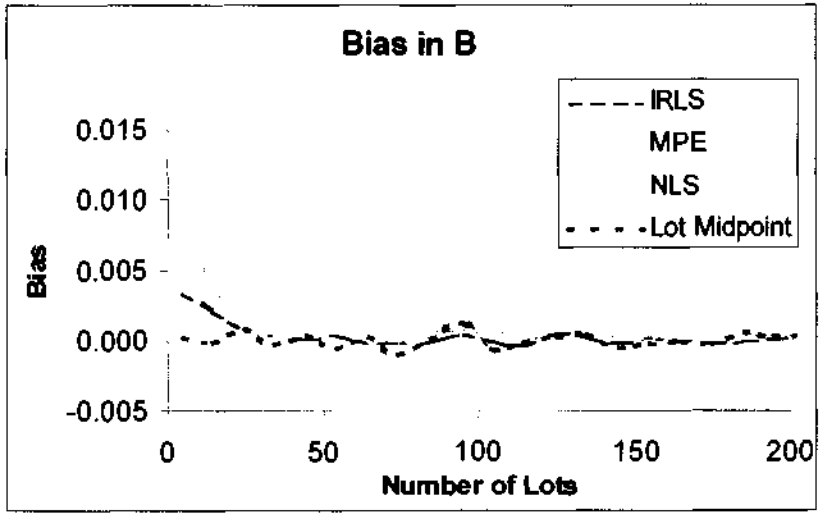
Figure 5.14. Effect of Slope on Percentage Error at Unit 1,000
(Results from 80% slope are shown on left; results from 90% slope are shown on right)

5.4 Simulation experiment 3: multiplicative, normal errors, learning slope = 80%, lot size = 10

This experiment is similar to Simulation Experiment 1, in that we restored the learning slope of 80%. However, we reduced the lot size from 50 units to only 10 units each. This experiment better corresponds to some aircraft manufacturing programs, whereas a lot size of 50 better corresponds to some missile programs.

Figures 5.15 through 5.19 show the relative errors for the different estimation methods. Because the individual lots are smaller, there is a greater concentration of data at lower cumulative quantities. Thus, the estimates of T_1 and the predictions of cost at lower quantities are more accurate, particularly when only a few lots have been observed (see Figures 5.20 and 5.21). However, because the data are concentrated at lower quantities, the predictions of cost at higher quantities are somewhat less accurate (see Figure 5.22).





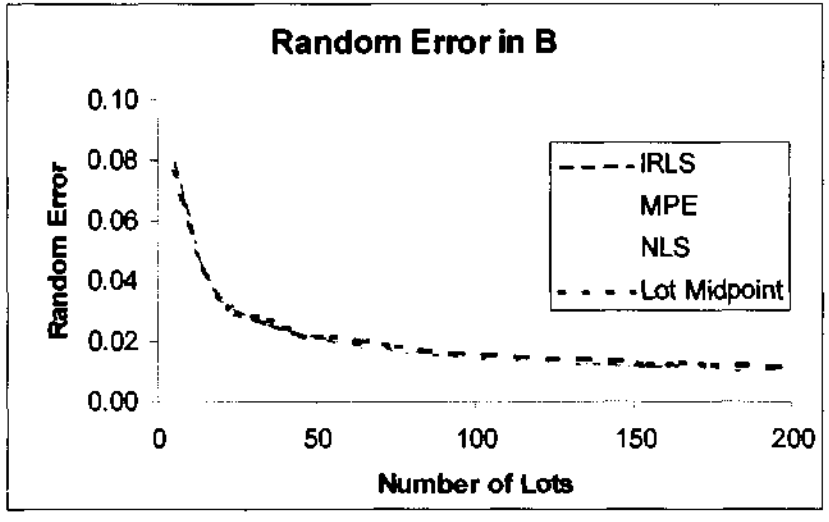


Figure 5.15. Simulation Experiment 3, Error in Slope Coefficient

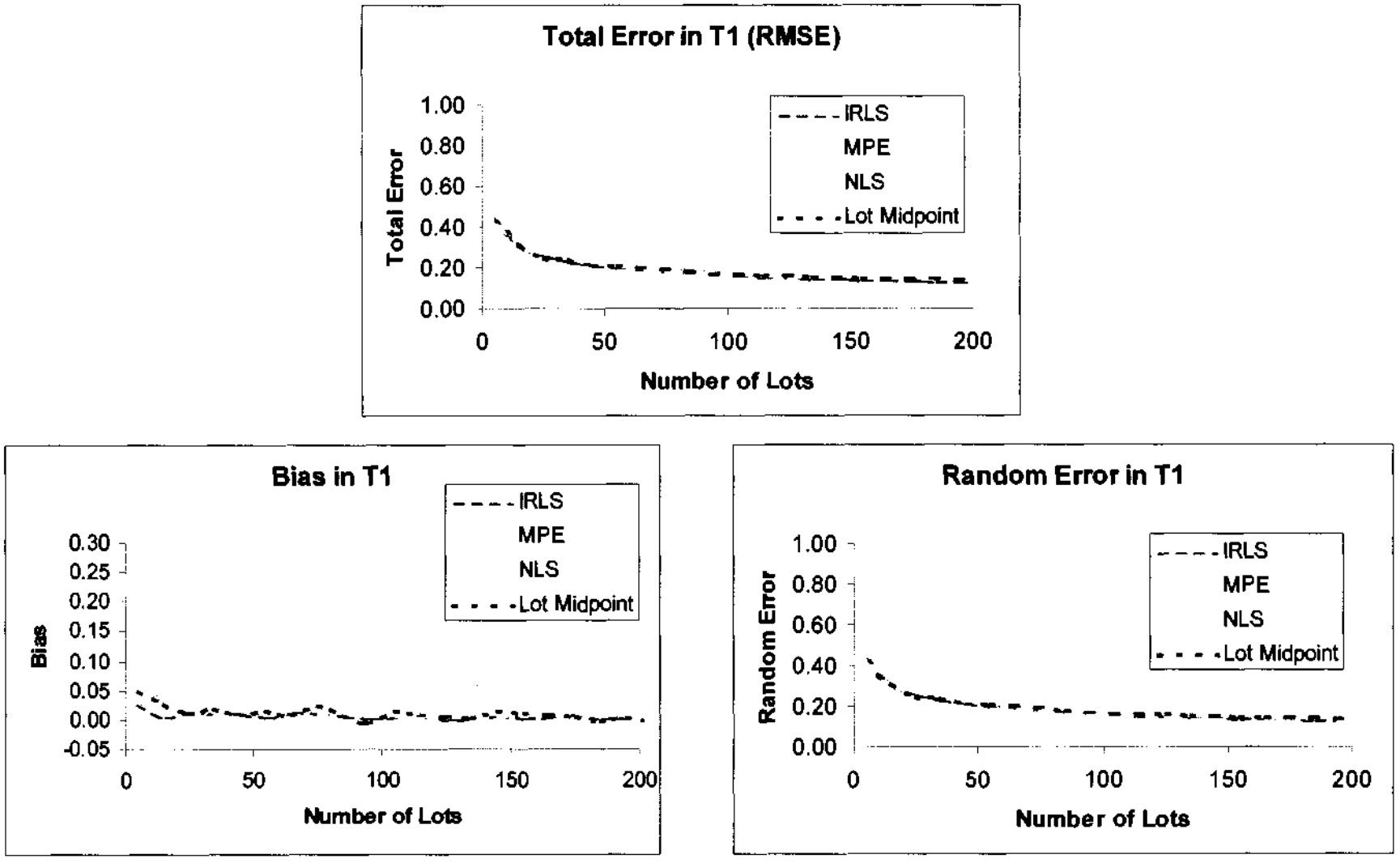
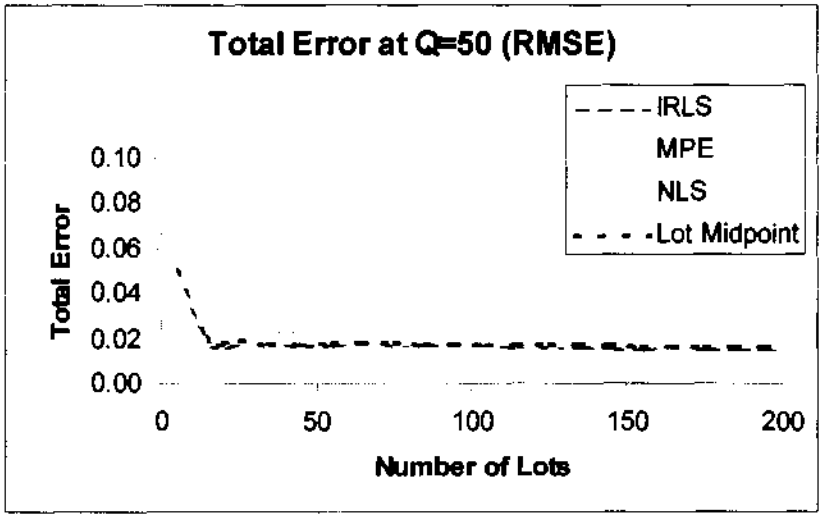
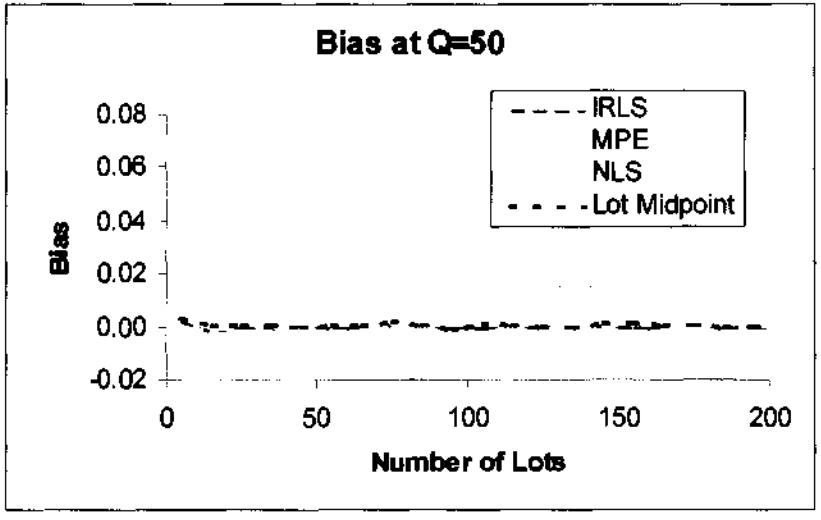


Figure 5.16. Simulation Experiment 3, Error in Intercept





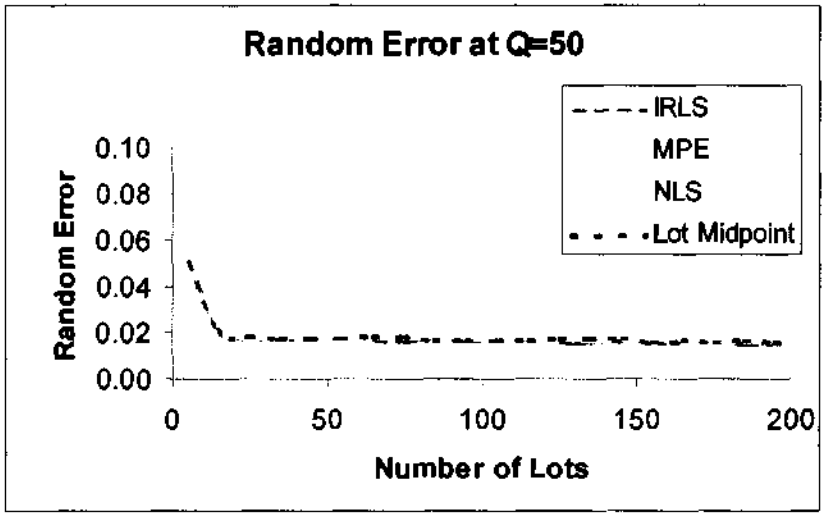
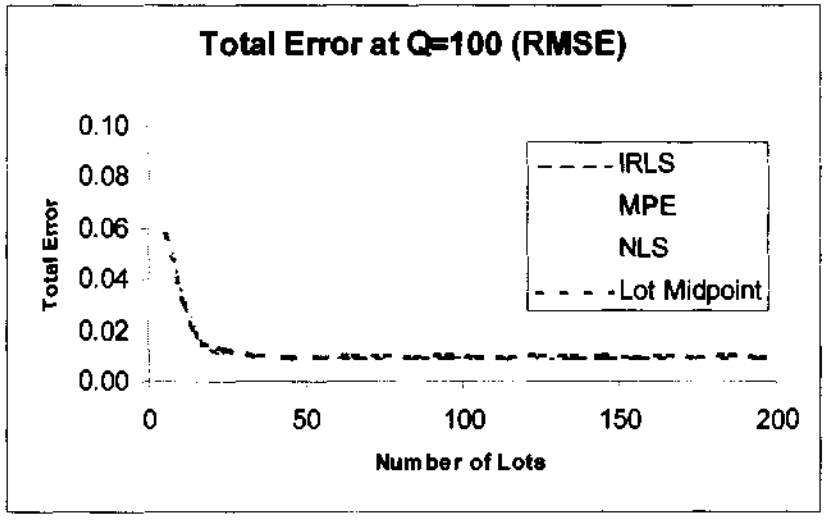
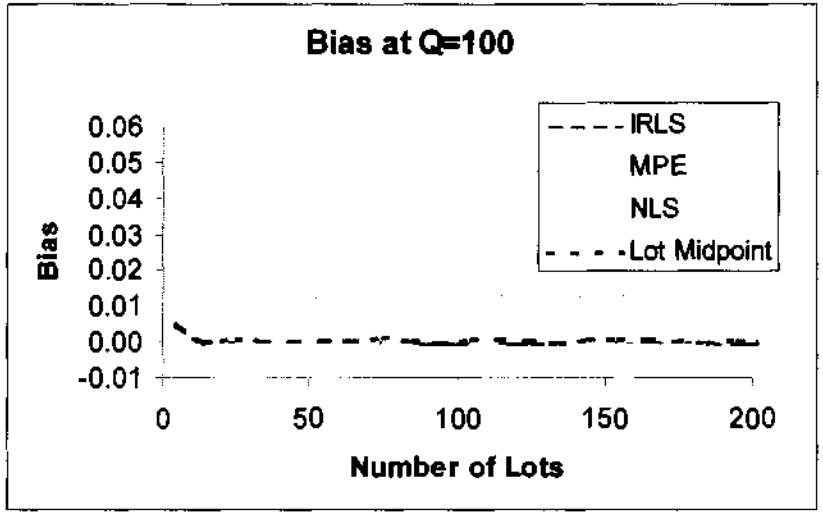


Figure 5.17. Simulation Experiment 3, Error in Predicting the Cost of Unit 50





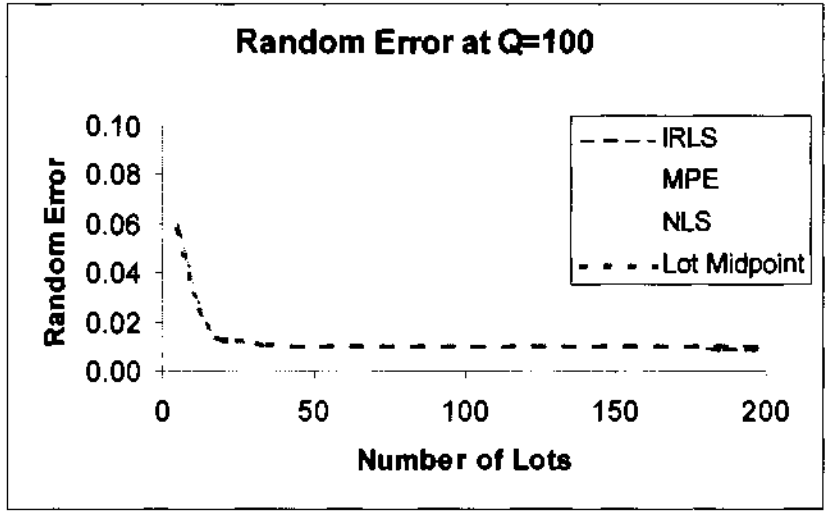
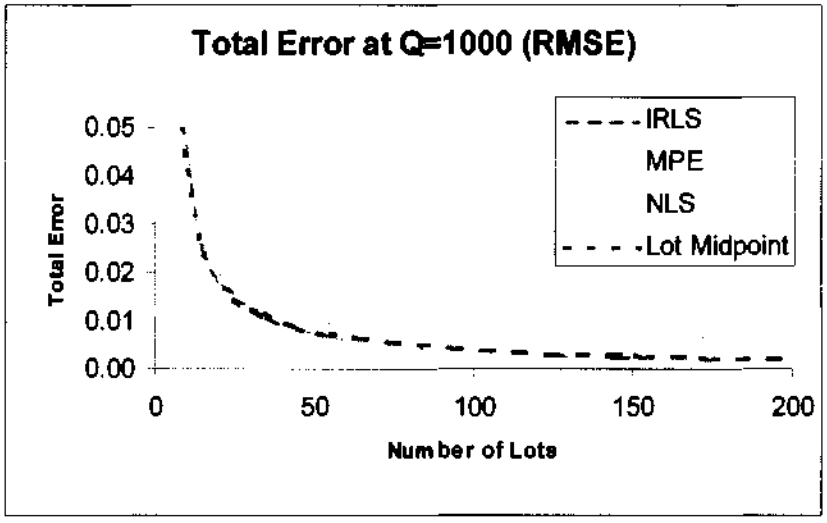
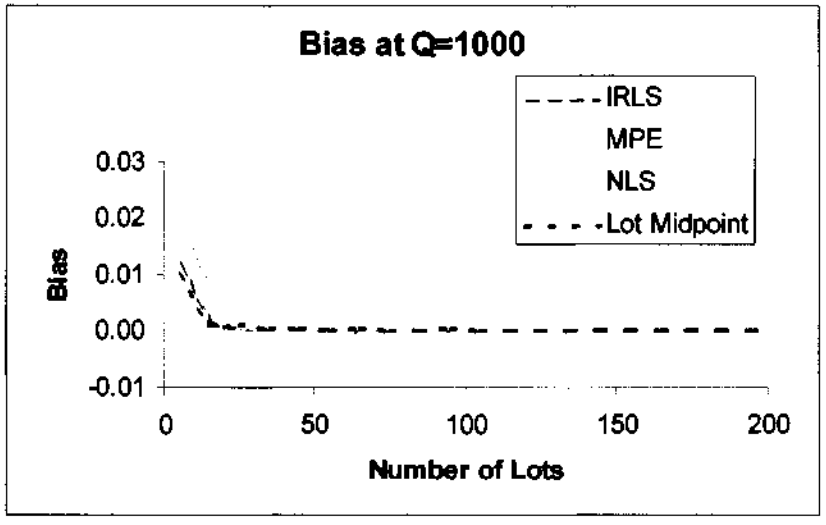


Figure 5.18. Simulation Experiment 3, Error in Predicting the Cost of Unit 100





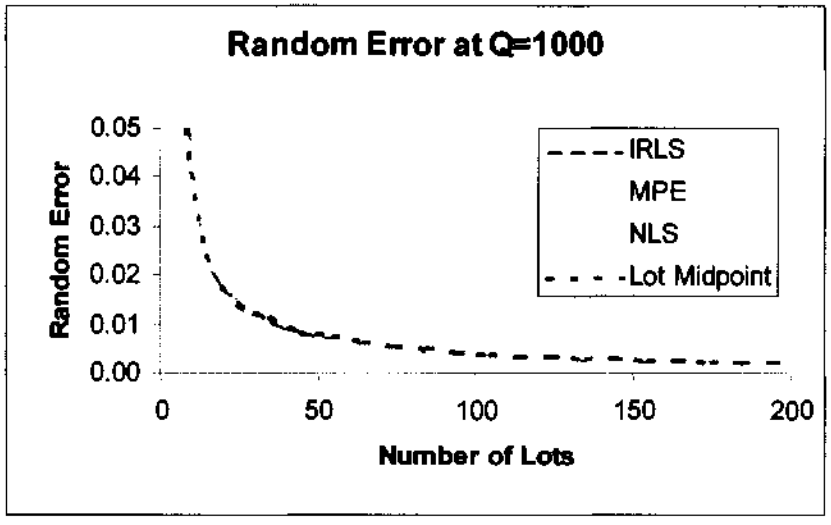


Figure 5.19. Simulation Experiment 3, Error in Predicting the Cost of Unit 1,000

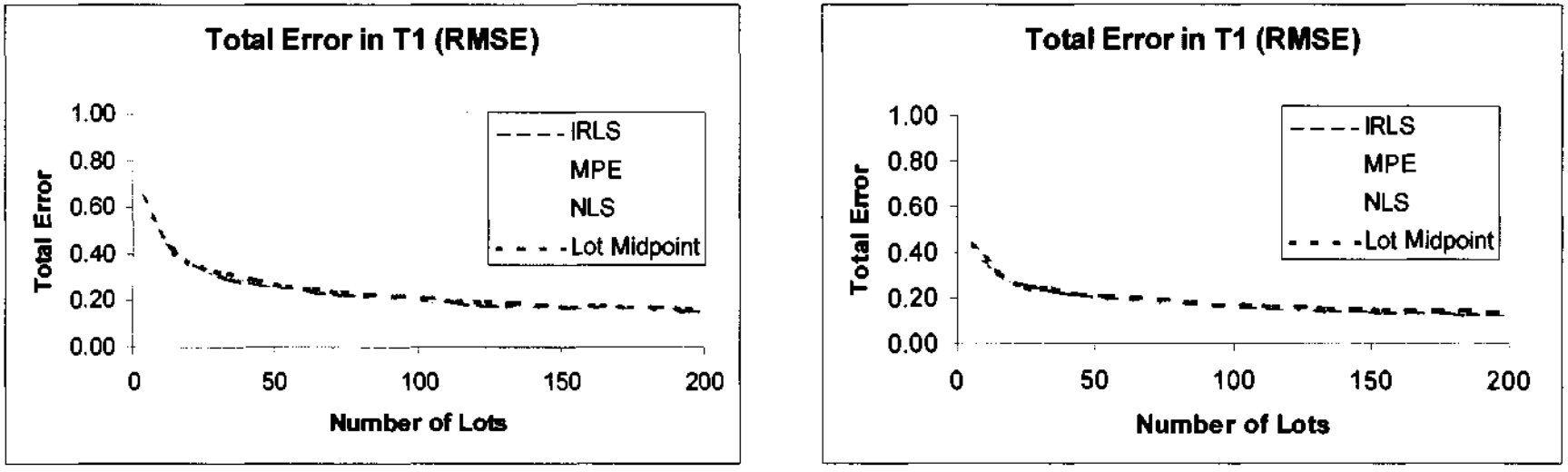


Figure 5.20. Effect of Lot Size on Prediction Errors for T_1
(Results from lots of size 50 are shown on left; results from lots of size 10 are shown on right)

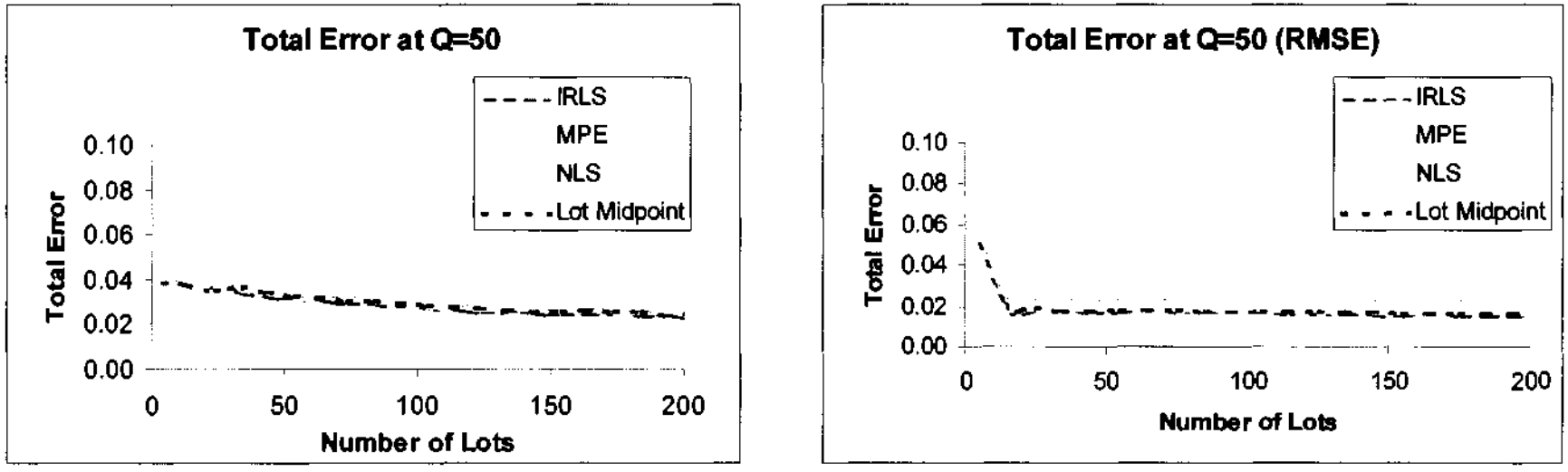
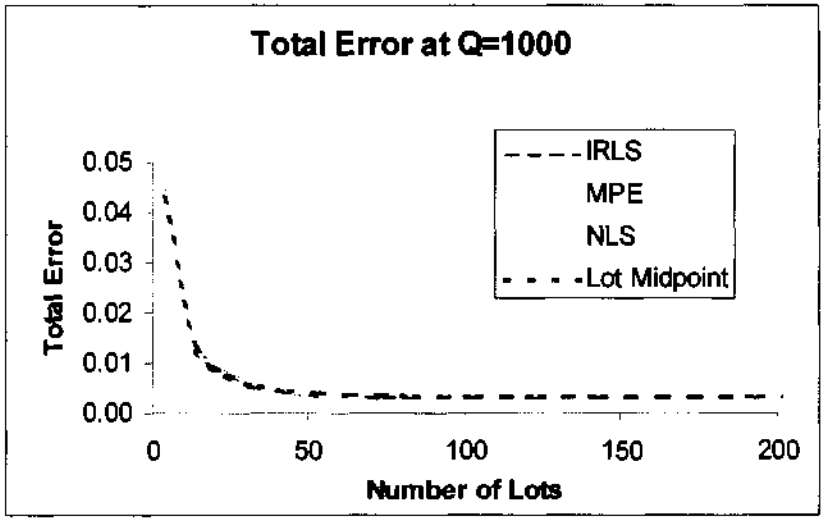


Figure 5.21. Effect of Lot Size on Estimated Cost of Unit 50
(Results from lots of size 50 are shown on left; results from lots of size 10 are shown on right)



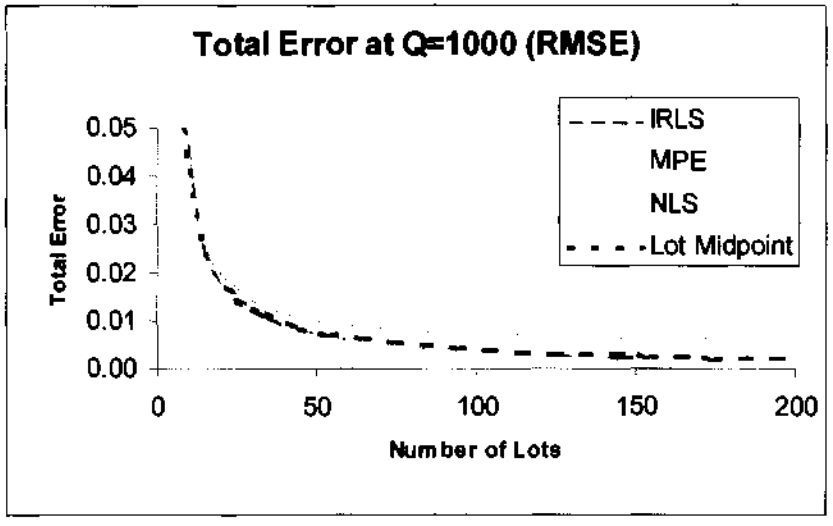


Figure 5.22. Effect of Lot Size on Estimated Cost of Unit 1,000
(Results from lots of size 50 are shown on left; results from lots of size 10 are shown on right)

5.5 Simulation experiment 4: multiplicative, normal errors, learning slope = 80%, sigma = 0.3

This experiment is identical to Simulation Experiment 1, except that we doubled the magnitude of the random error from a standard deviation of $\sigma = 0.15$ to $\sigma = 0.30$. With this one exception, we returned to the model parameters from Simulation Experiment 1, $b = -0.33$ and $T_1 = 1.8$. Hence, we generated the observed lot average cost as:

$$Obs\_LAC_i = LAC_i \times u_i, \tag{4.9}$$

where $u_i \sim N\left(1.0, 0.30^2\right)$.

Figure 5.23 compares the error distributions. The solid curve represents the baseline normal distribution with $\sigma = 0.15$, and the dashed curve represents a more dispersed normal distribution with $\sigma = 0.30$.

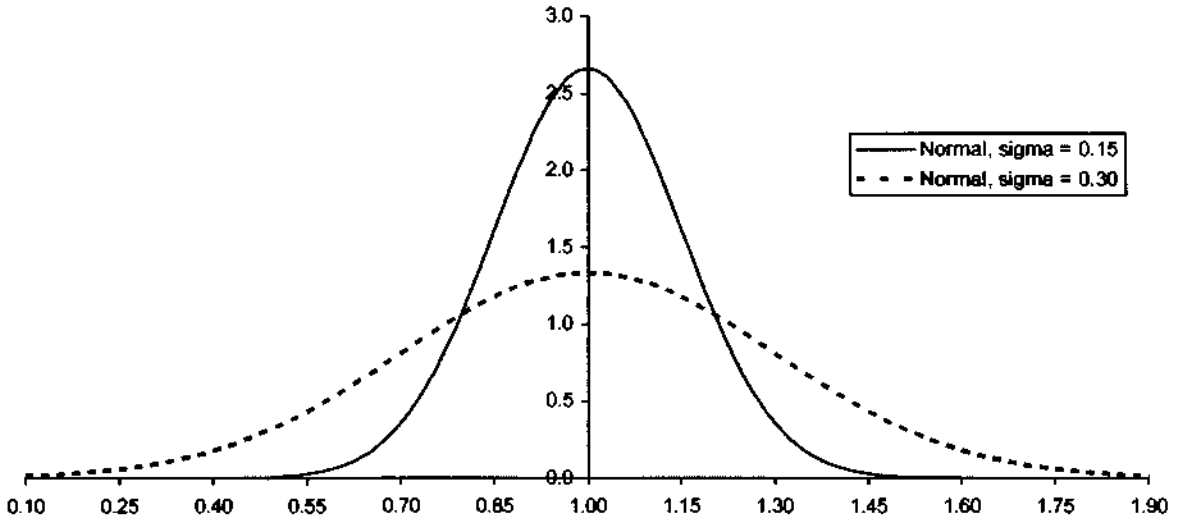


Figure 5.23. Comparison of Two Normal Distributions

Figures 5.24 through 5.28 show the relative errors for the different estimation methods. Perhaps the most striking finding is that MPE seems to perform better than the other methods when only a few lots have been observed. MPE appears, on the surface, to be less sensitive to the size of σ than are the other methods (see Figure 5.29). When the total error is decomposed, it becomes apparent that the random error component of the MPE estimate is less sensitive to σ than are the other methods (see Figure 5.30). However, while other methods remain

asymptotically unbiased, the bias in the MPE estimate is much worse under the larger value of σ.

That the bias in MPE is sensitive to the size of σ is not surprising. Recall that the parameters b and T_1 are solutions to the minimization problem:

$$\begin{pmatrix} b \\ T_1 \end{pmatrix} = \arg\min_{(b,T_1)} \sum_{i=1}^{n} \left(\frac{y_i - f[x_i,(b,T_1)]}{f[x_i,(b,T_1)]} \right)^2 . \tag{4.10}$$

There are two ways to minimize this function. The first is to make small prediction errors and thus make the numerator small. The second is simply to inflate the denominator by making $f[x_i,(b,T_1)]$ very large.

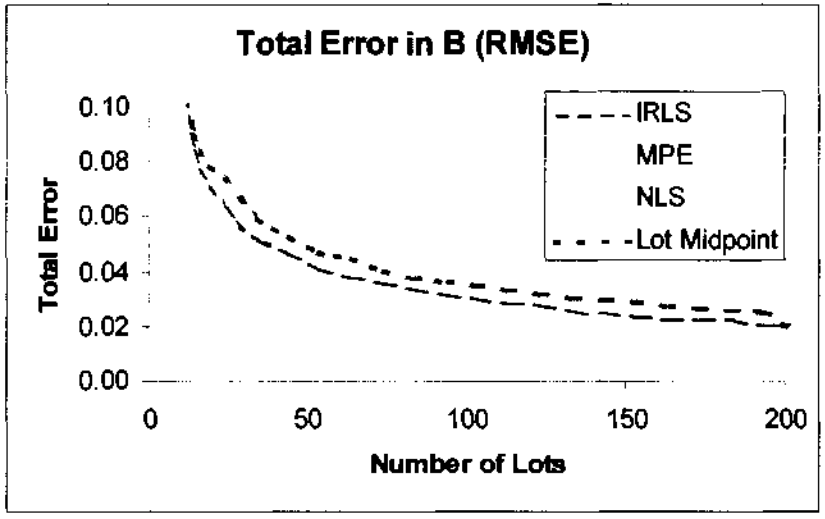
Recall that the magnitude of the random error, σ, is estimated by:

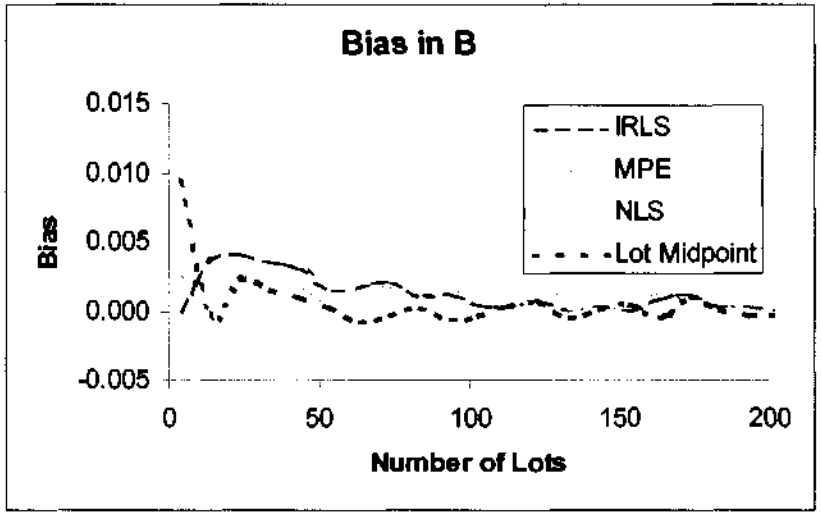
$$\hat{\sigma} = \left[\sum_{i=1}^{n} \left(y_i - f[x_i,(b,T_1)] \right)^2 \Big/ (n-2) \right]^{1/2} . \tag{4.11}$$

When σ increases, the numerator of MPE's minimization function (the right-hand side of equation (4.10)) necessarily grows larger. In that instance, the minimization algorithm is more likely to just increase the size of $f[x_i,(b,T_1)]$, amplifying the bias in the model predictions.

Between the two parameters b and T_1, the bias is more evident in T_1 because that parameter is strictly proportional to the model prediction, $f[Q_i,(b,T_1)] = T_1 \times Q_i^b$. A bias in estimating b would *tilt* the estimated learning curve, but would not uniformly inflate the model predictions that appear in the denominator of equation (4.10).

Finally, there is an ironic corollary in using MPE to predict unit cost. At unit 100, for example, we see in Figure 5.27 that the random error in MPE is small regardless of the number of lots used in estimation. Now examine Figure 5.31. Because the random error is small, the bias component dominates and the total error in MPE remains constant even for very large numbers of lots. Thus, our initial observation that MPE has small random error, even for large values of σ, is more than offset by a severe bias that does not diminish asymptotically.





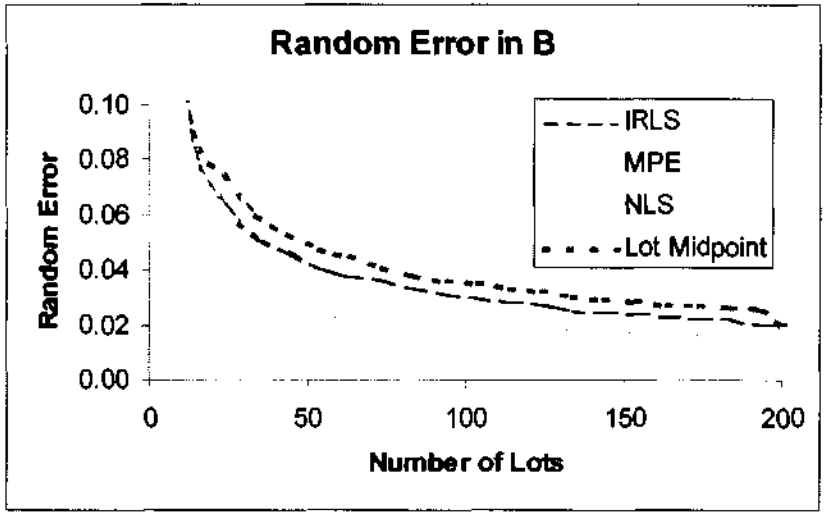
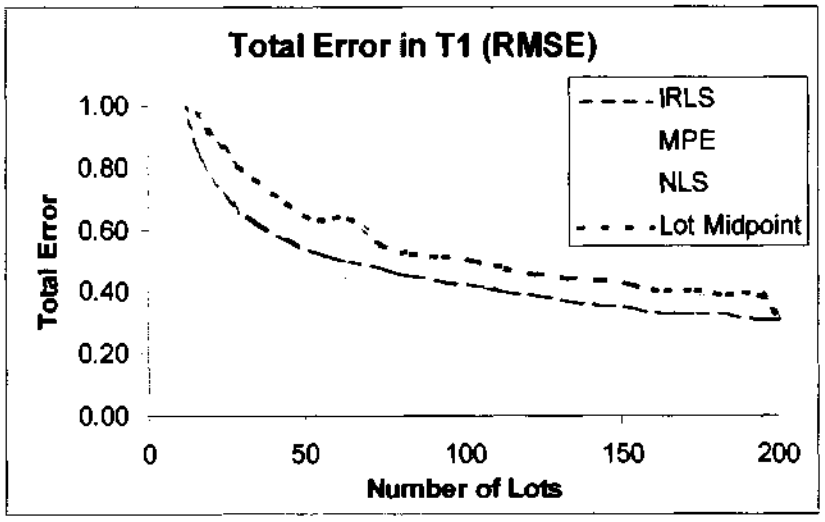
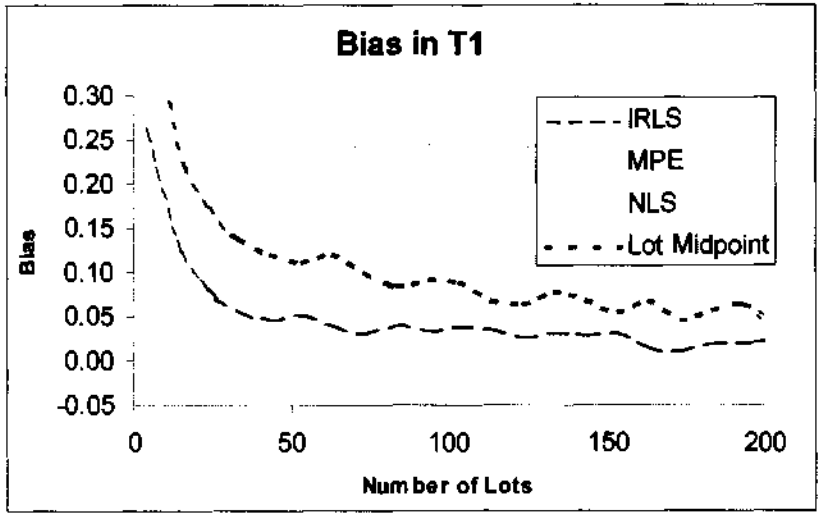


Figure 5.24. Simulation Experiment 4, Error in Slope Coefficient





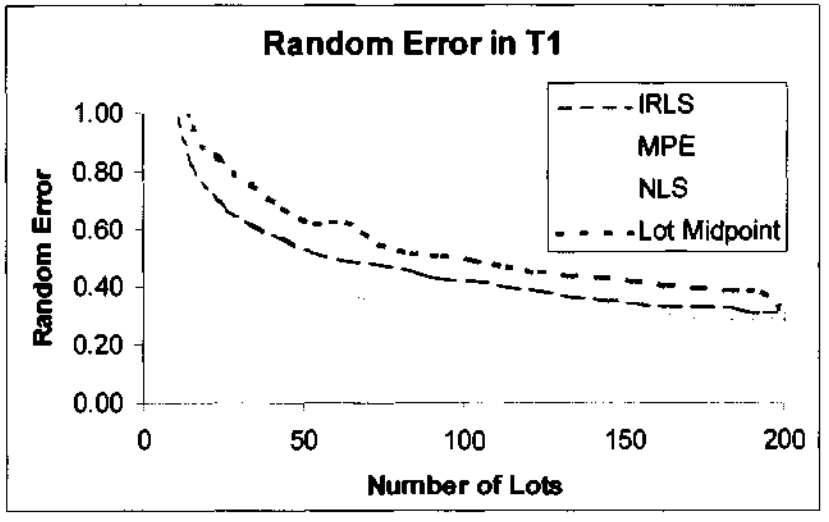
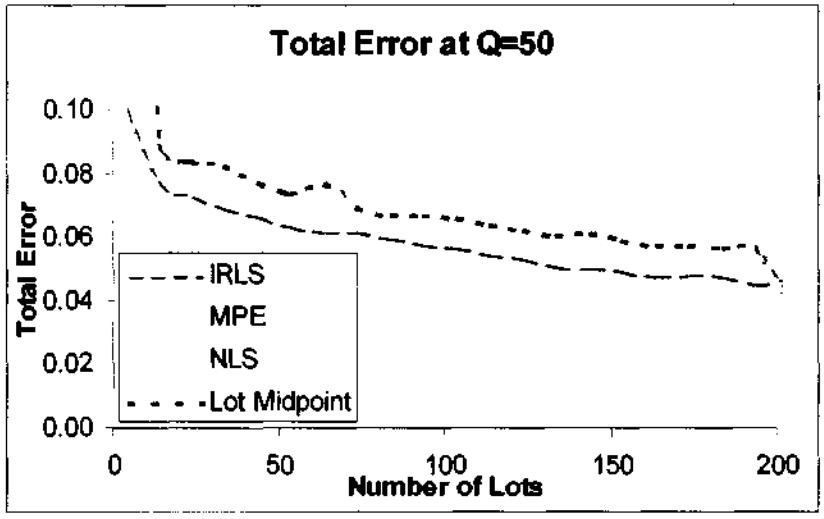
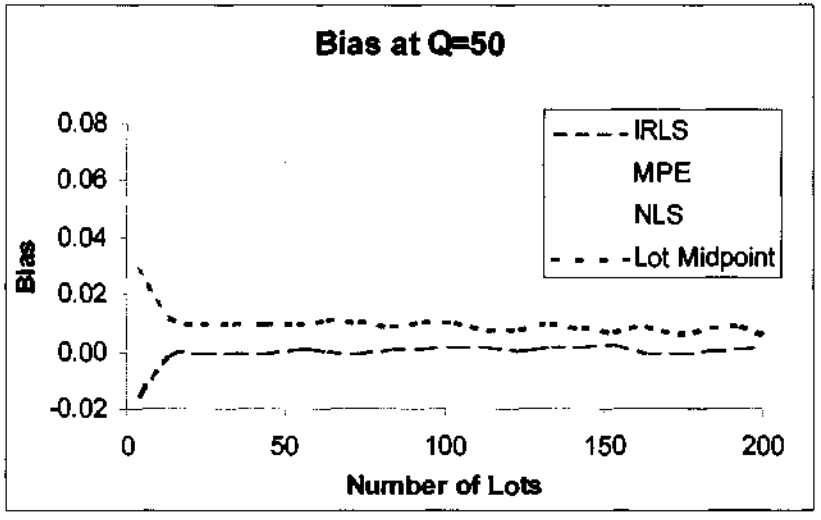


Figure 5.25. Simulation Experiment 4, Error in Intercept





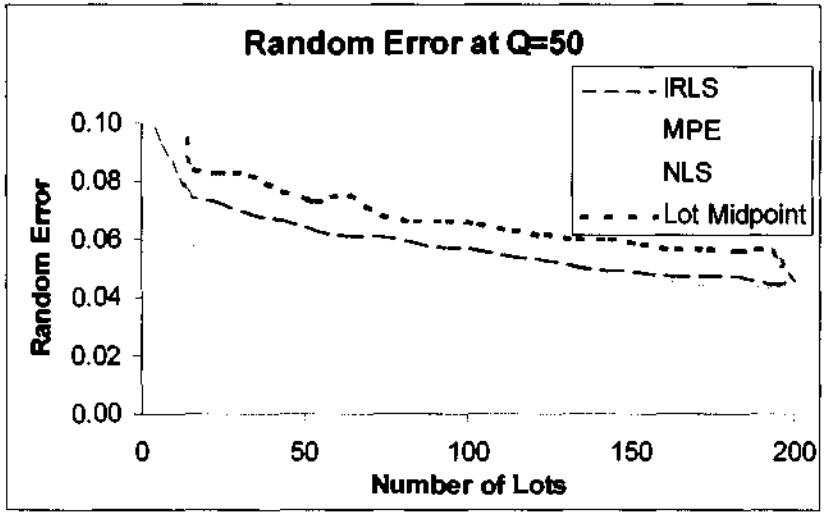
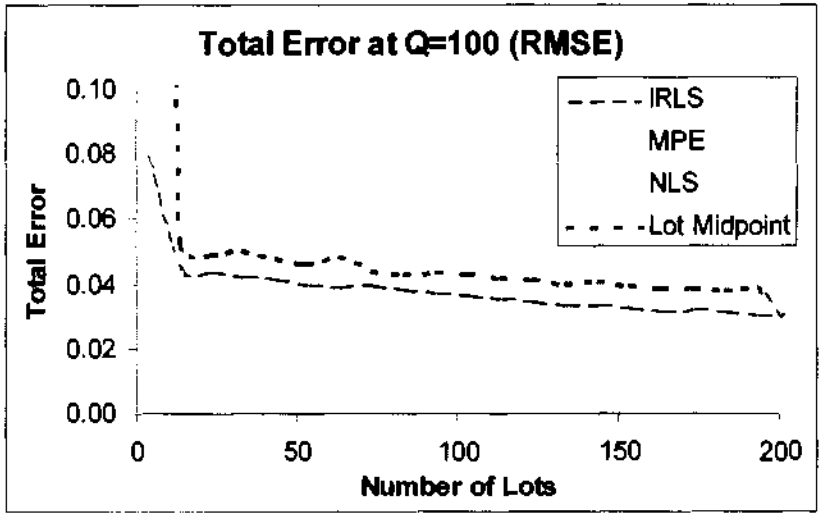
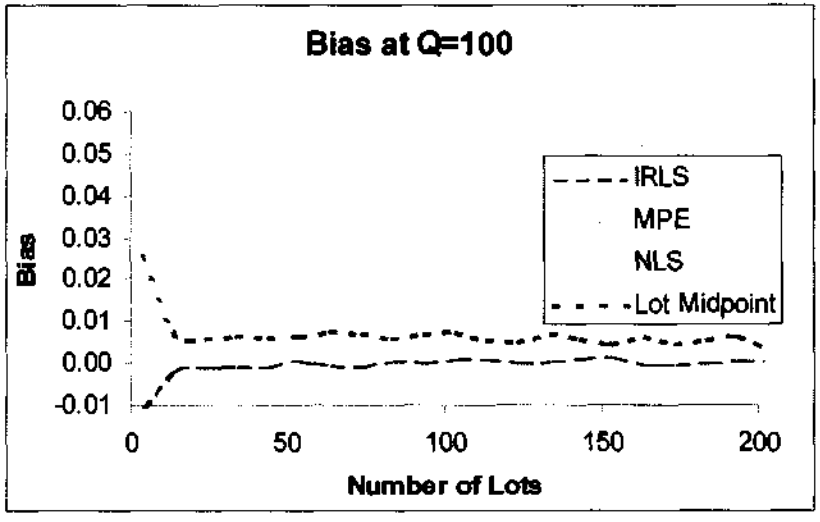


Figure 5.26. Simulation Experiment 4, Error in Predicting the Cost of Unit 50





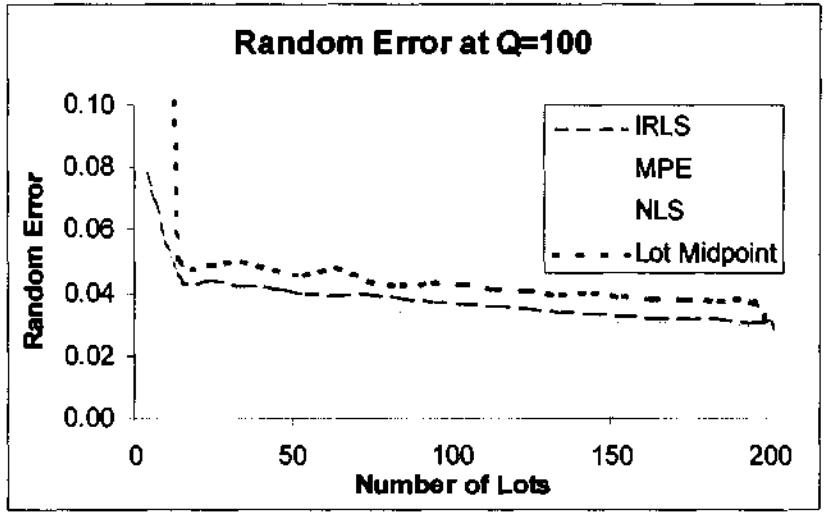
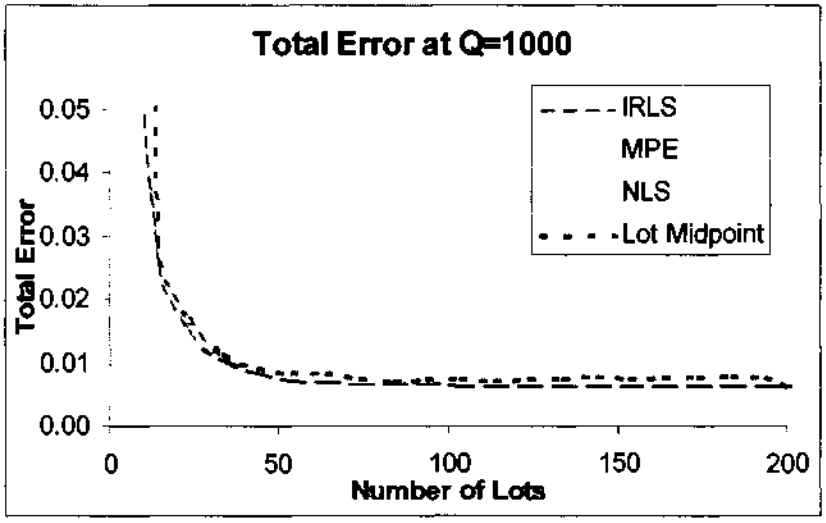
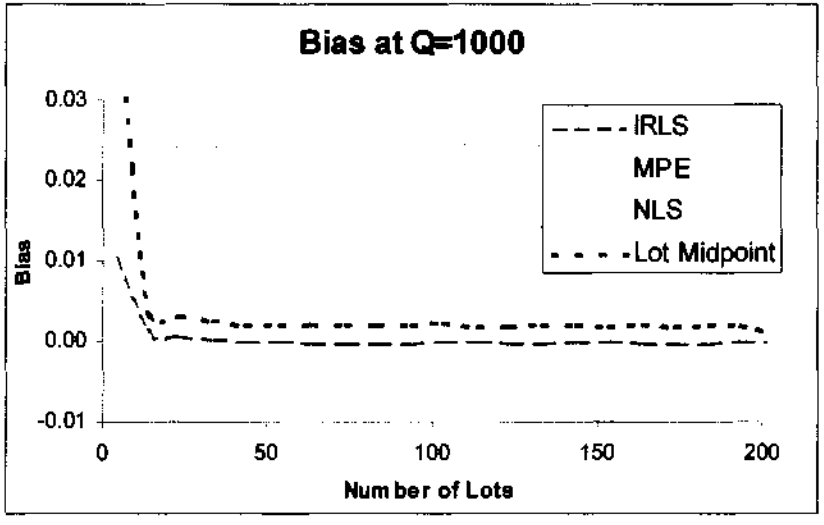


Figure 5.27. Simulation Experiment 4, Error in Predicting the Cost of Unit 100





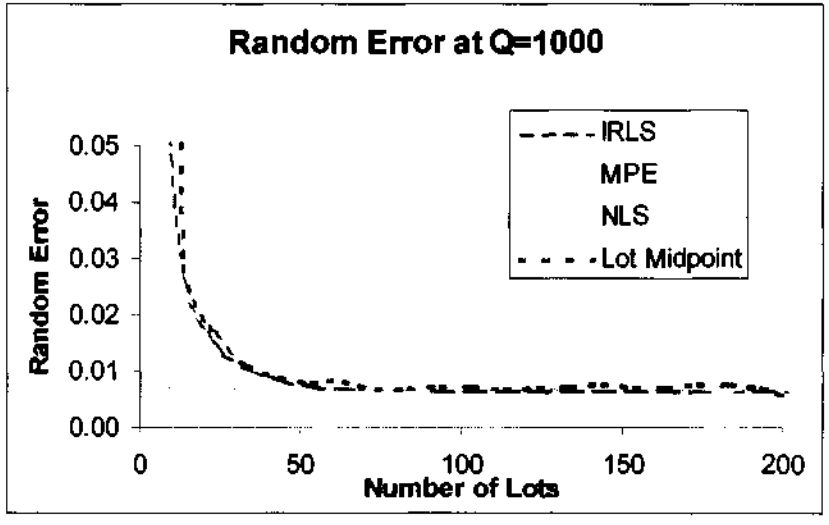


Figure 5.28. Simulation Experiment 4, Error in Predicting the Cost of Unit 1,000

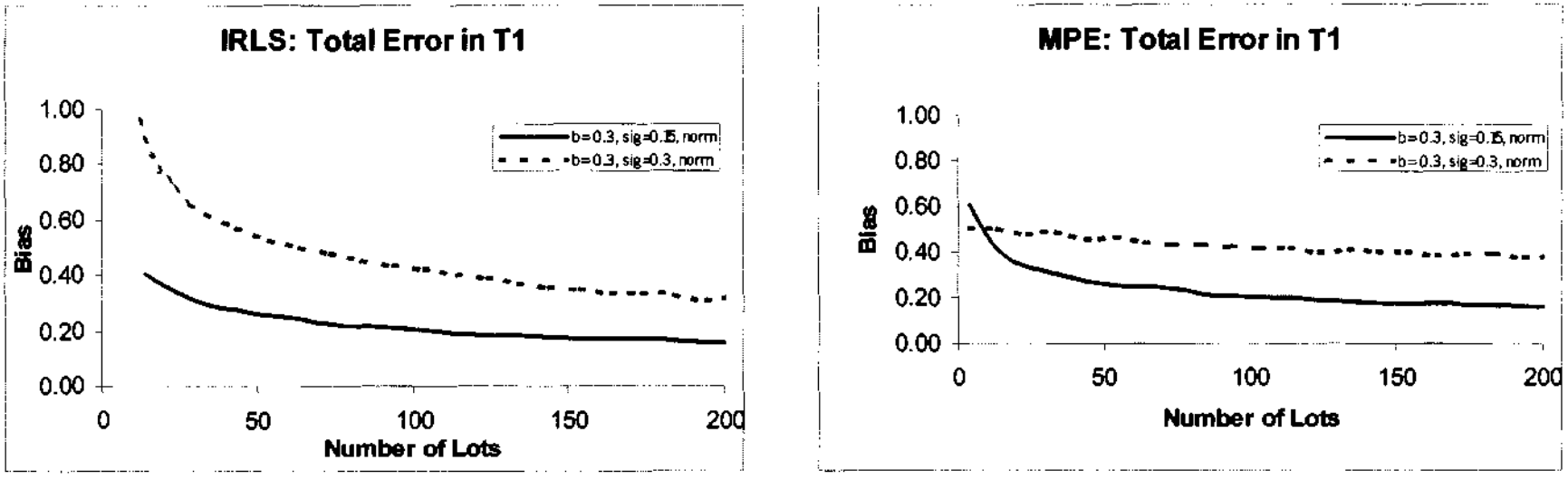
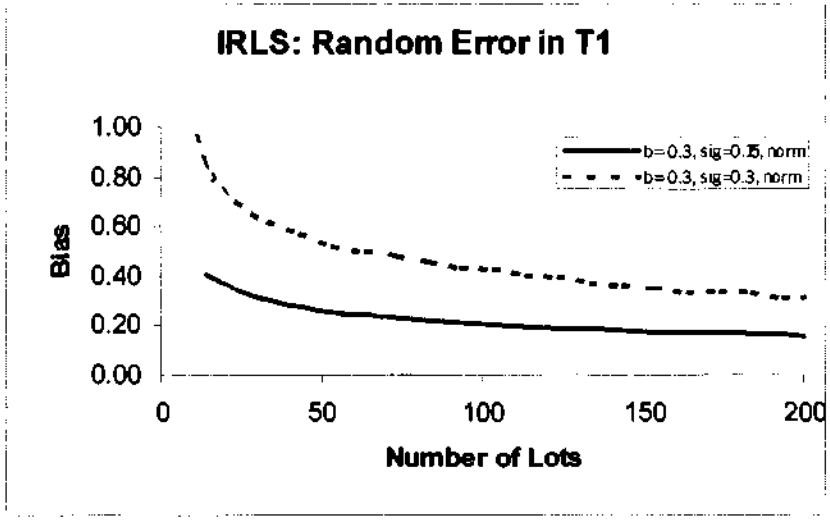
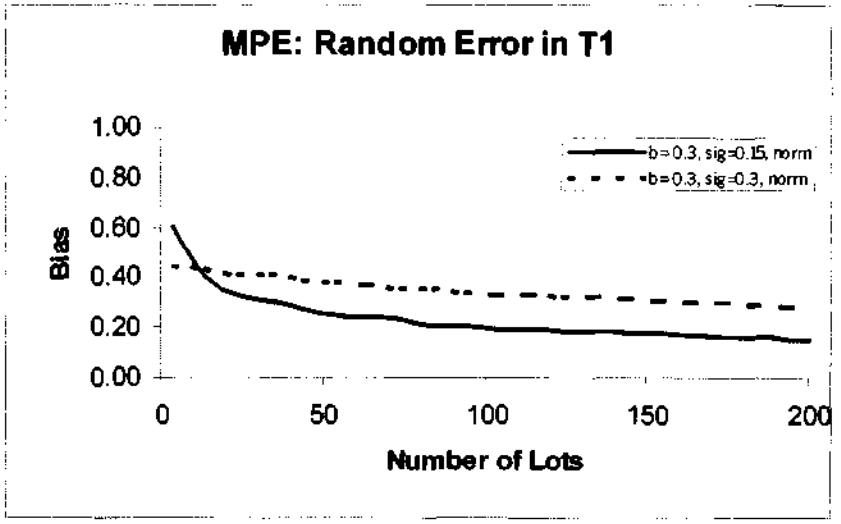


Figure 5.29. MPE Sensitivity to Standard Deviation in T_1 Total Error





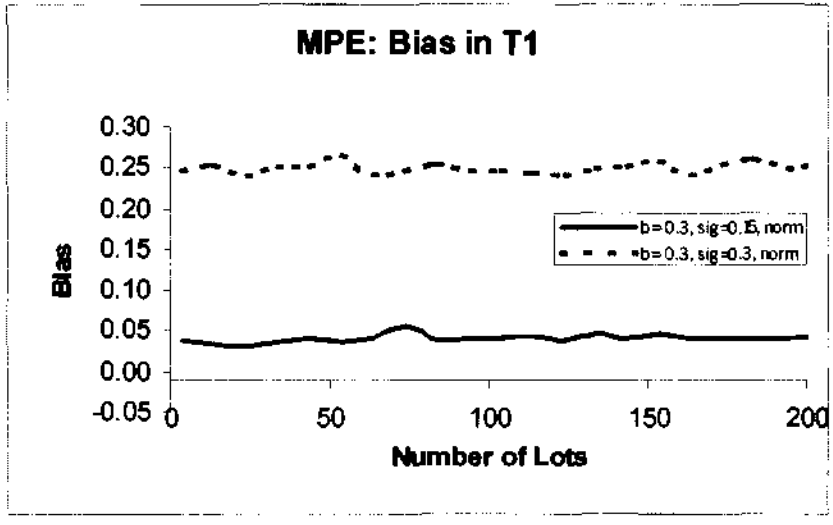


Figure 5.30. MPE Sensitivity to Standard Deviation in T_1 Error Components

5.6 Simulation experiment 5: multiplicative, uniform-distributed errors, learning slope = 80%

This experiment is identical to Simulation Experiment 1, except that we replaced the multiplicative normal error term with a uniform error term having the same mean and standard deviation. The uniform distribution does not correspond, even approximately, to the distributions assumed in the derivation of *any* of the four estimation methods under comparison. Because the uniform distribution does not have tails like the normal distribution, the uniform distribution generates many more extreme observations or outliers. The current experiment contrasts with the previous Experiment 4, which retained the shape of the error distribution but merely increased the magnitude of the error term.

To generate the uniform error term, first consider the canonical uniform distribution on the interval (0,1), denoted $U(0,1)$. This distribution has mean 0.5 and standard deviation $1/\sqrt{12}$. We can linearly transform this uniform random variable so that it becomes centered at a mean of 1.0 with a standard deviation of 0.15. The required transformation is as follows:

$$\begin{aligned} &Obs\_LAC_i = LAC_i \times u_i, \\ &\text{where } u_i \sim \left[0.30\times\sqrt{3}\times U(0,1)\right] + \left[1-0.15\times\sqrt{3}\right]. \end{aligned} \tag{4.12}$$

This transformed uniform error distribution has positive probability on an interval centered at its mean of 1.0: $1.0 \pm 0.15\times\sqrt{3} = 1.0 \pm 0.2598 = (0.7402, 1.2598)$. Figure 5.32 compares the two error distributions.

Figures 5.33 through 5.37 show the relative errors for the different estimation methods. Comparing these figures to Figures 5.1 through 5.5 from Simulation Experiment 1, we see that the results from uniform errors are virtually identical to those from normally distributed errors. All four estimation methods are robust in the face of data containing uniform errors.

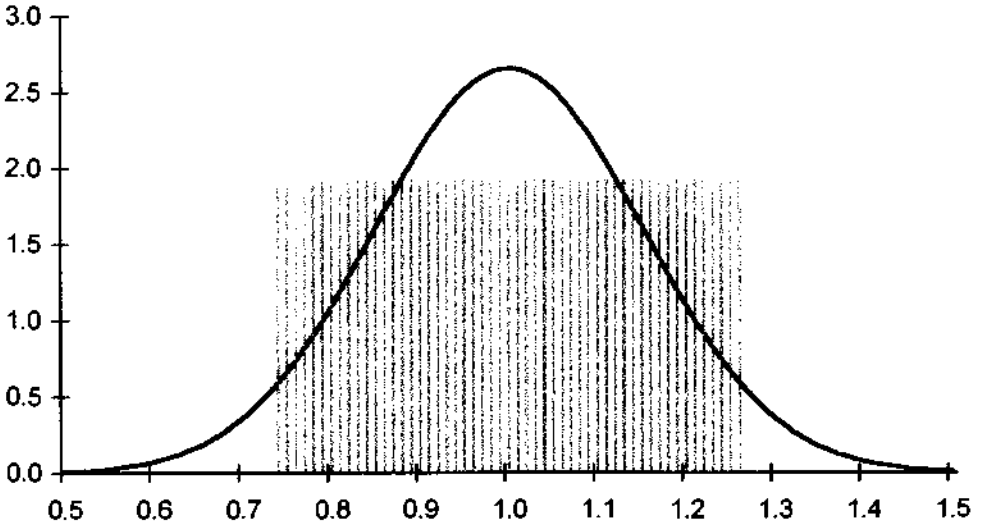
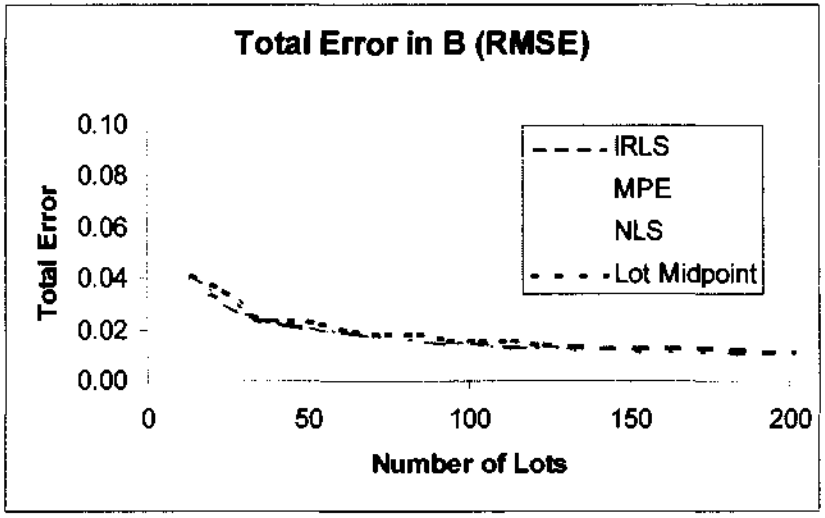
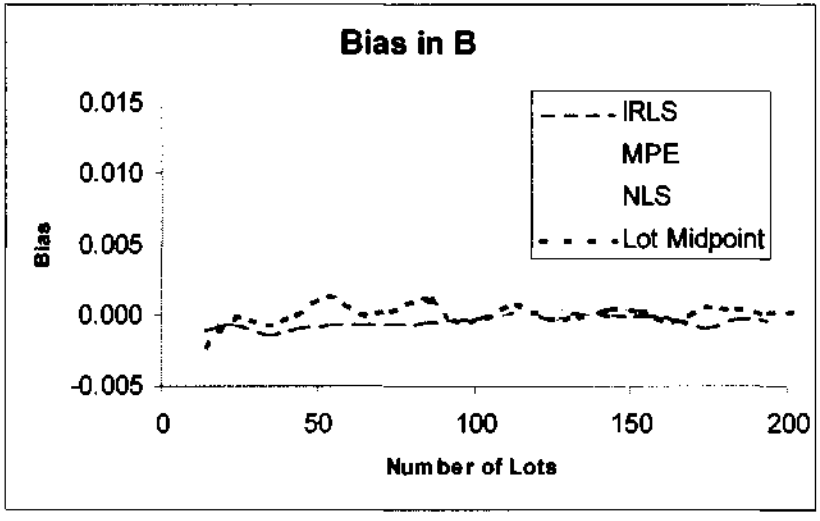


Figure 5.32. Comparison of Normal and Uniform Distributions





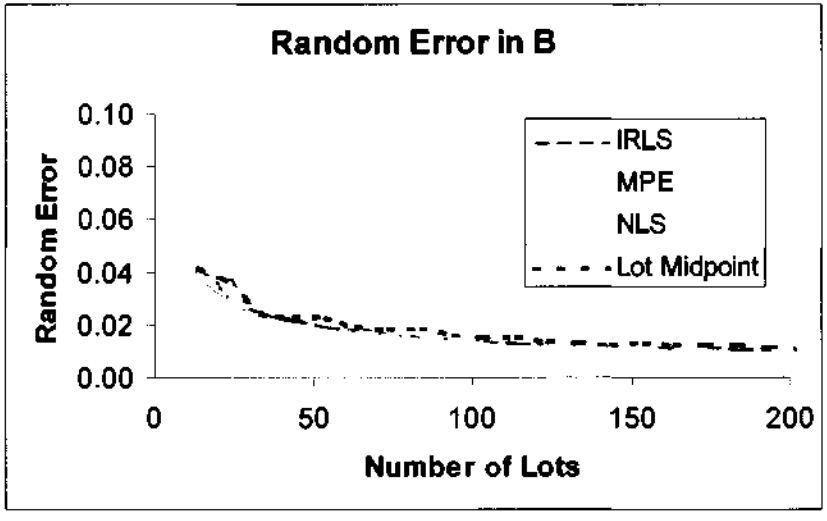
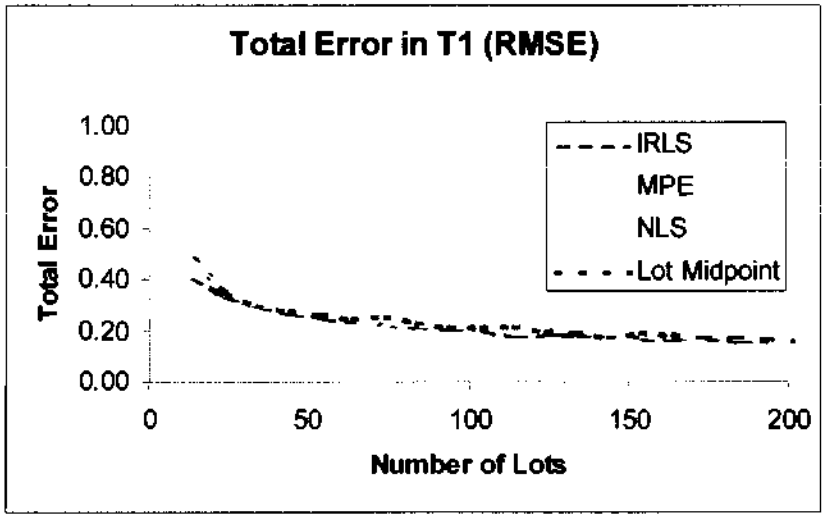
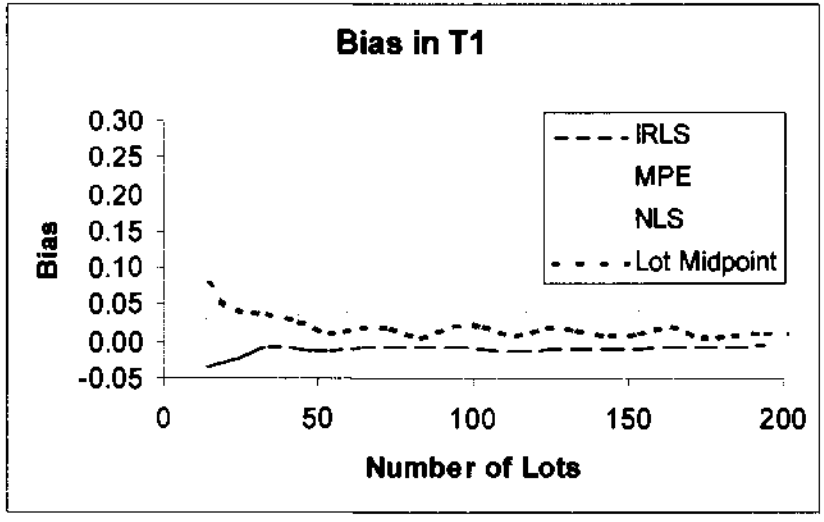


Figure 5.33. Simulation Experiment 5, Error in Slope Coefficient





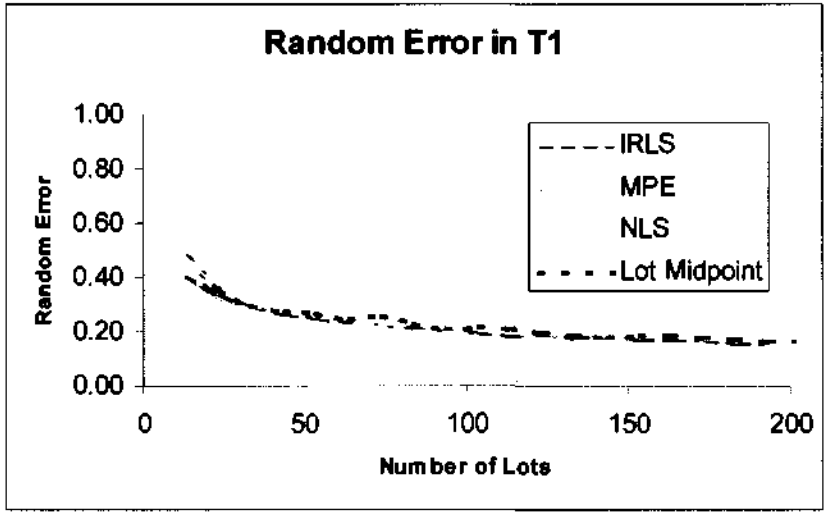
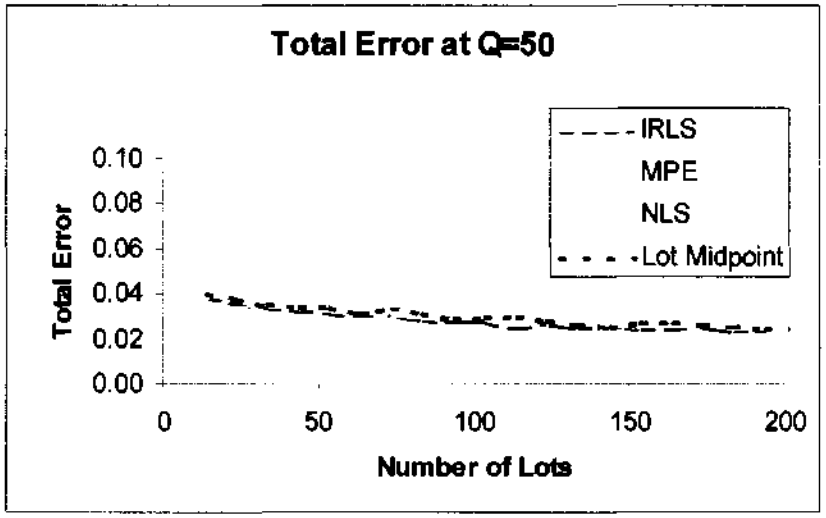
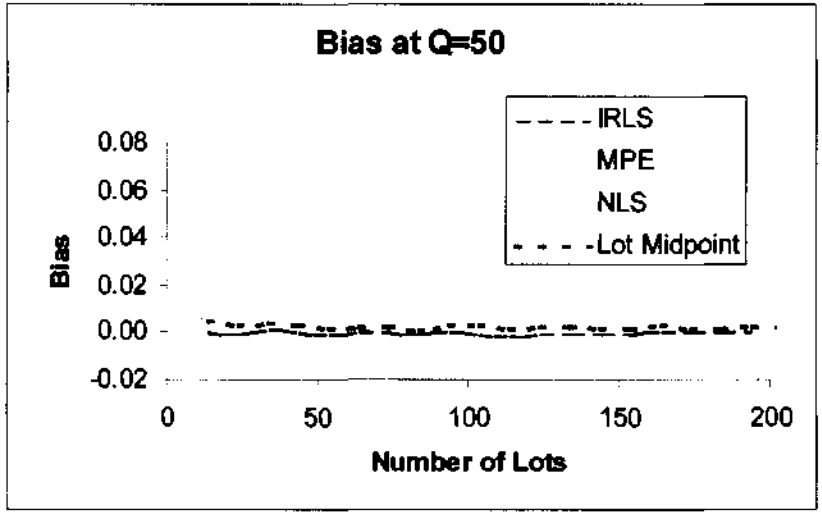


Figure 5.34. Simulation Experiment 5, Error in Intercept





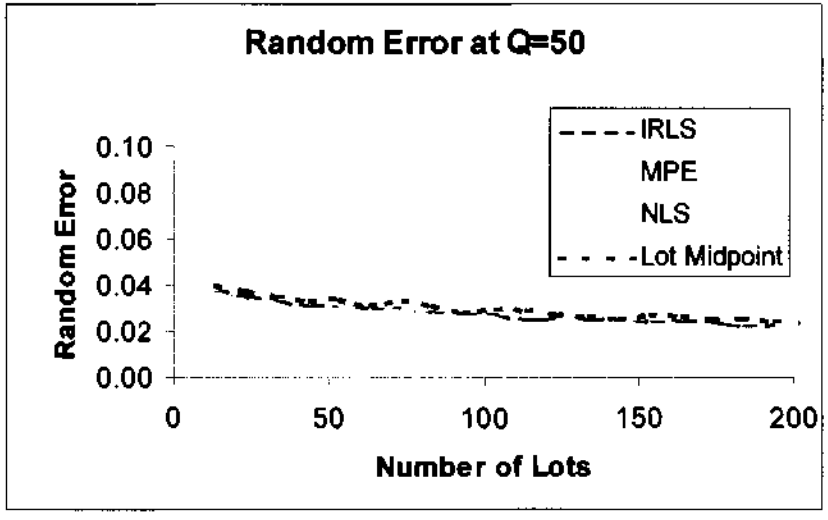
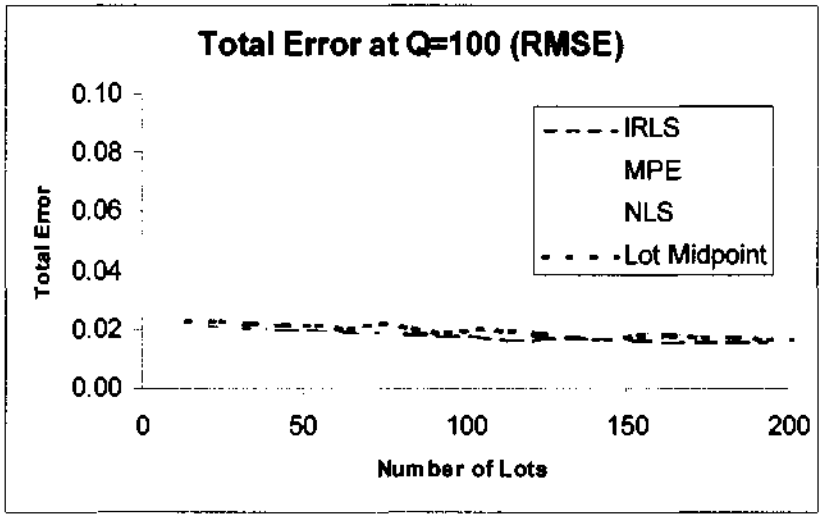
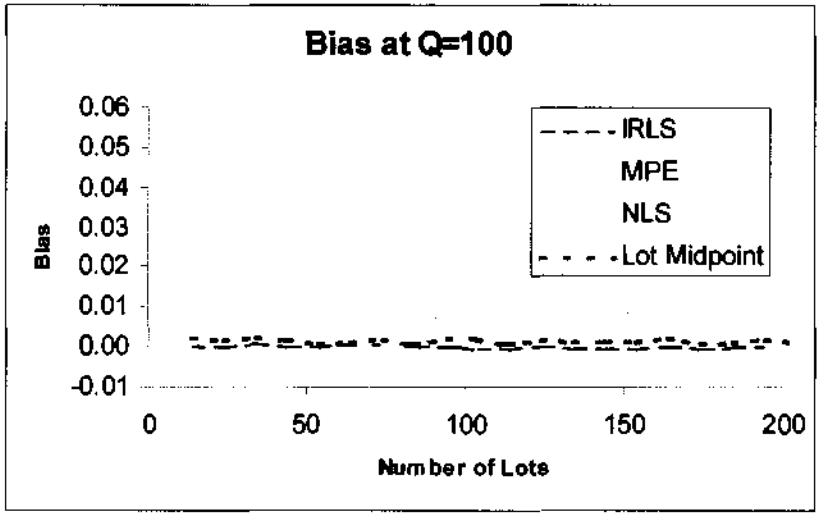


Figure 5.35. Simulation Experiment 5, Error in Predicting the Cost of Unit 50





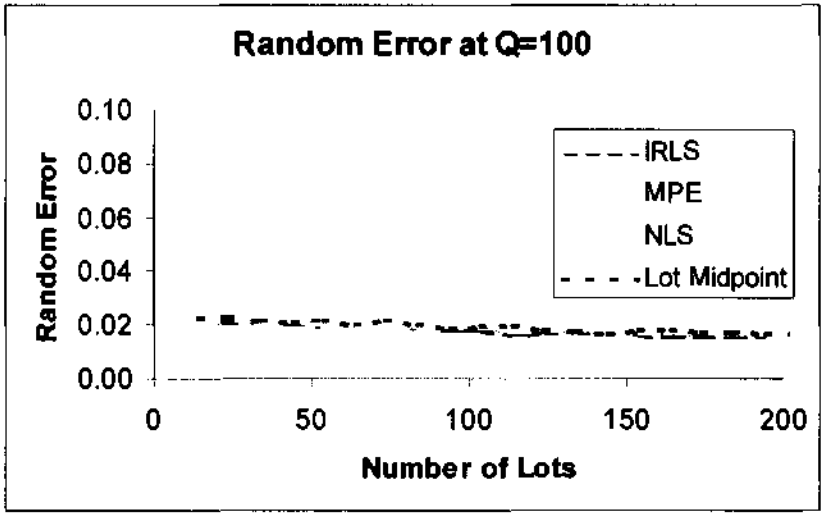
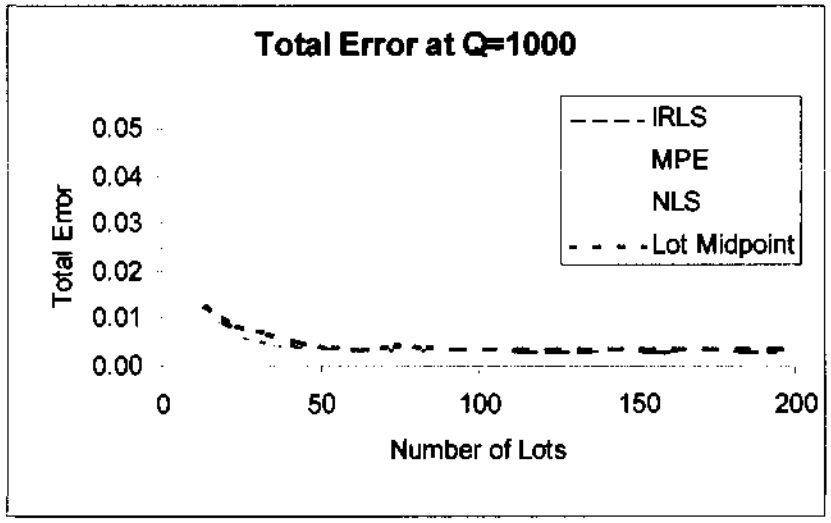
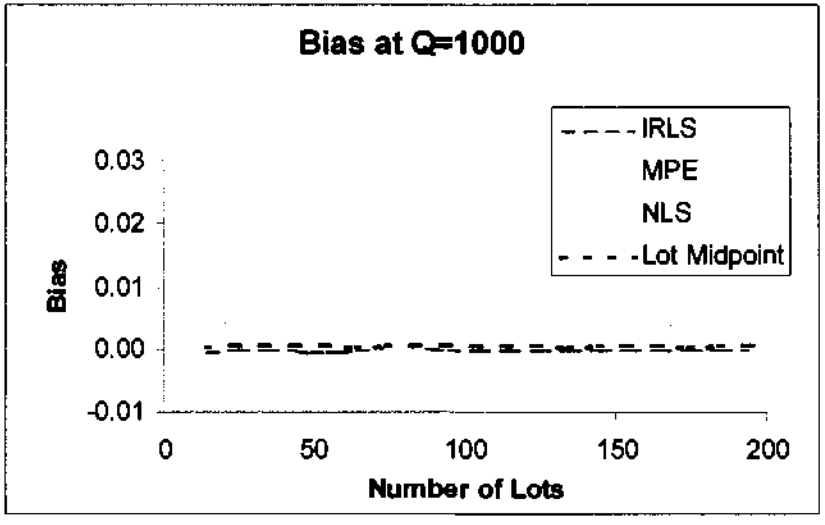


Figure 5.36. Simulation Experiment 5, Error in Predicting the Cost of Unit 100





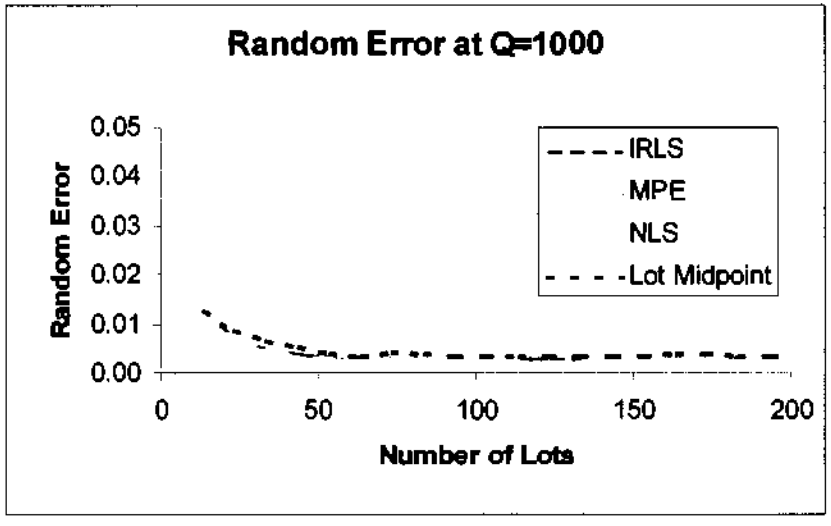


Figure 5.37. Simulation Experiment 5, Error in Predicting the Cost of Unit 1,000

5.7 Simulation experiment 6: multiplicative, *t*–distributed errors, learning slope = 80%

This experiment is identical to Simulation Experiment 1, except that we replaced the multiplicative normal error term with a *t*–distributed error term having the same standard deviation and an offset to yield a mean of 1.0. The *t*–distribution is indexed by the degrees-of-freedom parameter *df*, and approaches the normal distribution as $df \to \infty$. We chose a *t*–distribution with $df = 3$. This distribution has considerably thicker tails than does the normal distribution, thus generating more extreme observations or outliers.

The standard deviation of the *t*–distribution equals $\sqrt{df/(df-2)}$. Multiplying the error term by the factor $0.15 \times \sqrt{(df-2)/df}$ yields a distribution with standard deviation 0.15, comparable to the normal distribution from Simulation Experiment 1. With $df = 3$, this factor becomes $0.15/\sqrt{3}$. By normalizing the standard deviation to 0.15, we isolated the effect of the shape of the error distribution from that of the standard deviation. Recall that we already examined the latter effect in Simulation Experiment 4. Thus, we calculated the observed lot average cost as:

$$\begin{aligned} Obs\_LAC_i &= LAC_i \times (1+u_i), \\ \text{where } u_i &\sim t_{df=3} \times 0.15/\sqrt{3}. \end{aligned} \tag{4.13}$$

Figure 5.38 compares the two error distributions, both centered on a mean of zero. Note that the *t*–distribution has higher density than the normal distribution in both tails, specifically for error values $|u| > 0.42$ or $|u| > 2.80 \times \sigma$. For example, at the error value $u = \pm 0.50$ (or $u = \pm 3.33 \times \sigma$), the *t*–distribution is 2.81 times as high as the normal density having the same standard deviation. Thus, very large outlying errors are much more likely under the *t*–distribution than under the normal distribution.[51]

Figures 5.39 through 5.43 show the relative errors for the different estimation methods. The most noticeable feature of these results is the erratic behavior of the MPE estimates. The *t*–distribution has thicker tails than does the normal distribution. For the same reasons that MPE is sensitive to the size of the standard deviation, MPE is also sensitive to these outliers, occasionally leading to very large biases. Of course, a human

---

[51] The differences between the normal distribution and the normalized *t*–distribution are discussed in Johnson and Kotz (1970), Volume II, Chapter 27, p. 97. A more detailed, primary reference is Weir (1960).

analyst seeing these outlying lot costs could choose to exclude them from the data sample, thereby dampening their effect on the MPE estimates. Our simulation algorithm did not exclude any lots, so it could be argued that the errors shown here are larger than would be experienced in a real-life cost analysis. Nonetheless, there are borderline cases in which the cost analyst does not know whether a data point is truly an outlier, because it is not as extreme as some of those included in our simulated data. Given the typically small samples available to cost analysts, a conservative analyst might not be willing to discard any of these data points as outliers. These retained data points would influence MPE more than any of the other estimation methods. Below, we discuss specific examples of the effects of outliers on MPE.

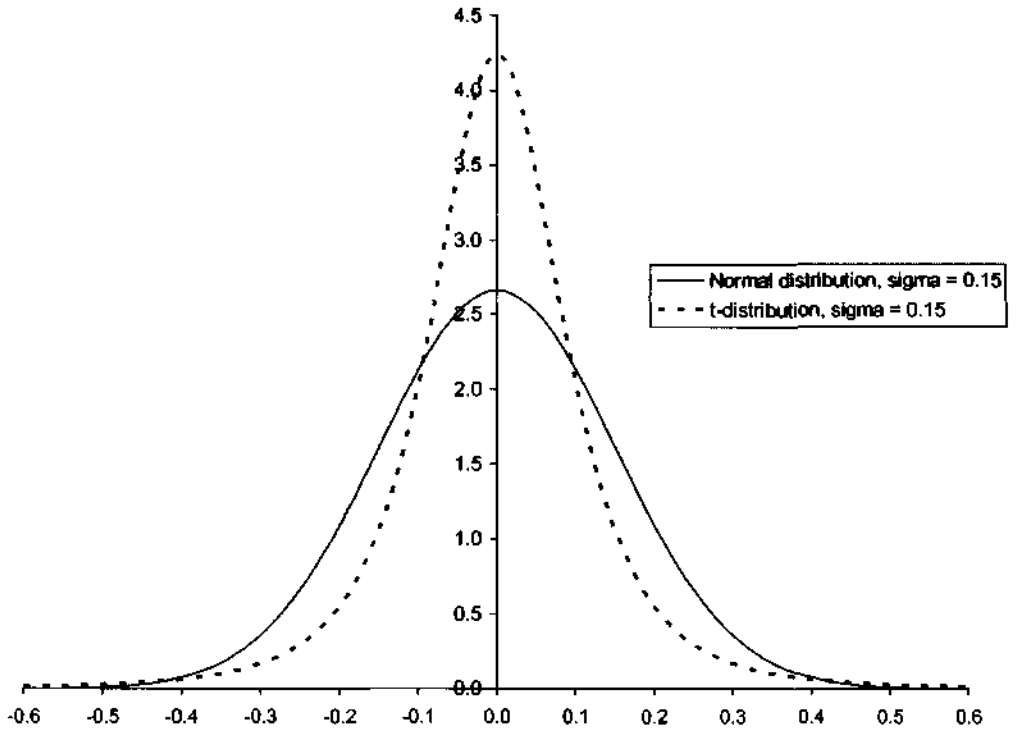
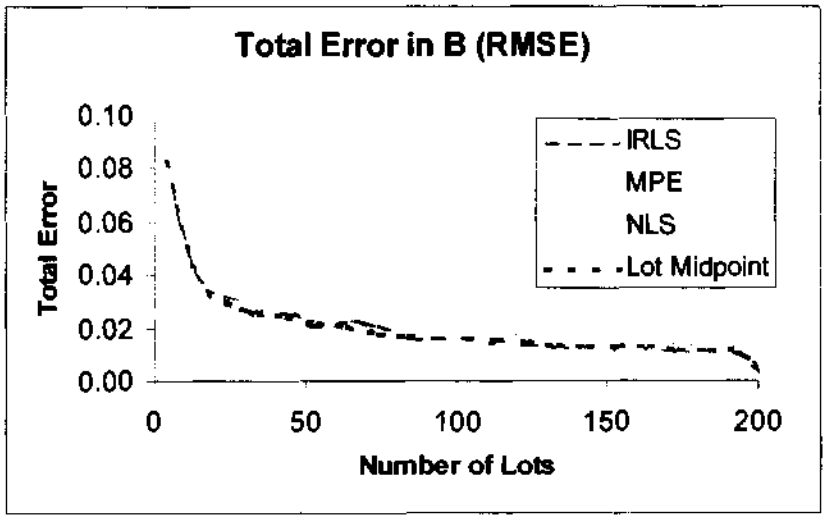
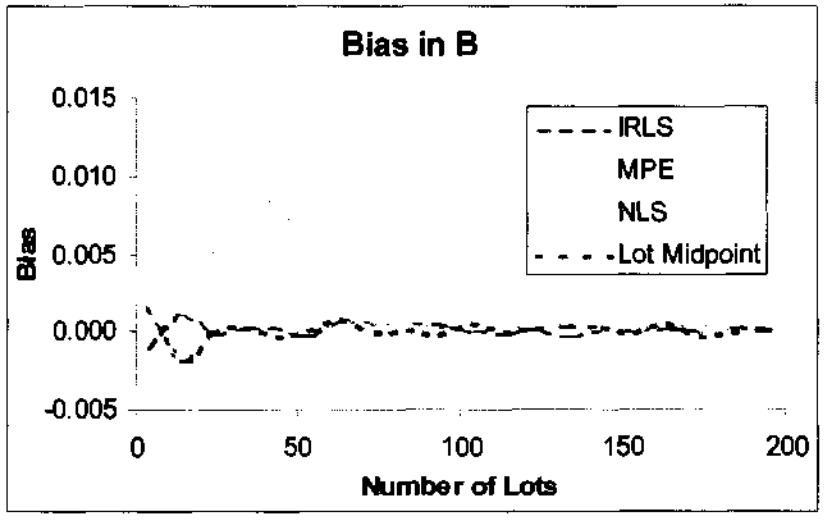


Figure 5.38. Comparison of Normal and t-distribution





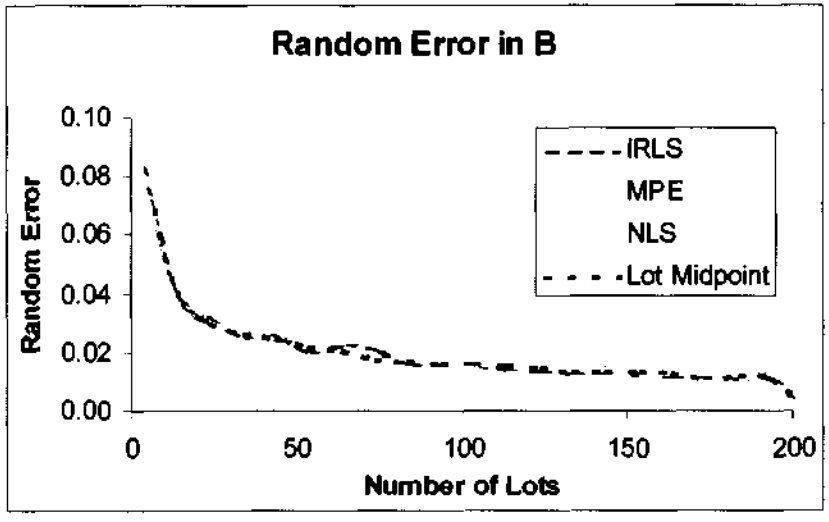
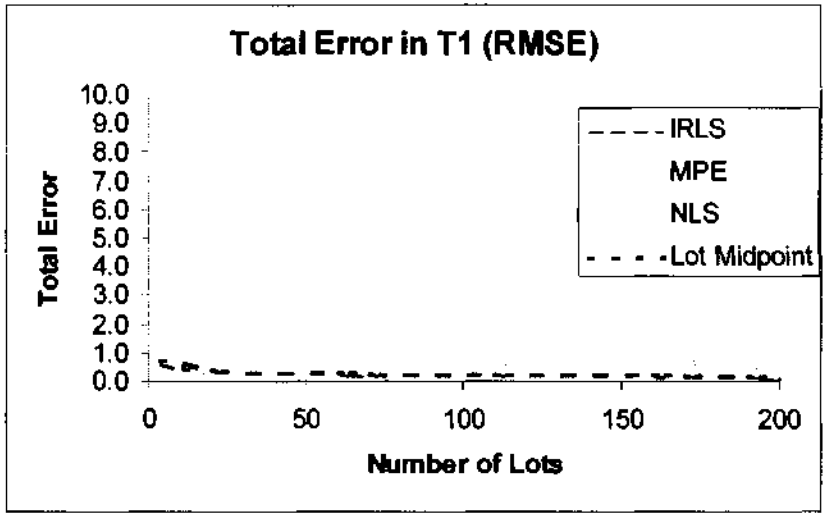
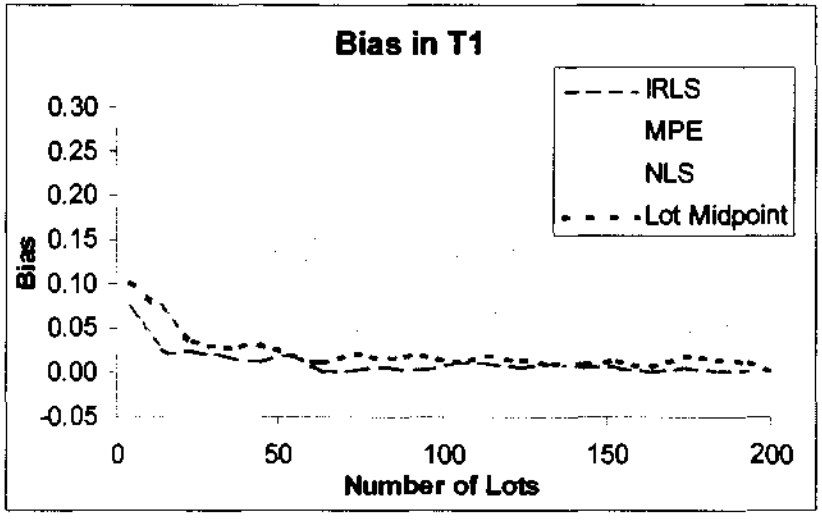


Figure 5.39. Simulation Experiment 6, Error in Slope Coefficient





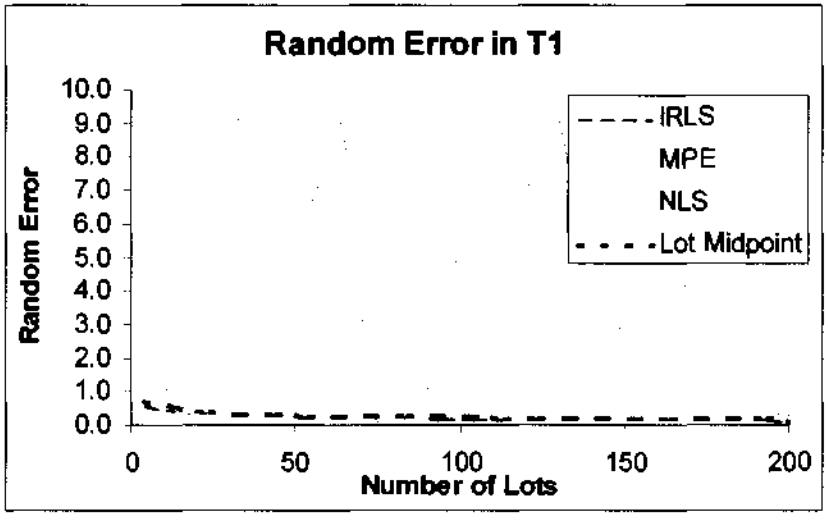
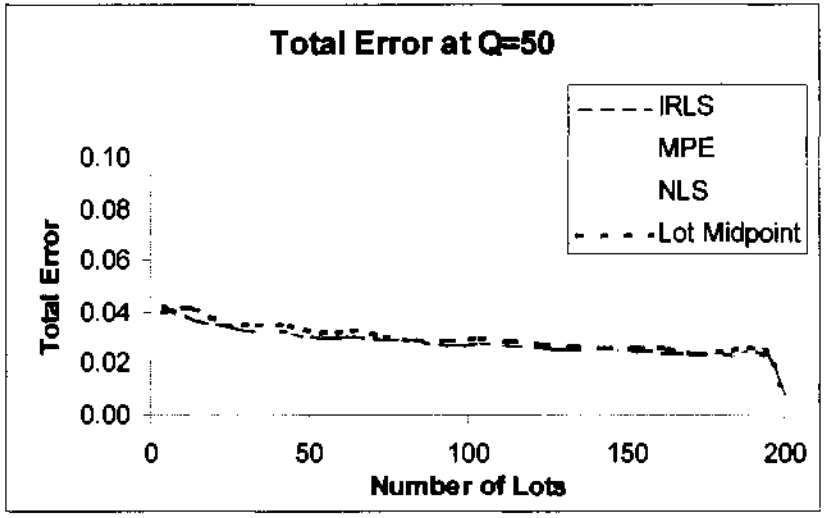
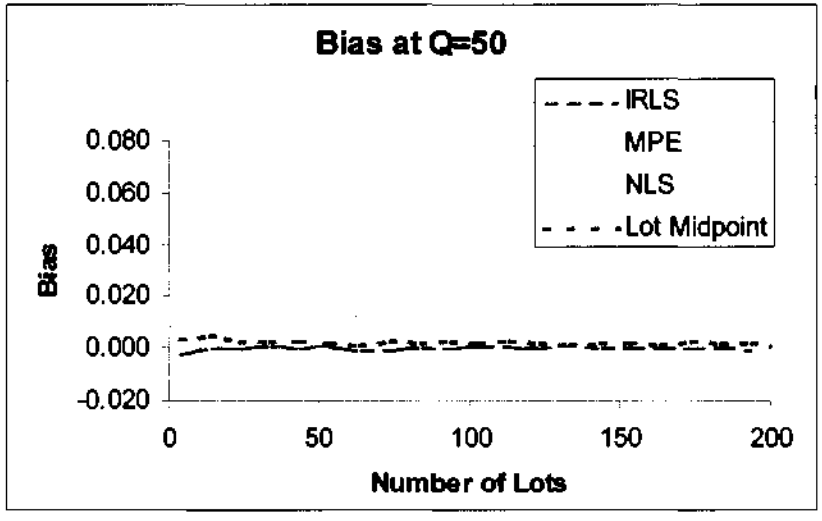


Figure 5.40. Simulation Experiment 6, Error in Intercept





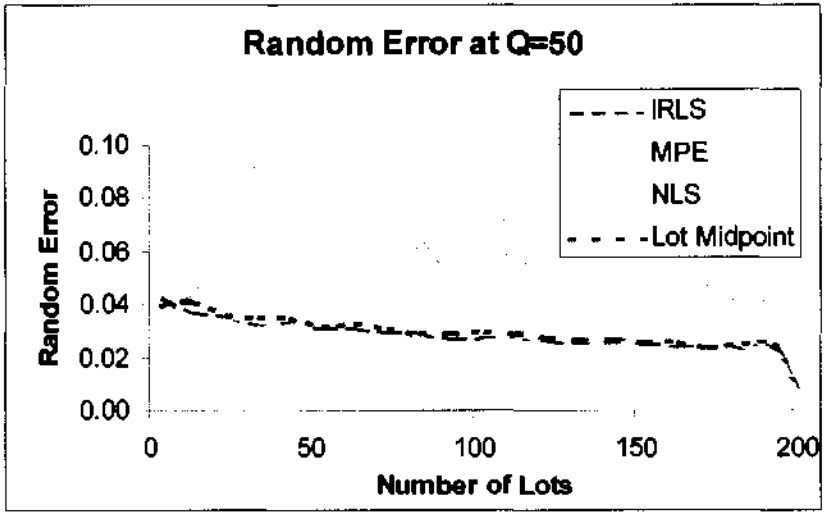
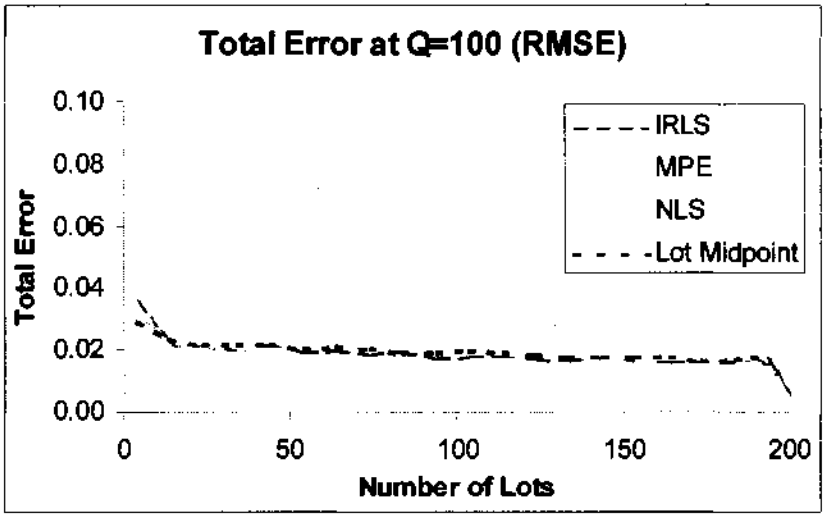
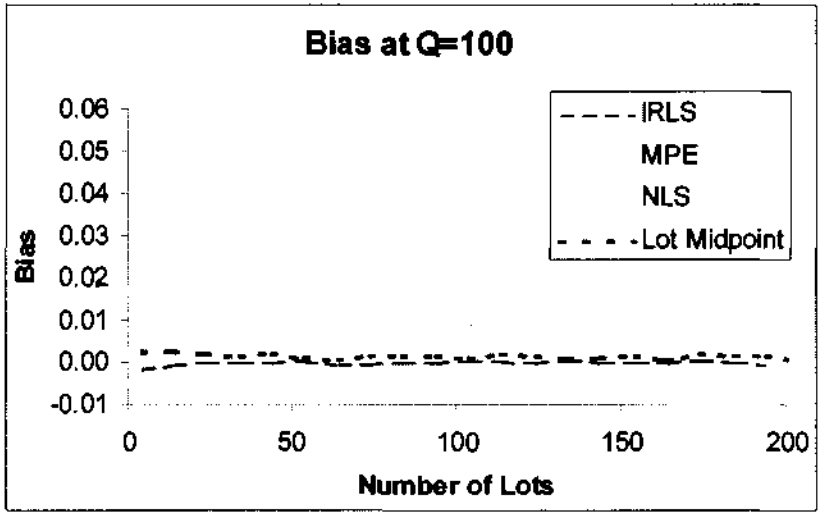


Figure 5.41. Simulation Experiment 6, Error in Predicting the Cost of Unit 50





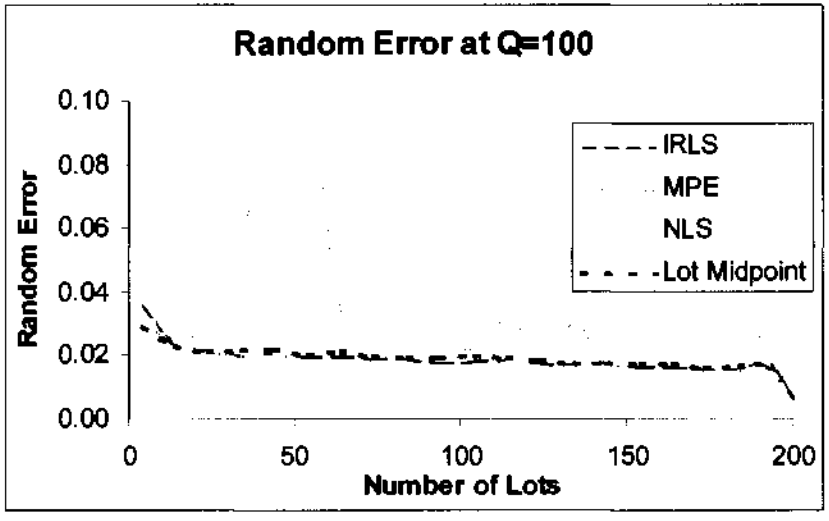
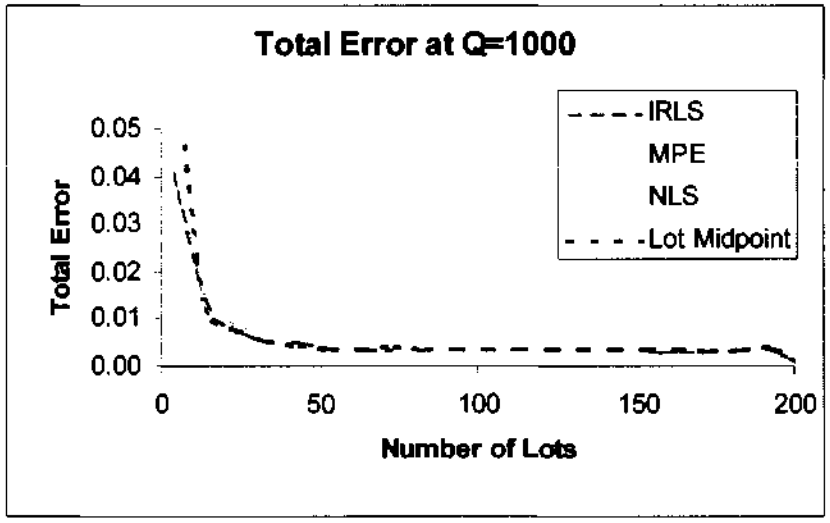
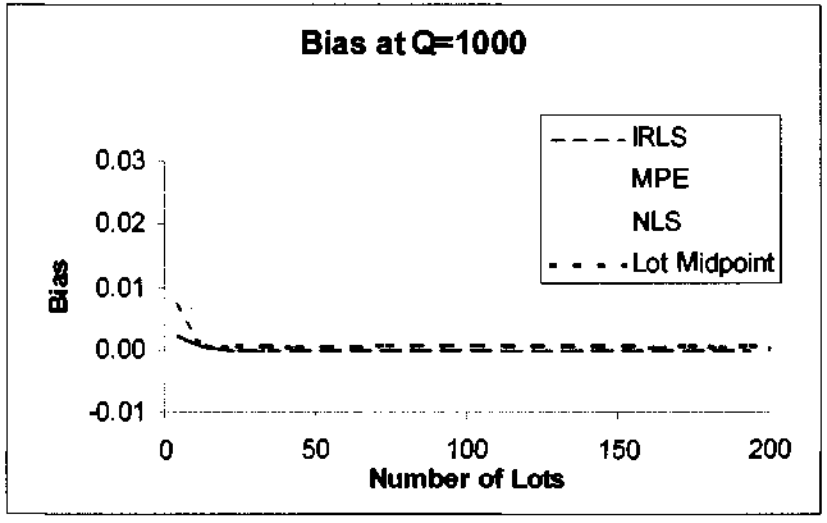


Figure 5.42. Simulation Experiment 6, Error in Predicting the Cost of Unit 100





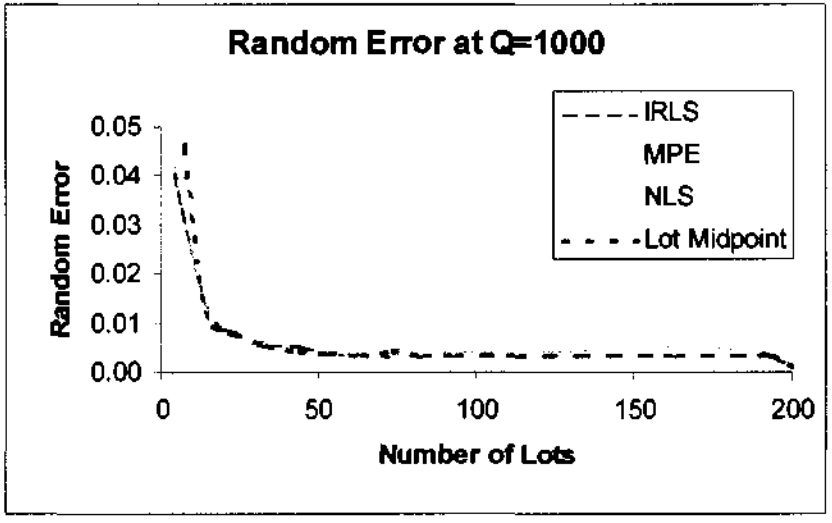


Figure 5.43. Simulation Experiment 6, Error in Predicting the Cost of Unit 1,000

We investigated which specific cases were leading to the extremely large biases in the MPE estimates of T_1 and, thus, in the predictions of unit cost for that method. We found that, in all cases, the results were driven by the costs of only one or two outlying lots. Although we ran the simulation on lots of size 50, the same results were found for lots of size 10 under *t*–distributed errors. For ease of interpretation, the examples below have lots of size 10.

Table 5.1 illustrates one case in which the MPE estimate of T_1 was biased high. The observed cost of the third lot is an extreme outlier. A cost analyst would likely try deleting this point. However, because our simulation did not have a decision rule for outliers, the third lot was included in the sample for all four estimation methods.

Table 5.1. Data with Outlier from Simulation

Lot start	Lot end	Observed average cost
1	10	1.0670
11	20	0.7521
21	30	2.3754
31	40	0.6086
41	50	0.4022
51	60	0.4123
61	70	0.4997
71	80	0.4200
81	90	0.3766
91	100	0.3463

Using the data in Table 5.1, MPE found $\hat{b} = -1.18$ and $\hat{T}_1 = 69.14$, and IRLS found $\hat{b} = -0.48$ and $\hat{T}_1 = 3.76$. Recall that the simulated data were generated using "true" parameter values $b = -0.33$ and $T_1 = 1.8$. Recall also that Lee (1997, p. 41) argued for the restriction $-1 < b \leq 0$. The MPE estimate violates that restriction, implying an implausibly steep 44% learning slope. However, the reason MPE estimated an implausible slope was to compensate for the even less plausible intercept, $\hat{T}_1 = 69.14$.

Figure 5.44 plots both the raw data and two fitted models. The MPE estimates have been pulled high by the outlying third lot, to a much greater extent than have the IRLS estimates. The reason for this difference can be seen by comparing the percentage errors for the MPE and IRLS estimates at convergence (see Table 5.2). The MPE method explicitly minimizes the sum-of-squared percentage errors. Thus, MPE would never tolerate the nearly 200% error that IRLS tolerates in lot #3 in order to better fit the other

(non-outlier) data points.[52] However, the price that MPE pays for fitting lot #3 is to tilt the learning curve toward very high levels at low cumulative quantities. This tilt is manifested in the much more severe overprediction of the cost of the first two lots (–94.2% and –73.7%), relative to the more moderate prediction errors under IRLS (–45.6% and –27.1%).

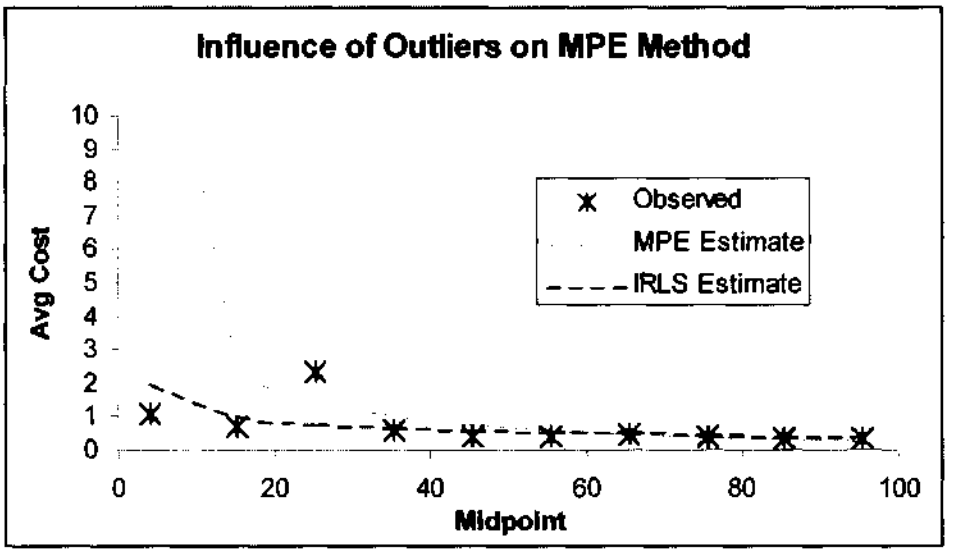


Figure 5.44. Influence of Outliers on Estimated Learning Curve

Table 5.2. Percentage Errors for Data with Outliers

Lot start	Lot end	Observed average cost	MPE estimate	Percent error for MPE	IRLS estimate	Percent error for IRLS
1	10	1.0670	18.3628	–94.2%	1.9608	–45.6%
11	20	0.7521	2.8565	–73.7%	1.0315	–27.1%
21	30	2.3754	1.5407	54.2%	0.8069	194.4%
31	40	0.6086	1.0343	–41.2%	0.6876	–11.5%
41	50	0.4022	0.7692	–47.7%	0.6103	–34.1%
51	60	0.4123	0.6075	–32.1%	0.5548	–25.7%
61	70	0.4997	0.4991	0.1%	0.5125	–2.5%
71	80	0.4200	0.4218	–0.4%	0.4789	–12.3%
81	90	0.3766	0.3641	3.4%	0.4513	–16.6%
91	100	0.3463	0.3195	8.4%	0.4280	–19.1%

[52] Recall that the percentage error is calculated as (*observed – predicted*) / *predicted*.

In case the reader doubts that the global minimum of MPE's criterion function occurred at $\hat{T}_1 = 69.14$, Figure 5.45 shows the minimized sum-of-squared percentage errors at various values of T_1. In each case, we found the best estimate $\hat{b}$ conditional on $\hat{T}_1$, and we calculated the criterion function at the parameter vector $(\hat{b}, \hat{T}_1)$. As is clearly shown in the figure, the minimum occurs when $\hat{T}_1$ is close to 70.

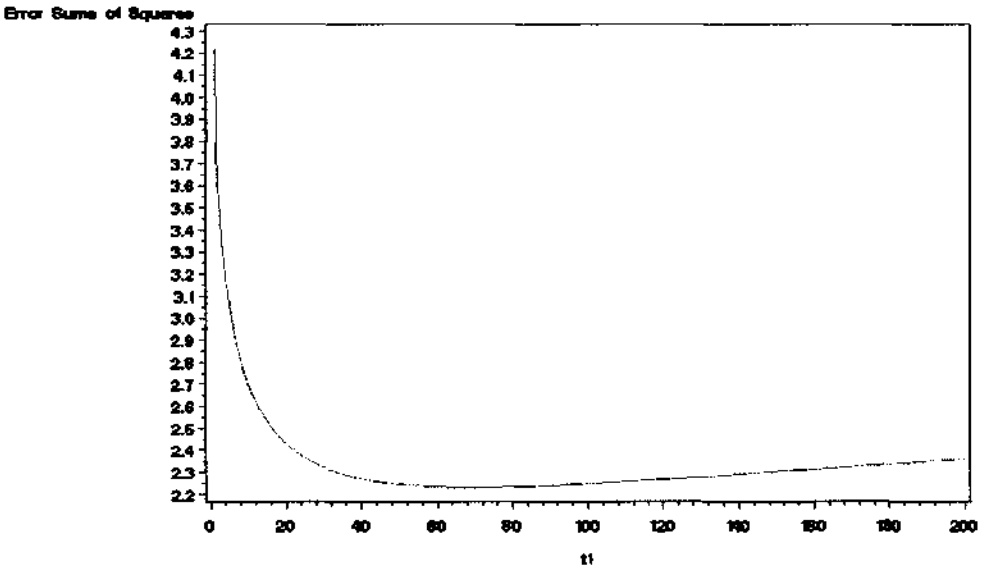


Figure 5.45. Minimum Squared Percentage Error as a Function of T_1

It could be argued that the example presented above is too extreme, and that any competent cost analyst would know to delete lot #3 from the analysis, thereby avoiding the extreme bias in $\hat{T}_1$. However, even in much less extreme cases, outliers continue to exert too much influence on the MPE estimates.

Example data are provided in Table 5.3. With this data set, MPE found $\hat{b} = -0.68$ and $\hat{T}_1 = 7.84$, and IRLS found $\hat{b} = -0.49$ and $\hat{T}_1 = 3.58$. Again, the simulated data were generated using "true" parameter values $b = -0.33$ and $T_1 = 1.8$. Figure 5.46 plots the true model, the simulated data containing t–distributed errors around the true model, and finally two fitted models. Unlike the previous example, it is difficult to say whether the outlier is lot #1 being too low or lot #2 being too high. In fact, the minimized sum-of-squared percentage errors is 0.21 when lot #1 is removed, and 0.11 when lot #2 instead is removed. MPE does not offer much guidance as to which of the two lots (if either) should be removed from the analysis. Yet, inclusion of both lot #1 and lot #2

yields an upward-biased estimate of $\hat{T}_1$. And again, MPE compensates with too steep a learning slope, 62% versus the "true" value of 80%.

Table 5.3. Data with Smaller Outlier from Simulation

Lot start	Lot end	Observed average cost
1	10	1.0586
11	20	1.7838
21	30	0.8029
31	40	0.5139
41	50	0.4909
51	60	0.4758
61	70	0.4549
71	80	0.4127
81	90	0.3916
91	100	0.3542

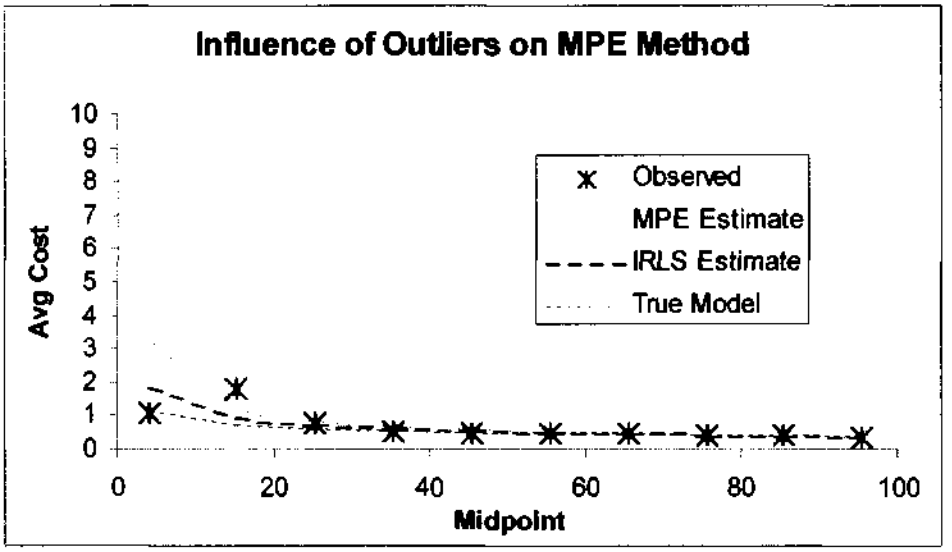


Figure 5.46. Influence of Smaller Outliers on Estimated Learning Curve

We conclude that MPE's sensitivity to outliers makes it a less reliable method than the other three examined.

5.8 Simulation experiment 7: multiplicative, normal errors with first-order serial correlation, learning slope = 80%

This experiment is identical to Simulation Experiment 1, except that we replaced the independent normal error term with one that exhibits first-order serial correlation. The errors still have a standard deviation of 0.15, but now the error for the i^{th} lot is highly dependent on the error for lot $i-1$. This situation may frequently arise in practice, because successive lots are likely produced by many of the same workers using mostly the same equipment.

Womer and Patterson (1983) found evidence of serial correlation, and devised special methods to efficiently estimate the learning curve in the face of this problem. They opined that (p. 268):

> Since learning is measured as successive units of output are produced, one should not be surprised at the presence of autocorrelation in the data. In many cases, this violation of the assumption of independent error terms is ignored or viewed as insignificant or unimportant. This is a careless oversight.

Although we agree with Womer and Patterson, our objective is different from theirs. The reality is that most cost analysts continue to apply estimation methods that were not specifically designed for serially correlated data. In their defense, it may be difficult to detect serial correlation in the small samples that typify cost analysis. Our objective in this section is to assess the robustness of the four estimation methods that were not designed for serial correlation, when they are applied to serially correlated data.[53]

For this analysis it is convenient to zero-out the mean of the error term. Thus we calculated the observed lot average cost as:

$$Obs\_LAC_i = LAC_i \times (1+u_i), \tag{4.14}$$

where:

$$u_i = \rho \times u_{i-1} + \sqrt{1-\rho^2} \times \varepsilon_i, \tag{4.15}$$

---

[53] Along these lines, Womer and Patterson found that maximum likelihood estimation of incremental lot cost is particularly sensitive to serially-correlated errors. We did not independently investigate maximum likelihood in this monograph, because it requires by far the most computation per iteration.

with $\rho = 0.5$ and $\varepsilon_i \sim N\left(0, 0.15^2\right)$.

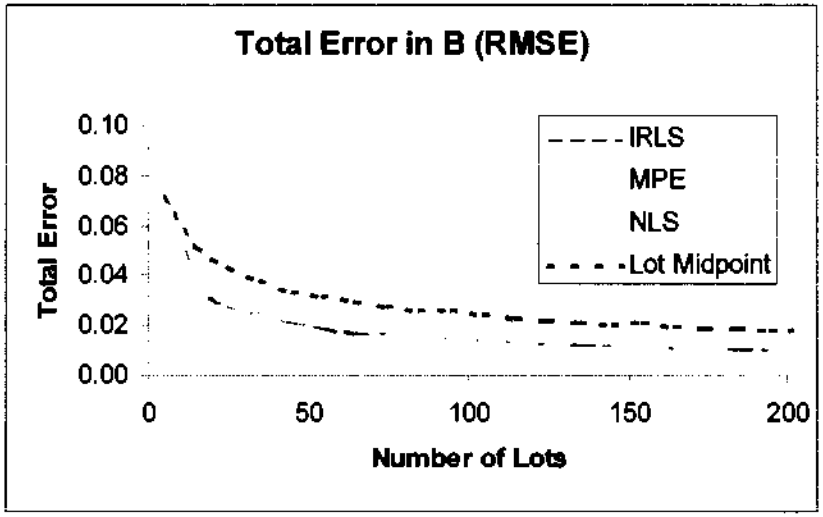
We generated the error terms $\{\varepsilon_i\}$ independently, without any serial correlation: $Corr(\varepsilon_i, \varepsilon_j) = 0.0$ for all $i \neq j$. However, the transformation in equation (4.15) induces serial correlation among the $\{u_i\}$: $Corr(u_i, u_{i-1}) = 0.5$.

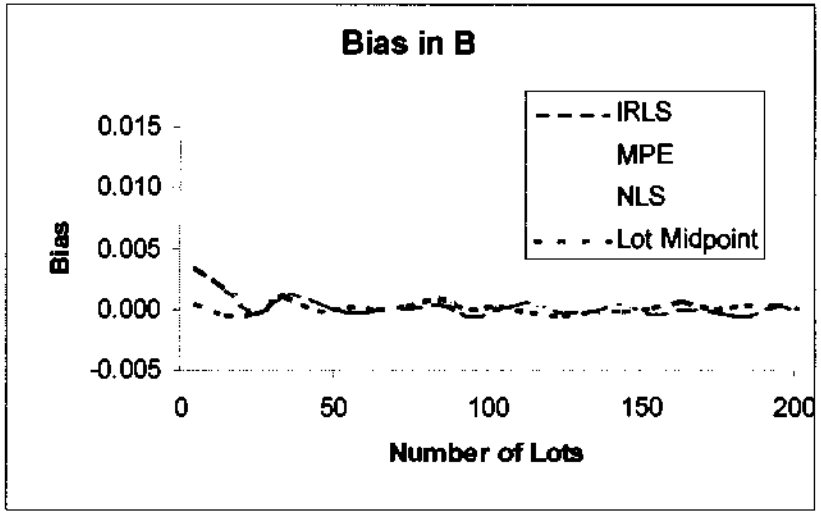
Note also that, by construction, the $\{u_i\}$ have the same standard deviation as the $\{\varepsilon_i\}$:

$$Var(u_i) = \rho^2 \times Var(u_{i-1}) + (1-\rho^2) \times Var(\varepsilon_i) . \qquad (4.16)$$

Because the $\{u_i\}$ are identically distributed, we have $Var(u_i) = Var(u_{i-1})$. Using this fact (and $|p| < 1$), we can solve equation (4.16) for $Var(u_i) = Var(\varepsilon_i)$. Thus, by first generating the $\{\varepsilon_i\}$ with $\sigma = 0.15$, and then applying the transformation in equation (4.15), we were able to obtain error terms $\{u_i\}$ having the same standard deviation.

Figures 5.47 through 5.51 show the relative errors for the different estimation methods. Of the four estimation methods compared, it appears that only lot-midpoint iteration is particularly sensitive to serial correlation. We see in Figure 5.52 that the loss of precision in lot-midpoint iteration due to serial correlation is almost as large as that caused by doubling the standard deviation. Because the theory behind lot-midpoint iteration is so poorly developed, we do not have a sound theoretical explanation for this result. Perhaps the iterative nature of lot-midpoint iteration serves to compound the errors as the procedure converges.





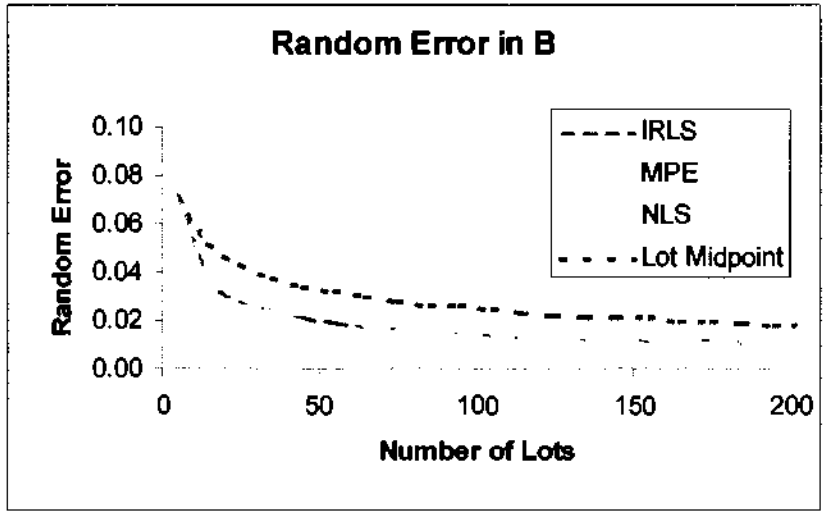
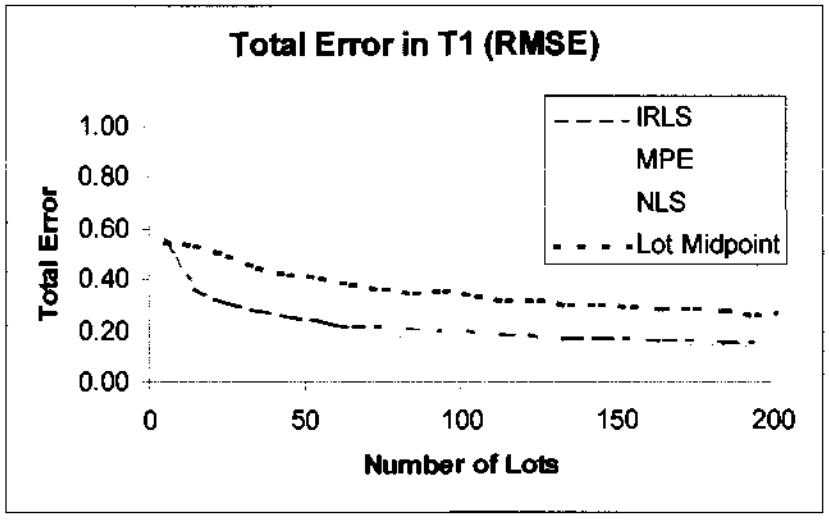
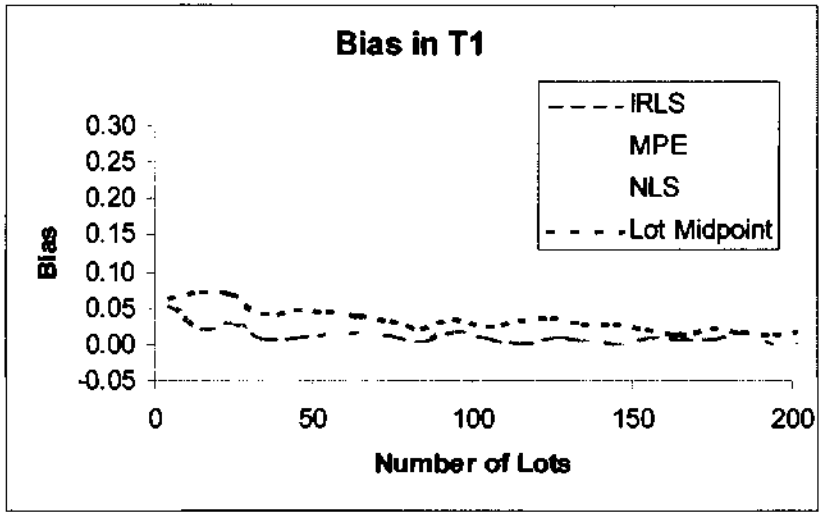


Figure 5.47. Simulation Experiment 7, Error in Slope Coefficient





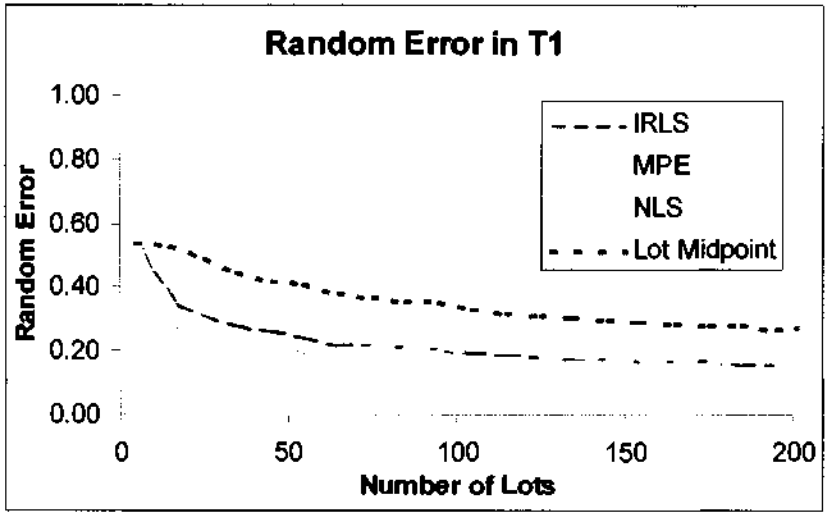
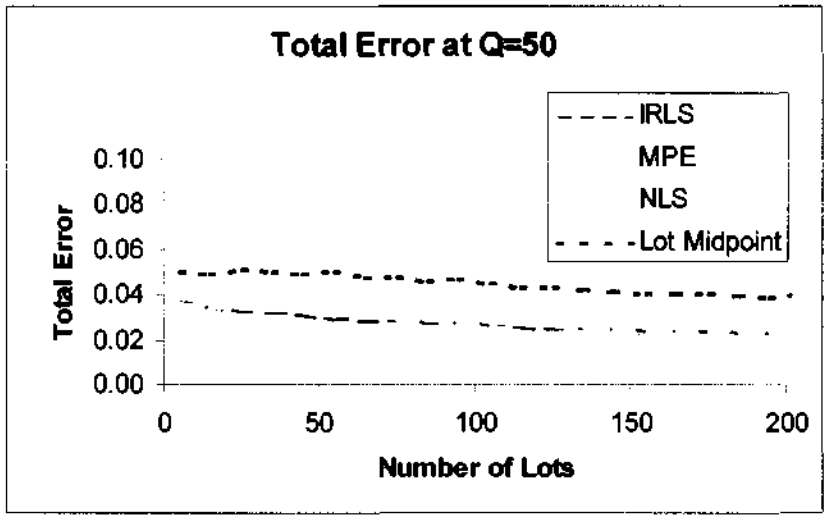
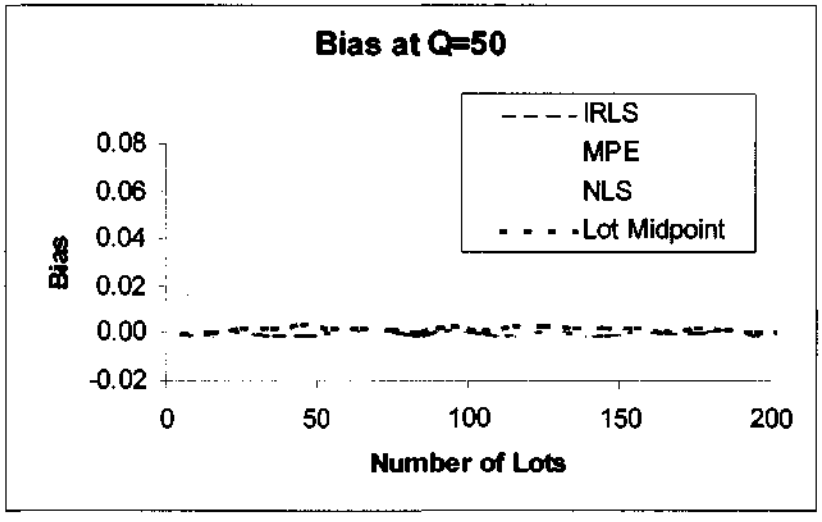


Figure 5.48. Simulation Experiment 7, Error in Intercept





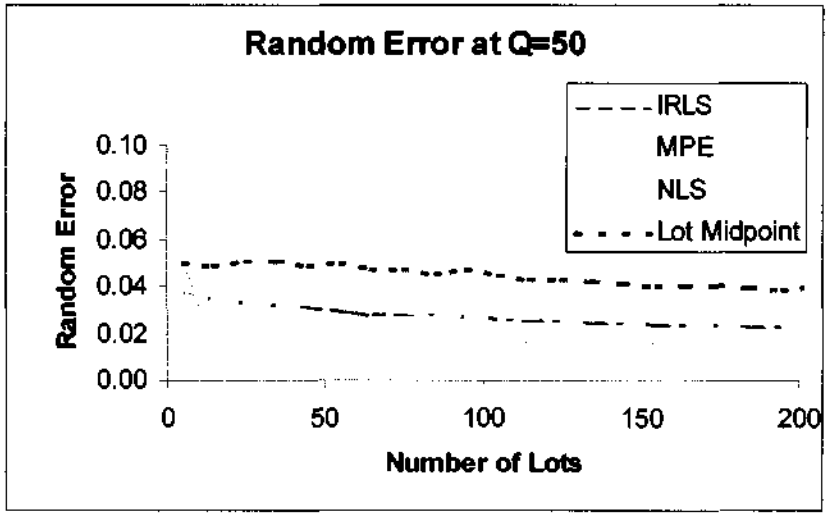
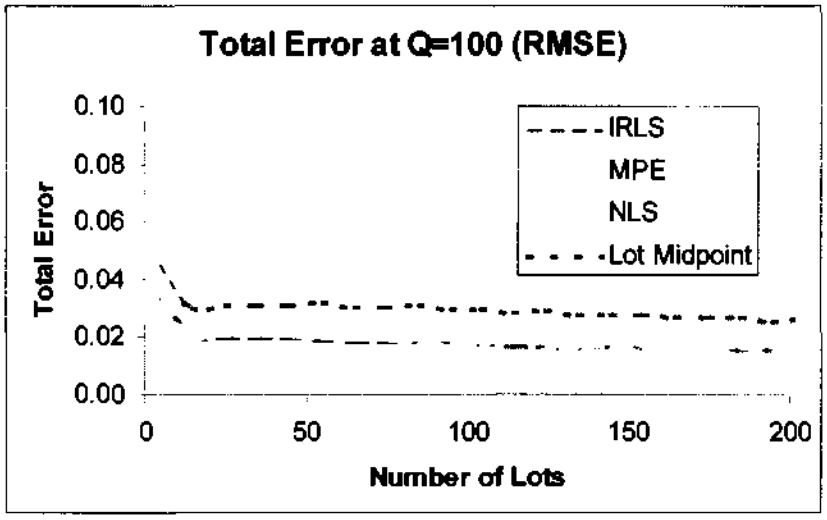
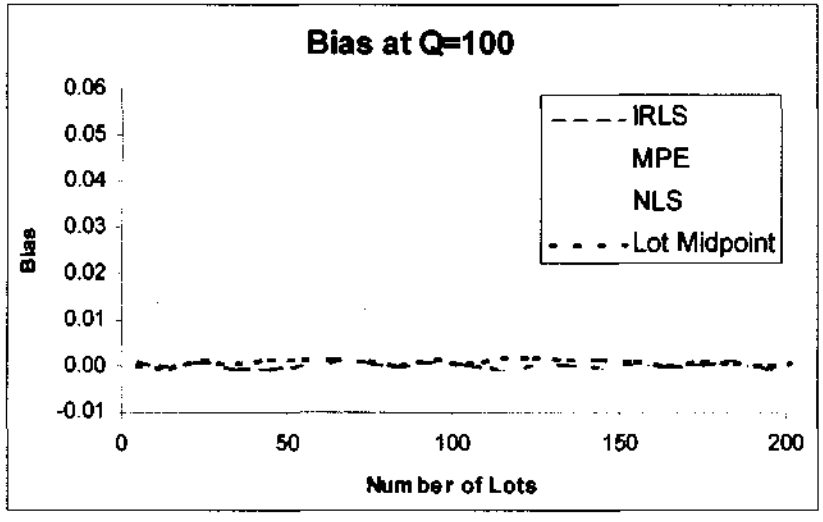


Figure 5.49. Simulation Experiment 7, Error in Predicting the Cost of Unit 50





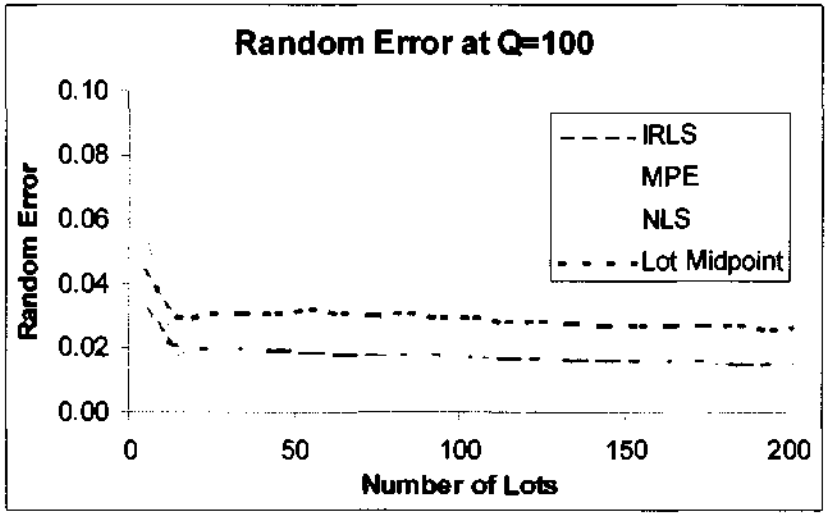
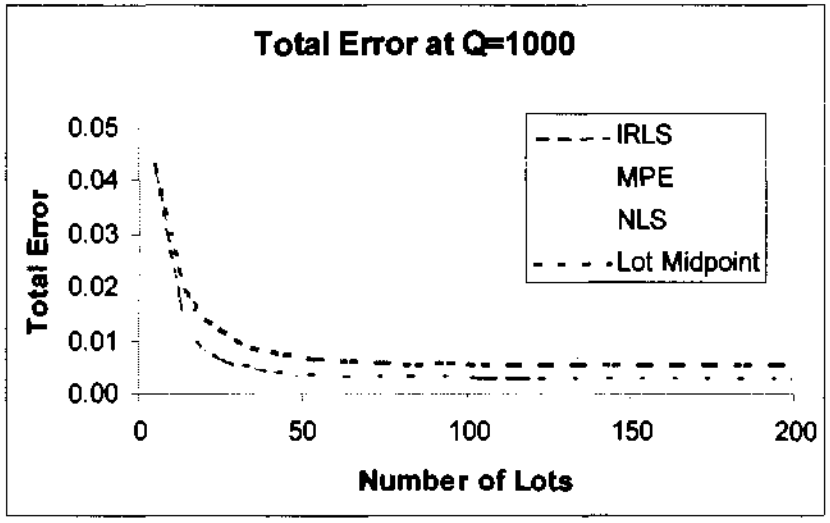
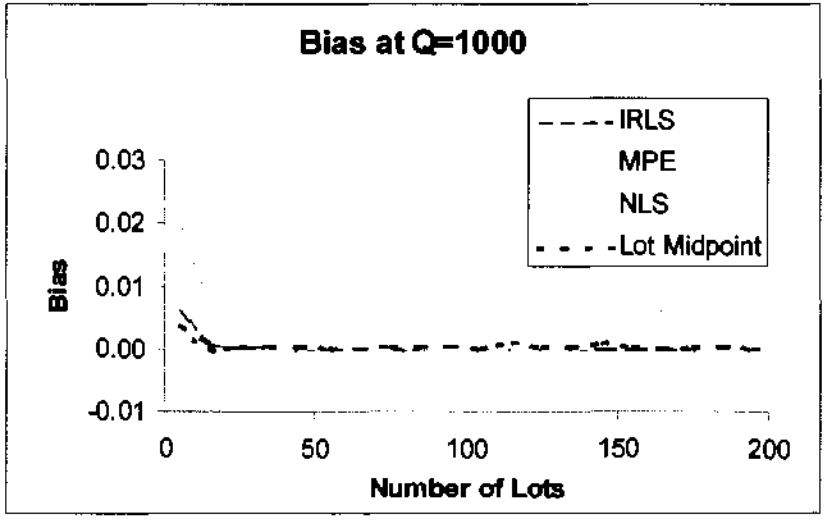


Figure 5.50. Simulation Experiment 7, Error in Predicting the Cost of Unit 100





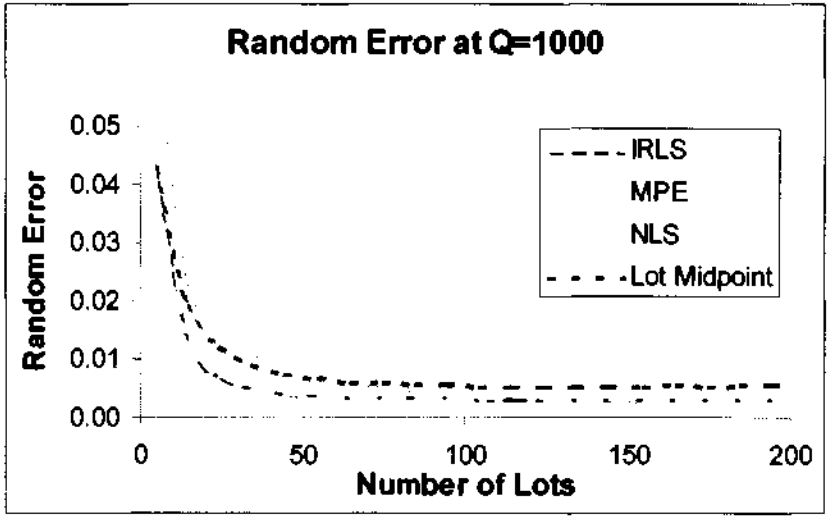
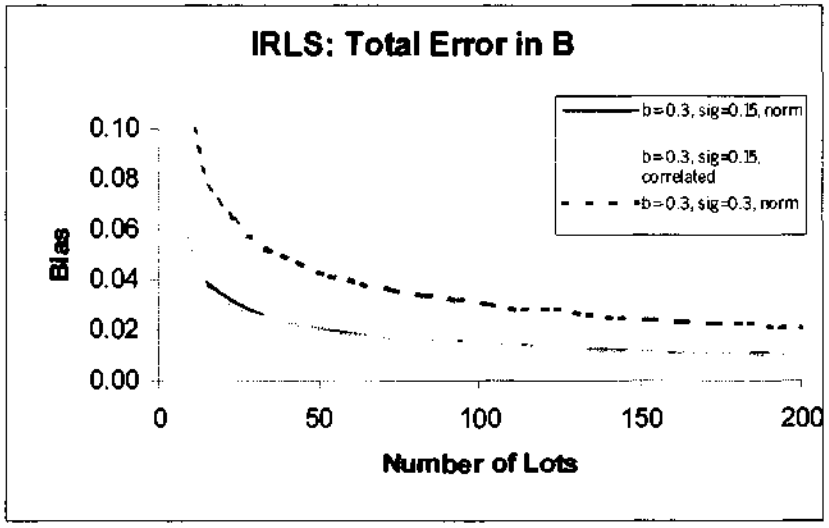
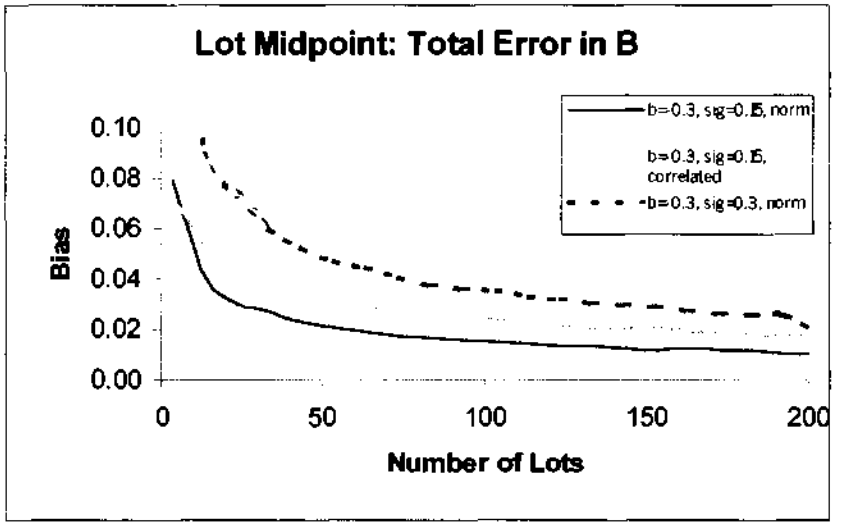
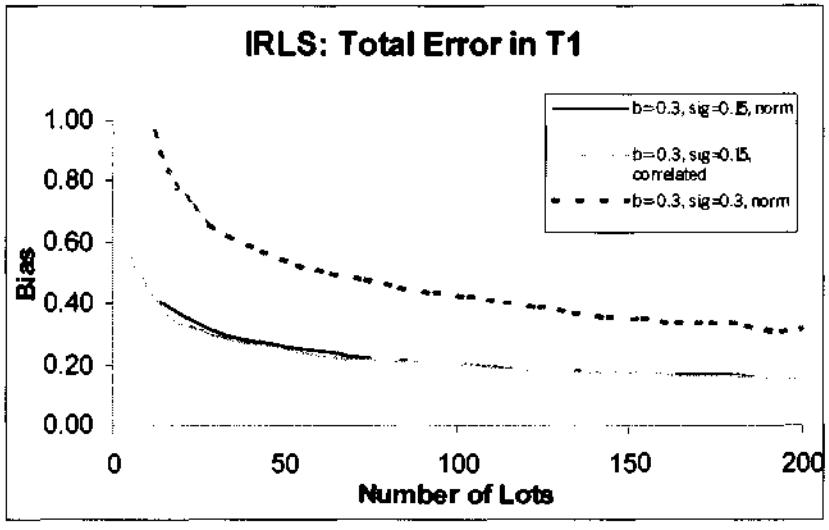


Figure 5.51. Simulation Experiment 7, Error in Predicting the Cost of Unit 1,000







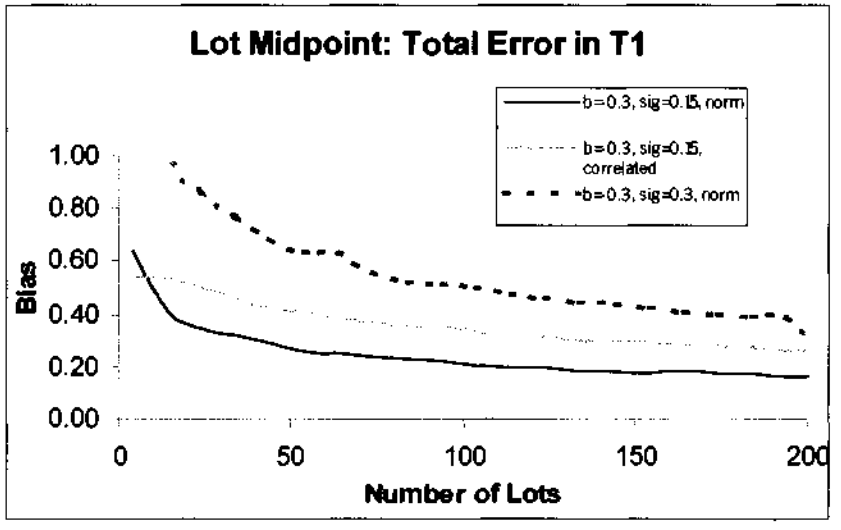


Figure 5.52. Sensitivity of Lot-Midpoint Iteration to Serial Correlation

5.9 Conclusions from the simulation experiments

Figures 5.53 through 5.56 compare the performance of each estimation method under various assumptions on the error term. We drew the following conclusions from the simulation experiments. First, both IRLS and lot-midpoint NLS are theoretically guaranteed to produce consistent estimates. However, these are general-purpose estimation methods, and their small-sample properties (such as bias) are not known for general predictor functions. In the particular case of learning curves, our simulation experiments suggest that IRLS and lot-midpoint NLS actually produce unbiased estimates even for small numbers of lots.

Most estimation methods are developed under a particular set of assumptions. Estimation methods are called *robust* if they continue to produce good estimates even when those assumptions are violated. None of the methods we compared rely on any particular assumption about the true learning slope, the number of units in a lot, or the standard deviation of the error term. However, it is still of interest to inquire whether the methods perform as well under a range of values for these parameters. Some of the methods rely on a particular distributional assumption, such as normally distributed errors. Thus, it is also of interest to inquire about the performance of the methods under alternative (non-normal) error distributions.

IRLS and lot-midpoint NLS continued to produce unbiased estimates under all of the simulation excursions. The performance of these two methods was essentially unaffected by the substitution of either uniform or t–distributed errors for the normal errors found in the baseline experiment. Naturally, however, the parameter estimates became less precise during Simulation Experiment 4 when we doubled the standard deviation of the error terms (see Figures 5.53 and 5.54). In addition, the predictions of unit cost became less precise when we replaced the baseline 80% learning slope with a shallower 90% slope. However, as explained in the discussion of Simulation Experiment 2, that loss of precision is not a bias, but rather an inevitable consequence of the pattern of data clustering under the shallower learning slope.

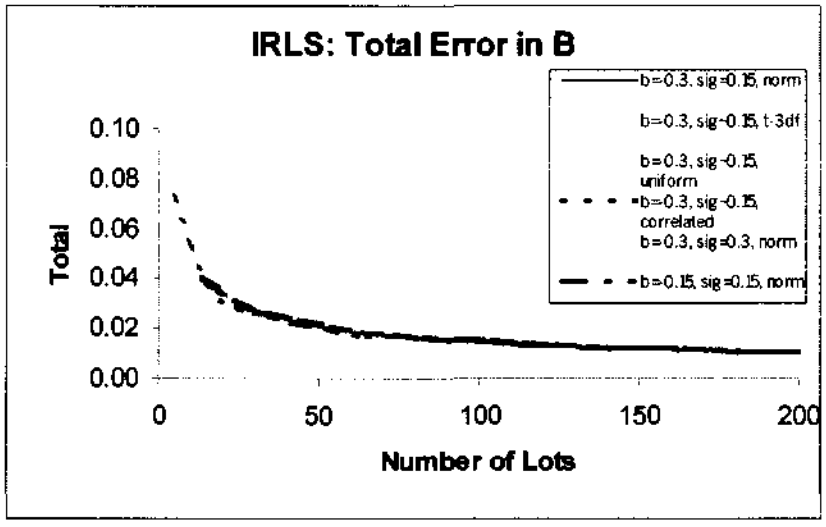
The estimates produced by lot-midpoint iteration and lot-midpoint NLS are numerically distinct. However, with just one exception, the numerical differences between the two sets of parameter estimates (e.g., between the estimated learning slopes) were essentially negligible. Consequently, both of these methods produced unbiased estimates even for small numbers of lots. The one exception is that the parameter estimates from lot-midpoint iteration (though not lot-midpoint NLS) became much less

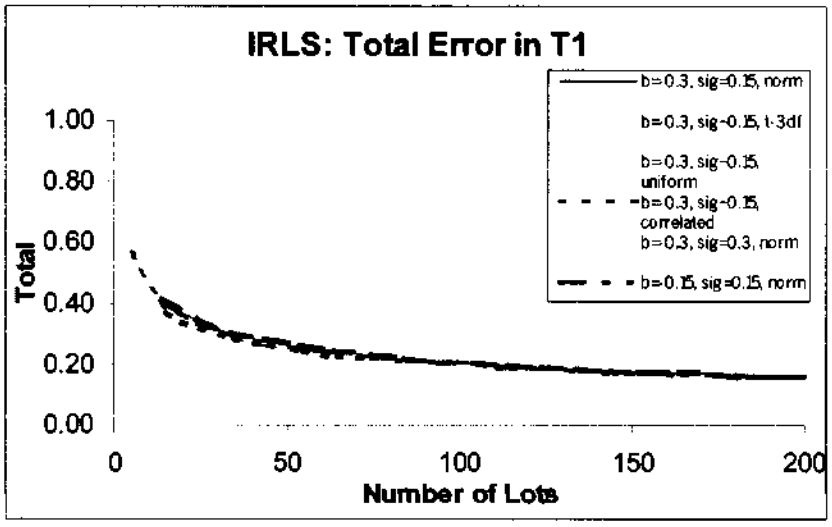
precise under first-order serial correlation (Simulation Experiment 7; see the summary in Figure 5.55). The introduction of serial correlation led to a drop in precision nearly equal to that engendered by doubling the standard deviation of the error terms (but without serial correlation). None of the other estimation methods exhibited any sensitivity to serial correlation. The difficulty is that serial correlation is not always detectable in the small samples that typify cost analysis. Thus, a cost analyst might inadvertently apply lot-midpoint iteration in a situation where it is rather imprecise. This imprecision could be avoided by applying other estimation methods (e.g., lot-midpoint NLS) that are robust to serial correlation.

Notwithstanding this case, the performance of lot-midpoint iteration was much better than we had expected. Prior to the simulation experiments, there was no theoretical basis for lot-midpoint iteration and little was known about the behavior of its estimates. We now know from Chapter 2 that lot-midpoint iteration does not minimize any continuously differentiable function. In a sense, that finding further undermines the theoretical basis for the method. Its apparently satisfactory performance characteristics, at least in the absence of serial correlation, remain a theoretical mystery.

The MPE estimates of T_1 were biased high, even in large samples, under every one of the simulation excursions. Similarly, the MPE predictions of unit cost were also biased high. Moreover, the biases increased both when we doubled the standard deviation of the normal errors, and (unique to this method) when we substituted t–distributed errors for the normal errors (Simulation Experiment 6; see the summary in Figure 5.56). The latter result illustrates that the performance of MPE degrades when there are more outlier observations (in statistical parlance, the error distribution has "thicker tails") than would be expected under a normal error distribution. Because of these biases and sensitivities, we recommend against the use of MPE.

In light of the latter result, as well as the sensitivity of lot-midpoint iteration to serial correlation, we recommend either IRLS or lot-midpoint NLS as the estimation methods of choice. NLS is already available as an option in most statistical software packages. IRLS is becoming increasingly available as a built-in feature in many statistical packages, and the equivalent method of quasi-likelihood can be programmed quite easily using any computational software or even a simple spreadsheet. There is no longer any excuse for cost analysts to use methods that produce inconsistent parameter estimates.





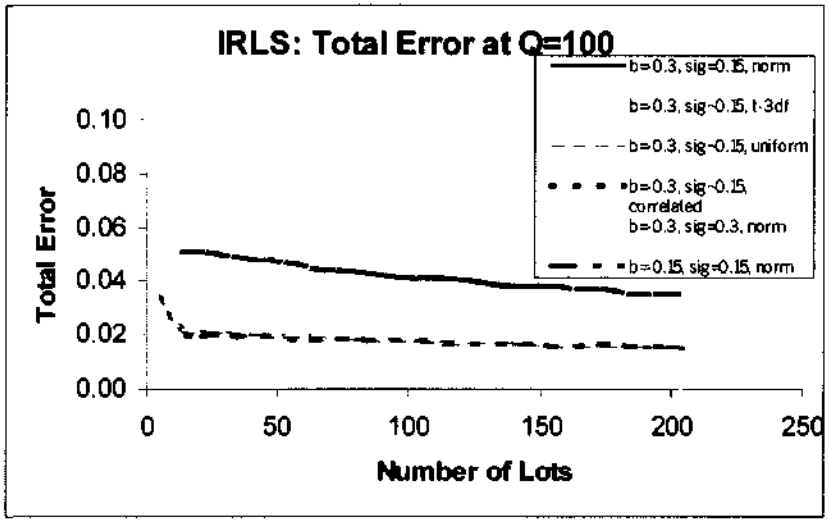
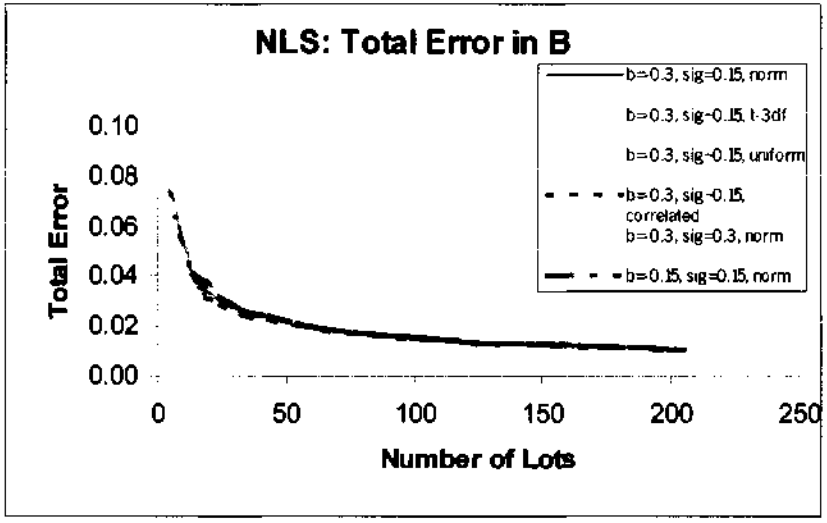
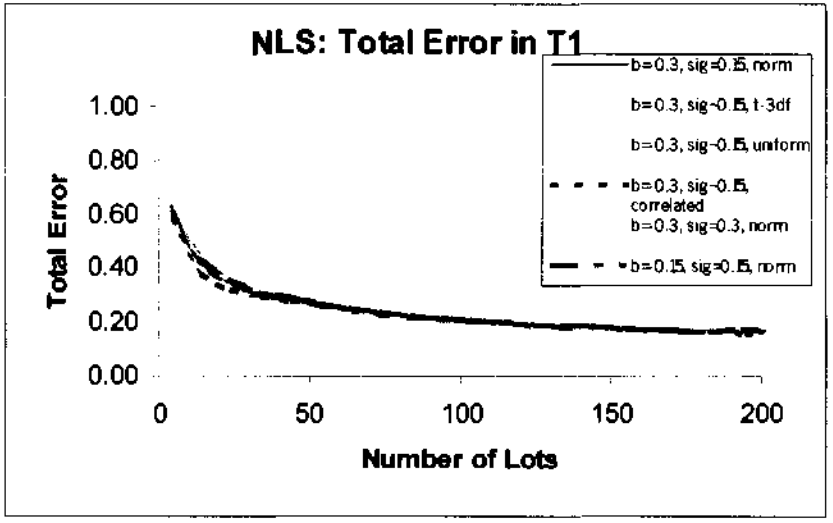


Figure 5.53. Robustness of IRLS





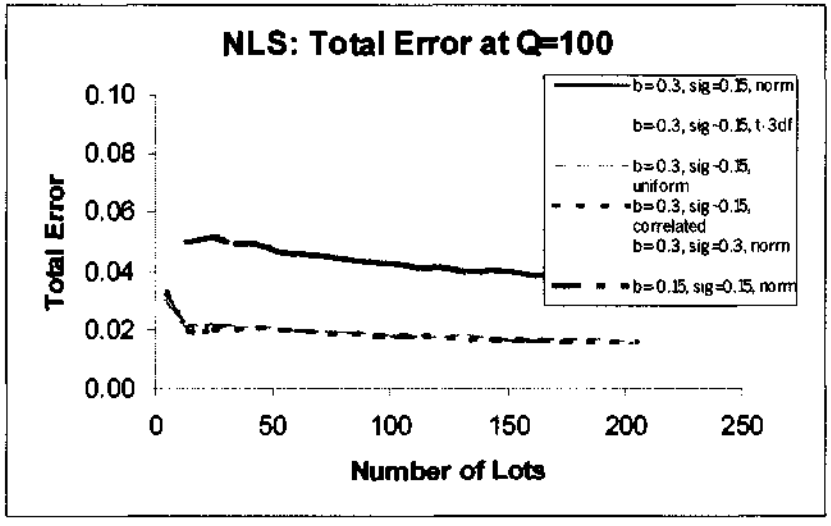
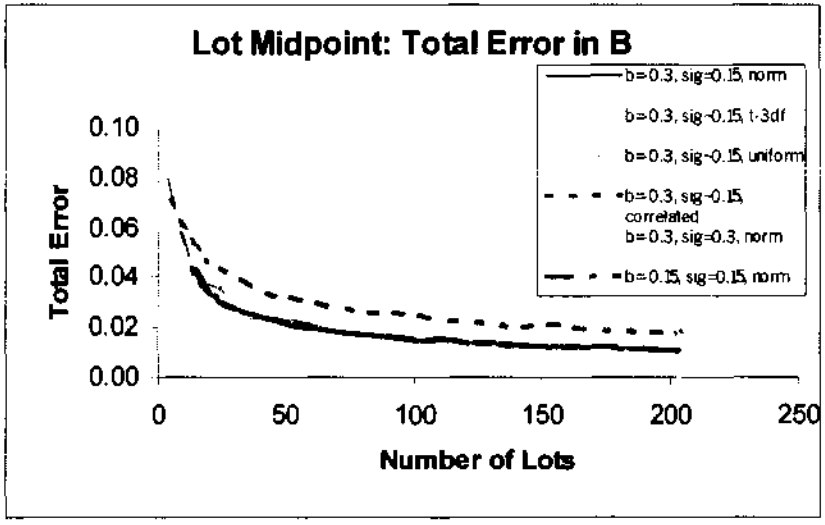
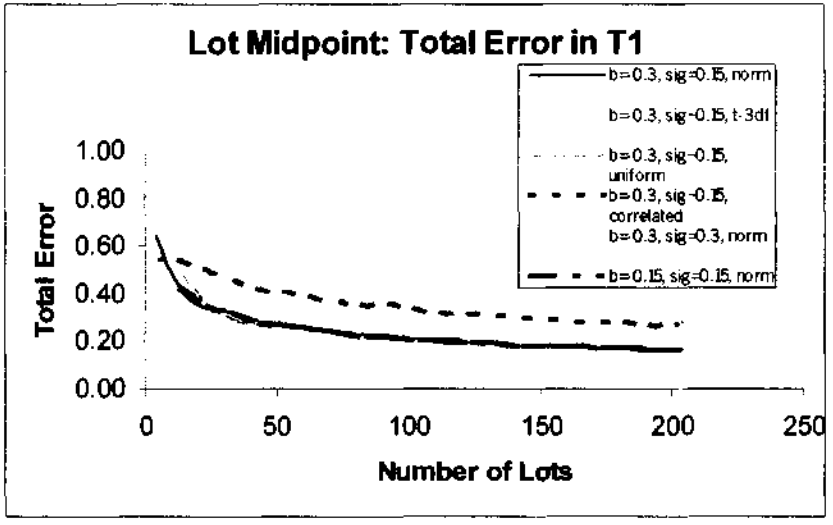


Figure 5.54. Robustness of NLS





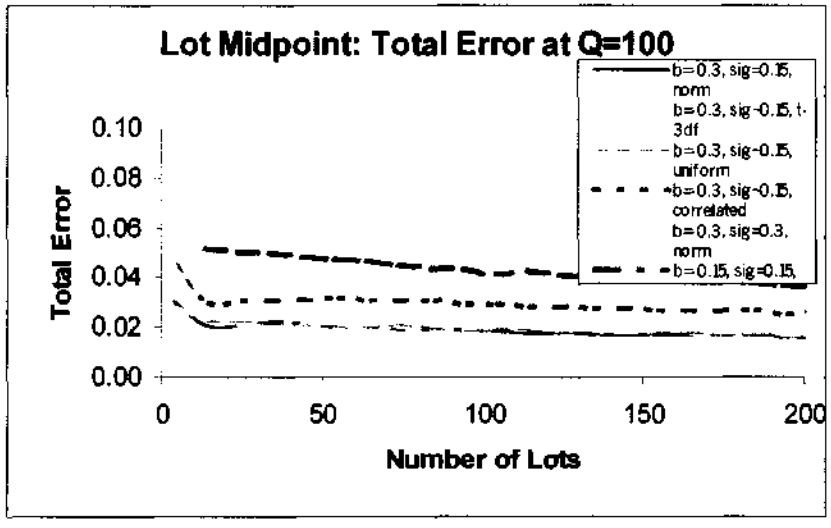
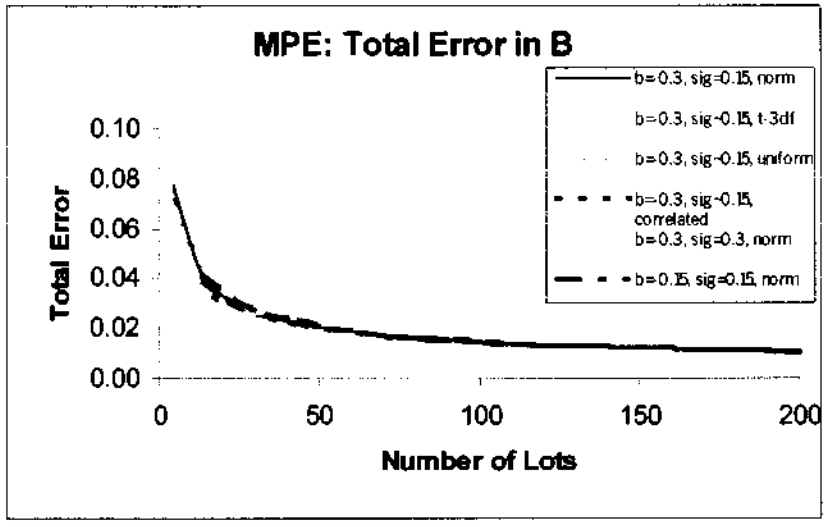
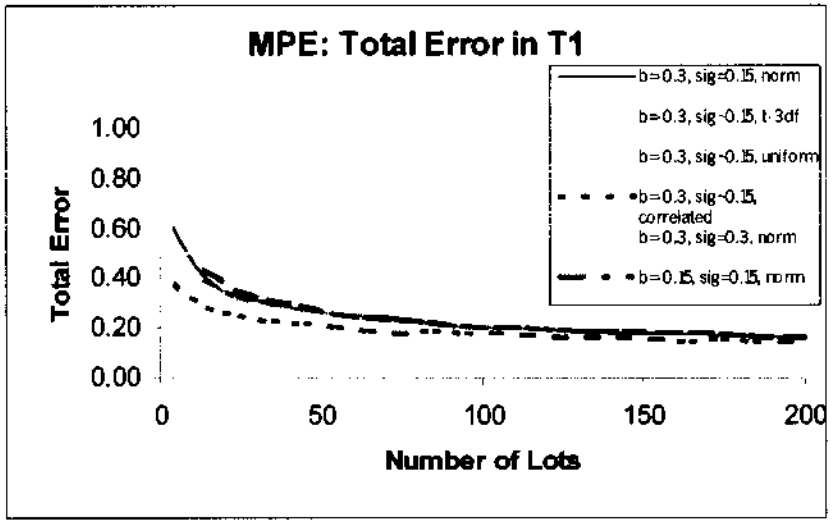


Figure 5.55. Robustness of Lot Midpoint Iteration





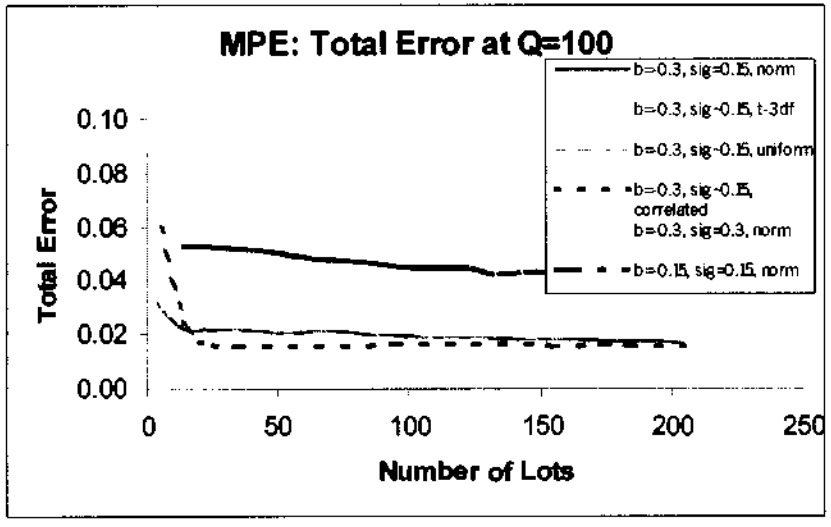


Figure 5.56. Robustness of MPE

APPENDIX: CONVERGENCE THEORY FOR LOT-MIDPOINT ITERATION

In this appendix, we investigate the existence of a solution to lot-midpoint iteration, the uniqueness of any solution, and convergence to that solution.

The regression sum-of-squares is given by:

$$\sum_{i=1}^{n}\left(\ln(LAC_i) - \ln(T_1) - b\ln[\overline{Q}_i(b)]\right)^2 . \tag{A.1}$$

Begin with an initial estimate of b, denoted $b^{(0)}$. Fix $b=b^{(0)}$ in the definition of the lot midpoint, $\overline{Q}_i(b^{(0)})$, and minimize the regression sum-of-squares with respect to b as the *regression coefficient only*. The minimum occurs at a new estimate, $b^{(1)}$. Now fix $b=b^{(1)}$ in the definition of the lot midpoint $\overline{Q}_i(b^{(1)})$ and again minimize with respect to b as the regression coefficient only. In general, estimate the following sequence of regressions:

$$\ln(LAC_i) = \ln(T_1) + b^{(p+1)}\ln[\overline{Q}_i(b^{(p)})] + v_i , \tag{A.2}$$

for $p=0,1,2,\ldots$. Finally, the lot-midpoint estimator is defined as the limit of the sequence:

$$b_1 = \lim_{p\to\infty} b^{(p)} , \tag{A.3}$$

when the limit exists. In practice, the lot-midpoint estimator is taken where the sequence converges within a pre-specified numerical tolerance

It can be shown that lot-midpoint iteration is numerically distinct from lot-midpoint NLS. Lot-midpoint iteration does not minimize the regression sum-of-squares when the functional dependence of $\overline{Q}_i(b)$ on b is acknowledged. In fact, lot-midpoint iteration does not minimize *any* continuously differentiable function. Letting $a=\ln(T_1)$, at any iteration $p=0,1,2,\ldots$, the parameter estimates satisfy the two normal equations for *linear* least squares:

$$a: \quad 0 = \sum_{i=1}^{n}\left(\ln(LAC_i) - a^{(p+1)} - b^{(p+1)} \ln[\bar{Q}_i(b^{(p)})]\right), \tag{A.4}$$

and

$$b: \quad 0 = \sum_{i=1}^{n}\ln[\bar{Q}_i(b^{(p)})]\times\left(\ln(LAC_i) - a^{(p+1)} - b^{(p+1)} \ln[\bar{Q}_i(b^{(p)})]\right). \tag{A.5}$$

At convergence, however, the value of b used to define the lot midpoint ($b^{(p)}$) is identical to the OLS regression coefficient that multiplies the logarithm of the lot midpoint ($b^{(p+1)}$), or $b^{(p)}=b^{(p+1)}$. Thus, equations (A.4) and (A.5) reduce to:

$$a: \quad 0 = \sum_{i=1}^{n}\left(\ln(LAC_i) - a - b \ln[\bar{Q}_i(b)]\right), \tag{A.6}$$

and

$$b: \quad 0 = \sum_{i=1}^{n}\ln[\bar{Q}_i(b)]\times\left(\ln(LAC_i) - a - b \ln[\bar{Q}_i(b)]\right). \tag{A.7}$$

If the solution (a,b) represented an interior optimum of some continuously differentiable function on an open set, then the gradient of that function would vanish at (a,b). In fact, equations (A.6) and (A.7) would be precisely those gradient conditions. Thus, there would exist a parent objective function $F(a,b)$ such that the right-hand side of equation (A.6) equals $F_a(a,b)$, and the right-hand side of equation (A.7) equals $F_b(a,b)$. Because a continuously differentiable function has a symmetric Hessian matrix, existence of a parent function would further require that the cross-partial derivatives be equal.[54] However, the partial derivative of equation (A.6) with respect to b is equal to:

$$\partial F_a(a,b)/\partial b: \quad -\sum_{i=1}^{n}\left(\ln[\bar{Q}_i(b)] + b\frac{\partial \ln[\bar{Q}_i(b)]}{\partial b}\right), \tag{A.8}$$

and the partial derivative of equation (A.7) with respect to a is equal to:

---

[54] This is the exactness condition for differential forms; it is both necessary and sufficient. See Kaplan (1958, pp. 44–48).

$$\partial F_b(a,b)/\partial a: \quad -\sum_{i=1}^{n} \ln[\bar{Q}_i(b)]. \tag{A.9}$$

These two expressions cannot generally be equal because the location of the lot midpoint depends on the value of b. Thus, equations (A.6) and (A.7) *cannot* be integrated back to a parent objection function, $F(a,b)$.[55]

It is not currently known from theory whether a value of b always exists that balances equation (A.2); whether such a value, if it exists, is always unique; or whether lot-midpoint iteration is guaranteed to converge to such a value. The situation would be particularly problematic if there were multiple, distinct values of b that balance equation (A.2). In a maximization problem, we can always compare the value of the objective function at two distinct local maxima, disposing of the smaller value because it cannot be the global maximum. But because lot-midpoint iteration does not maximize any continuous objective function, we have no basis to choose between two distinct values of b that both balance equation (A.2).

We demonstrate that the existence, uniqueness, and convergence of lot-midpoint iteration depend upon the slopes of certain functions being less than 1.0 in absolute value. Before examining this condition more formally, we present a simple example to illustrate the problem in two dimensions.

Consider the following two functions: $f(y) = y - y^3$ and $g(y) = (1+\varepsilon)\times\sin(y)$, where we restrict our attention to the interval $0.0 \le y \le 0.9$. If we choose $\varepsilon = \left[(\pi/4)/\sin(\pi/4)\right] - 1 \approx 0.1107$, then the function $g(y)$ has a fixed point at $\pi/4$ (≈ 0.7854): $g(\pi/4) = \pi/4$. The fixed point is illustrated by the intersection of $g(y)$ with the 45–degree line in Figure A.1. The function $g(y)$ actually intersects the 45–degree line twice for non-negative values of y. The slope $g'(\pi/4) = \pi/4 \approx 0.7854$ at the fixed point already identified. In addition, there is a second fixed point at the origin, $g(0) = 0$ and $g'(0) = 1+\varepsilon \approx 1.1107$. Finally, the function $f(y)$ has a single fixed point at the origin, $f(0) = 0$ and $f'(0) = 1.0$.

---

[55] In particular, lot-midpoint iteration does not maximize the likelihood function for any continuous probability density. Despite a superficial resemblance, lot-midpoint iteration is *not* an example of an EM algorithm because the latter always converges to a stationary point (local or global maximum, or saddle point) of the likelihood function. On the latter property of the EM algorithm, see McLachlan and Krishnan (1977), especially chapter 3.

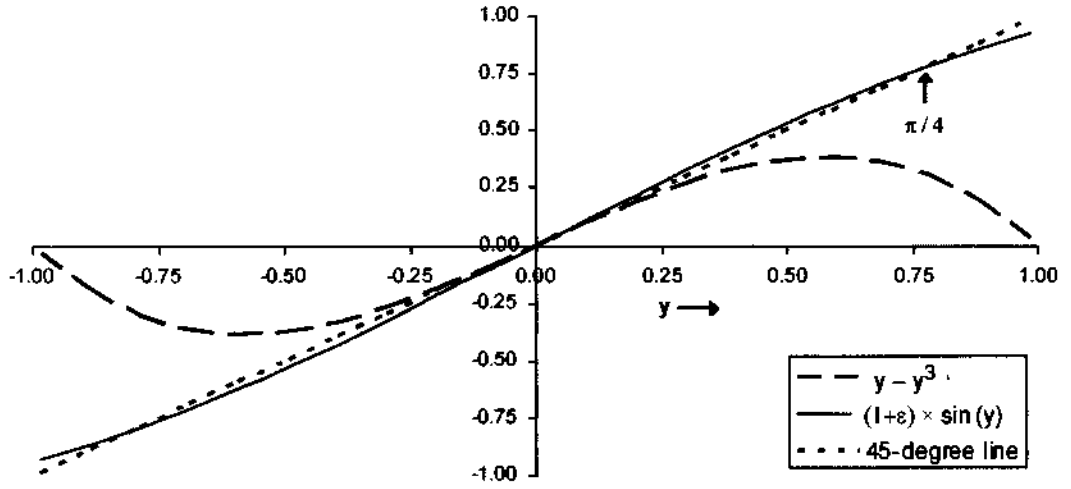


Figure A.1. Fixed Points of Cubic and Trigonometric Functions

Now consider an iterative scheme such as $y^{(p+1)} = f\left(y^{(p)}\right)$ for $p = 0,1,2,\ldots$. This scheme will converge to the fixed point at the origin, albeit slowly, even for starting values $y^{(0)}$ lying to the right of the peak of the cubic (which occurs at $y = 1/\sqrt{3} \approx 0.5774$). However, let us turn to the less well-behaved trigonometric function $g(y)$ and consider the iterative scheme $y^{(p+1)} = g\left(y^{(p)}\right)$. This scheme will converge to the fixed point at $\pi/4$ [where $g'(\pi/4) < 1$] from any starting value $0 < y^{(0)} < \pi$; it will *never* converge to the second fixed point at the origin [where $g'(0) > 1$] from any such starting value. Figure A.2 illustrates the convergence to $\pi/4$ from a starting value of 0.9, as well as from a starting value of 0.1 which is much closer to the origin.

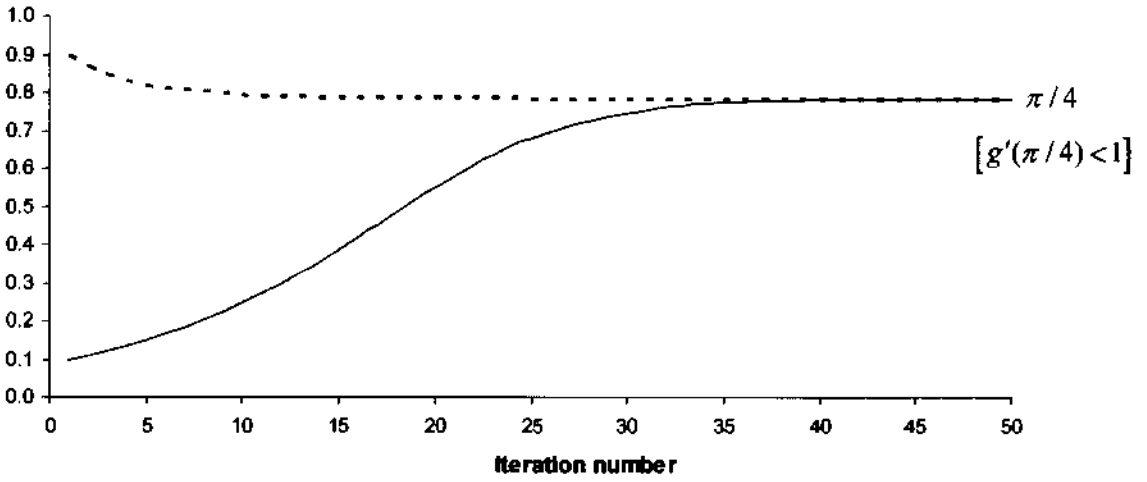


Figure A.2. Convergence of Iteration on Trigonometric Function

It appears from this example that fixed points tend to attract (repel) iterative schemes if the absolute value of the slope is less than (greater than) 1.0. Although this basic conclusion is sound, it will not carry over exactly to higher-dimensional problems. We show that a bounded gradient (the multi-dimensional extension of the concept of a bounded slope), while sufficient for convergence, is not actually necessary. Indeed, we present an example of a lot-midpoint iteration that converges despite having an absolute gradient slightly greater than 1.0 at a starting value near the (apparently) unique fixed point.

These issues of existence, uniqueness, and convergence may be explored more formally using the advanced mathematics of contraction mappings. Lot-midpoint iteration induces a mapping from the current estimates, $T_1^{(p)}$ and $b^{(p)}$, to the new estimates, $T_1^{(p+1)}$ and $b^{(p+1)}$. Consider the 2×2 Jacobian matrix of that mapping:

$$J = \begin{pmatrix} \partial T_1^{(p+1)} / \partial T_1^{(p)} & \partial T_1^{(p+1)} / \partial b^{(p)} \\ \partial b^{(p+1)} / \partial T_1^{(p)} & \partial b^{(p+1)} / \partial b^{(p)} \end{pmatrix}. \tag{A.10}$$

By Ostrowski's theorem on contraction mappings, if the eigenvalues of J are all less than 1.0 in absolute value throughout a region of parameter space (or, equivalently, if the maximum absolute eigenvalue is less than 1.0 throughout the region), then quite remarkably:

- there exists a pair of values T_1 and b in the region that balance equation (A.2);
- the pair T_1 and b is unique in the region; and
- iteration, starting from any point in the region, generates a sequence that converges to the unique root.[56]

In our situation, the Jacobian will reduce to a 1×1 matrix (i.e., a scalar) because the definition of the lot midpoint (equation (2.17)) depends on b but not T_1. Thus, we need only consider the absolute value of the derivative $\partial b^{(p+1)} / \partial b^{(p)}$. A change in $b^{(p)}$ affects the lot midpoints via equation (2.17), in turn affecting the updated estimate $b^{(p+1)}$ via the regression normal equations. We now show that, by theory alone, the absolute derivative *cannot* be bounded above by 1.0. In Chapter 4 we gave a numerical example in which the absolute derivative actually exceeds 1.0, yet lot-midpoint iteration nonetheless

---

[56] See Ortega and Rheinboldt (1970), theorems 5.1.3, 10.1.3, and 12.1.2. These theorems require that the iteration map a closed parameter set into itself.

converges. We obtain convergence in that example because the eigenvalue condition is sufficient for convergence, but not necessary. Importantly, however, with failure of the eigenvalue condition there is no theoretical guarantee that, even when lot-midpoint iteration converges, the root is unique. Thus, alternative starting values could conceivably lead the algorithm to converge to a different root. Again, because lot-midpoint iteration does not maximize any continuous objective function, if two distinct solutions are located we have no basis to choose between them.

In simple linear regression, of which any step of equation (A.2) is an example, the slope is given by:

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \tag{A.11}$$

and the intercept by:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}. \tag{A.12}$$

In this general regression notation, the Jacobian matrix becomes:

$$J = \begin{pmatrix} \partial a^{(p+1)}/\partial a^{(p)} & \partial a^{(p+1)}/\partial b^{(p)} \\ \partial b^{(p+1)}/\partial a^{(p)} & \partial b^{(p+1)}/\partial b^{(p)} \end{pmatrix}. \tag{A.13}$$

Now $b^{(p)}$ acts on $a^{(p+1)}$ and $b^{(p+1)}$ via the definition of the lot midpoints, but $a^{(p)}$ (i.e., $T_1^{(p)}$) has no such effect. Moreover, it follows from equation (A.12) that $\partial a^{(p+1)}/\partial b^{(p)} = -\bar{x}\,\partial b^{(p+1)}/\partial b^{(p)}$. Thus, the Jacobian matrix simplifies to:

$$J = \begin{pmatrix} 0 & -\bar{x}\,\partial b^{(p+1)}/\partial b^{(p)} \\ 0 & \partial b^{(p+1)}/\partial b^{(p)} \end{pmatrix}. \tag{A.14}$$

The matrix J is asymmetric, but the eigenvalues are nonetheless defined. Because J is singular, one eigenvalue is zero. It is simple to show that the second eigenvalue is $\partial b^{(p+1)}/\partial b^{(p)}$ with corresponding eigenvector $(-\bar{x} \quad 1)$. Thus, as claimed earlier, the convergence condition amounts to showing that $\left|\partial b^{(p+1)}/\partial b^{(p)}\right| < 1$.[57]

A change in $b^{(p)}$ affects the logarithmic midpoint of every single lot $\left\{\ln[\overline{Q}_i(b)] \middle| i = 1,\ldots,n\right\}$, or in the current notation $\left\{x_i \middle| i = 1,\ldots,n\right\}$. Moreover, differentiation of equation (A.11) yields:[58]

$$\partial \hat{b}/\partial x_i = \frac{(y_i - \bar{y}) - 2\hat{b}(x_i - \bar{x})}{S_{xx}}. \tag{A.15}$$

Thus, we have:

$$\frac{\partial b^{(p+1)}}{\partial b^{(p)}} = \sum_{i=1}^{n} \frac{\partial b^{(p+1)}}{\partial x_i} \times \frac{\partial x_i}{\partial b^{(p)}} = \sum_{i=1}^{n} \left[(y_i - \bar{y}) - 2\hat{b}(x_i - \bar{x})\right] \times \frac{\partial x_i}{\partial b^{(p)}} \Big/ S_{xx}. \tag{A.16}$$

The derivative $\partial x_i/\partial b^{(p)}$ is, in principle, computable from the definition of the lot midpoint, equation (2.17). However, inserting this information into equation (A.16), it is not at all obvious that $\left|\partial b^{(p+1)}/\partial b^{(p)}\right|$ is bounded above by 1.0. In fact, in Chapter 4 we give a numerical example of an apparently well-conditioned problem (i.e., no obvious data anomalies) in which this expression exceeds 1.0 (although lot-midpoint iteration nonetheless converges). If it can be verified in a particular example that $\left|\partial b^{(p+1)}/\partial b^{(p)}\right| < 1$, then existence, uniqueness, and convergence are guaranteed by Ostrowski's theorem. However, our numerical counterexample proves that there can be no *universal* guarantee of existence, uniqueness, or convergence; the structure of the lot-midpoint problem does not automatically satisfy the condition $\left|\partial b^{(p+1)}/\partial b^{(p)}\right| < 1$.

---

57 Equivalently, one could substitute equation (A.12), $\ln(\hat{T}_1) = \hat{a} = \bar{y} - \hat{b}\bar{x}$, into equation (A.2), thereby eliminating the intercept from the problem and reducing the iteration to a univariate mapping from $b^{(p)}$ to $\partial b^{(p+1)}$. The derivative that we have been studying, $\left|\partial b^{(p+1)}/\partial b^{(p)}\right|$, is the slope of that mapping.

58 A similar result is found in the statistics of outliers; see Chatterjee and Hadi (1988, p. 151).

At this point, we have learned the following:

- The standard sufficient conditions that guarantee existence, uniqueness, and convergence may or may not hold in the lot-midpoint problem — there is no universal guarantee;
- Even if the sufficient conditions fail in a particular example, lot-midpoint iteration may nonetheless converge — the bounded-eigenvalue condition is sufficient for convergence, but not necessary.

It remains to reconcile the ability of lot-midpoint iteration to converge when $\left|\partial b^{(p+1)}/\partial b^{(p)}\right| > 1$ with the geometric intuition that we developed in Figures A.1 and A.2. In Figure A.1, the origin is an inflection point for both the cubic and trigonometric functions. An iterative scheme could converge to the inflection point of the cubic function because $f'(0)=1$, but is repelled from the inflection point of the trigonometric function because $g'(0)>1$. From this example, it may appear that a bounded gradient is necessary as well as sufficient for convergence (i.e., violation of the bound prevents convergence to a particular root).

This low-dimensional example is actually somewhat misleading because, for continuously differentiable functions of a single variable, the derivative has the same value whether approached from the left or from the right. By contrast, for continuously differentiable functions of several variables, a particular gradient element (i.e., the first-partial derivative with respect to one of the function arguments) generally depends upon *all* of the function arguments. Thus, although the gradient element will have the same value whether approached from the left or from the right (i.e., from the west or from the east), it may have a different value when approached from the north or the south, or from any other direction.

To illustrate these points, consider an iterative scheme designed to locate the fixed point of the pair of functions $f(T)$ and $g(T,b)$. The iterative scheme takes the form $T^{(p+1)}=f\left(T^{(p)}\right)$ and $b^{(p+1)}=g(T^{(p)},b^{(p)})$ for $p=0,1,2,\ldots$. We restrict our attention to the unit circle $T^2+b^2\le 1$. We assume that $f(T)$ has a fixed point at an infinitesimal positive value, $f(T^*)=T^*$ where $0<T^*\ll 1$. We also assume that $f(T)$ has a bounded gradient $|f'(T)|<1$ for all $|T|<1$.

We assume that $g(T,b)$ has the following form:

$$g(T,b) \;=\; (3b/4) + (3b/2\pi)\times\arctan(b/T) - (3T/4\pi)\times\ln\left[\left(T^2+b^2\right)/T^2\right], \qquad \text{(A.17)}$$

with partial derivative:

$$\partial g(T,b)/\partial b \;=\; 0.75 \;+\; (1.5/\pi)\times \arctan(b/T). \tag{A.18}$$

We note that $g(T^*,0)=0$, so that the pair of functions $f(T)$ and $g(T,b)$ has a fixed point at $(T,b)=(T^*,0)$. Also, the Jacobian matrix has the form:

$$J \;=\; \begin{pmatrix} f'(T) & 0 \\ \partial g(T,b)/\partial T & \partial g(T,b)/\partial b \end{pmatrix}. \tag{A.19}$$

The eigenvalues of J are precisely its diagonal elements. Moreover, given our assumption that $|f'(T)|<1$, any difficulties with the maximum eigenvalue are entirely confined to the southeast diagonal element, $\partial g(T,b)/\partial b \;=\; 0.75 \;+\; (1.5/\pi)\times \arctan(b/T)$.

Figure A.3 depicts the gradient element $0.75 \;+\; (1.5/\pi)\times \arctan(b/T)$ along the unit circle; the gradient values at selected points along the unit circle are labeled 0.7500, 0.9375, 1.0000, and so on. Because the gradient element depends on T and b only through their ratio, the gradient element is constant along any diameter of the unit circle. The function $g(T,b)$ is increasing in b throughout the unit circle, $\partial g(T,b)/\partial b \geq 0$ (with strict inequality $\partial g(T,b)/\partial b > 0$ almost everywhere), so the fixed point of the pair of functions at $(T,b)=(T^*,0)$ is neither a maximum nor a minimum of $g(T,b)$. It might appear that convergence to the fixed point will occur only from starting points within the unit circle where the gradient element $\partial g(T,b)/\partial b$ is less than 1.0 (i.e., the two lighter-shaded sub-regions). However, this supposition will prove false because the gradient element may change in magnitude along a particular iterative path.

We make the example a bit more concrete by choosing the function $f(T)=T^*$ for all $|T|<1$, where $0<T^*\ll 1$. This function has the desired property $|f'(T)|<1$ (in fact, $f'(T)=0$) for all $|T|<1$, as well as a fixed point at $T=T^*$. Note that the iterative scheme $T^{(p+1)}=f\left(T^{(p)}\right)$ converges to the fixed point in a single iteration from any starting value $|T^{(0)}|<1$. We further definitize the example by choosing $T^*=10^{-4}$. Regarding the second of the pair of functions, the iteration $b^{(p+1)}=g(T^{(p)},b^{(p)})$ reduces to $b^{(p+1)}=g(T^*,b^{(p)})$ for $p=1,2,3,\ldots$, effectively a univariate iteration along the vertical line $T=T^*$.

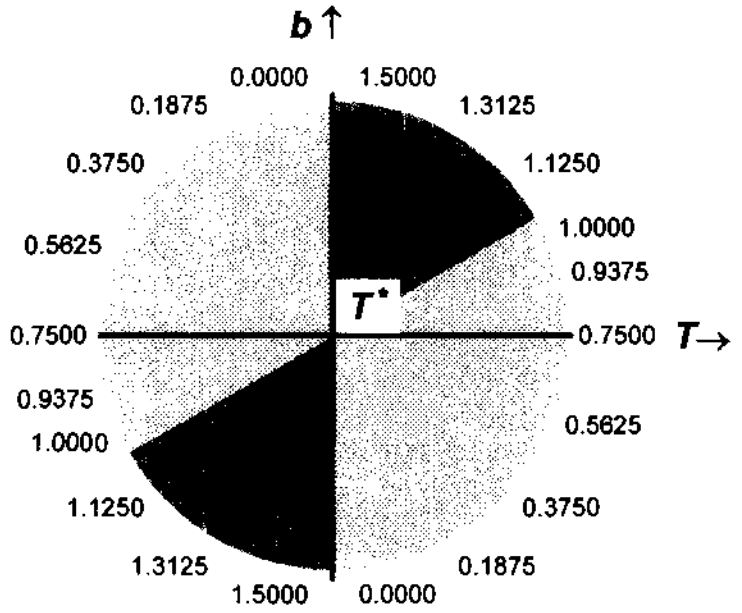


Figure A.3. Gradient Values Along the Unit Circle

Figure A.4 illustrates one type of divergence. The iterative scheme begins in the lighter-shaded sub-region in the northwest, where the gradient element $\partial g(T,b)/\partial b$ equals 0.3750. However, the first iteration jumps to the point $(T,b)=(T^*,0.316)$ at which $\partial g(T,b)/\partial b$ equals 1.4998. From that point the iteration diverges northward, exiting the unit circle by the fourth iteration. The gradient element $\partial g(T,b)/\partial b$ remains approximately equal to 1.5 as the iteration diverges.

Figure A.5 illustrates one type of convergence. The iterative scheme begins in the darker-shaded sub-region in the southwest, where the gradient element $\partial g(T,b)/\partial b$ equals 1.1250. Nonetheless, the first iteration jumps to the point $(T,b)=(T^*,-0.745)$ at which $\partial g(T,b)/\partial b$ equals 0.0001. From that point the iteration converges to the fixed point, effectively reaching it by the fourth iteration. The gradient element $\partial g(T,b)/\partial b$ increases along the convergence path as the angle $\theta=\arctan(b/T^*)$ sweeps counterclockwise from $1.50004\times\pi$ (once reaching $T=T^*$ after the first iteration) to 2π, but remains bounded above by 0.75 (the value along the equator of the circle) and never again exceeds 1.0.

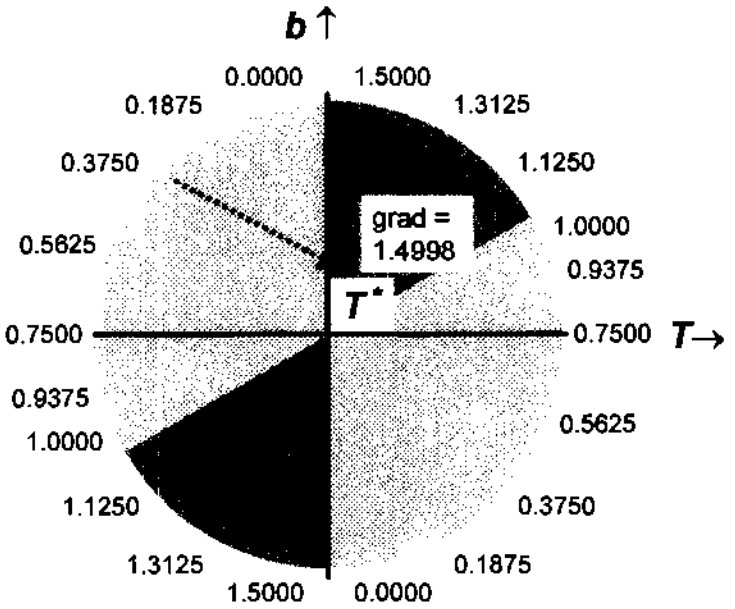


Figure A.4. Example of Divergent Iteration

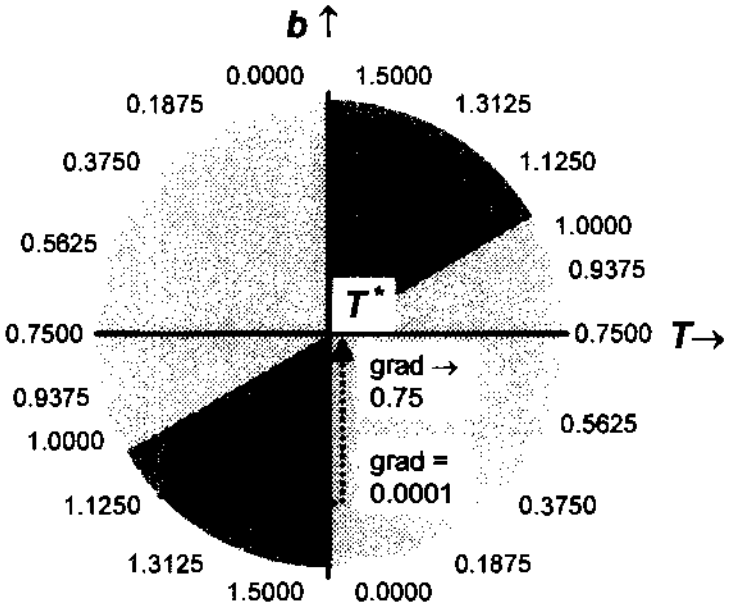


Figure A.5. Example of Convergent Iteration

It turns out that, for this problem, the iteration converges from any starting point on the unit circle in the southern hemisphere (including the equator), $\pi \leq \theta \leq 2\pi$. The iteration diverges from any starting point on the unit circle in the northern hemisphere, $0 < \theta < \pi$. The important lesson is that the value of the maximum eigenvalue at the starting point (i.e., the shading of the sub-region) does not necessarily predict the convergence or divergence of the iterative scheme. In particular, the second example (Figure A.5) illustrates that the iterative scheme may converge to the fixed point by simply jumping over the sub-region in which the maximum eigenvalue exceeds 1.0 in absolute value. The condition of bounded eigenvalues throughout an entire region, although sufficient for convergence, is far from necessary.

REFERENCES

Asher, Harold, "Cost-Quantity Relationships in the Airframe Industry," RAND Corporation, Report 291, July 1956.

Balut, Stephen, Thomas Gulledge, and N. Keith Womer, "A Method of Repricing Aircraft Procurement," *Operations Research*, vol. 37, 1989, pp. 255–265.

Bard, Yonathan, *Nonlinear Parameter Estimation*, New York: Academic Press, 1974.

Book, Stephen A., and N. Y. Lao, "Deriving Minimum-Percentage-Error CERs under Zero-Bias Constraints" (briefing slides), The Aerospace Corporation, June 1996.

Book, Stephen A., and Philip H. Young, "General-Error Regression for USCM–7 CER Development" (briefing slides), The Aerospace Corporation, July 1995.

Book, Stephen A., and Philip H. Young, "General-Error Regression for Deriving Cost-Estimating Relationships," *Journal of Cost Analysis*, Fall 1997, pp. 1–28.

Bradley, E. L., "The Equivalence of Maximum Likelihood and Weighted Least Squares Estimates in the Exponential Family," *Journal of the American Statistical Association*, vol. 68, 1973, pp. 199–200.

Bradu, Dan, and Yair Mundlak, "Estimation in Lognormal Linear Models," *Journal of the American Statistical Association*, vol. 65, no. 329, March 1970, pp. 198–211.

Camm, Jeffrey D., James R. Evans, and N. Keith Womer, "The Unit Learning Curve Approximation of Total Cost," *Computers and Industrial Engineering*, vol. 12, June 1987, pp. 205–213.

Charnes, A., E. L. Frome, and P. L. Yu, "The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family," *Journal of the American Statistical Association*, vol. 71, 1976, pp. 169–171.

Chatterjee, Samprit, and Ali S. Hadi, *Sensitivity Analysis in Linear Regression*, New York: John Wiley, 1988.

Conway, R. W., and Andrew Schultz, Jr., "The Manufacturing Progress Function," *Journal of Industrial Engineering*, vol. 10, January 1959, pp. 39–54.

Davison, A.C., and D.V. Hinkley, *Bootstrap Methods and Their Application*, New York: Cambridge University Press, 1997.

Dennis, John E., and Robert B. Schnabel, *Numerical Methods for Unconstrained Maximization and Nonlinear Equations*, Philadelphia: SIAM, 1996.

Dhrymes, Phoebus J., *Econometrics: Statistical Foundations and Applications*, New York: Springer-Verlag, 1974.

Donaldson, Janet R., and Robert B. Schnabel, "Computational Experience with Confidence Regions and Confidence Intervals for Nonlinear Least Squares," *Technometrics*, vol. 29, no. 1, February 1987, pp. 67–82.

Duan, Naihua, "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, vol. 78, no. 383, September 1983, pp. 605–610.

Duan, Naihua, Willard G. Manning, Carl N. Morris, and Joseph P. Newhouse, "A Comparison of Alternative Models for the Demand for Medical Care," RAND Corporation, R-2754-HHS, January 1982.

Dutton, J. M., A. Thomas, and J. E. Butler, "The History of Progress Functions as a Management Technology," *Business History Review*, vol. 58, 1984, pp. 204–233.

Efron, Bradley and Robert J. Tibshirani, *An Introduction to the Bootstrap*, New York: Chapman and Hall, 1993.

Eskew, Henry L., "Learning Curves, Lot Midpoints, and Statistical Estimation: An Expository Note," CNA Corporation, unpublished notes, 2000.

Eskew, Henry L., and Kletus S. Lawler, "A Study of the Predictive Accuracy of Alternatively Estimated Statistical Cost Models," CNA Corporation, CRM 93-228, December 1993.

Eskew, Henry L., and Kletus S. Lawler, "Correct and Incorrect Specifications in Statistical Cost Models," *Journal of Cost Analysis*, Spring 1994, pp. 105–124.

Feller, William, *An Introduction to Probability Theory and Its Applications* (Volume II), New York: John Wiley, 1971.

Gallant, A. Ronald, *Nonlinear Statistical Models*, New York: John Wiley, 1987.

Gallant, A. Ronald, and J. J. Goebel, "Nonlinear Regression with Autocorrelated Errors," *Journal of the American Statistical Association*, vol. 71, 1976, pp. 961–967.

Goldberger, Arthur S., *Econometric Theory*, New York: John Wiley, 1964.

Goldberger, Arthur S., "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica*, vol. 36, 1968, pp. 464–472.

Goldfeld, Stephen M., and Richard E. Quandt, *Nonlinear Methods in Econometrics*, Amsterdam: North Holland, 1972.

Green, P. J., "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and Some Robust and Resistant Alternatives," *Journal of the Royal Statistical Society, Series B*, vol. 46, 1984, pp. 149–192.

Harmon, Bruce, Anduin Touw, and James Woolsey, "Estimating an Augmented Cost Progress Function for Tactical Aircraft" (briefing slides), Institute for Defense Analyses, January 2000.

Heien, Dale M., "A Note on Log-Linear Regression," *Journal of the American Statistical Association*, vol. 63, no. 323, September 1968, pp. 1034–1038.

Henderson, James M., and Richard E. Quandt, *Microeconomic Theory: A Mathematical Approach*, third edition, New York: McGraw Hill, 1980.

Jennrich, R. I., and R. H. Moore, "Maximum Likelihood Estimation by Means of Non-Linear Least Squares," *American Statistical Association, Proceedings of the Statistical Computing Section*, 1975, pp. 57–65.

Johnson, Norman L., and Samuel Kotz, *Distributions in Statistics: Continuous Univariate Distributions* (Volumes I and II), New York: John Wiley, 1970.

Johnston, John, *Econometric Methods*, New York: McGraw Hill, 1972.

Jorgensen, B., "Maximum Likelihood Estimation and Large-Sample Inference for Generalized Linear and Nonlinear Regression Models," *Biometrika*, vol. 70, 1983, pp. 19–28.

Jorgensen, B., "The Delta Algorithm and GLIM," *International Statistical Review*, vol. 52, 1984, pp. 283–300.

Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee, *The Theory and Practice of Econometrics*, New York: John Wiley, 1985.

Kadiyala, K. R., "A Transformation Used to Circumvent the Problem of Autocorrelation," *Econometrica*, vol. 36, 1968, pp. 93–96.

Kaplan, Wilfred, *Ordinary Differential Equations*, Reading, Massachusetts: Addison-Wesley, 1958.

Large, Joseph P. et al., "Production Rate and Production Cost," RAND Corporation, R-1609-PA&E, December 1974.

Lee, David A., *The Cost Analyst's Companion*, Logistics Management Institute, 1997.

Loerch, Andrew G., "Incorporating Learning Curve Costs in Acquisition Strategy Optimization," *Naval Research Logistics*, vol. 46, no. 3, April 1999, pp. 255–271.

Lurie, Philip M., Matthew S. Goldberg, and Mitchell S. Robinson, "A Handbook of Cost-Risk Analysis Methods," Institute for Defense Analyses, P-2734, April 1993.

McCullagh, P., and J. A. Nelder, *Generalized Linear Models*, London: Chapman and Hall, 1989.

McLachlan, Geoffrey J., and Thriyambakam Krishnan, *The EM Algorithm and Extensions*, New York: John Wiley, 1997.

Miller, D. M., "Reducing Transformation Bias in Curve Fitting," *American Statistician*, vol. 38, 1984, pp. 124–126.

Ortega, J. M. and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, New York: Academic Press, 1970.

RAND Corporation, "Military Equipment Cost Analysis, A Report Prepared for the Office of the Secretary of Defense (Systems Analysis)," 1971.

Rao, C. R., *Linear Statistical Inference and Its Applications*, second edition, New York: John Wiley, 1973.

Schmidt, Peter, *Econometrics*, New York: Marcel Dekker, 1976.

Seber, G. A. F. and C. J. Wild, *Nonlinear Regression*, New York: John Wiley, 1989.

Seiler, F. A., "Error Propagation for Large Errors," *Risk Analysis*, vol. 7, no. 4, 1987.

Sewell, W. P., "Least-Squares, Conditional Predictions and Estimator Properties," *Econometrica*, vol. 37, 1969, pp. 39–43.

Taylor, Angus E., and W. Robert Mann, *Advanced Calculus*, Lexington, Mass.: Xerox College Publishing, 1972.

Varian, Hal R., *Microeconomic Analysis*, third edition, New York: Norton, 1992.

Weir, J. B. de V., "Standardized *t*," *Nature*, vol. 185, 1960.

White, Halbert, "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, vol. 48, 1980, pp. 817–838.

Womer, N. Keith, "Learning Curves, Production Rate, and Program Costs," *Management Science*, vol. 25, no. 4, April 1979, pp. 312–319.

Womer, N. Keith, "Estimating Learning Curves from Aggregate Monthly Data," *Management Science*, vol. 30, no. 8, August 1984, pp. 982–992.

Womer, N. Keith, and J. Wayne Patterson, "Estimation and Testing of Learning Curves," *Journal of Business and Economic Statistics*, vol. 1, no. 4, October 1983, pp. 265–272.

Yelle, L. E., "The Learning Curve: Historical Review and Comprehensive Survey," *Decision Science*, vol. 10, April 1979, pp. 302–328.

Young, Philip H., "The Meaning of the Standard Error of the Estimate in Logarithmic Space," *Journal of Cost Analysis and Management*, Winter 1999, pp. 59–65.

AUTHOR INDEX

SUBJECT INDEX

03/12/2003
CIM D0006870.A3 1Rev