1990

Biased Estimation in the Context of the Hedonic Pricing Model for Housing.

John Ross Knight
Louisiana State University and Agricultural & Mechanical College

Order Number 9112242

Biased estimation in the context of the hedonic pricing model for housing

Knight, John Ross, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

BIASED ESTIMATION IN THE CONTEXT OF THE HEDONIC PRICING MODEL FOR HOUSING

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Interdepartmental Program in Business Administration

by
John R. Knight
B.A., Tulane University of Louisiana, 1969
M.B.A., Louisiana State University, 1978
August 1990

ACKNOWLEDGEMENTS

I am grateful to a number of people who have prepared me to write this dissertation and who have provided help, encouragement and guidance along the way. I am especially indebted to my major and minor professors; C. F. Sirmans for his interest in the topic and his helpful suggestions, and R. Carter Hill for four semesters of econometrics and his generous contribution of time. Professor Hill allowed me to use his programs for many of the estimators evaluated in this study. I modified these for the purpose of the dissertation, and any programming errors are mine.

I also wish to thank the members of the committee; Michael Irwin, James Shilling, John Glascock and John Howe. Glenn Boyle was also very generous to me with his time and ideas.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF TABLES v

ABSTRACT ix

Chapter

1. INTRODUCTION 1

The Economic Problem

The Econometric Problem

An Application to Housing

Overview

2. A COMPARISON OF ESTIMATORS USING TIME SERIES AND CROSS-SECTIONAL DATA 11

Introduction

Review of the Literature

The Statistical Model

Multicollinearity and Data Extrapolation

Biased Estimators

The Data and the Models

Results

Conclusions

Chapter

3. COMPARING ESTIMATORS OF HEDONIC MODELS: A MONTE CARLO EXPERIMENT 61

Introduction

Monte Carlo Experiments

Methodology

Monte Carlo Results

4. SUMMARY AND CONCLUSIONS 98

REFERENCES 105

Appendix

A. PREDICTION WITH THE HEDONIC PRICING MODEL: THE EFFECT OF SIMULTANEITY 112

B. MULTICOLLINEARITY 116

C. SUPPLEMENTAL TABLES 119

D. SAS PROGRAM FOR THE MONTE CARLO EXPERIMENT . . . 152

LIST OF TABLES

Table		Page
1.	Summary of Compared Estimators	41
2.	Hedonic Model for Baton Rouge Data	43
3.	Hedonic Model for American Housing Survey Data	44
4.	Summary Statistics, Baton Rouge	45
5.	Ordinary Least Squares Parameter Estimates, Baton Rouge, Semilog Form	46
6.	Ordinary Least Squares Parameter Estimates, Baton Rouge, Linear Form	47
7.	Summary of Prediction Performance, Baton Rouge, Semilog Form	48
8.	Summary of Prediction Performance, Baton Rouge, Linear Form	49
9.	Parameter Estimates of the Compared Estimators, Baton Rouge, Semilog Form, Period 5	50
10.	Summary Statistics, American Housing Survey	51
11.	Ordinary Least Squares Parameter Estimates, American Housing Survey, Semilog Form	53
12.	Ordinary Least Squares Parameter Estimates, American Housing Survey, Linear Form	55
13.	Summary of Prediction Performance, American Housing Survey, Semilog Form	57
14.	Summary of Prediction Performance, American Housing Survey, Linear Form	58

Table		Page
15.	Parameter Estimates of the Compared Estimators, American Housing Survey, Boston	59
16.	Specification of the "True" Parameter Vector American Housing Survey, Monte Carlo Experiment	85
17.	Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error of In-Sample Prediction, First Subsample	86
18.	Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error, First Subsample	87
19.	Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Ensemble Mean Square Error, First Subsample . . .	88
20.	Parameter Estimates of the Compared Estimators, Baton Rouge, Period 1	121
21.	Parameter Estimates of the Compared Estimators, Baton Rouge, Period 2	122
22.	Parameter Estimates of the Compared Estimators, Baton Rouge, Period 3	123
23.	Parameter Estimates of the Compared Estimators, Baton Rouge, Period 4	124
24.	Parameter Estimates of the Compared Estimators, American Housing Survey, Chicago	125
25.	Parameter Estimates of the Compared Estimators, American Housing Survey, Dallas	127
26.	Parameter Estimates of the Compared Estimators, American Housing Survey, Detroit	129

Table Page

27. Parameter Estimates of the Compared Estimators, American Housing Survey, New York 131

28. Parameter Estimates of the Compared Estimators, American Housing Survey, San Diego 133

29. Relative Performance of Compared Estimators Under Mean Absolute Percentage Error, First Subsample 135

30. Relative Performance of Compared Estimators Under Mean Absolute Percentage Error, Second Subsample 137

31. Relative Performance of Compared Estimators Under Mean Absolute Percentage Error, Third Subsample 139

32. Relative Performance of Compared Estimators Under Mean Absolute Percentage Error, Fourth Subsample 141

33. Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error of In-Sample Prediction, Second Subsample 143

34. Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error of In-Sample Prediction, Third Subsample 144

35. Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error of In-Sample Prediction, Fourth Subsample 145

36. Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error, Second Subsample 146

37. Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error, Third Subsample 147

38. Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Mean Square Error, Fourth Subsample 148

Table		Page

Table		Page
39.	Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Ensemble Mean Square Error, Second Subsample . . .	149
40.	Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Ensemble Mean Square Error, Third Subsample . . .	150
41.	Relative Performance of LINDLEY, STEINRLS, and EBAYES Estimators Under Ensemble Mean Square Error, Fourth Subsample . . .	151

ABSTRACT

This dissertation compares eight biased estimators as alternatives to Ordinary Least Squares estimation in the context of predicting residential real estate prices. It considers ridge rule estimation and principal components regressions, techniques that have previously been proposed for this application. It also introduces the use of Stein-like rules for predicting housing prices. The study examines relative performance of these estimators in three data settings and under four separate assumptions regarding loss criteria.

The first test of the estimators uses Multiple Listing Service (MLS) data for Baton Rouge, Louisiana between 1984 and 1989 to examine relative predictive effectiveness in a time series framework with highly descriptive data. Next, American Housing Survey (AHS) data for six metropolitan areas is employed to compare the estimators in a cross-sectional context with the type of data typically used to create housing price indexes. Finally, the AHS data is used as the basis for a Monte Carlo experiment that compares estimator performance in numerically simulated repeated samples.

The partitioned Stein-like estimators do well in all three data environments. Two of them provide especially

impressive performance. Under quadratic loss, in the Monte Carlo experiment, these estimators outperform all compared alternatives across the entire range of generated samples.

CHAPTER ONE

INTRODUCTION

I. The Economic Problem

In order for scarce resources to be allocated efficiently in an economy, the relative value of those resources must be accurately determined and quickly communicated to market participants. Value determination presents little problem in a market for homogeneous products, as the ability to observe their relative prices is sufficient for economic choices among these goods.

Valuation in a market for differentiated products is not so straightforward. While the price of the goods can still be observed, one must now be also concerned with the ways and the extent to which the products differ. Frequently with these heterogeneous goods, a large investment of time and other resources is involved in their production or purchase. Therefore, the ability to predict values is especially important for directing economic activity into or out of these markets.

Lancaster [1966] and Muth [1966] provide the early development of a model designed to deal with the pricing of heterogeneous goods. The motivation for these models is the notion that consumers value a product, not for the product

itself, but rather for the utility which they derive from various product characteristics. The explicit price of the product is a result of the quantities and implicit prices of the individual characteristics contained in the product. Rosen [1974] develops the theoretical market foundation for implicit pricing of attributes.

Hedonic pricing models are used in connection with heterogeneous goods in two ways: first, observations of the prices of these goods are used to infer the implicit prices of the attributes and to estimate the demand for these attributes; and second, hedonic prices of the attributes are used to predict the value of a heterogeneous product based on the extent to which the attributes are contained in the product. New insight into the second of these applications is the contribution of this dissertation.

The ability to predict residential housing values holds considerable practical importance for appraisers, prospective home buyers, and especially for tax assessors. In guiding the resource allocation decision, the relative ability of estimators to predict house prices is especially important, but has received very little attention. This dissertation fills this void in the hedonic pricing literature, providing a comprehensive evaluation of the predictive performance of a variety of estimators in both real and simulated settings.

II. The Econometric Problem

The hedonic pricing model can be expressed mathematically as

$$Y = X\beta + \epsilon$$

where X is a matrix of the quantities of utility-bearing attributes of the products, and β is the parameter vector of prices of these attributes implied by Y, the price vector of the composite product. ϵ is the vector of errors resulting from estimating the model.

There are several econometric difficulties encountered in implementing this model. Specification of the utility-providing attributes and measurement of these attributes are among the most important and most fundamental. To be properly specified, the model must include all characteristics that provide utility to the user of the product. When attributes are omitted that are also correlated with other included attributes, bias is introduced into the estimation. But with a complex product, it is difficult to know what utility-bearing characteristics the product provides.

Even in the event that the model is perfectly specified, there remains the problem of measuring the included attributes. The characteristics from which utility is derived are not directly observable, and can only be measured by proxies for those characteristics. Often, some of these proxies are highly correlated with each other, adding the problem of multicollinearity to the existing econometric difficulties.

It must also be remembered that the supply and demand for the differentiated products are being simultaneously determined in the market for those goods. Likewise, the supply and demand for the characteristics which comprise any single product are contemporaneously observed. Thus, the companion problems of simultaneity and identification are important econometric considerations in applying the hedonic pricing model.

All of these econometric problems are of considerable theoretical importance. However, the nature and extent of the problem depends on the orientation and application of the research. In this dissertation, the hedonic pricing model is employed to obtain predictions of prices of heterogeneous goods. In this context, the problems of specification and collinearity among attributes are especially troublesome. Simultaneity, of paramount concern in other applications of the hedonic model, is not considered problematic as the model is constructed and applied in this dissertation.[1]

III. An Application to Housing

Estimating the value of residential real estate is one of the primary applications of the hedonic pricing model for differentiated products. One of several differentiated

---

[1] Appendix A discusses the treatment of this issue in the hedonic pricing literature and explains why the problem is not considered a major concern in this present study.

products for which the model is appropriate[2], real estate comprises the largest percentage of world wealth and residential real estate must be considered by every economic unit that consumes housing services. Typically, home equity makes up a large share of the home owner's investment asset portfolio, and the flow of housing services provided by this asset occupies a predominant position in the individual's budget. Also, the production of residential housing is costly, both in terms of time and other scarce resources. This makes accurate prediction of house values especially important.

The literature has responded prolifically to the importance of housing as an application of the hedonic model.[3] Three distinct orientations have emerged. One segment of the literature focuses on model specification. These studies attempt to identify the attributes which influence house prices and to estimate the demand for those specific attributes. Location and structural characteristics are included as measured attributes in almost all hedonic models of the

---

[2] The model has also been used extensively in the transportation sector (Brown and Mendelsohn [1984], Smith and Kaoru [1987]), and in the construction of hedonic wage functions (Knieser and Leeth [1988], Biddle and Zarkin [1989]). Such diverse products as automobiles (Hartman [1987], Dinopoulos and Kreinin [1988]), alfalfa hay (Pardew [1988], and CAT (computed axial tomography) scanners (Trajtenberg [1989]) are priced with the hedonic model. Recently, it has been proposed as an alternative to the Capital Asset Pricing Model and the Arbitrage Pricing Model as the theoretical model for pricing financial assets (Ibbotson [1984]).

[3] Other real estate applications include office markets and shopping centers.

housing market, while many studies concentrate on neighborhood amenities and disamenities or attempt to determine the positive or negative valuation of a single attribute[4].

Another line of literature has been concerned with the econometric problems associated with the estimation of the parameters in the hedonic regression. The simultaneity problem has received early and continued treatment, and remains an important theoretical consideration in interpreting empirical results from hedonic pricing models.[5] In the studies that concentrate on a single attribute or on the precise estimation of individual parameters (See n. 4), the simultaneity issue is of greatest concern.

The third segment of the literature, and that most closely aligned with the present work, is oriented toward predicting house values from the hedonic model. Much of this applications oriented research has been aimed at dealing with the collinearity problem which commonly accompanies the use of a hedonic model. When the parameter estimates obtained from the model are used to predict housing prices

---

[4] Abelson [1979], Krumm [1980], and Linneman [1981] study the positive and negative features of a neighborhood which are valued or avoided by consumers. Examples of these characteristics are crime, and Thaler [1978]; pollution, Harrison and Rubinfeld [1978], Nelson [1978], and Ridker and Henning [1967]; zoning, Jud [1980]; noise, McMillan, Reid and Dilen [1980], and Miezskowski and Saper [1978]; and, shoreline, Brown and Pollakowski [1977].

[5] Follain and Jimenez [1985] discuss three distinct facets of simultaneity imbedded within the hedonic pricing model. Discussions of the issue and proposals for dealing with the problem are offered by Rosen [1974], Brown and Rosen [1982], Bartik [1987], Dale-Johnson [1982] and Epple [1987].

from another sample, the imprecise estimates which result from highly correlated characteristics provide unstable predictions. Biased estimation techniques, which trade bias in the parameter estimates for a reduction in the variance with which the parameters are estimated, have been proposed as a logical remedy.

IV. Overview

This dissertation accomplishes three things: First, it fills the empirical gap left by Varian [1974], who recognizes the potential benefits of improved prediction in the context of residential real estate; second, it is the first comprehensive performance evaluation of the methods available which use hedonic pricing models to predict residential housing prices; and third, the dissertation introduces the application of Stein-like rules for incorporating non-sample information into the problem of predicting house prices.

This dissertation focuses on the prediction of residential housing prices using alternative econometric models and estimators and on predictor performance comparison as measured by several alternative loss functions. It is composed of two essays that are devoted to two separate approaches to evaluating predictor performance. The first involves the use of "real" data from two separate sources. The second provides a controlled environment for studying the comparative estimator performance, using the Monte Carlo technique to simulate housing data.

Essay One.

The first essay uses two distinct data sets to compare the performance of two alternative econometric models and nine alternative estimators as measured by mean square error of prediction. This essay describes the models and the estimators that are used throughout the dissertation as the subjects of comparison. Also, the analysis of the data and the estimators gives the reader an appreciation for the econometric difficulties commonly encountered in the real estate setting and an appreciation for the potential importance of this work.

Another area receives special attention in this first essay. It is the careful analysis of the pattern of multicollinearity in the samples used for estimation and the samples held out for comparing prediction performance. Belsley, Kuh and Welsh [1980] propose the use of condition numbers and indices for establishing the presence and measuring the severity of collinearity in data. Their method will be explained and used in the analysis.

The results of essay one are mixed. While the relative performance among the estimators is consistent for the two data sets, no clear choice among the estimators emerges from these initial tests. Two of the Stein-like rules introduced here do, however, appear to perform generally better.

Essay Two.

The first essay provides a means of evaluating the various biased estimation techniques under "real" conditions. However, it is difficult to make inferences about the relative performance of the predictors because a) the "true" parameter values are unknown so the loss measures may be misleading, and b) the performances are sensitive to the manner in which the prediction data from which predictions are made changes with respect to the data from which parameter estimates are obtained. The second essay is aimed at reducing the performance uncertainty occasioned by (a) and (b).

Essay two evaluates the same estimators described in essay one but employs Monte Carlo simulations to systematically make changes in the length of the parameter vector being estimated. Observing predictor performance under these controlled conditions permits some intuition about the conditions under which some estimators are superior to others. Also, the impact of specification changes on predictive performance can be assessed.

While still unknown, within the context of the Monte Carlo experiment, the "truth" can be specified. This allows for evaluation of the forecasting power of the estimators under four different loss functions. These are a) mean square error, b) mean square error of in-sample prediction, and c) mean square error of out-of-sample prediction, and d)

mean absolute percentage error. One other loss function, "ensemble" mean square error, focuses on estimator performance with respect to a small group of variables within the model.

It is generally true that imparting additional information improves the precision of estimation, regardless of the quality of the added information. However, as the new information gets far from the truth, the cost of introducing bias may override the benefit of reduced variance. Monte Carlo simulations permit judgment of the way in which the "quality" of prior information affects prediction, since this methodology allows the imposition of prior information which can become systematically farther from or closer to the "truth".

The results of the Monte Carlo experiment favor three of the Stein-like rules impressively. Except for one loss criterion, these estimators achieve the best performance consistently over the entire range of generated samples.

CHAPTER TWO

A COMPARISON OF ESTIMATORS USING TIME SERIES AND CROSS-SECTIONAL DATA

I. Introduction

As a tool for predicting residential housing prices, the hedonic pricing model is used in several ways. It is employed by appraisers either implicitly or explicitly when they use "comparables" as an input in the determination of fair market value for a house[1]. In some areas it is utilized extensively by tax assessors in forming the base value of homes and other real property for the purpose of property taxation.[2] The hedonic pricing model is also used for creating indexes of housing prices and rental values.[3]

This chapter tests the abilit" of a variety of econometric estimators, in two separate data environments, to predict prices of residential housing. It compares estimators that have been previously suggested for real

---

[1] See, for example, Church [1975], Isakson [1986], and Case and Quigley [1989].

[2] See Thibodeau and Vandell [1982] and Kowalski and Colwell [1986], for example.

[3] Examples of this use of hedonic models are Follain and Malpezzi [1980], Malpezzi, Ozanne, and Thibodeau [1980], Blackley, Follain and Lee [1986] and Thibodeau [1989]. Gruenstein and de Silva [1989] employ the model to estimate an index for commercial buildings.

estate applications, as well as proposing Stein-like estimation as a method of improving the precision of housing price predictions. The two contexts in which estimator effectiveness is compared are both familiar, but differ considerably in purpose, in the quality of the available data, and in the precision of the hedonic model employed.

The results of this study, while generally supportive of Stein-like estimation for improving prediction, are mixed and inconclusive. The estimators do demonstrate consistent relative effectiveness in the two trial settings, but none achieves dramatically superior performance.

The remainder of this chapter is organized as follows. Section 2 reviews the literature using the hedonic model to predict housing prices. Section 3 introduces the statistical model and assumptions and describes the method used to measure predictive performance. Section 4 discusses the problems created by multicollinearity and various methods for diagnosing its presence and its degree of severity. Section 4 also deals with the more general problem of moving from in-sample estimation to out-of-sample prediction. The notion that multicollinearity is not a problem as long as the pattern of interdependencies remains stable will be explored. In Section 5, the biased estimators used for comparison are described. Also, introduced in this section is the use of Stein-like rules for adding non-sample information to the data in a risk-

reducing way. Section 6 describes the data and the results and conclusions are in Section 7.

II. Review of the Literature

The use of regression analysis as a tool for predicting housing prices in a mass assessment context is not new, nor is the recognition of the benefits to be derived from improved prediction. Mark and Goldberg [1988] provide a survey of the literature in this area, as well as a comparison of the methods available. Their intent is more of description than of evaluation. Varian [1974] studies the mass appraisal problem and suggests a Bayesian approach, using data based and non-data based prior information to increase the precision of estimation.

Hedonic regressions are also used extensively in constructing indexes of housing prices from city to city and for the entire United States. Using Annual Housing Survey data, Thibodeau [1989] recently created hedonic price indexes for sixty metropolitan areas in the period 1974 through 1983. Other recent use of the hedonic price model for index production has been done by Blackley, Follain and Lee [1986].

In applying the hedonic pricing model to prediction, by far the most troublesome aspect is the frequent high correlation among the independent variables. Multicollinearity in hedonic models of real estate is well documented. In fact, Belsley, Kuh, and Welsch [1980] use

the Harrison and Rubinfeld [1978] study of the effect of air pollution on housing prices as a textbook case of the problem for demonstrating their diagnostic technique. Because it offers the potential for a favorable tradeoff between bias and variance when collinearity is severe, biased estimation techniques have been proposed for the hedonic pricing model for residential housing as an alternative to ordinary least squares estimation.

Most studies that propose biased estimation as a cure for the collinearity advocate the use of one biased estimator. These studies then compare the performance of that one estimator with the performance of ordinary least squares in arriving at "reasonable" parameter estimates. Ridge regression, principal components regression, factor analysis, and stepwise regression are examples of biased estimation techniques with have been put forward as solutions to the data problem.

Anderson [1981], for example, applies the ridge regression technique (Hoerl and Kennard [1970a]) in a housing context and demonstrates the capacity of this method to transform "improperly signed" estimates to the proper sign. Several other studies provide further support for this method (See for example, Moore, Reichert and Cho [1984], and Ferreira and Sirmans [1988]).

In a comparative study of estimation and prediction performance, Gau and Kohlhepp [1978] compare OLS, a stepwise

regression technique, and principal components regressions. They find that the principal components method which retains all components provides the best forecasts, but the superiority is marginal. Only one holdout sample is used in this study.

III. The Statistical Model

The model considered here is the linear statistical model

$$Y = X\beta + e \qquad 2.1$$

where Y is a T X 1 vector of the dependent variable; X is a T X K data matrix comprised of T observations on the K explanatory variables; β is the K X 1 vector of parameters to be estimated; and e is a T X 1 vector of normally distributed random error terms with mean zero and variance σ^2. It is assumed that X is of full rank K and uncorrelated with the error term.

Since β is unknown, there is a loss associated with using an estimator of β, say δ, when the true value of the parameter is β. Risk can then be generally expressed as the expected value of that loss:

$$R[\delta,\beta,Q] = E[(\delta - \beta)'Q(\delta - \beta)] \qquad 2.2$$

where Q is a weighting matrix, typically chosen to be either the identity matrix of rank K, or the correlation matrix of the independent variables. The choice of the identity matrix for Q yields mean square error, while X'X yields the mean square error of in-sample prediction.

Since the purpose of this study is to gauge the out of sample prediction performance of a variety of estimators, define the statistical model

$$Y_p = X_p\beta + e_p \qquad 2.3$$

where the "p" subscript indicates that the m > k observations on the variables occur outside the sample from which the parameters are estimated, and e_p is distributed $N(0,\sigma^2 I_m)$. It is further assumed that the in-sample and out-of-sample errors are uncorrelated, i.e. $E[ee_p'] = 0$. This model provides a measure of out-of-sample mean square error of prediction. Using $X_p'X_p$ as the weighting matrix Q, the resulting risk of the predictor $y_p = X_p\delta$ is

$$\begin{aligned} R(\beta,\delta,X_p'X_p) &= E[(\delta-\beta)'X_p'X_p(\delta-\beta)] \\ &= E[(X_p\delta - X_p\beta)'(X_p\delta - X_p\beta)] \\ &= E[(y_p - Ey_p)'(y_p - Ey_p)]. \end{aligned} \qquad 2.4$$

The properties of Ordinary Least Squares(OLS) for the classical linear statistical model are familiar. The OLS estimator, $b=(X'X)^{-1}X'Y$, is the best linear unbiased estimator for 2.1, and $\hat{\sigma}^2 = (y-Xb)'(y-Xb)/(T-K) = s/(T-K)$ is minimum variance unbiased estimator for σ^2. The risk of b under loss measure 2.2 is $\sigma^2 tr(X'X)^{-1}$.

It is convenient at this point to represent 2.1 in its principal components form, as this form is useful for discussing multicollinearity, measures of data extrapolation, and some of the biased estimation techniques introduced in Section 3. Since X'X is a positive definite,

symmetric matrix, there exists a transformation matrix P such that $P'P=I_k$ and $P'X'XP=\Lambda$, the diagonal matrix of the characteristic roots of the data matrix. P is the matrix of characteristic vectors, the ith column of which is the ith characteristic vector.

The principal components regression model is of the form

$$y = X\beta + e = XPP'\beta + e = Z\theta + e \qquad 2.5$$

$Z = XP$ is the matrix of principal components and $\theta = P'\beta$ is the parameter vector of the transformed model.

IV. Multicollinearity and Data Extrapolation

A. Multicollinearity

Model 2.1 assumes that the X is fixed in repeated samples, non-stochastic, and of full column rank. In a strict sense, multicollinearity is a violation of the rank condition which allows one independent variable to be expressed as an exact scalar multiple of one of the others. The problem of collinearity is one of degree. At one extreme, orthogonality, each explanatory variable shows only independent variation. At the other extreme, the model is not of full rank. Neither of these extremes is normally encountered however, and the problem becomes one of assessing the extent to which collinearity is present in data, and the severity of the difficulty its presence creates.

Several methods have been proposed for evaluating the seriousness of collinearity[5], none of which avoid the problem of being reduced to a final subjective judgment or "rule of thumb" upon which to base the severity diagnosis. Except for the method of Belsley, Kuh, and Welsch [1980], which is discussed below, they also are unable to detect interrelationships among more than two variables.

Belsey, Kuh, and Welsh [1980] propose a condition index as a method of detecting the presence and inferring the severity of multicollinearity. The eigenvalues of X'X, where X is the matrix of explanatory variables, provide insight into the linear dependencies among the variables and are the basis of this index. The presence of extremely large eigenvalues implies that some axes of the sample space can be rotated in such a way that much of the variation within the sample occurs along these axes. Conversely, the presence of extremely small eigenvalues implies very small amounts of variation in the direction of the corresponding axes. If the eigenvalues are ordered such that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$, then the square root of the ratio of the largest to the smallest eigenvalue of X'X provides a condition number

$$k(X) = (\lambda_1/\lambda_k)^{1/2} \qquad 2.6$$

[5] Appendix B provides a review of some popular methods of diagnosing multicollinearity.

which measures the sensitivity(elasticity) of b to changes in X'Y or X'X. Belsey, Kuh, and Welsh [1980] provide results from numerical experiments which suggest that index numbers larger than 30 indicate a moderate to severe condition, while index numbers between 5 and 10 are relatively free from the collinearity problem.

If multicollinearity is deemed a problem in the sample at hand, its primary statistical consequence is that of imprecise parameter estimation. The resulting model is extremely sensitive to small changes in the composition of the data matrix. Moreover, the covariation among the explanatory variables can mask the true relationship between the dependent variable and independent variables which the model is intended to describe.

B. Extrapolation

Further complicating the analysis when the focus is on prediction, as with this study, is the fact that while multicollinearity may be extremely severe, still it may pose no problem for prediction as long as the pattern of collinearity in the data for prediction is the same as the pattern of collinearity in the data from which the parameter estimates are obtained. There is the problem of assessing the importance of collinearity, even if it is deemed severe.

An appreciation for the manner in which data extrapolation can affect prediction can be gleaned by returning to 2.4 using OLS as the estimator. In this case

$$R[y_p,\ \hat{y}_p=X_pb] = \sigma^2\ tr[X_p(X'X)^{-1}X_p]$$
$$= \sigma^2\ tr[X_p'X_p(X'X)^{-1}] \qquad 2.7$$

This can be written in deviation from mean form as:

$$R[y_p,\ \hat{y}_p] = \sigma^2[m/t + tr[X_p'X_p(X^{*\prime}X^*)^{-1} + md'(X^{*\prime}X^*)^{-1}d] \qquad 2.8$$

where the asterisk indicates the centering of the data and $d=(x_p-x)$, the distance between in-sample and out-of sample variable means.

The prediction risk can then be represented in its principal components form as:

$$R[y_p,\ \hat{y}_p] = \sigma^2[m/t + tr[P'P_p\Lambda_pP_p'P\Lambda^{-1}] + md'P'\Lambda^{-1}Pd] \qquad 2.9$$

This representation is the most illuminating as it clearly shows prediction risk to be directly proportional to 1) σ^2, the variance of the regression function, 2) m, the number of observations for which predictions must be made, 3) $tr(\Lambda_p)$, a measure of the collinearity of the out-of-sample data[6], and 4) d'd, the squared distance between means vectors. Prediction risk is inversely proportional to 1) T, the number of observations from which parameter estimates are obtained, and 2) $tr(\Lambda^{-1})$, a measure of the orthogonality of the in-sample data.

The risk of predicting out-of-sample with the OLS estimator is also affected by 1) the rotation of the out-of-sample data relative to the in-sample data as measure by the relative direction and length of the vectors in P and P_p,

[6] The lower bound of $tr(\Lambda_p)$ equals k in the case of orthogonal data.

and 2) the orientation of d'd relative to the in-sample data scatter.

The reader is reminded that 2.9 is only an appropriate measure of prediction risk under the assumptions of 2.1 and when OLS is the estimator. Still, this equation suggests three direct measures of the differences between in-sample and out-of-sample data. First, of course, is a measure of distance, the graphical representation of which is familiar in its univariate form. The Euclidean measure,

$$\text{DISTANCE} = (d'd)^{1/2} \qquad 2.10$$

is used here.

Second, samples differ in the amount of total variation they exhibit and the extent to which this variation is collinear. X_p can be said to be variationally equivalent to X if $\Lambda_p = \Lambda$ or $\Lambda_p\Lambda^{-1} = I_k$. In this paper the measure of variation is

$$\text{SCATTER} = \text{tr}[I(k) - \Lambda_p\Lambda^{-1}] = K - \Lambda_p\Lambda^{-1}. \qquad 2.11$$

Finally, rotation can be measured by the difference in orientation of P and P_p. In sample and out-of-sample data are rotationally equivalent if $P'P_p = I_k$. Here

$$\text{ROTATION} = \det[I(k)-P'P_p] \qquad 2.12$$

which takes the value zero when the data sets are rotationally equivalent.

In examining data extrapolation, it is also useful to incorporate all of the sources of prediction risk into one overall representation. Two such measures that are used in

this study are

$$\text{RISKHAT} = \hat{\sigma}^2 \; tr(X_p(X'X)^{-1}X_p') \tag{2.13}$$

and

$$\text{RISKHAT2} = \hat{\sigma}^2 \; tr(I(k)+X_p'X_p(X'X)^{-1}) \tag{2.14}$$

Section 5 - Biased Estimators

The properties of Ordinary Least Squares for the classical linear statistical model are familiar. OLS is the best linear unbiased estimator and the minimum variance unbiased estimator. Under the conditions frequently encountered in hedonic pricing of residential real estate, however, there may be a beneficial tradeoff between bias and variance possible from the use of biased estimators.

The biased estimators considered here, whose performances are compared with OLS, fall into two basic categories: 1) Ridge regression estimators, and 2) Stein-like rules. These estimators share one common methodology. They are designed to "shrink" one or more of the parameter coefficient estimates and thereby to improve the precision of estimation. The estimators differ in their motivations and in the manner in which they accomplish the shrinkage. Table 1 provides a summary of the nine estimators compared in this dissertation.[7]

[7] Tables appear at the end of the chapter.

A. Ridge Rules

Ridge rules are designed to deal with problems associated with imprecise parameter estimation. In this paper, three of the ridge rule estimators are considered. First is the ordinary ridge estimator, RIDGE, introduced by Hoerl and Kennard [1970a]:

$$b_{RIDGE}(k) = (X'X + kI)^{-1} X'Y \qquad 2.15$$

Here, the shrinkage factor k is directly proportional to the bias and inversely proportional to the variance of $b_{RIDGE}(k)$. Note that b_{RIDGE} is a function of k, so that the vector of parameter estimates changes depending on the selection of k. The question, of course, is how one selects k? In their companion paper, Hoerl and Kennard [1970b] suggest that one method might be a visual inspection of the ridge trace, which plots changes in the elements of the parameter vector against changes in k values. The value of k at which the estimates stabilize and exhibit the proper signs can be chosen from this visual examination.

Hoerl and Kennard [1970a] show that there always exists a value of k such that the mean square error of $b_{RIDGE}(k)$ is less than that of the OLS estimate b, and that a sufficient condition for this is $k < K\sigma^2/\theta^2_{max}$, where θ_{max} is the largest element of the vector $\theta = P'\beta$. One natural choice for k which follows from this proof is $k = K\hat{\sigma}^2/b'b$. While k, thus chosen is a function of the data and thus will vary from

sample to sample, it is an objective manner of choosing k, and is the one used for estimating b_{RIDGE} in this paper.

The second ridge rule, ITHKB, is an adaptive version of the first, using an iterative technique to determine the "optimal" value of the shrinkage parameter k. Hoerl, Kennard, and Baldwin [1975] propose a numerical optimization routine which chooses values of k_{ITHKB} and b_{ITHKB} to minimize the mean square error of in-sample prediction.

The third ridge rule considered, STRAW, is one proposed by Strawderman [1978]. It is a generalized rule, dependent on the data, and designed to achieve minimaxity over a specified range. Strawderman's estimator:

$$\delta(b,s) = [\, I + \frac{asQ^{-1}(X'X)}{b'X'Xb + gs + h}]^{-1}\, b \qquad 2.16$$

where s is the sum of squared errors from OLS estimation, $h \geq 0$, and $g \geq 2K/(T-K+2)$. In 2.16, a is a scalar which controls the amount of shrinkage of the OLS estimate, b, and when

$$0 \leq a \leq \frac{2(K-2)}{(T-K+2)} \; \frac{1}{\lambda_{max}(Q^{-1}(X'X))} = a_{max} \qquad 2.17$$

then 2.16 is minimax. λ_{max} here is the largest characteristic root of the matrix in braces.

B. Stein-like Rules

Stein-like rules are rules for combining sample and non-sample information in a risk-improving way. These rules take the general form

$$b_{STEIN} = (1 - c/u)(b-b^{*}) + b^{*} \qquad 2.18$$

where c is a constant proportional to the number of observations, the number of variables, and the number of restrictions imposed on the model; u is the likelihood ratio test statistic for the restrictions; and b^* the estimate of the parameter vector if the restrictions are true. If the sample data support the restrictions, u will be small and the Stein-like estimator will move toward the restricted least squares estimator, b^*.

The Stein-like rules differ mainly in the manner in which the restrictions are derived. Four different Stein-like rules are considered here. First is the LINDLEY estimator which hypothesizes that the slope coefficients are zero and shrinks the parameter vector toward the grand mean ($\bar{y}$).

STEINRLS and EBAYES partition the parameter coefficient vector based on the judgment of the experimenter into $\beta_1 \mid \beta_2$, where β_2 represents the elements of the slope coefficient vector hypothesized to be zero. These estimators take advantage of the relationships among the explanatory variables in-sample and shrink the estimator toward the restricted least squares estimator, b^* depending on the degree to which the hypothesis ($\beta_2 = 0$) is true. The estimator is

$$b_{STEIN}(b,s) = \left[1 - \frac{as}{(r-Rb)'[R(X'X)^{-1}R']^{-1}(r-Rb)} \right] (b-b^*)+b^* \qquad 2.19$$

where b is the OLS estimator, b_{STEIN} the Stein-like estimator, s the in-sample estimate of sum of squared errors, and a is a shrinkage factor chosen to assure minimaxity.

Examining the numerator of the ratio in 2.19, as a or s approach zero, the term in brackets approaches one, and the Stein rule estimator converges on the OLS estimator. This is appealing because a large sample or a small in-sample prediction error would provide greater faith in the OLS estimator. The other way for the Stein estimator to approach the OLS estimator is for the denominator of the ratio in 2.19 to become large. The more untrue are the restrictions, the larger becomes the denominator, and the closer is the Stein estimator to OLS. Conversely, if the restrictions imposed on the model are nearly true, the denominator becomes very small, the bracketed result approaches zero and the Stein rule estimator approaches the restricted least squares estimator. A combination of small samples, large in-sample prediction error, and true or nearly true restrictions can in fact generate a negative value for b_{STEIN} in which case the restricted least squares estimator is chosen.

The fourth Stein-like rule, PCSTEIN, draws from principal components analysis, restricting the transformed model to retain only those principal components which are supported by hypothesis testing. The restrictions implied by principal components analysis are that the parameter

coefficients for the principal components with low eigenvalues are zero. If the implied restrictions are true, then the Stein rule shrinks the estimator toward the principal components estimator, reduces the effect of those parameters which are imprecisely estimated, and improves the predictive ability of the resulting model.

If the eigenvalues are arranged from largest to smallest and P is partitioned $P_1 | P_2$, then the restriction $\theta_2 = P'_2\beta = 0$ can be tested as $R\beta = r$ where r is a vector of zeroes. One customary criterion for deciding which principal components to retain is to add eigenvectors of the small eigenvalues to P_2 as long as $\theta_2 = 0$ is not rejected. This is the criterion which is adopted in this dissertation.

VI. Description of the Data and the Models

Two very different sets of data are used to compare the predictive ability of the estimators described above. The first consists of five samples collected over a four year period in one geographic market, Baton Rouge, Louisiana. This set of time series data from one location is highly descriptive of the physical characteristics and produces a hedonic model with excellent explanatory power.

The second set of data, cross-sectional in nature, is taken from the American Housing Survey of 1983 for six major metropolitan areas. These data, and the hedonic model constructed from them, differ from the Baton Rouge data set in several ways that are detailed below.

A. Time-Series Data

The time-series data utilized in this dissertation consist of Multiple Listing Service (MLS) Sold Data for Baton Rouge, Louisiana from October 1984 through June 1989. Contained in this data set are the structural characteristics, details of financing, and neighborhood location information for every residence sold in this area during the period.

For the sake of manageability this data is reduced to 975 observations based on location. The remaining data consist of residences in the various neighborhoods and subdivisions along Highland Road. Highland road is a major traffic artery in Baton Rouge, running south through the middle of the Louisiana State University campus. The homes along this road are very heterogeneous in terms of size, age, and structural quality.

The Baton Rouge data is then segmented into time periods and parameter estimation is performed on five successive samples of nine months each. In each case, the immediately ensuing three month period is used as the out-of-sample test of predictive performance. The model 2.1 is estimated for each of the nine-month sample periods using two functional forms, linear and semilog. The dependent variables for these forms are selling price of the home and the natural log of the sales price respectively.

The hedonic regressors are chosen to represent the characteristics from which home buyers derive utility. Seventeen independent variables describe the structural characteristics, locational specifics and financing arrangements typically used in the literature.[8] The chosen sample is assumed to be homogeneous in terms of neighborhood amenities and disamenities. Therefore, no attempt is made to measure such utility-affecting attributes as crime, air pollution and school quality.

Table 2 defines the variables used in the hedonic model constructed from the Baton Rouge data.

B. Cross-Sectional Data

Empirical studies of hedonic pricing in real estate markets are frustrated by the sensitivity of these models to changing data. It is not unusual for researchers using similar methodologies and models to arrive at diametrically opposite conclusions about the hedonic price of an attribute or the demand for that attribute. Similarly, the estimators being tested here may perform quite differently in other settings and with other data.

To partially overcome the argument against inferences to be made from the Baton Rouge results, the estimators are also compared using data from the 1983 American Housing Survey. The most important advantage of this data is its

---

[8] See, for example, the review by Miller [1982].

ready familiarity and availability to others studying various aspects of the housing markets in this country. In the context of this study, it also provides a considerably different set of data and resulting hedonic model upon which to base the comparison of predictive ability.

The American Housing Survey (formerly the Annual Housing Survey) is taken by the Bureau of Census and consists of information gathered on residences and their occupants throughout the United States. Completed survey questionnaires provide over 300 variables describing the location, structural characteristics and quality, and impressions of neighborhood quality by both the interviewer and respondent. Each survey is conducted over a nine month period, and includes rental houses, apartments, and privately owned homes. Of interest in this study are single family dwellings occupied by the owner.

The data is hampered by two major shortcomings in terms of model usefulness for predictive purposes. First, the dependent variable is subjective in nature. Unlike MLS data which contains the actual observed sales prices of residences, the AHS measure of the dependent variable is the owner's opinion of house value.[9] Furthermore, the information provided is not a point estimate of value, but rather a range of values. The second major shortcoming of

---

[9] Thibodeau [1989] notes that since 1981 the dependent variable is the actual sales price of the home, if the sales transaction took place within the preceding twelve months.

AHS data relative to MLS data is the absence of a precise measure of living space provided by the dwelling.[10]

From the 1983 American Housing Survey, data are taken for six geographically diverse Metropolitan Statistical Areas (MSA's). Boston, Chicago, Dallas, Detroit, New York, and San Diego are selected based on data sufficiency, geographical separation, and the fact that only one central city is contained in each of these statistical areas.

The hedonic models constructed for these cities vary slightly, primarily because of differences in construction occasioned by the different climatic conditions. Steam heat, for instance, is not a measured attribute for Dallas or San Diego. The variables used follow as closely as possible those itemized for owner units in Table 1 of Blackley, Follain and Lee [1986].[11] A description of these variables is contained in Table 3. Generally, these variables describe the structural characteristics, age, length of tenancy of the owner, location (central city or suburban), and the owner's opinion of the neighborhood quality.

---

[10] This shortcoming is less severe than the first since the attribute being measured might be adequately proxied by the number of rooms, number of bathrooms, and the age of the dwelling, characteristics that are available in the survey.

[11] The authors state this model is a derivative of that used in Malpezzi, Ozanne, and Thibodeau [1980]. It is also quite similar to the model of Follain and Malpezzi [1980] and resembles Thibodeau's [1989] model.

Observations from the six cities were divided randomly into estimation samples and test samples, with roughly three fourths of the observations being allotted to the former.[12] Parameter estimation, using the various estimators described in the preceding section, was then performed on the estimation samples of each of the six cities.

VII. Results

A. Time-Series Data

The results from the Baton Rouge sample data are provided in Tables 4 through 9. Table 4 presents the summary statistical information for the five sample periods used and reports the condition numbers and measures of extrapolation described in Section IV. The model used is significant in each period at the 1% level and R-squares are impressive, being in no period less than .85. Adjusted R-squares, which penalize for the large number of variables in the model, are never below 0.83. By the Belsley, Kuh, and Welsch benchmark, multicollinearity is moderate and approaching severe in each period. Condition index numbers range from 26.67 in period 1 to 28.92 in period 5.

Continuing to look at Table 4, a comparison of the in-sample and out-of-sample data from the five sample periods

[12] The Chicago, Detroit, and New York samples were too large for some of the processing required in the parameter estimation using SAS PROC MATRIX. These data sets were first reduced by one third before being processed in the same manner as the Boston, Dallas, and San Diego data.

reveals that the data are very nearly rotationally equivalent. The measures of distance and variation from in-sample to out-of-sample are not remarkable. RISKHAT and RISKHAT2, the two estimates of the overall prediction risk for the periods, indicate that the risk caused by data extrapolation is highest in period 5 and lowest in periods 2 and 3.

Table 5 and 6 give the OLS parameter estimates for the semilog and linear functional forms of the model, respectively. T-statistics of these estimates for each of the periods are included. The only consistently significant variable at either the 1% or 5% level is LIVAREA. Based on the linear model results, estimates of the price of an additional square foot of living area range from $53 to $74, which is reasonable for the area and the time frame of the study. Each of the other variables, with the exception of the two financing variables, CASH and SPECFIN, and the dummy variable for corporate-owned homes, DCORP, is significant at the 5% level in at least one of the periods.

AGE, DISTLSU, and BR are consistently negatively signed and AGE is significant in 4 of the 5 periods. DISTLSU is significant in three periods for the linear model and four periods for the log-linear model. DPOOL is consistently positive and is significant at the 10% level in all but period 4. All of the other variables show the instability of changing signs which might characterize ill-conditioned

data or model misspecification. One very puzzling result is the large negative value for the intercept term in four of the five cases for the linear functional form.

The main story of this chapter is told in tables 7 and 8, which summarize the out-of-sample performance of the various estimators in each of the periods. In these tables, the performance relative to OLS using

$$\text{Performance Index} = \frac{\text{RMSE (BIASED ESTIMATOR)}}{\text{RMSE (OLS)}} \qquad 2.20$$

Most of the estimators compared offer modest gains over OLS in some periods and are slightly inferior to OLS in other periods. There are two exceptions to this. First, the iterative ridge rule is subject to wide swings in performance, providing gains of 15% and 9% in periods 4 and 5, while losing 17% and 53% in periods 1 and 3. This estimator also appears to be the most sensitive to functional form specification.

Second, the most impressive performance based on a visual inspection of Tables 7 and 8 is achieved by STEINRLS and EBAYES. These two judgmentally motivated Stein-like rules have lower out-of-sample prediction loss than OLS in four of the five periods for the linear model and three of the five periods for the semilog model. They outperform OLS for the period identified *ex ante* as having the greatest

extrapolation risk (Period 5). They also outperform OLS for the periods of lowest risk (Periods 2 and 3).[13]

Table 9 provides the parameter estimates yielded by each of the compared estimators at the loss minimizing iteration of the estimation process.[14] This table reports the results for time period five.[15] Two aspects of this table warrant comment. First, note the manner in which the partitioned Stein-like estimators, STEINRLS and EBAYES, degenerate to the Restricted Least Squares estimator.[16]

---

[13] Hill, Cartwright, and Arbaugh [1989] report impressive benefits from "overshrinking" the parameter estimates in their marketing model study. While minimaxity is no longer guaranteed under these circumstances, in the case of their data set, the benefits far outweigh the costs. Experimentation with overshrinkage yielded mixed results in this study. Slight gains were noted for some estimators in some periods from using a shrinkage factor of 3. In no case were gains observed from shrinking by more than a factor of 9. The results reported in this study were obtained by using the maximum shrinkage which still provides minimaxity.

[14] An iterative process was used to vary the value of the shrinkage constant between 0.5 and 75.0 in eleven steps(0.5, 1, 3, 5, 7, 9, 15, 30, 45, 60, and 75). Minimaxity is assured for the Stein-like rules for values of the constant between 0.5 and 1.0. At each of these steps, parameter vectors consistent with the shrinkage constant of that step were estimated. Root mean squared error was then calculated for each parameter vector. The iterations were then searched to find the shrinkage factor and parameter estimates which minimized RMSE. These loss minimizing estimates are the ones reported in Table 5.

Of course, knowledge of the "best" estimates and estimators is ex post information. Determining the conditions under which a given estimator or a given degree of shrinkage works best is one of the main quests of this study.

[15] Results for the other time periods are tabled in Appendix C.

[16] This occurred four of five time periods for the linear model and two of five for the semilog model.

This implies that the data is highly supportive of the restrictions, i.e. that the coefficients of 11 of the 17 parameters are zero.

Second, note that the principal components rule yields estimates identical to those of the OLS estimator. This occurs in every case for the semilog model. In this model the independent variables are tasked with describing the variation in the log of the sales price, which is considerably less than the variation in the sales price itself. All principal components are retained when this functional form is used because the variation in even the principal component with least variation is large relative to the variation in the dependent variable. The principal components estimator which retains all variables is just a linear transformation of the OLS estimator.

B. Cross-Sectional Results

Table 10 contains the summary statistics and measures of data extrapolation for the cross-sectional data. The hedonic model using AHS data contains more than double the number of explanatory variables, yet explains sales price variation not nearly as well as the model constructed from MLS data. Part of the low explanatory power is surely caused by the problems in measuring the dependent variable discussed earlier. Part is a result of the low quality of independent variables available, and the subjectivity present in the measurement of some of these.

Scatter, rotation, and the overall measures of extrapolation risk appear quite similar among cities. However, there is a very large difference in in-sample to out-of-sample distance for Boston as compared with Dallas and San Diego. The two latter cities exhibit over 100 times the Euclidean distance when estimation samples are compared with test samples. It is not surprising under these circumstances that OLS is the superior predictor for Boston while the Stein-like rules achieve their most impressive predictive performance relative to OLS in Dallas and San Diego.

A look at the Ordinary Least Squares parameter estimates and the associated T-statistics in Tables 11 and 12 gives substance to the data weaknesses discussed earlier. In the absence of a precise measure of living area, several variables compete for this role. Note, for example, the size and significance of the BATHMORE, ROOMMORE, and ROOM5 variables.

Location in the central city is significant for three of the central cities, being a negative feature in Chicago and Detroit, but positive for Dallas. AGE is not significant in this model, probably at least partly explained by its high correlation with AGE_SQ and AGE_CUBE.[17] The significance and signing of the attributes

[17] Including these variables and TEN_SQ along with TEN is puzzling. Any benefit from compensating for the non-linear nature of these variables appears to be outweighed by the

vary greatly from city to city. While some of this is simply cross-sectional variation in the value placed on characteristics, much is symptomatic of the aforementioned overspecification and collinearity problems. How else can one explain a negative value for a Chicago home owner's evaluation of his neighborhood as good.[18]

Tables 13 and 14 indicate that the comparative performance of the estimators on this cross-sectional data is remarkably similar to that on the data for Baton Rouge. This is so in spite of the rather large differences in the quality of the data and in spite of the disparity in explanatory power of the two models. Here again, the iterative ridge rule is characterized by substantial variation in performance. Again, most of the estimators compare modestly favorably with OLS. And especially of interest, once again the partitioned Stein-like rules achieve the most impressive results. There appear to be penalties associated with the use of these estimators only when extrapolation risk is very low.

VIII. Conclusions

This chapter has introduced a variety of biased estimators and compared their predictive performance in two "real" settings as measured by root mean square error of

---

collinearity which is introduced.

[18] The situation is reversed in Detroit, where a bad neighborhood has a significant and positive value.

prediction. Of course, the important questions to be answered are how one knows a priori which estimator to use, if and how much to shrink the parameter vector, and how much is gained by shrinkage or overshrinkage. Since the true parameter values are unknown, this question unfortunately eludes a conclusive answer.

This chapter does, however, provide some insights into the relative effectiveness of the compared estimators from several perspectives. The most interesting result is the similarity in performance among the estimators, relative to each other, over the entire spectrum of data, models, and functional form specifications. There is almost no change in the ordering of estimator performance in moving from the semilog to the linear model in any given time period or for any given city.

None of the estimators is clearly dominant as a predictor of house prices in either of the data contexts. However, each of them demonstrates the same general pattern of performance in both environments. The iterative ridge rule is the most volatile, while the improvements in prediction of all but STEINRLS and EBAYES are almost always within 5% of that of OLS.

STEINRLS and EBAYES generally achieve the best and most frequent improvements among the alternatives considered. Furthermore, the potential benefit from biased estimation using these two alternatives appears greater with the

American Housing Survey data, where model explanatory power is lower. Still, the mixed nature of the outcomes and the limited number of comparisons prevent the conclusive choice of a preferred estimator for predicting housing prices with a hedonic model.

TABLE 1

SUMMARY OF COMPARED ESTIMATORS

ESTIMATOR NAME	PARAMETER ESTIMATE COMPUTATION	VARIABLE DEFINITIONS [1]
OLS Ordinary Least Squares	$b = (X'X)^{-1}X'y$	X = data matrix of T rows and K columns. y = dependent variable. In this study, the log of house price.
RIDGE Basic Ridge Rule	$b_{ridge} = (X'X + kI)^{-1}X'y$	$k = K\hat{\sigma}^2 / b'b$ Where $\hat{\sigma}^2$ = OLS variance estimate.
ITHKB Iterative Ridge Rule	$b_{ithkb} = (X'X + k_i\ I)^{-1}X'y$	$k_i = K\hat{\sigma}^2 / b_{ridge}'b_{ridge}$ 'Optimal' k_i determined iteratively
STRAW Strawderman Generalized Ridge Rule	$b_{straw} = \left[\frac{as}{b'(X'X)^{-1}b + gs} * I + I(K)\right]^{-1} * b$	a = 2(K-2)/(T-K+2) s = OLS Sum of Squared Errors g = 2K/(T-K+2)
STEIN Basic Stein Rule	$b_{stein} = c * b$	$c = \left[1 - \frac{as}{b'(X'X)^{-1}b}\right]$

[1] Unless redefined, variable definitions hold for all subsequent equations. Computation of "a" varies slightly for LINDLEY, STEINRLS, and EBAYES. Appendix D provides the precise programming description for these computations.

TABLE 1
(Continued)

SUMMARY OF THE COMPARED ESTIMATORS

ESTIMATOR NAME	PARAMETER ESTIMATE COMPUTATION	VARIABLE DEFINITIONS[1]
PCSTEIN Principal Components Stein Rule	$b_{pcstein} = c * b + (1-c) * b^*$	$c = \left[1 - \frac{as}{b'R'[R(X'X)^{-1}R']^{-1}Rb}\right]$ R = Last J rows of eigenvector matrix of X'X. J = Number of omitted variables, iteratively determined. b^* = Restricted least squares (principal components) estimator.
LINDLEY Lindley Stein Rule	$b_{lindley} = c * b + (1-c) * b^*$	R = Last (K-1) rows of an Identity Matrix of Rank K. $b^* = [b_1 \vdots 0]'$ $b_1 = \bar{y}$
STEINRLS Restricted Least Squares Stein Rule	$b_{steinrls} = c * b + (1-c) * b^*$	R = Last (K-7) rows of an Identity Matrix of Rank K. $b_1 = (X_1'X_1)^{-1}X_1y$
EBAYES Empirical Bayes Stein Rule	$b_{ebayes} = c * b + (1-c) * b^*$	

[1] Unless redefined, variable definitions hold for all subsequent equations. Computation of "a" varies slightly for LINDLEY, STEINRLS, and EBAYES. Appendix D provides the precise programming description for these computations.

TABLE 2

HEDONIC MODEL
MULTIPLE LISTING SERVICE DATA
BATON ROUGE, LOUISIANA

Variable	Description	Note
LIVAREA	Square Feet of Living Area	
AGE	Age of Residence in Years	
DPOOL	Residence has Swimming Pool	*
DISTLSU	Distance from Louisiana State University, in Miles	
CASH	Dummy Variable Indicating Cash Sale	*
BR	Number of Bedrooms Minus 3	
FULLBATH	Number of Full Bathrooms Minus 2	
NETAREA	Total Sq. Ft. Minus LIVAREA	
LOTSIZE	Size of Lot in Acres	
GARAGE	Number of Cars Accommodated by Garage	
COVTPRCH	Sq. Ft. of Covered Porch Area	
CARPORT	Number of Cars Accommodated by Carport	
DBRICK	Residence is of Brick Construction	*
DFIREPL	Residence has at least one Fireplace	*
SPECFIN	Dummy Variable Indicating Assumption or Owner Financing	
DCORP	Residence Corporate Owned	*
DSBP	Residence Sold Before Processing in MLS Book	*

* Binary variable. One if true, zero otherwise.

TABLE 3

HEDONIC MODEL
1983 AMERICAN HOUSING SURVEY DATA

Variable	Description	Note
CENCITY	Central City Location.	*
BATHMORE	More than two bathrooms.	*
ROOMMORE	Number of other rooms if > 5.	
AGE	Age of structure.	
STRUCTUR	Poor structural features.	Combination of binary variables
EXCNBHD	Neighborhood regarded as excellent by owner.	*
BATH1	One and a half bathrooms.	*
BATH2	Two bathrooms.	*
ROOM4	Four other rooms.	*
ROOM5	Five other rooms.	*
BED1	One bedroom.	*
BED2	Two bedrooms.	*
BED4	Four bedrooms.	*
BEDMORE	Number of bedrooms if > 4.	
ATTCHD	Single-family attached home.	*
GARAGE	Garage present.	*
BASEMENT	Basement present.	*
AGE_SQ	Square of home's age.	
AGE_CUBE	Cube of home's age.	
PRIOR40	Structure built prior to 1940.	*
ROOMHEAT	Wall or room heater.	*
STMHEAT	Steam or hot water heat.	*
ELECHEAT	Electric heat.	*
ROOMAIR	Room air-conditioning.	*
CENAIR	Central air-conditioning.	*
RMWOHT	Rooms without heat.	*
NOPRIV	No privacy.	*
RMWOELEC	Rooms without electric outlets.	*
TEN	Length of tenure.	
TEN_SQ	Length of tenure squared.	
OLDTEN	Moved in prior to 1950.	*
PPERROOM	Persons per room.	
BLACKHD	Black household head.	*
SPANISHD	Spanish household head.	*
GOODNBHD	Neighborhood regarded as good by owner.	*
POORNBHD	Neighborhood regarded as poor by owner.	*
ABANDON	Abandoned housing on street.	*

* Binary variable. One if true, zero otherwise.

TABLE 4

BATON ROUGE DATA
SUMMARY STATISTICS
AND
DATA EXTRAPOLATION MEASURES

A. SEMILOG MODEL

STATISTIC	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	PERIOD 5
R-SQ.	0.88	0.93	0.90	0.90	0.88
ADJUSTED R-SQ.	0.87	0.92	0.89	0.89	0.86
# IN-SAMPLE OBS.	192.00	155.00	178.00	210.00	140.00
# OUT-SAMPLE OBS.	77.00	57.00	47.00	59.00	53.00
CONDITION INDEX	26.67	28.46	27.62	28.29	28.92
RISKHAT	0.18	0.13	0.11	0.13	0.29
RISKHAT2	0.60	0.40	0.50	0.62	0.75
DISTANCE	100.52	31.18	120.36	88.69	80.29
SCATTER	11.00	11.42	13.84	12.51	8.28
ROTATION	4.9E-02	-2.5E-23	1.5E-04	-1.0E-23	1.1E-20

B. LINEAR MODEL

STATISTIC	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	PERIOD 5
R-SQ.	0.85	0.90	0.88	0.86	0.86
ADJUSTED R-SQ.	0.83	0.89	0.87	0.85	0.85
# IN-SAMPLE OBS.	192.00	155.00	178.00	210.00	140.00
# OUT-SAMPLE OBS.	77.00	57.00	47.00	59.00	53.00
CONDITION INDEX	26.67	28.46	27.62	28.29	28.92
RISKHAT	3.7E+09	2.5E+09	2.0E+09	2.6E+09	4.3E+09
RISKHAT2	1.2E+10	7.9E+09	9.4E+09	1.2E+10	1.1E+10
DISTANCE	100.52	31.18	120.36	88.69	80.29
SCATTER	11.00	11.42	13.84	12.51	8.28
ROTATION	4.9E-02	-2.5E-23	1.5E-04	-1.0E-23	1.1E-20

TABLE 5

ORDINARY LEAST SQUARES
PARAMETER ESTIMATES
(T-Statistics in Parentheses)

BATON ROUGE DATA
SEMILOG MODEL

VARIABLE	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	PERIOD 5
INTERCEPT	10.3152	10.4339	10.3441	10.4345	10.3693
	(134.53)	(141.02)	(125.22)	(118.62)	(107.82)
LIVAREA	0.0005	0.0004	0.0005	0.0006	0.0004
	(14.95)	(12.81)	(14.95)	(16.43)	(11.60)
AGE	-0.0046	-0.0069	-0.0009	-0.0051	-0.0095
	(-3.86)	(-4.70)	(-0.64)	(-4.36)	(-5.01)
DPOOL	0.0490	0.0670	0.0714	-0.0037	0.2074
	(1.06)	(1.26)	(1.62)	(-0.08)	(3.32)
DISTLSU	-0.0098	-0.0099	-0.0126	-0.0184	-0.0039
	(-2.68)	(-2.96)	(-3.55)	(-4.82)	(-0.90)
CASH	-0.0217	0.0243	-0.0328	0.0028	0.0340
	(-0.72)	(0.8)	(-0.66)	(0.06)	(0.60)
BR	-0.0384	-0.0579	-0.0230	-0.0262	0.0444
	(-1.21)	(-2.21)	(-0.76)	(-0.89)	(1.19)
FULLBATH	0.0062	0.0488	-0.0587	-0.0252	-0.0219
	(0.19)	(1.83)	(-1.62)	(-0.81)	(-0.50)
NETAREA	0.0001	0.0003	0.0002	0.0001	0.0006
	(0.66)	(3.24)	(2.17)	(1.39)	(4.54)
LOTSIZE	0.1389	0.3383	0.0539	0.1440	0.2156
	(1.15)	(3.09)	(0.52)	(1.29)	(1.46)
GARAGE	0.1787	0.0024	0.0257	-0.0932	-0.0559
	(3.37)	(0.05)	(0.46)	(-1.52)	(-0.62)
COVTPRCH	0.0002	0.0000	0.0001	0.0000	-0.0005
	(1.38)	(-0.47)	(0.44)	(-0.36)	(-2.97)
CARPORT	0.0744	-0.0158	0.0177	-0.0346	-0.0281
	(3.05)	(-0.76)	(0.65)	(-1.17)	(-0.69)
DBRICK	-0.0164	-0.0066	-0.0334	-0.0201	-0.0863
	(-0.66)	(-0.29)	(-1.41)	(-0.82)	(-2.57)
DFIREPL	0.0122	0.0621	0.0635	0.1091	-0.0752
	(0.43)	(2.10)	(2.19)	(3.52)	(-1.75)
SPECFIN	0.0104	-0.0604	-0.0333	-0.0171	0.0639
	(0.29)	(-1.92)	(-0.83)	(-0.40)	(1.49)
DCORP	0.0176	-0.1497	-0.0155		
	(0.15)	(-2.00)	(-0.24)		
DSBP	-0.0096	-0.0095	0.0830	-0.1927	
	(-0.12)	(-0.13)	(1.55)	(-3.56)	

TABLE 6

ORDINARY LEAST SQUARES
PARAMETER ESTIMATES
(T-Statistics in Parentheses)

BATON ROUGE DATA
LINEAR MODEL

VARIABLE	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	PERIOD 5
INTERCEPT	-32795.4	-11345.7	-24868.6	-5356.6	-15157.8
	(-2.97)	(-1.1)	(-2.19)	(-0.44)	(-1.29)
LIVAREA	74.2	53.9	65.5	72.8	53.4
	(14.60)	(11.1)	(15.04)	(15.58)	(11.5)
AGE	-762.2	-1127.2	-200.4	-726.5	-1112.5
	(-4.47)	(-5.46)	(-1.03)	(-4.43)	(-4.83)
DPOOL	11440.4	12662.7	15286.1	3513.6	34025.3
	(1.73)	(1.7)	(2.52)	(0.53)	(4.48)
DISTLSU	-881.5	-1054.0	-1142.0	-1930.2	-243.9
	(-1.68)	(-2.24)	(-2.33)	(-3.63)	(-0.46)
CASH	925.4	2950.0	-6206.8	-2076.7	1108.8
	(0.21)	(0.69)	(-0.9)	(-0.34)	(0.16)
BR	-12132.7	-4508.1	-7203.4	-7022.2	-1972.3
	(-2.65)	(-1.23)	(-1.73)	(-1.71)	(-0.44)
FULLBATH	-289.3	8700.2	-4137.5	-3646.1	4824.2
	(-0.06)	(2.33)	(-0.83)	(-0.85)	(0.9)
NETAREA	-17.8	6.6	16.9	9.5	64.0
	(-1.23)	(0.53)	(1.33)	(0.73)	(3.93)
LOTSIZE	23347.3	36747.0	-20892.2	-26851.9	6054.8
	(1.34)	(2.4)	(-1.47)	(-1.74)	(0.34)
GARAGE	21837.3	19381.7	9316.9	-10578.4	-2417.6
	(2.86)	(3.12)	(1.22)	(-1.24)	(-0.22)
COVTPRCH	46.4	25.3	22.6	7.1	-41.1
	(2.59)	(1.75)	(1.34)	(0.47)	(-1.95)
CARPORT	8816.7	1928.8	680.9	-4971.3	-4685.7
	(2.52)	(0.66)	(0.18)	(-1.21)	(-0.94)
DBRICK	-6675.6	387.9	-1478.4	-4516.9	-9896.4
	(-1.87)	(0.12)	(-0.45)	(-1.32)	(-2.41)
DFIREPL	-6600.6	676.4	-334.1	3999.3	-12967.9
	(-1.63)	(0.16)	(-0.08)	(0.93)	(-2.48)
SPECFIN	2664.9	-4026.0	-1174.0	-1486.8	8886.6
	(0.52)	(-0.91)	(-0.21)	(-0.25)	(1.7)
DCORP	-8630.7	-8898.1	995.8		
	(-0.52)	(-0.85)	(0.11)		
DSBP	-1809.8	-7350.4	13614.6	-21705.0	
	(-0.16)	(-0.71)	(1.85)	(-2.88)	

TABLE 7

SUMMARY OF PREDICTION PERFORMANCE
ROOT MEAN SQUARED ERROR OF PREDICTION
BATON ROUGE, SEMILOG MODEL

ESTIMATOR	PERIOD 1		PERIOD 2		PERIOD 3	
	RMSE	% OF OLS	RMSE	% OF OLS	RMSE	% OF OLS
OLS	14805.6	100.00%	20833.9	100.00%	22550.2	100.00%
RIDGE	14751.6	99.64%	20425.8	98.04%	22902.6	101.56%
ITHKB	17377.5	117.37%	19988.2	95.94%	34478.2	152.90%
LINDLEY	14997.6	101.30%	20310.4	97.49%	22879.5	101.46%
STRAW	15001.1	101.32%	20297.3	97.42%	22887.4	101.50%
PCSTEIN	14805.6	100.00%	20833.9	100.00%	22550.2	100.00%
STEIN	14810.3	100.03%	20824.4	99.95%	22555.8	100.02%
STEINRLS	15602.9	105.39%	20397.5	97.91%	21639.0	95.96%
EBAYES	15487.2	104.60%	20375.0	97.80%	21687.9	96.18%

ESTIMATOR	PERIOD 4		PERIOD 5	
	RMSE	% OF OLS	RMSE	% OF OLS
OLS	25671.0	100.00%	31770.1	100.00%
RIDGE	24674.0	96.12%	30222.2	95.13%
ITHKB	21893.5	85.28%	28810.3	90.68%
LINDLEY	25187.5	98.12%	31287.5	98.48%
STRAW	25173.1	98.06%	31283.4	98.47%
PCSTEIN	25671.0	100.00%	31770.1	100.00%
STEIN	25661.0	99.96%	31768.5	99.99%
STEINRLS	26451.8	103.04%	29562.8	93.05%
EBAYES	26352.1	102.65%	29724.3	93.56%

TABLE 8

SUMMARY OF PREDICTION PERFORMANCE
ROOT MEAN SQUARED ERROR OF PREDICTION
BATON ROUGE, LINEAR MODEL

ESTIMATOR	PERIOD 1 RMSE	PERIOD 1 % OF OLS	PERIOD 2 RMSE	PERIOD 2 % OF OLS	PERIOD 3 RMSE	PERIOD 3 % OF OLS
OLS	16352.7	100.00%	21509.3	100.00%	24485.4	100.00%
RIDGE	15669.1	95.82%	21670.4	100.75%	24955.6	101.92%
ITHKB	16437.9	100.52%	22185.6	103.14%	32756.7	133.78%
LINDLEY	16282.3	99.57%	21086.9	98.04%	24944.4	101.87%
STRAW	16282.4	99.57%	21081.3	98.01%	24948.9	101.89%
PCSTEIN	16747.9	102.42%	18480.4	85.92%	23373.4	95.46%
STEIN	16364.5	100.07%	21390.3	99.45%	24585.5	100.41%
STEINRLS	17400.8	106.41%	19439.2	90.38%	23354.3	95.38%
EBAYES	17249.0	105.48%	19585.4	91.06%	23394.3	95.54%

ESTIMATOR	PERIOD 4 RMSE	PERIOD 4 % OF OLS	PERIOD 5 RMSE	PERIOD 5 % OF OLS
OLS	22547.2	100.00%	34764.0	100.00%
RIDGE	22027.7	97.70%	33523.1	96.43%
ITHKB	23523.1	104.33%	29914.0	86.05%
LINDLEY	22464.6	99.63%	34046.9	97.94%
STRAW	22463.5	99.63%	34046.0	97.93%
PCSTEIN	21365.9	94.76%	34545.1	99.37%
STEIN	22408.8	99.39%	34732.3	99.91%
STEINRLS	21436.4	95.07%	32315.6	92.96%
EBAYES	21475.2	95.25%	32477.8	93.42%

TABLE 9

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
BATON ROUGE DATA
RESULTS FOR TIME PERIOD 5

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.36930	10.39290	10.65650	10.54060	10.57520	10.36930	10.36280	10.47740	10.47740
LIVAREA	0.00044	0.00042	0.00026	0.00038	0.00036	0.00044	0.00044	0.00054	0.00054
AGE	-0.00948	-0.00877	-0.00527	-0.00810	-0.00782	-0.00948	-0.00947	-0.00883	-0.00883
DPOOL	0.20739	0.20008	0.15783	0.17729	0.17121	0.20739	0.20726	0.20144	0.20144
DISTLSU	-0.00389	-0.00477	-0.00812	-0.00332	-0.00321	-0.00389	-0.00388	-0.00528	-0.00528
CASH	0.03398	0.03598	0.04638	0.02905	0.02806	0.03398	0.03396	-0.03806	-0.03806
BR	0.04438	0.05779	0.10304	0.03793	0.03664	0.04438	0.04435	0.00000	0.00000
FULLBATH	-0.02189	-0.01095	0.06877	-0.01872	-0.01808	-0.02189	-0.02188	0.00000	0.00000
NETAREA	0.00061	0.00050	0.00027	0.00052	0.00050	0.00061	0.00061	0.00000	0.00000
LOTSIZE	0.21561	0.23111	0.31443	0.18431	0.17799	0.21561	0.21547	0.00000	0.00000
GARAGE	-0.05594	0.00271	0.07500	-0.04782	-0.04618	-0.05594	-0.05590	0.00000	0.00000
COVTPRCH	-0.00051	-0.00038	0.00001	-0.00044	-0.00042	-0.00051	-0.00051	0.00000	0.00000
CARPORT	-0.02812	-0.00452	0.01179	-0.02403	-0.02321	-0.02812	-0.02810	0.00000	0.00000
DBRICK	-0.08634	-0.07439	-0.02290	-0.07381	-0.07128	-0.08634	-0.08629	0.00000	0.00000
DFIREPL	-0.07515	-0.05848	0.01455	-0.06424	-0.06204	-0.07515	-0.07511	0.00000	0.00000
SPECFIN	0.06390	0.05452	0.00195	0.05462	0.05275	0.06390	0.06386	0.00000	0.00000

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	-15157.80	-12067.20	21602.10	23090.70	15190.60	-3053.52	-14914.90	-7782.69	-7782.69
LIVAREA	53.36	50.10	28.44	37.65	40.90	49.86	52.51	60.94	60.94
AGE	-1112.47	-1023.29	-556.06	-785.00	-852.64	-1148.31	-1094.64	-971.01	-971.01
DPOOL	34025.30	32829.90	24072.20	24009.50	26078.30	33762.70	33480.10	30954.50	30954.50
DISTLSU	-243.93	-368.16	-856.81	-172.12	-186.95	-668.43	-240.02	-583.90	-583.90
CASH	1108.84	1468.20	3192.24	782.44	849.86	3077.23	1091.07	-6338.35	-6338.35
BR	-1972.28	182.25	9025.93	-1391.72	-1511.63	974.19	-1940.68	0.00	0.00
FULLBATH	4824.19	5919.72	12861.00	3404.13	3697.44	6366.07	4746.89	0.00	0.00
NETAREA	64.01	52.77	28.62	45.16	49.06	52.37	62.98	0.00	0.00
LOTSIZE	6054.78	8966.50	25623.20	4272.48	4640.61	-2770.69	5957.76	0.00	0.00
GARAGE	-2417.60	4361.64	11684.20	-1705.95	-1852.94	4260.68	-2378.87	0.00	0.00
COVTPRCH	-41.14	-26.95	9.92	-29.03	-31.53	-26.02	-40.48	0.00	0.00
CARPORT	-4685.73	-1975.58	-190.55	-3306.43	-3591.32	-1916.16	-4610.65	0.00	0.00
DBRICK	-9896.42	-8426.88	-2012.21	-6983.29	-7584.99	-8451.70	-9737.85	0.00	0.00
DFIREPL	-12967.90	-10799.80	-205.06	-9150.62	-9939.06	-13410.30	-12760.10	0.00	0.00
SPECFIN	8886.57	7655.80	113.17	6270.70	6810.99	6719.51	8744.17	0.00	0.00

TABLE 10

SUMMARY STATISTICS AND
DATA EXTRAPOLATION MEASURES
1983 AMERICAN HOUSING SURVEY

A. SEMILOG MODEL

STATISTIC	BOSTON	CHICAGO	DALLAS	DETROIT	NEW YORK	SAN DIEGO
R-SQ.	0.55	0.58	0.69	0.70	0.46	0.51
ADJ. R-SQ.	0.45	0.52	0.60	0.65	0.38	0.35
# IN OBS.	177.00	303.00	156.00	264.00	288.00	129.00
# OUT OBS.	59.00	101.00	52.00	88.00	96.00	43.00
COND. IND.	279.64	226.48	170.64	276.03	269.36	184.47
RISKHAT	1.54	1.88	2.52	1.91	1.87	1.56
RISKHAT2	5.10	5.83	9.24	7.00	6.17	4.98
DISTANCE	20.22	261.40	2692.37	762.49	619.31	3835.96
SCATTER	24.61	23.82	26.89	27.44	26.34	24.13
ROTATION	5.2E-09	1.5E-08	-1.7E-20	2.5E-19	1.3E-06	-1.2E-20

TABLE 10
(Continued)

SUMMARY STATISTICS AND
DATA EXTRAPOLATION MEASURES
1983 AMERICAN HOUSING SURVEY

B. LINEAR MODEL

STATISTIC	BOSTON	CHICAGO	DALLAS	DETROIT	NEW YORK	SAN DIEGO
R-SQ.	0.54	0.61	0.66	0.66	0.42	0.47
ADJ. R-SQ.	0.44	0.55	0.57	0.60	0.34	0.31
# IN OBS.	177.00	303.00	156.00	264.00	288.00	129.00
# OUT OBS.	59.00	101.00	52.00	88.00	96.00	43.00
COND. IND.	279.64	226.48	170.64	276.03	269.36	184.47
RISKHAT	1.7E+10	1.3E+10	1.6E+10	7.5E+09	1.9E+10	2.6E+10
RISKHAT2	5.5E+10	4.1E+10	5.9E+10	2.7E+10	6.3E+10	8.4E+10
DISTANCE	20.22	261.40	2692.37	762.49	619.31	3835.96
SCATTER	24.61	23.82	26.89	27.44	26.34	24.13
ROTATION	5.2E-09	1.5E-08	-1.7E-20	2.5E-19	1.3E-06	-1.2E-20

TABLE 11

ORDINARY LEAST SQUARES
PARAMETER ESTIMATES
1983 AMERICAN HOUSING SURVEY DATA
A. SEMILOG MODEL

VARIABLE	BOSTON	CHICAGO	DALLAS	DETROIT	NEW YORK	SAN DIEGO
INTERCEPT	10.8582	10.5933	10.5899	10.0738	10.9504	11.6596
	(24.35)	(34.93)	(11.93)	(30.55)	(33.91)	(25.31)
CENCITY	-0.1924	-0.2018	0.4379	-0.3730	-0.0007	-0.0228
	(-1.58)	(-3.37)	(3.98)	(-4.11)	(-0.01)	(-0.34)
BATHMORE	0.3855	0.2671	0.7441	0.4418	0.3172	0.1318
	(3.49)	(3.23)	(4.34)	(4.75)	(4.21)	(0.88)
ROOMMORE	0.0741	0.1212	0.0461	0.0775	0.0698	0.0528
	(2.29)	(4.19)	(0.95)	(2.02)	(2.87)	(1.68)
AGE	-0.0536	0.0220	0.0242	-0.0198	-0.0162	0.0257
	(-1.01)	(0.78)	(0.62)	(-0.53)	(-0.5)	(0.7)
STRUCTUR	0.0015	0.0059	0.0267	-0.0720	0.1072	0.0296
	(0.04)	(0.16)	(0.35)	(-2.15)	(2.88)	(0.43)
EXCNBHD	0.2373	-0.1485	0.3643	0.5110	-0.0445	0.0979
	(0.93)	(-0.94)	(0.65)	(4.32)	(-0.25)	(0.28)
BATH1	0.0581	0.0401	0.0249	0.0840	0.0977	-0.1870
	(0.86)	(0.69)	(0.15)	(1.32)	(1.65)	(-1.26)
BATH2	0.1231	0.0885	0.3533	0.2780	0.2649	-0.0490
	(1.23)	(1.54)	(2.63)	(2.99)	(4.15)	(-0.4)
ROOM4	0.2693	0.5363	0.1341	0.2907	0.0635	0.0315
	(1.07)	(2.55)	(0.37)	(1.11)	(0.31)	(0.13)
ROOM5	0.3698	0.6630	0.3566	0.4423	0.3802	0.2231
	(1.66)	(3.46)	(1.11)	(1.84)	(2.22)	(1.02)
BED1		0.0766		0.5344	0.1878	
		(0.31)		(1.50)	(1.00)	
BED2	0.0230	0.0205	-0.0641	-0.1141	0.0156	-0.0466
	(0.25)	(0.29)	(-0.49)	(-1.48)	(0.18)	(-0.38)
BED4	-0.0384	-0.0619	-0.1203	-0.0399	-0.0418	-0.0498
	(-0.5)	(-0.9)	(-0.94)	(-0.47)	(-0.72)	(-0.53)
BEDMORE	-0.0164	-0.0565	0.0132	0.0020	-0.0058	
	(-0.51)	(-1.98)	(0.19)	(0.04)	(-0.27)	
ATTCHD	-0.0701	-0.1165	0.0571	-0.0952	0.0867	0.0795
	(-0.62)	(-0.98)	(0.32)	(-1.01)	(1.24)	(0.47)
GARAGE	0.1900	0.0628	-0.0137	0.1898	0.0681	-0.0795
	(3.28)	(0.97)	(-0.12)	(3.13)	(1.2)	(-0.61)
BASEMENT	0.0191	0.1173		0.0267	0.0851	0.2542
	(0.17)	(2.27)		(0.37)	(1.45)	(1.21)
AGE_SQ	0.0028	-0.0017	-0.0013	0.0004	0.0000	-0.0019
	(1.04)	(-1.14)	(-0.61)	(0.22)	(-0.02)	(-0.97)
AGE_CUBE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(-1.04)	(1.32)	(0.4)	(-0.09)	(0.36)	(1.03)
PRIOR40	-0.3614	-0.1066	0.0634	-0.4497	-0.1541	0.0145
	(-1.21)	(-0.63)	(0.25)	(-2.09)	(-0.79)	(0.06)

TABLE 11
(Continued)

ORDINARY LEAST SQUARES
PARAMETER ESTIMATES
1983 AMERICAN HOUSING SURVEY DATA

A. SEMILOG MODEL

VARIABLE	BOSTON	CHICAGO	DALLAS	DETROIT	NEW YORK	SAN DIEGO
ROOMHEAT	-0.0765	-0.3260	-0.3251	0.1255	-0.2284	-0.1574
	(-0.3)	(-1.21)	(-1.55)	(0.78)	(-1.23)	(-1.39)
STMHEAT	0.0412	0.1248		-0.0322	-0.0273	
	(0.67)	(1.82)		(-0.34)	(-0.51)	
ELECHEAT	0.1531			-0.1691	0.0211	
	(0.71)			(-0.86)	(0.11)	
ROOMAIR	0.0075	0.0436	-0.2231	-0.0066	0.0715	0.0354
	(0.14)	(0.72)	(-0.91)	(-0.11)	(1.48)	(0.28)
CENAIR	0.0203	0.1224	-0.1193	0.0751	0.2732	0.2393
	(0.17)	(2.05)	(-0.39)	(1.16)	(2.9)	(2.56)
RMWOHT	-0.0743	-0.0803	0.0917	0.0748	-0.0177	0.0455
	(-1.27)	(-1.5)	(1.05)	(1.17)	(-0.34)	(1.01)
NOPRIV	-0.1085	-0.0228	-0.1449	0.2493	-0.0295	-0.0680
	(-0.76)	(-0.18)	(-0.89)	(2.04)	(-0.3)	(-0.59)
RMWOELEC		0.0356	-0.0125	0.2251	0.0303	0.3641
		(0.21)	(-0.03)	(0.98)	(0.15)	(0.98)
TEN	0.0036	-0.0125	-0.0453	0.0105	-0.0233	-0.0140
	(0.28)	(-1.21)	(-2.26)	(0.89)	(-2.32)	(-0.88)
TEN_SQ	-0.0002	0.0004	0.0014	-0.0005	0.0007	0.0002
	(-0.43)	(1.28)	(1.89)	(-1.39)	(2.1)	(0.52)
OLDTEN	-0.0288	-0.1940	-0.2394	-0.0752	-0.1394	
	(-0.24)	(-1.48)	(-0.87)	(-0.52)	(-1.5)	
PPERROOM	-0.1486	-0.0778	0.1304	-0.0609	0.2162	-0.3863
	(-1.03)	(-0.81)	(0.67)	(-0.55)	(2.57)	(-2.49)
BLACKHD		-0.1651	-0.2910	-0.1080	-0.3049	-0.2276
		(-2.38)	(-2.03)	(-1.24)	(-3.95)	(-1.27)
SPANISHD		-0.0902	-0.2706	-0.1938	-0.0500	0.0834
		(-0.52)	(-1.4)	(-0.49)	(-0.37)	(0.79)
GOODNBHD	0.0621	-0.3356	0.2187	0.4060	-0.2043	0.0273
	(0.24)	(-2.15)	(0.39)	(3.55)	(-1.17)	(0.08)
POORNBHD	0.0216	-0.3892	0.1855	0.3057	-0.2916	0.0866
	(0.08)	(-2.26)	(0.33)	(2.48)	(-1.58)	(0.22)
ABANDON	-0.2649	-0.0484	-0.5075	-0.2022	-0.3061	
	(-1.93)	(-0.45)	(-3.21)	(-2.48)	(-2.37)	

TABLE 12

ORDINARY LEAST SQUARES
PARAMETER ESTIMATES
1983 AMERICAN HOUSING SURVEY DATA
LINEAR MODEL

VARIABLE	BOSTON	CHICAGO	DALLAS	DETROIT	NEW YORK	SAN DIEGO
INTERCEPT	31227.50	3418.69	60967.00	14749.30	53222.90	122263.00
	(0.67)	(0.13)	(0.86)	(0.71)	(1.64)	(2.04)
CENCITY	-17893.70	-16730.40	51500.10	-7985.98	-2475.75	-5650.38
	(-1.41)	(-3.33)	(5.84)	(-1.4)	(-0.43)	(-0.65)
BATHMORE	48925.20	29344.70	65448.40	38144.40	32516.80	16590.50
	(4.25)	(4.23)	(4.76)	(6.54)	(4.29)	(0.86)
ROOMMORE	9172.22	14183.60	5777.83	7182.54	6590.28	8406.07
	(2.72)	(5.84)	(1.48)	(2.99)	(2.69)	(2.06)
AGE	-4767.27	2224.00	2893.07	-1703.80	-1323.07	2448.97
	(-0.87)	(0.94)	(0.92)	(-0.73)	(-0.4)	(0.51)
STRUCTUR	-230.50	306.36	12361.80	-3055.13	8259.67	4041.83
	(-0.06)	(0.1)	(2.05)	(-1.45)	(2.21)	(0.46)
EXCNBHD	17425.50	-13489.50	-17322.20	20978.50	918.76	6907.82
	(0.65)	(-1.02)	(-0.39)	(2.83)	(0.05)	(0.15)
BATH1	2226.03	-1846.98	-4258.03	5116.29	8899.07	-20664.10
	(0.32)	(-0.38)	(-0.33)	(1.28)	(1.49)	(-1.07)
BATH2	20450.10	9024.12	19891.90	12803.30	24876.00	-17194.80
	(1.97)	(1.87)	(1.85)	(2.2)	(3.87)	(-1.09)
ROOM4	40747.30	79954.40	35114.70	38079.80	16650.80	20813.20
	(1.55)	(4.53)	(1.21)	(2.32)	(0.8)	(0.64)
ROOM5	51847.00	84884.70	42918.60	43726.40	40570.30	40105.50
	(2.23)	(5.27)	(1.67)	(2.91)	(2.35)	(1.42)
BED1		-2833.05		43012.50	10308.90	
	(0)	(-0.13)	(0)	(1.92)	(0.55)	(0)
BED2	2912.42	1087.94	-14772.10	-7673.68	811.88	-3928.03
	(0.3)	(0.18)	(-1.4)	(-1.59)	(0.09)	(-0.25)
BED4	-5115.19	-9559.72	-7814.51	-2822.27	-3870.11	-11627.00
	(-0.64)	(-1.66)	(-0.76)	(-0.53)	(-0.67)	(-0.94)
BEDMORE	-2320.43	-8399.87	600.54	-2155.31	744.54	
	(-0.69)	(-3.51)	(0.11)	(-0.75)	(0.34)	(0)
ATTCHD	-4129.66	-13561.60	686.30	-12325.60	10435.60	-3238.46
	(-0.35)	(-1.36)	(0.05)	(-2.09)	(1.48)	(-0.15)
GARAGE	17042.40	3710.75	-11385.50	5864.04	2615.48	-8870.11
	(2.82)	(0.69)	(-1.22)	(1.54)	(0.46)	(-0.53)
BASEMENT	6148.62	11205.80		2981.54	9655.22	29657.40
	(0.51)	(2.59)	(0)	(0.65)	(1.64)	(1.08)
AGE_SQ	228.29	-171.45	-215.49	24.11	-23.87	-190.61
	(0.82)	(-1.38)	(-1.21)	(0.2)	(-0.14)	(-0.74)
AGE_CUBE	-3.20	3.10	3.76	0.12	1.22	3.24
	(-0.77)	(1.63)	(1.32)	(0.07)	(0.47)	(0.78)
PRIOR40	-31717.60	-6489.04	28887.10	-29144.20	-13247.20	2411.64
	(-1.02)	(-0.46)	(1.4)	(-2.16)	(-0.68)	(0.08)

TABLE 12
(Continued)

ORDINARY LEAST SQUARES
PARAMETER ESTIMATES
1983 AMERICAN HOUSING SURVEY DATA

LINEAR MODEL

VARIABLE	BOSTON	CHICAGO	DALLAS	DETROIT	NEW YORK	SAN DIEGO
ROOMHEAT	-2853.49	-9194.97	-16699.30	8458.73	-20178.20	-26191.60
	(-0.11)	(-0.41)	(-1)	(0.84)	(-1.08)	(-1.78)
STMHEAT	2851.06	17577.20		2331.29	-1575.67	
	(0.44)	(3.06)	(0)	(0.39)	(-0.29)	(0)
ELECHEAT	16590.60			-16847.70	9831.77	
	(0.74)	(0)	(0)	(-1.36)	(0.53)	(0)
ROOMAIR	2649.54	726.99	-14757.70	-172.89	6022.30	-3937.39
	(0.46)	(0.14)	(-0.76)	(-0.05)	(1.24)	(-0.24)
CENAIR	-3511.50	10627.60	5836.11	7433.68	25093.50	33201.60
	(-0.29)	(2.12)	(0.24)	(1.83)	(2.65)	(2.73)
RMWOHT	-4945.50	-2223.88	1480.30	5872.22	-3669.54	8364.49
	(-0.81)	(-0.5)	(0.21)	(1.47)	(-0.7)	(1.42)
NOPRIV	-15068.50	-5619.92	-387.00	11645.70	-1845.41	-17195.40
	(-1.01)	(-0.54)	(-0.03)	(1.52)	(-0.18)	(-1.14)
RMWOELEC		-3516.59	-7059.20	20.42	-1282.26	35429.80
	(0)	(-0.25)	(-0.23)	(0)	(-0.06)	(0.73)
TEN	251.17	-796.76	-3822.82	279.16	-2802.12	-1058.07
	(0.19)	(-0.92)	(-2.38)	(0.38)	(-2.78)	(-0.51)
TEN_SQ	-6.08	31.76	109.89	-19.30	84.70	18.34
	(-0.14)	(1.12)	(1.91)	(-0.79)	(2.51)	(0.26)
OLDTEN	-1750.81	-16542.70	-22747.20	-6417.78	-17751.80	
	(-0.14)	(-1.51)	(-1.04)	(-0.71)	(-1.89)	(0)
PPERROOM	-17996.00	-13197.80	14751.80	-5347.05	22863.10	-42788.40
	(-1.19)	(-1.63)	(0.94)	(-0.78)	(2.7)	(-2.13)
BLACKHD		-12073.10	-36149.70	-9552.49	-25860.90	-15306.90
	(0)	(-2.07)	(-3.16)	(-1.75)	(-3.33)	(-0.66)
SPANISHD		-1516.87	-28816.90	-21100.00	-7704.87	7313.42
	(0)	(-0.1)	(-1.87)	(-0.85)	(-0.56)	(0.53)
GOODNBHD	2008.80	-29559.90	-29295.00	12347.00	-13895.40	-4310.00
	(0.08)	(-2.26)	(-0.65)	(1.73)	(-0.79)	(-0.09)
POORNBHD	4811.51	-32077.10	-33779.80	7879.07	-18557.60	-10299.30
	(0.17)	(-2.22)	(-0.75)	(1.02)	(-1)	(-0.2)
ABANDON	-21796.20	-5773.22	-17385.70			
	(-1.52)	(-0.64)	(-1.37)	(0)	(0)	(0)

TABLE 13

SUMMARY OF PREDICTION PERFORMANCE
ROOT MEAN SQUARED ERROR OF PREDICTION
AMERICAN HOUSING SURVEY, SEMILOG MODEL

ESTIMATOR	BOSTON		CHICAGO		DALLAS	
	RMSE	% OF OLS	RMSE	% OF OLS	RMSE	% OF OLS
OLS	31781.3	100.00%	22472.4	100.00%	34039.2	100.00%
RIDGE	32008.6	100.72%	21996.6	97.88%	33424.9	98.20%
ITHKB	41540.9	130.71%	23782.3	105.83%	38052.3	111.79%
LINDLEY	36006.6	113.29%	22259.0	99.05%	35152.0	103.27%
STRAW	34139.4	107.42%	22186.6	98.73%	34532.9	101.45%
PCSTEIN	31781.3	100.00%	22472.4	100.00%	34039.2	100.00%
STEIN	31870.9	100.28%	22450.8	99.90%	33939.1	99.71%
STEINRLS	35134.8	110.55%	21909.4	97.49%	29268.4	85.98%
EBAYES	36095.4	113.57%	22376.1	99.57%	29044.0	85.33%

ESTIMATOR	DETROIT		NEW YORK		SAN DIEGO	
	RMSE	% OF OLS	RMSE	% OF OLS	RMSE	% OF OLS
OLS	19164.3	100.00%	37712.3	100.00%	34805.9	100.00%
RIDGE	20080.9	104.78%	36921.1	97.90%	33662.4	96.71%
ITHKB	23539.6	122.83%	37712.3	100.00%	34805.9	100.00%
LINDLEY	19576.0	102.15%	37662.7	99.87%	35182.5	101.08%
STRAW	19423.6	101.35%	37358.4	99.06%	33540.2	96.36%
PCSTEIN	19164.3	100.00%	37712.3	100.00%	34805.9	100.00%
STEIN	19184.8	100.11%	37699.6	99.97%	34627.0	99.49%
STEINRLS	19420.7	101.34%	36500.0	96.79%	31553.8	90.66%
EBAYES	19713.3	102.86%	36570.1	96.97%	31553.8	90.66%

TABLE 14

SUMMARY OF PREDICTION PERFORMANCE
ROOT MEAN SQUARED ERROR OF PREDICTION
AMERICAN HOUSING SURVEY

LINEAR MODEL

ESTIMATOR	BOSTON		CHICAGO		DALLAS	
	RMSE	% OF OLS	RMSE	% OF OLS	RMSE	% OF OLS
OLS	32720.4	100.00%	27589.5	100.00%	35456.6	100.00%
RIDGE	32623.2	99.70%	26205.4	94.98%	35791.8	100.95%
ITHKB	38805.3	118.60%	24096.7	87.34%	37898.0	106.89%
LINDLEY	34688.2	106.01%	26117.3	94.66%	34735.2	97.97%
STRAW	33499.3	102.38%	26386.1	95.64%	34564.2	97.48%
PCSTEIN	36103.7	110.34%	27589.5	100.00%	34313.3	96.78%
STEIN	33412.9	102.12%	27078.0	98.15%	33844.2	95.45%
STEINRLS	34872.0	106.58%	24932.2	90.37%	31439.8	88.67%
EBAYES	34872.0	106.58%	24724.6	89.62%	31418.0	88.61%

ESTIMATOR	DETROIT		NEW YORK		SAN DIEGO	
	RMSE	% OF OLS	RMSE	% OF OLS	RMSE	% OF OLS
OLS	19799.0	100.00%	37486.3	100.00%	35430.8	100.00%
RIDGE	20395.9	103.01%	36924.7	98.50%	34133.4	96.34%
ITHKB	22496.3	113.62%	37486.3	100.00%	35430.8	100.00%
LINDLEY	19285.1	97.40%	37665.1	100.48%	37759.1	106.57%
STRAW	19311.7	97.54%	37253.9	99.38%	35063.3	98.96%
PCSTEIN	20441.4	103.24%	36592.1	97.61%	35857.8	101.21%
STEIN	19812.6	100.07%	37200.9	99.24%	33464.9	94.45%
STEINRLS	19346.2	97.71%	36389.6	97.07%	33058.8	93.31%
EBAYES	19916.0	100.59%	36532.8	97.46%	33058.8	93.31%

TABLE 15

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
AMERICAN HOUSING SURVEY DATA
RESULTS FOR BOSTON

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.8582	10.7453	11.2468	11.3139	10.9103	10.8582	10.8566	10.8855	10.8964
CENCITY	-0.1924	-0.1957	-0.0479	0.0000	-0.1704	-0.1924	-0.1924	-0.2272	-0.2411
BATHMORE	0.3855	0.4050	0.0825	0.0000	0.3414	0.3855	0.3855	0.4303	0.4482
ROOMMORE	0.0741	0.0618	0.0055	0.0000	0.0656	0.0741	0.0741	0.0546	0.0469
AGE	-0.0536	-0.0011	0.0004	0.0000	-0.0474	-0.0536	-0.0535	-0.0284	-0.0184
STRUCTUR	0.0015	0.0009	-0.0020	0.0000	0.0013	0.0015	0.0015	-0.0070	-0.0104
EXCNBHD	0.2373	0.1910	0.0401	0.0000	0.2102	0.2373	0.2373	0.2342	0.2330
BATH1	0.0581	0.0594	0.0034	0.0000	0.0514	0.0581	0.0581	0.0332	0.0232
BATH2	0.1231	0.1240	0.0165	0.0000	0.1090	0.1231	0.1231	0.0703	0.0493
ROOM4	0.2693	0.1949	-0.0392	0.0000	0.2385	0.2693	0.2693	0.1538	0.1078
ROOM5	0.3698	0.2947	-0.0264	0.0000	0.3275	0.3698	0.3697	0.2112	0.1480
BED2	0.0230	0.0177	-0.0159	0.0000	0.0203	0.0230	0.0229	0.0131	0.0092
BED4	-0.0384	-0.0248	0.0199	0.0000	-0.0340	-0.0384	-0.0384	-0.0219	-0.0154
BEDMORE	-0.0164	-0.0105	0.0035	0.0000	-0.0145	-0.0164	-0.0164	-0.0094	-0.0066
ATTCHD	-0.0701	-0.0692	-0.0039	0.0000	-0.0621	-0.0701	-0.0701	-0.0400	-0.0281
GARAGE	0.1900	0.1915	0.0511	0.0000	0.1682	0.1900	0.1900	0.1085	0.0760
BASEMENT	0.0191	0.0044	0.0066	0.0000	0.0169	0.0191	0.0191	0.0109	0.0077
AGE_SQ	0.0028	0.0001	0.0000	0.0000	0.0025	0.0028	0.0028	0.0016	0.0011
AGE_CUBE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRIOR40	-0.3614	-0.0975	-0.0213	0.0000	-0.3200	-0.3614	-0.3613	-0.2064	-0.1446
ROOMHEAT	-0.0765	-0.0888	-0.0172	0.0000	-0.0677	-0.0765	-0.0764	-0.0437	-0.0306
STMHEAT	0.0412	0.0426	0.0153	0.0000	0.0365	0.0412	0.0412	0.0235	0.0165
ELECHEAT	0.1531	0.1391	0.0560	0.0000	0.1356	0.1531	0.1531	0.0875	0.0613
ROOMAIR	0.0075	0.0131	0.0051	0.0000	0.0066	0.0075	0.0074	0.0043	0.0030
CENAIR	0.0203	0.0177	0.0286	0.0000	0.0179	0.0203	0.0203	0.0116	0.0081
RMWOHT	-0.0743	-0.0712	-0.0205	0.0000	-0.0658	-0.0743	-0.0743	-0.0424	-0.0297
NOPRIV	-0.1085	-0.0949	-0.0165	0.0000	-0.0961	-0.1085	-0.1085	-0.0620	-0.0434
TEN	0.0036	-0.0006	-0.0002	0.0000	0.0032	0.0036	0.0036	0.0021	0.0015
TEN_SQ	-0.0002	0.0000	0.0000	0.0000	-0.0002	-0.0002	-0.0002	-0.0001	-0.0001
OLDTEN	-0.0288	-0.0514	-0.0085	0.0000	-0.0255	-0.0288	-0.0288	-0.0164	-0.0115
PPERROOM	-0.1486	-0.1532	-0.0401	0.0000	-0.1316	-0.1486	-0.1486	-0.0849	-0.0595
GOODNBHD	0.0621	0.0229	-0.0332	0.0000	0.0550	0.0621	0.0621	0.0355	0.0249
POORNBHD	0.0216	-0.0192	-0.0304	0.0000	0.0192	0.0216	0.0216	0.0124	0.0087
ABANDON	-0.2649	-0.2354	-0.0491	0.0000	-0.2346	-0.2649	-0.2649	-0.1513	-0.1060

TABLE 15
(Continued)

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR BOSTON

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
ONE	31227.5	27222.7	82220.0	90579.1	38227.2	21913.2	30546.1	41860.1	45014.8
CENCITY	-17893.7	-17681.5	-4387.5	0.0	-15783.4	-12249.5	-17503.3	-18720.0	-18965.2
BATHMORE	48925.2	51313.6	10820.5	0.0	43155.2	32743.3	47857.6	53531.4	54898.1
ROOMMORE	9172.2	7178.7	676.5	0.0	8090.5	7811.0	8972.1	5979.5	5032.2
AGE	-4767.3	-279.4	29.7	0.0	-4205.0	-1726.0	-4663.2	-2121.9	-1337.1
STRUCTUR	-230.5	-328.9	-95.2	0.0	-203.3	906.7	-225.5	-470.8	-542.1
EXCNBHD	17425.5	13627.6	4267.1	0.0	15370.4	15475.0	17045.3	18909.3	19349.6
BATH1	2226.0	2476.0	-662.5	0.0	1963.5	1185.9	2177.5	1078.3	737.7
BATH2	20450.1	20659.2	2928.0	0.0	18038.3	12642.7	20003.9	9906.0	6777.5
ROOM4	40747.3	28020.5	-3968.9	0.0	35941.7	25784.2	39858.1	19737.9	13504.3
ROOM5	51847.0	38972.2	-2636.9	0.0	45732.4	33282.1	50715.6	25114.6	17182.9
BED2	2912.4	2290.0	-1635.7	0.0	2568.9	3984.8	2848.9	1410.8	965.2
BED4	-5115.2	-2802.8	2580.8	0.0	-4511.9	-1648.6	-5003.6	-2477.8	-1695.3
BEDMORE	-2320.4	-1299.2	572.1	0.0	-2046.8	-876.4	-2269.8	-1124.0	-769.0
ATTCHD	-4129.7	-4520.9	327.3	0.0	-3642.6	-3284.5	-4039.6	-2000.4	-1368.6
GARAGE	17042.4	17024.7	5436.8	0.0	15032.5	19019.1	16670.6	8255.3	5648.1
BASEMENT	6148.6	4576.0	1223.2	0.0	5423.5	7093.5	6014.4	2978.4	2037.8
AGE_SQ	228.3	3.6	0.6	0.0	201.4	82.3	223.3	110.6	75.7
AGE_CUBE	-3.2	0.1	0.0	0.0	-2.8	-1.2	-3.1	-1.5	-1.1
PRIOR40	-31717.6	-7825.0	-1599.7	0.0	-27977.0	-17291.8	-31025.5	-15364.0	-10511.7
ROOMHEAT	-2853.5	-4187.8	-2624.3	0.0	-2517.0	-1791.2	-2791.2	-1382.2	-945.7
STMHEAT	2851.1	2741.2	1576.8	0.0	2514.8	7066.0	2788.8	1381.1	944.9
ELECHEAT	16590.6	16043.3	5961.5	0.0	14634.0	10332.2	16228.6	8036.5	5498.4
ROOMAIR	2649.5	3241.3	747.4	0.0	2337.1	3326.8	2591.7	1283.4	878.1
CENAIR	-3511.5	-3231.2	2261.4	0.0	-3097.4	-1395.3	-3434.9	-1701.0	-1163.8
RMWOHT	-4945.5	-4254.0	-1581.3	0.0	-4362.3	-9906.4	-4837.6	-2395.6	-1639.0
NOPRIV	-15068.5	-14016.5	-2364.7	0.0	-13291.4	-9732.2	-14739.7	-7299.2	-4994.0
TEN	251.2	-77.6	-6.3	0.0	221.6	-17.1	245.7	121.7	83.2
TEN_SQ	-6.1	5.3	0.4	0.0	-5.4	5.5	-6.0	-2.9	-2.0
OLDTEN	-1750.8	-3662.6	-438.5	0.0	-1544.3	-374.9	-1712.6	-848.1	-580.2
PPERROOM	-17996.0	-19410.3	-5712.6	0.0	-15873.6	-10330.8	-17603.3	-8717.2	-5964.2
GOODNBHD	2008.8	-1334.9	-3697.7	0.0	1771.9	-669.3	1965.0	973.1	665.8
POORNBHD	4811.5	1101.8	-2488.5	0.0	4244.1	2718.1	4706.5	2330.7	1594.6
ABANDON	-21796.2	-18769.5	-4601.9	0.0	-19225.6	-14136.4	-21320.5	-10558.0	-7223.6

CHAPTER THREE

COMPARING ESTIMATORS IN HEDONIC PRICING MODELS: A MONTE CARLO EXPERIMENT

I. Introduction

Because econometric estimators perform differently in a relative sense in different data environments, with different model specifications, and under different loss criteria, the choice of a preferred estimator for a particular application is difficult. This chapter investigates the relative effectiveness of several alternative estimators in the context of the hedonic pricing model for housing. Using the Monte Carlo method, it provides some insight into choosing a preferred estimator for predicting housing prices.

While the empirical results of Chapter 2 are generally supportive of the use of Stein-like rules for predicting house prices, these results compare the estimators only five times in a time-series setting and six times in a cross-sectional setting. The eleven estimator comparisons permit general observations regarding relative effectiveness, but do not invite inferences outside the experiment and may be an inadequate basis for choice. Selecting the best estimator requires knowing the relative performance among

alternatives in a repeated samples context. The Monte Carlo experiment of this chapter permits such an evaluation.

The design of the experiment allows systematic control over model specification and the degree to which imposed restrictions depart from the "truth". A pattern of superiority of one predictor over another under these controlled but changing conditions provides the basis for inferences and decisions. Numerical simulations of housing data permit evaluation of each of the estimators under four separate loss criteria in a repeated samples sense.

The findings of this study are very favorable to the Stein-like estimators introduced in this dissertation. The three partitioned Stein rules, LINDLEY, STEINRLS, and EBAYES completely dominate the other estimators considered in all simulation samples and at all specified levels of model R-square. Furthermore, these estimators achieve superior performance under all examined loss criteria except one, Mean Absolute Percentage Error (MAPE). The effectiveness of these estimators conveys obviously favorable information about the benefit of Stein rule estimation in hedonic pricing applications. It also provides insight into prediction losses from model overspecification.

The remainder of this chapter is organized as follows. Section II provides a description of Monte Carlo methodology and discusses its recent applications in finance and econometrics. Section III describes in detail the Monte

Carlo methodology of this dissertation. In this section, the estimators, data, models, and loss criteria for the numerical simulations are developed. Finally, Section IV presents the results. Tables and figures supplement the analysis.

II. Monte Carlo Experiments

A. Descriptions and Applications in Business

Monte Carlo experiments are used extensively to provide solutions to problems that either cannot be solved, or are too costly and cumbersome to be solved analytically.[1] The random sampling and replication characteristic of this technique are particularly well suited to systems simulations. Using Monte Carlo experiments, statistical inferences can be made about system outputs conditional on system inputs. The methodology's value is especially high when the historical data for a classical approach is unavailable, to costly to collect, or inadequate in size.

Because of its appeal on both theoretical and practical grounds, the Monte Carlo technique has been applied in

---

[1] Intriligator [1978] and Kalos and Whitlock [1986] provide examples of Monte Carlo experimentation as an alternative to analytical solutions. Intriligator [1978] gives as an example the approximation of the area of a complex two dimensional figure by enclosing the figure within a square, then randomly selecting coordinates representing points within the square. The number of observations falling within the complex figure divided by the total number of observation approaches in probability the ratio of the area of the figure to the area of the square. Kalos and Whitlock [1986] provide a similar example of approximating the value of pi using random sampling, replication, and the known relationship between the area of a circle and the area of a square.

several fields of study. Physical scientists and engineers are among the earliest and most frequent practitioners of the methodology[2], but numerical simulation is finding important applications outside these fields. Computer simulations are applied to a wide range of problems in finance, including portfolio selection (Clarkson and Meltzer [1960]), market systems operations (Barish and Siff [1969]), and options pricing (Boyle [1977], Brennan and Schwartz [1977], Jarrow and Rudd [1982] and Parkinson [1977]).

Recently, Monte Carlo experiments have been used in studies closely related to this dissertation. Ohsfeldt and Smith [1985] employ this methodology to establish the magnitude of exogenous price variation required for accurate estimation of the structural parameters in a representative hedonic model of housing. Pace and Gilley [1990] use numerical simulations to establish the superiority of inequality restricted least squares (IRLS) as an alternative to ordinary least squares for estimating a hedonic model of housing. These two studies are obviously connected to this dissertation by the subject, hedonic models for housing. Another important relationship, however, is the use of Monte Carlo experiments to answer econometric questions.

---

[2] Kalos and Whitlock [1986] provide a history of Monte Carlo applications and references to usage of the technique in statistical physics, quantum mechanics, chemistry and many other fields.

B. Econometric Applications

One way to understand how Monte Carlo experiments are used in econometrics is to compare empirical work with the simulations work of the econometrician. In empirical work, the "true" model parameters are not known. Econometric estimators are applied to economic models to estimate these unknown parameters. By contrast, in Monte Carlo simulations the true parameter values are specified, the econometric estimators are applied to the economic models, and the resulting estimates are then compared to the "truth". Simply put, empirical work uses estimators to get information about unknown parameter values. Monte Carlo experimentation uses specified parameter values to get information about the unknown properties of econometric estimators.

Econometricians use Monte Carlo experiments in two ways: 1) to learn more about the small-sample properties of estimators, and 2) to compare the performance of alternative estimators in a variety of settings. Sowey [1973] classifies the usage of the Monte Carlo methodology in econometrics for the period 1948-1972. A large share of these studies are concerned with the small-sample properties of estimators, providing numerical approximations for the bias and covariance matrices under known conditions or providing comparisons of numerical approximations with analytical derivations.

Many of the Monte Carlo studies that compare alternative estimator performance are concerned with estimators for simultaneous equation models.[3] Within the context of single equation estimation, the performance of biased estimators has been studied using Monte Carlo methods.[4] For example, McDonald and Galarneau [1975] employ Monte Carlo methodology to evaluate a number of ridge-type estimators[5] with respect to ordinary least squares. Using Mean Square Error as the performance criteria, they find that none of the ridge rules generally outperforms OLS.

III. METHODOLOGY

This section describes the Monte Carlo experiment performed as a part of this dissertation. The relative performance of the estimators is a function of a) the data, b) the hedonic model, c) the properties of the estimators compared and d) the loss criteria. Each of these four items receives separate discussion. Before beginning the detailed description, however, an outline of the general step-by-step procedure is presented. The SAS program written in PROC

---

[3] See for example Mikhail [1975], Nagar [1960] or Wagner [1958].

[4] Hendry [1983] reviews the application of the Monte Carlo technique to econometric problems. He also describes the procedure for designing a Monte Carlo experiment.

[5] The ridge-type estimators examined by McDonald and Galarneau [1975] differ in the method for selecting "k", the shrinkage parameter.

MATRIX language appears as Appendix D, providing both the details of the Monte Carlo procedure and the programming descriptions of the estimators compared.

A. Procedure

The seven basic steps of the procedure that are performed for each subset of data are:

Step 1: Specify the TxK matrix of the independent variables, X.

Step 2: Specify the "true" Kx1 parameter vector, β.

Step 3: Generate a Tx1 vector of error terms, ϵ.

Step 4: Construct the Tx1 vector $Y = X\beta + \epsilon$.

Step 5: Calculate estimates $\hat{\beta}$ of the Kx1 parameter vector, β.

Step 6: Calculate the losses from using $\hat{\beta}$ rather than knowing β.

Step 7: Repeat Steps 3 through 6 1000 times.

These seven steps receive elaboration in the subsections which follow.

B. Data

Data from the tape of the 1983 American Housing Survey is used as the basis for the Monte Carlo experiments. The data consists of surveyed housing characteristics for six geographically diverse Selected Metropolitan Statistical Areas (SMSA's): Boston, Chicago, Dallas, Detroit, New York,

and San Diego. In order to provide a greater number of comparison opportunities for the estimators, the data for each of the six cities is divided into four subsets of equal size. Each of these 24 data sets is then used as the basis for a Monte Carlo simulation of 1000 replications.

The design matrix, X, remains fixed for each Monte Carlo sample within a given data set, as does the specification of the true parameter vector β. The changes in estimation which occur on each pass are a result of changes in the error term, ϵ. These error terms are generated by the SAS RANNORM function under the assumption that $\epsilon \sim (0,\theta^2)$ where $\theta^2 = s/(t-k)$ is the variance estimate for the full set of city data. "s" is the sum of squared errors from OLS estimation of the model. With non-stochastic X and β, Y is also distributed normally with zero mean and variance θ^2.

C. Model

In Chapter 2, the hedonic model formed from American Housing Survey data contains 38 variables. One cost of dividing each city's data into four subsets is a reduction in the number of variables available for explaining house price variation. Those variables that do not appear in all 24 subsets of data are eliminated from the model.[6] The resulting hedonic model is specified for all cities as:

[6] Also, AGE_SQ, AGE_CUBE, and TEN_SQ are removed.

$$Y = \beta_0 + \beta_1 CENCITY + \beta_2 BATHMORE + \beta_3 ROOMMORE + \beta_4 AGE + \beta_5 STRUCTUR + \beta_6 EXCNBHD + \beta_7 BATH1 + \beta_8 BED2 + \beta_9 GARAGE + \beta_{10} CENAIR + \beta_{11} NOPRIV + \beta_{12} PPERROOM \qquad 3.1$$

where Y = Log of House Price

β_0 = Intercept

β_i, (i=1,...,12) = Slope coefficients of the housing attributes

CENCITY = One if Central City Location, zero otherwise

BATHMORE = One if more than two bathrooms, zero otherwise

ROOMMORE = Number of other rooms if more than five

AGE = Age of Structure

STRUCTUR = Linear combination of dummy variables denoting poor structural features.

EXCNHBD = One if neighborhood regarded as excellent by homeowner, zero otherwise.

BATH1 = One if one and a half or fewer baths, zero otherwise.

BED2 = One if two bedrooms, zero otherwise.

GARAGE = One if dwelling has garage, zero otherwise.

CENAIR = One if dwelling has central heating and air conditioning, zero otherwise.

NOPRIV = One if dwelling arrangement does not offer privacy for occupants, zero otherwise.

PPERROOM = Number of persons per room.

D. Estimators

All of the estimators described in Section V of Chapter Two are compared in the Monte Carlo experiment with the exception of PCSTEIN. The combined effects of a semilog functional form and the sensitivity of principal components techniques to scaling results in a principal components

model which is just a linear transformation of the OLS model (i.e. all principal components are retained in the selection process). Under these circumstances, the performance of PCSTEIN is always identical with that of OLS.

As in the empirical study of the preceding chapter, the Stein-like rules are used for adding variance-reducing non-sample information into the estimation procedure. The level of improvement realized depends on the degree of truth in the additional information imparted. The approach is very flexible and can accommodate a variety of types of information that might be available for predicting residential property prices. For instance, Stein-like rules could incorporate replacement cost information, expert opinion, and/or information from earlier published results to improve the precision of parameter estimation and predictive performance.

The information that is added here via the Stein-like rules is that a number of the variables included in the hedonic model for explaining price variation, may be of little value in <u>predicting</u> price variation in another sample. It is known that the hedonic model for housing suffers from major econometric problems.[7] Because of this, the signal-to-noise ratio for many of the variables in the model may be quite low. Eliminating those variables from

---

[7] Data collinearity, misspecification, and simultaneity are the problems most frequently cited.

the predictive model might, in itself, improve prediction. With the Stein-like rules, the decision to eliminate or to reduce the impact of those variables on prediction, is made corresponding to the degree to which the data support the model partitioning.

E. Loss Criteria

Just as the relative performance of econometric estimators differs with changes in data environments and model specifications, so does performance vary with the standards by which the estimators are judged. The econometrician specifies loss criteria for an experiment that reflect the cost of inaccuracy in estimating parameter values or in predicting the level or variation of the dependent variable. The criteria can be made as specific as desired[8], but at some point a tradeoff exists between accurately representing the penalties and the mathematical tractability of the chosen criteria.

Loss can be measured in several ways and the choice of loss functions can have a profound impact on the relative performance of estimators in a predictive role. In general,

$$\text{Loss} = L[\delta, \beta, Q] \qquad 3.2$$

where δ is the vector of parameter estimates, β is the vector of true parameter values, and Q is a weighting matrix

[8] Leamer [1983], for example, suggests that the loss function explicitly consider the benefits in adding data as weighed against the costs of imprecision from omitting the data.

for expressing the variable cost of inaccuracy among the individual elements.

Several loss functions are chosen here for comparing the econometric estimators. The first three measure comparative performance in estimating the full hedonic model, equation 3.1. These loss functions are mean square error, mean square error of in-sample prediction, and mean absolute percentage error.

The mean square error (MSE) loss functions[9]

$$MSE\_I = (\delta - \beta)'I(\delta - \beta) \qquad 3.3$$

where I is a KxK identity matrix, and

$$MSE\_XTX = (\delta - \beta)'X'X(\delta - \beta) \qquad 3.4$$

are quadratic in form, imposing increasingly severe penalties as the estimate gets farther from the true value. Since this aspect has intuitive appeal in many situations, MSE is the most common measure of loss used in comparative studies. MSE is particularly appropriate for this study, because it explicitly recognizes that total loss from estimation is composed of two parts: 1) Loss caused from bias in the estimates, and 2) Loss caused from imprecision in estimation.

---

[9] These MSE loss functions are comparisons of the estimators in the first and second weak sense, as described by Fomby, Hill, and Johnson [1984]. An estimator δ is superior to an estimator θ in the strong mean square error sense iff

$$E[(\delta - \beta)'Q(\delta - \beta)] \leq E[(\theta - \beta)'Q(\theta - \beta)]$$

for any positive definite matrix Q. The first and second weak sense comparisons that are used here are Q=I (Mean Square Error) and Q=X'X (Mean Square Error of In-Sample Prediction).

The third comparison criteria for the full hedonic model is mean absolute percentage error (MAPE). This loss function

$$MAPE = \left[\frac{1}{T}\sum_{t=1}^{T}\frac{|\,x_t\delta - x_t\beta\,|}{x_t\beta}\right] x \;\; 100 \qquad 3.5$$

has the advantages of being dimensionless and easy to interpret, but it penalizes only bias in estimation. It is simply the average error in predicting house price expressed as a percentage of true house price. It should be noted that this loss function favors underprediction[10], an aspect that might appeal in many house price prediction contexts. For example, an appraiser valuing a property for a lending institution might desire a criteria that penalized overvaluation more than undervaluation. MAPE might also be appropriate for a tax assessor more concerned with lost voters than with lost tax revenues.

The intercept and first six variables of 3.1 are considered the focus variables of the hedonic model in terms of their predictive ability, a notion explicitly accounted for by the STEINRLS and EBAYES estimators. It is therefore of interest to determine which of the estimators form better estimates of these individual parameters. To do this, mean square error of the individual parameter elements is calculated

$$MSE_i = (\delta_i - \beta_i)^2 \qquad 3.6$$

---

[10] 100% is the maximum possible underprediction while overprediction is unbounded.

as is an ensemble MSE for the seven variables of primary interest

$$EMSE = \sum_{i=0}^{6} MSE_i \qquad 3.7$$

The risk of using a given estimator is the expected loss of using that estimator.

$$R(\delta,\beta,Q) = E[L(\delta,\beta,Q)] \qquad 3.8$$

Within this Monte Carlo experiment, each of the losses discussed above is calculated for each estimator 1000 times for each subset of city data. Repeated sampling allows evaluation of the risk of each estimator as the mean of the losses calculated on each iteration of the experiment.

$$R = \bar{L} = \frac{1}{1000} \sum_{i=1}^{1000} L_i \qquad 3.9$$

Similarly, the variance of the risk for an estimator is the average squared deviation from the mean loss.

$$Var(R) = \frac{1}{1000} \sum_{i=1}^{1000} (L_i - \bar{L})^2 \qquad 3.10$$

F. Specification of the "Truth"

In order to compare the estimators, it is necessary to assign values for the unknown parameter vector elements. This is accomplished as follows:

1) Assumptions are made about the expected sign of each of the measured attributes (positive, negative, uncertain).

2) Ordinary Least Squares estimation is performed on the full sample of city data for each of the six cities.

3) If the OLS estimate agrees with the assumption, that value is assigned to the corresponding element in the β vector. Otherwise the simple average of the coefficient for the six cities is used as the parameter value.[11]

The values assigned for each of the housing characteristics are presented in table 16.

One criticism of the Monte Carlo Methodology is that results are specific to the true parameter specification. Generalizations about the relative estimator performance under untested parameter specifications are not well founded. For this reason, performance is compared over four separate specifications of the true parameter vector. Estimators are compared at 50%, 100%, 200%, and 500% of the "truth" for β described above.

The method for assigning error terms remains unchanged as β is respecified. Thus, the signal to noise ratio for explanatory variables is lowest at 50% and increases as the assigned parameter vector becomes longer. As this ratio increases, the R-square of the model corresponding to "truth" specification rises, and one would expect the benefit from using alternatives to OLS to decline. However, if the restrictions imposed by STEINRLS and EBAYES are

---

[11] In one case, STRUCTUR, the expected sign is negative since this is a linear combination of negative structural characteristics of the dwelling. However, among the six cities, the average of this coefficient was positive. Therefore, this element in the "true" parameter vector is assigned a value of zero.

nearly true, there may still be advantages to using these rules for prediction.

IV. Monte Carlo Results

This section reports the results from the Monte Carlo experiment. Under one loss criterion, Mean Absolute Percentage Error, none of the estimators outperforms OLS. Also, many of the compared estimators deliver results that are either consistently unfavorable or consistently unremarkable with respect to OLS. These two findings, along with some general observations about the overall performance of the various estimators, are discussed first.

On a more favorable note, three of the estimators exhibit consistently strong showings over a wide range of model specifications and for most of the loss criteria evaluated. These three, LINDLEY, STEINRLS, and EBAYES receive separate treatment which includes: 1) analysis of performance differences coincident to differences in data characteristics among the cities, and 2) examination of changes in estimator performance in response to changes in the signal-to-noise ratio of the model.

Finally, conclusions and inferences from these results are offered.

A. General Results

For one of the four loss measure considered, Mean Absolute Percentage Error (MAPE), there is no reduction in

risk to be achieved by choosing an alternative to OLS to estimate the model. This is not a surprising result, since OLS is the only <u>unbiased</u> estimator considered, and since MAPE measures only the loss from bias in estimation without considering the loss from variance of the estimates. Because none of the alternative estimators considered offers improvement under this loss measure, specific results for MAPE are reported in Appendix C.

For the other three loss measures considered, four of the estimators offer very little or no gain over OLS. The first two of these, the basic RIDGE estimator and its iterative version, ITHKB, show almost identical performance throughout the test. Except for one 3% improvement[12], these two estimators <u>never</u> showed better than a 1% gain over OLS. This is true over all four loss measures, over all four subsamples of each of the six cities, and over all four signal-to-noise specifications of the "true" parameter vector.

Worse than offering no substantial risk reduction, the ridge estimators considered here expose the user to potentially significant additional risk. The risk of using these estimators, as compared with OLS, generally declines as the ratio of explained to unexplained variation increases, and under MAPE criterion there is very little

---

[12] This occurred for Detroit's second subsample at the lowest signal-to-noise specification of the parameter vector.

difference. Under MSE loss, however, the additional risk of using the ridge estimators can be quite high. These estimators are almost always worse than OLS under the three MSE criteria considered, and their risk ranges up to 178% of that of OLS under ensemble MSE loss. The poor showing for the ridge rules is an interesting result, in that ridge rules are frequently proposed as a superior alternative to OLS for predicting real estate prices with a hedonic model.[13]

Two other estimators offer little improvement over OLS. They are the Strawderman and the basic Stein rule estimators. While their performance is not unfavorable, neither is it very exciting. The risk of using these estimators is never significantly higher or lower than that of OLS. In fact, over all loss measures and all parameter specifications, the difference in risk between these two estimators and OLS is never greater that 1% and is only once greater than .5%. The rather large size of the subsamples on which estimation and prediction is performed explains these results.[14]

---

[13] See, for example, Anderson [1981] and Ferreira and Sirmans [1988]. It should be noted that the settings in which the ridge rules have been proposed are somewhat different from the setting of the Monte Carlo experiment of this study.

[14] Subsample size ranges from 43 observations for San Diego, to 101 observations for Chicago.

B. Results of the Partitioned Stein-like Estimators

The remaining three estimators considered in this study provide risk that is dramatically different from that of OLS. These estimators have in common a partitioned shrinkage of the parameter vector. Two of them, STEINRLS and EBAYES almost always perform better, and at times impressively so. LINDLEY, on the other hand, provides the most significant risk reduction in some circumstances and increases risk substantially in other circumstances.

Tables 17, 18, and 19, and figures 1 through 18 summarize the striking achievements of these estimators. As the relative effectiveness of the estimators is quite similar across subsamples of data, only the results of the first of the four subsamples is presented in these tables. The same information for the other three subsamples is tabled in Appendix C.

Figures 1 through 18 present the information of tables 17 through 19 in graphic form. These figures illustrate the information on a city by city basis, and visually display the consistently superior performance, as compared with OLS, of LINDLEY, STEINRLS, and EBAYES over the entire range of the parameter space tested.

LINDLEY

The LINDLEY estimator imposes the restriction that the intercept term in the model is the only element of the parameter vector with predictive content. When the data

support this restriction, the estimator simply becomes the grand mean of the observations on the dependent variable. The LINDLEY estimator always has less risk than OLS under MSE and Ensemble MSE loss.

Its most impressive achievements are obtained under MSE loss, where the loss weighting matrix, Q, is the identity matrix. For this criterion, at the lowest signal-to-noise ratio, LINDLEY risk is less than 50% of OLS risk in 22 of 24 observations (six cities X four subsamples) and less than 75% of OLS risk on the other two occasions. As the signal-to-noise ratio increases, the relative effectiveness of the estimator declines. However, even when the length of the "true" parameter vector is doubled, LINDLEY risk is still less than 95% of OLS risk 23 of 24 times. It is only when the signal-to-noise ratio improves by a multiple of five that LINDLEY fails to achieve at least a 5% reduction in risk.

LINDLEY performs almost as well under the Ensemble MSE criterion, where loss is measured as the sum of the mean square errors of estimates of the first seven elements of the parameter vector. Here, LINDLEY risk is less than 50% of OLS risk 14 of 24 times when the "true" parameter vector length is halved. As with MSE loss, the estimator achieves at least a 5% reduction in risk 23 of 24 occasions when the signal-to-noise ratio is doubled.

The performance of the LINDLEY estimator under MSE of

in-sample prediction is erratic. The estimator is occasionally effective at the lowest signal-to-noise ratio (at least a 5% risk reduction on 6 of 24 opportunities), but exposes the user to risks higher than those of OLS as the ratio of explained to unexplained variation rises. Curiously, in Dallas, New York, and San Diego, the riskiness of LINDLEY peaks when the length of the "true" parameter vector is left unchanged. This phenomenon is not yet understood.

STEINRLS and EBAYES

STEINRLS and EBAYES are quite similar in motivation and in computation. Both of these estimators partition the parameter vector and they impose a similar restriction on the model; that the intercept and first six model variables are important for predicting housing prices while the other explanatory variables are not. Not surprisingly, the performance of these two estimators is very similar over the complete range of the Monte Carlo experiment.

The results for these two estimators are even more impressive than those for LINDLEY. Under MSE loss, at the lowest signal-to-noise ratio, STEINRLS and EBAYES had risks less than 50% of that of OLS on 24 and 23 occasions respectively. Even at the highest signal-to-noise ratio, each of these estimators achieved at least a five percent reduction in risk on four occasions.

Under the ensemble MSE criterion, LINDLEY performs

slightly better at low R-square levels for the model, but STEINRLS and EBAYES compensate by showing markedly better results at the other end of the parameter space spectrum. Each of these estimators achieves lower risk and more often than LINDLEY when the length of the parameter vector is doubled and quintupled.[15]

It is under MSE of in-sample prediction that the superiority of STEINRLS and EBAYES stands out. Where LINDLEY was less risky than OLS only six times over the entire parameter space, all 24 observations of STEINRLS and EBAYES performance at the lowest r-square specification of the model were at least a 10% improvement over OLS, and 19 of these were better than 25% improvements. These two estimators are almost always better than OLS until the length of the "true" parameter vector is doubled. At this and higher R-square levels, their performance becomes mixed.

C. Conclusions

The conclusions to be drawn from this Monte Carlo experiment are readily apparent from the foregoing review of the results. STEINRLS and EBAYES show markedly superior performance among the estimators considered in all cities

---

[15] STEINRLS is at least 10% better than OLS on eight occasions and at least 5% better on 14 occasions when the signal-to-noise ratio is increased by a factor of two. At the highest r-square tested, this estimator achieves a five percent risk reduction once. EBAYES also 22 instances of at least five percent reduction in risk at double the parameter length, and two of these are better than 25% risk reductions. At the highest signal-to-noise ratio, EBAYES is at least 5% better than OLS four times.

and under all evaluated loss criteria with the exception of Mean Absolute Percentage Error. LINDLEY competes well with these two Stein-like estimators except under Mean Square Error of In-Sample Prediction, where its use for prediction is potentially quite risky.

The superior performance of those three estimators that restrict the model by partitioning the parameter space supports two important conclusions. First, of course, is the notion that these Stein-like rules are an excellent method for adding non-sample information to the estimation process, and thereby to improve the quality of housing price predictions. Other estimators that have been proposed for this application, specifically the ridge rule estimators, offer no improvement over OLS in a predictive context.

The second conclusion is implied by the motivation for the partitioned Stein-like estimators as they are used here. These estimators essentially hypothesize that the hedonic model is overspecified when prediction is the purpose for which estimates are intended. The notable effectiveness of LINDLEY, STEIN, and EBAYES implies that there may be a considerable degree of truth to this hypothesis.

A final observation is with respect to the sensitivity of the estimator performance to the criteria by which losses are evaluated. Although the partitioned Stein-like estimators are markedly better than OLS under all MSE criteria considered, it is important to note that <u>none</u> of

the estimators beats OLS if MAPE is the appropriate measure of loss. The performance of an estimator in a predictive context is a function of both the properties of the estimator and the selection of loss criteria. Thus, careful consideration of the manner in which penalties for prediction errors are imposed should precede the choice of estimator.

Notwithstanding this caution, the clear conclusion from this study is that Stein-like rules offer exciting potential for improving the prediction of housing prices with hedonic models.

TABLE 16

SPECIFICATION OF THE "TRUE" PARAMETER VECTOR
1983 AMERICAN HOUSING SURVEY
MONTE CARLO EXPERIMENT

PARAMETER	BOSTON	CHICAGO	DALLAS	DETROIT	NEW YORK	SAN DIEGO
INTERCEPT	10.877	10.876	10.665	10.543	10.879	11.665
CENCITY	-0.210	-0.238	0.309	-0.610	0.000	-0.019
BATHMORE	0.433	0.342	0.438	0.431	0.246	0.294
ROOMMORE	0.030	0.029	0.028	0.025	0.036	0.021
AGE	-0.003	-0.003	-0.007	-0.001	-0.003	-0.003
STRUCTUR	-0.016	0.000	-0.006	-0.041	0.000	-0.016
EXCNBHD	0.185	0.214	0.096	0.187	0.220	0.037
BATH1	0.005	0.073	-0.220	0.059	-0.003	-0.152
BED2	0.065	-0.056	-0.163	-0.119	0.032	-0.200
GARAGE	0.232	0.062	0.070	0.248	0.070	0.002
CENAIR	0.054	0.167	0.558	0.125	0.125	0.236
NOPRIV	-0.149	-0.090	-0.467	-0.021	-0.027	-0.086
PPERROOM	-0.070	-0.097	-0.189	-0.072	-0.055	-0.231

TABLE 17

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER
MEAN SQUARE ERROR OF IN-SAMPLE PREDICTION
LOSS CRITERION*

FIRST SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	1.278	1.277	1.260	1.277
	LINDLEY	1.009	1.261	1.050	1.008
	STEINRLS	0.692	0.881	0.961	0.995
	EBAYES	0.699	0.921	1.001	1.004
CHICAGO	OLS	1.452	1.465	1.452	1.444
	LINDLEY	1.014	1.026	1.010	1.004
	STEINRLS	0.698	0.880	0.951	0.994
	EBAYES	0.706	0.933	0.993	1.003
DALLAS	OLS	3.026	3.054	3.054	3.080
	LINDLEY	0.989	1.031	1.010	1.003
	STEINRLS	0.781	0.961	0.990	0.994
	EBAYES	0.784	0.983	0.997	0.995
DETROIT	OLS	1.933	1.917	1.944	1.916
	LINDLEY	1.020	1.011	1.004	1.000
	STEINRLS	0.711	0.905	0.966	0.994
	EBAYES	0.719	0.937	0.995	1.000
NEW YORK	OLS	1.811	1.862	1.806	1.830
	LINDLEY	0.772	1.115	1.058	1.011
	STEINRLS	0.589	0.680	0.836	0.966
	EBAYES	0.592	0.671	0.883	0.995
SAN DIEGO	OLS	1.377	1.384	1.396	1.362
	LINDLEY	0.981	1.704	1.268	1.042
	STEINRLS	0.644	0.841	1.021	1.004
	EBAYES	0.655	0.836	1.011	1.001

* Table shows actual calculation of Mean Square Error of In-Sample Prediction for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 18

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER MEAN SQUARE ERROR
LOSS CRITERION*

FIRST SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.323	0.315	0.308	0.322
	LINDLEY	0.295	0.628	0.877	0.977
	STEINRLS	0.421	0.622	0.862	0.972
	EBAYES	0.431	0.626	0.857	0.969
CHICAGO	OLS	0.158	0.163	0.159	0.162
	LINDLEY	0.406	0.752	0.926	0.988
	STEINRLS	0.397	0.625	0.854	0.975
	EBAYES	0.405	0.611	0.834	0.969
DALLAS	OLS	0.902	0.909	0.936	0.932
	LINDLEY	0.321	0.678	0.903	0.981
	STEINRLS	0.435	0.707	0.892	0.974
	EBAYES	0.457	0.743	0.906	0.977
DETROIT	OLS	0.211	0.208	0.215	0.213
	LINDLEY	0.537	0.823	0.952	0.992
	STEINRLS	0.411	0.642	0.853	0.974
	EBAYES	0.429	0.664	0.862	0.975
NEW YORK	OLS	0.202	0.208	0.200	0.203
	LINDLEY	0.238	0.550	0.836	0.968
	STEINRLS	0.339	0.462	0.719	0.940
	EBAYES	0.345	0.434	0.681	0.912
SAN DIEGO	OLS	0.487	0.478	0.496	0.475
	LINDLEY	0.202	0.526	0.805	0.964
	STEINRLS	0.369	0.624	0.884	0.975
	EBAYES	0.392	0.633	0.891	0.977

* Table shows actual calculation of Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 19

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER
ENSEMBLE MEAN SQUARE ERROR LOSS CRITERION*

FIRST SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.153	0.151	0.147	0.154
	LINDLEY	0.414	0.735	0.910	0.981
	STEINRLS	0.688	0.790	0.932	0.981
	EBAYES	0.692	0.791	0.930	0.979
CHICAGO	OLS	0.073	0.076	0.076	0.075
	LINDLEY	0.535	0.815	0.945	0.992
	STEINRLS	0.614	0.734	0.897	0.982
	EBAYES	0.620	0.707	0.866	0.974
DALLAS	OLS	0.357	0.377	0.371	0.377
	LINDLEY	0.310	0.675	0.903	0.982
	STEINRLS	0.629	0.804	0.934	0.980
	EBAYES	0.649	0.841	0.948	0.983
DETROIT	OLS	0.088	0.087	0.088	0.087
	LINDLEY	0.751	0.904	0.978	0.995
	STEINRLS	0.665	0.787	0.907	0.983
	EBAYES	0.677	0.792	0.908	0.983
NEW YORK	OLS	0.091	0.093	0.086	0.091
	LINDLEY	0.402	0.746	0.924	0.982
	STEINRLS	0.638	0.687	0.844	0.964
	EBAYES	0.643	0.664	0.758	0.902
SAN DIEGO	OLS	0.193	0.192	0.201	0.186
	LINDLEY	0.243	0.588	0.825	0.972
	STEINRLS	0.661	0.768	0.902	0.986
	EBAYES	0.676	0.781	0.919	0.989

* Table shows actual calculation of Ensemble Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

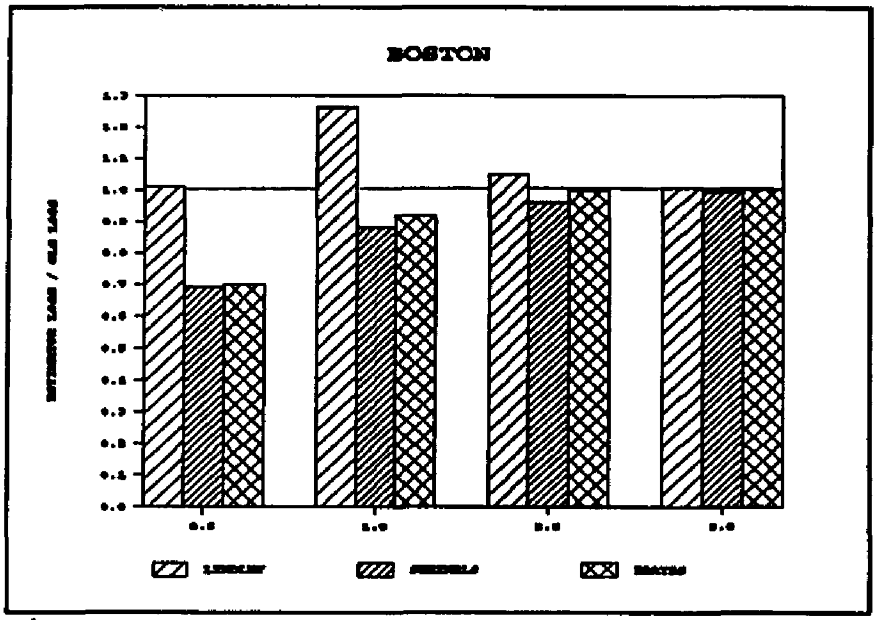

Figure 1 Mean Square Error of In-Sample Prediction. Boston. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

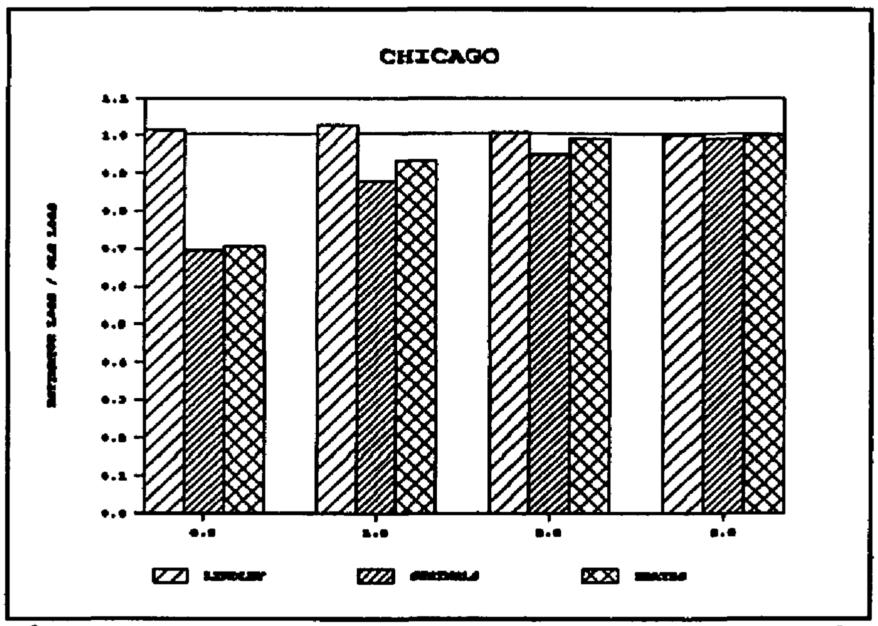

Figure 2 Mean Square Error of In-Sample Prediction. Chicago. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

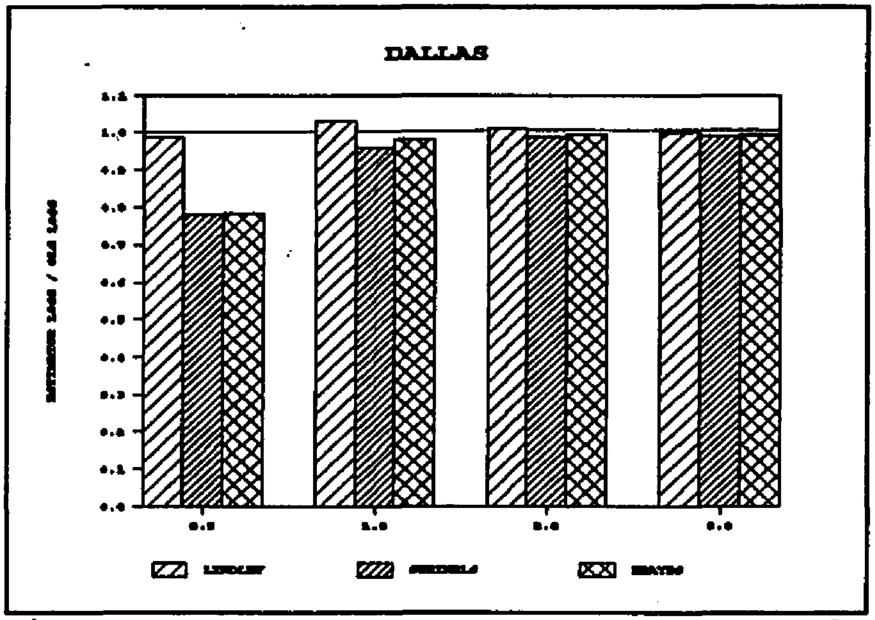

Figure 3. Mean Square Error of In-Sample Prediction. Dallas. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

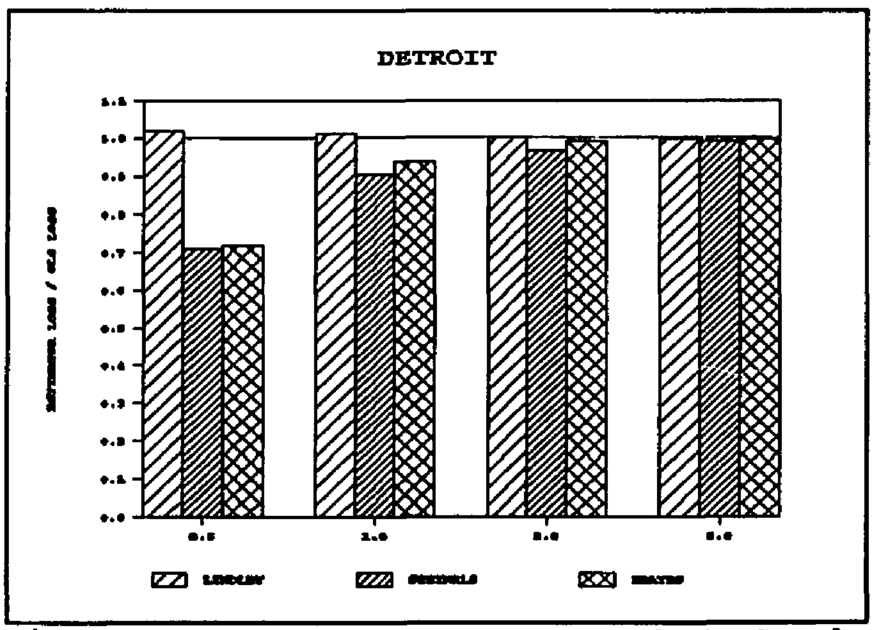

Figure 4. Mean Square Error of In-Sample Prediction. Detroit. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

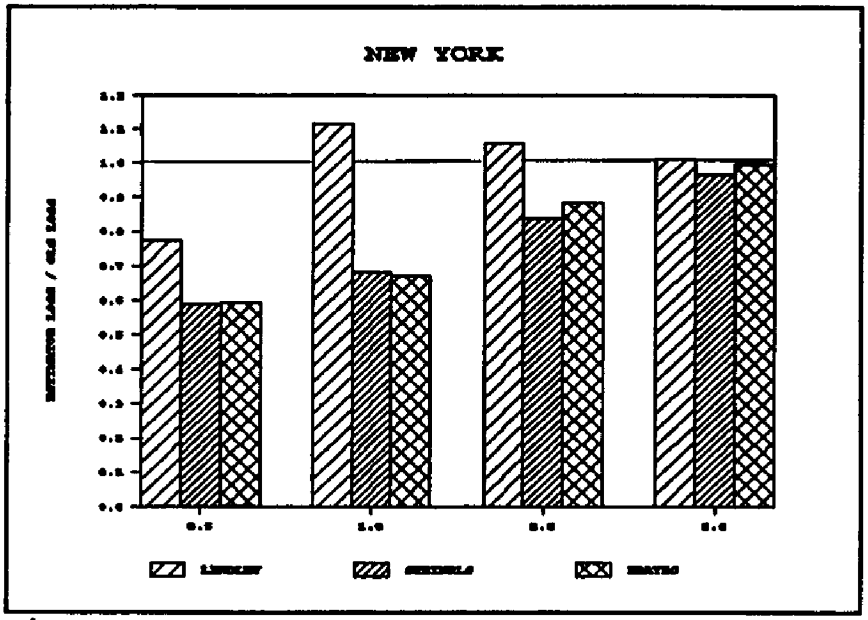

Figure 5. Mean Square Error of In-Sample Prediction. New York. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

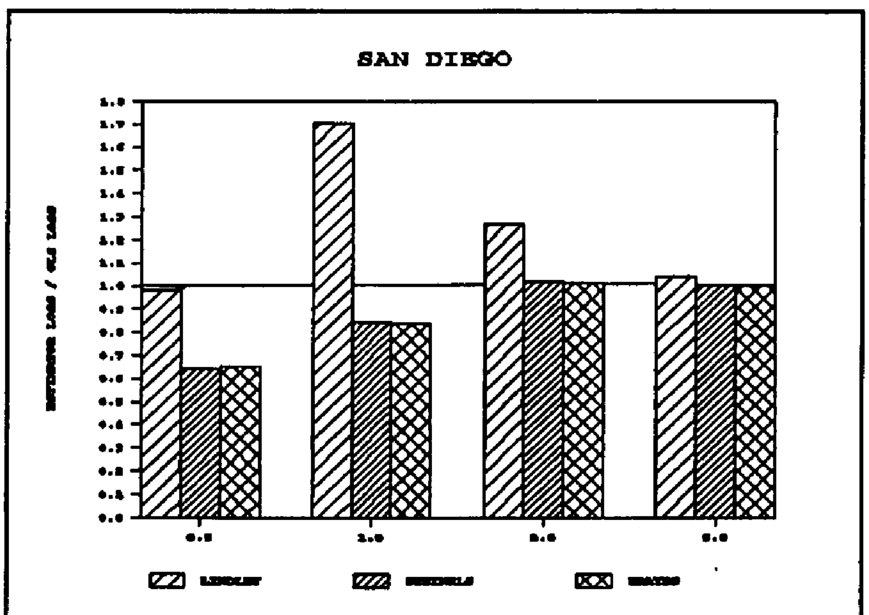

Figure 6. Mean Square Error of In-Sample Prediction. San Diego. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

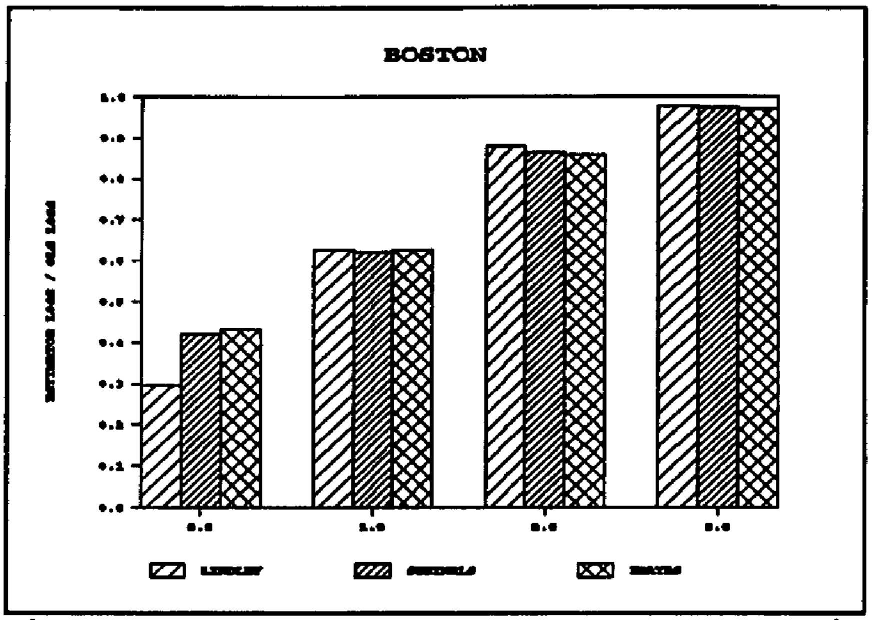

Figure 7. Mean Square Error. Boston. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

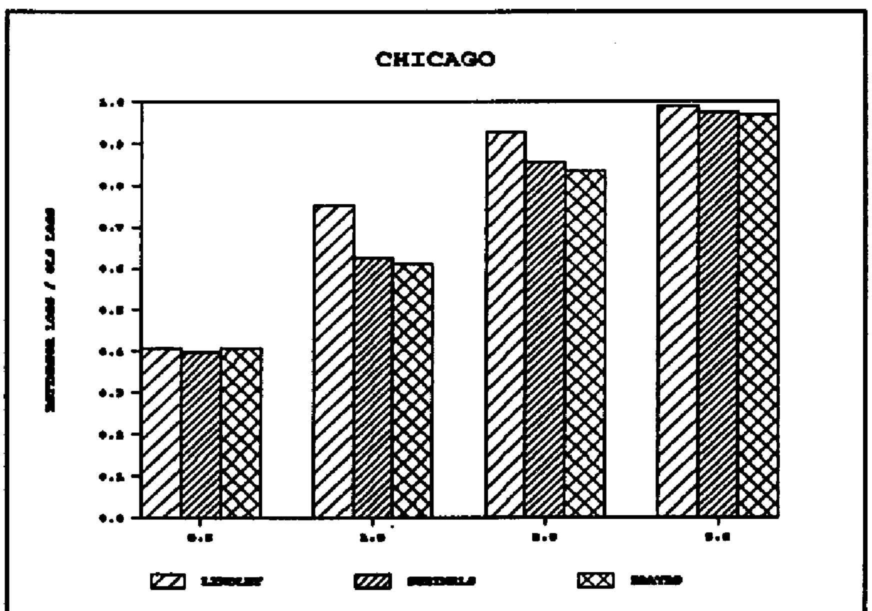

Figure 8. Mean Square Error. Chicago. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

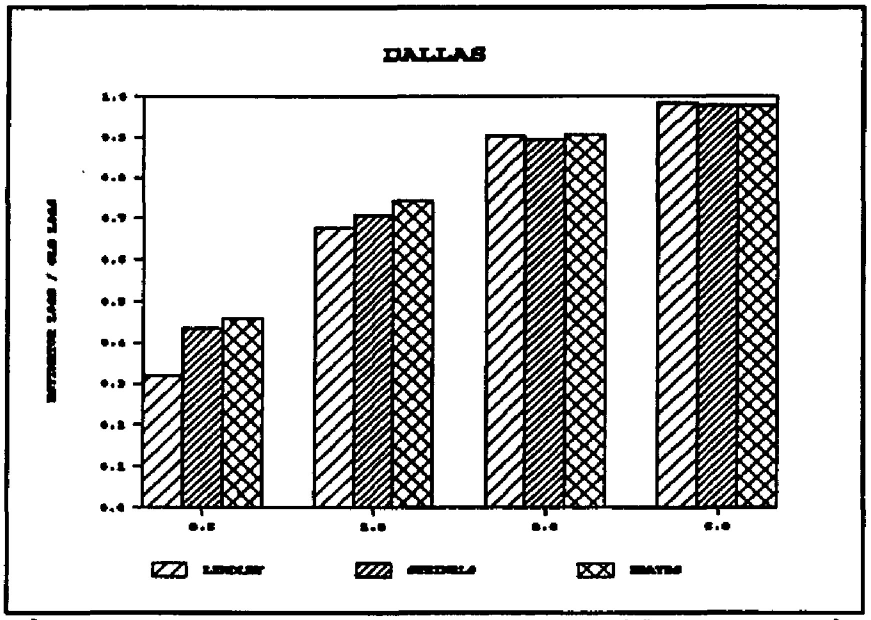

Figure 9. Mean Square Error. Dallas. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

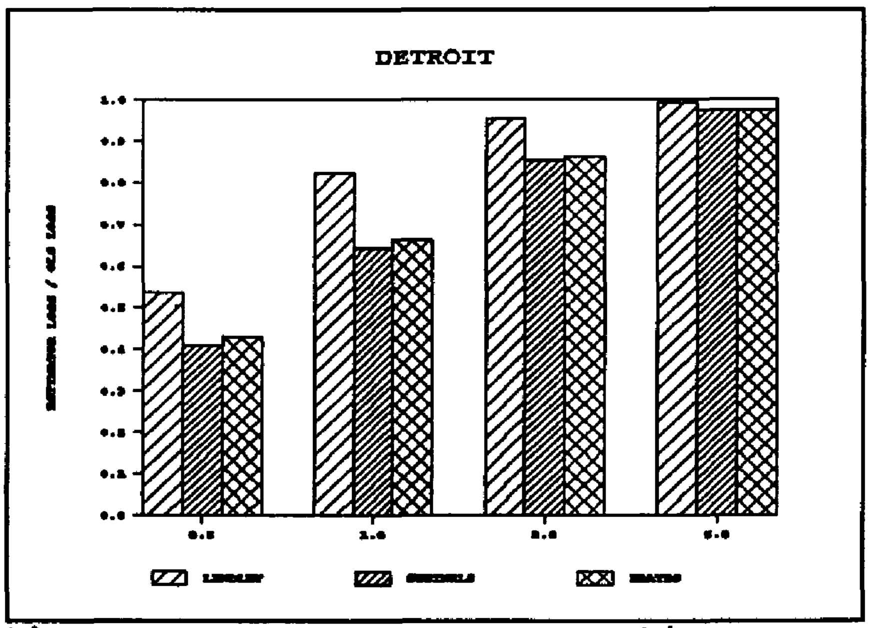

Figure 10. Mean Square Error. Chicago. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

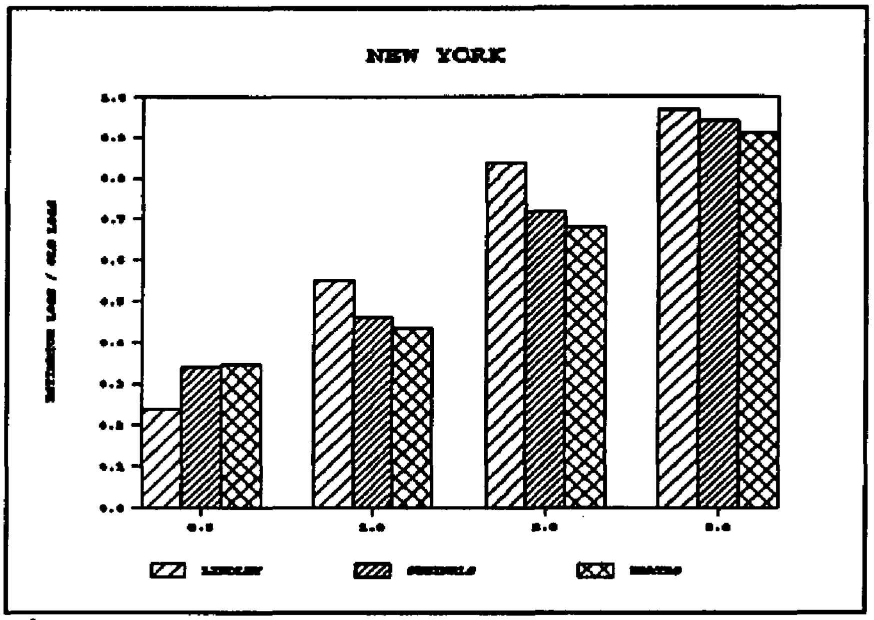

Figure 11. Mean Square Error. New York. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

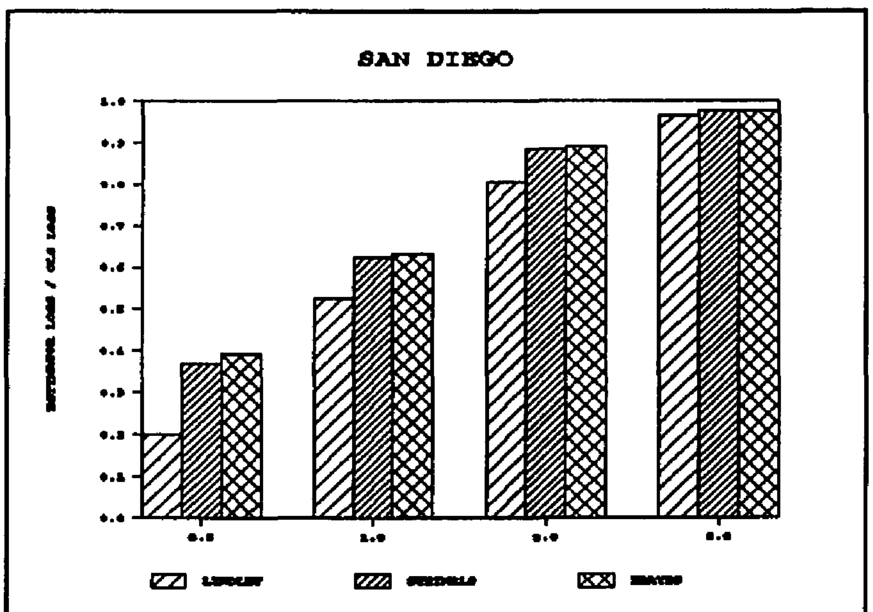

Figure 12. Mean Square Error. San Diego. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

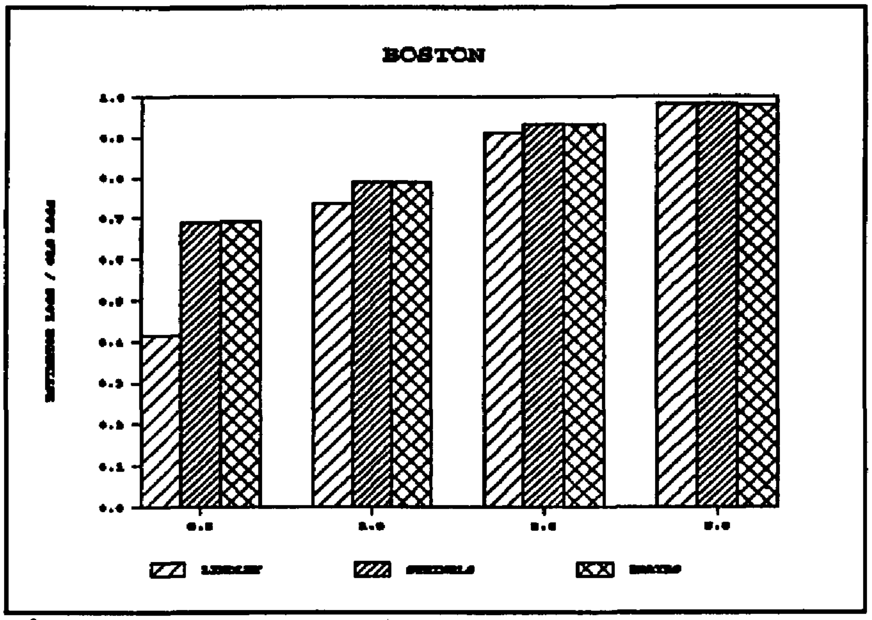

Figure 13 Ensemble Mean Square Error. Boston. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

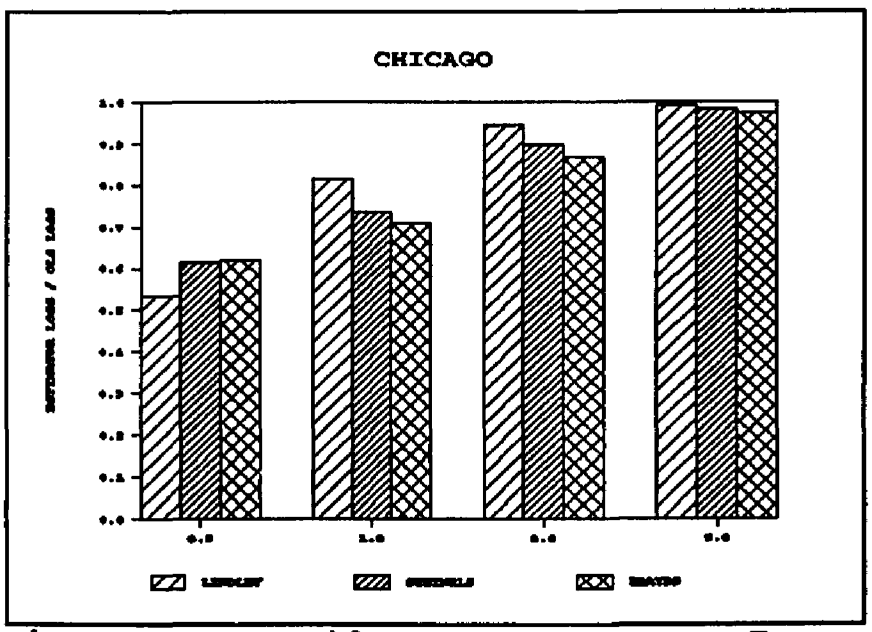

Figure 14 Ensemble Mean Square Error. Chicago. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

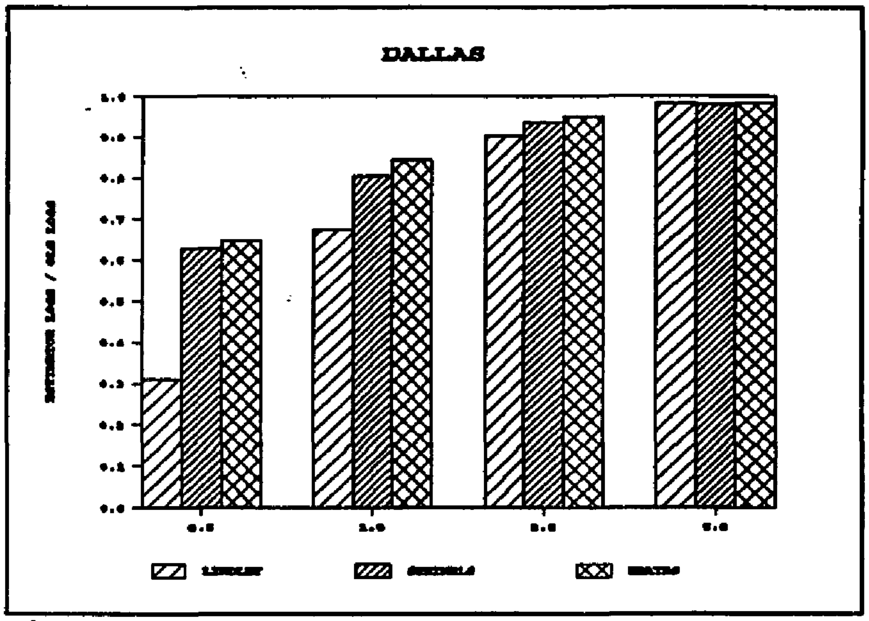

Figure 15 Ensemble Mean Square Error. Dallas. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

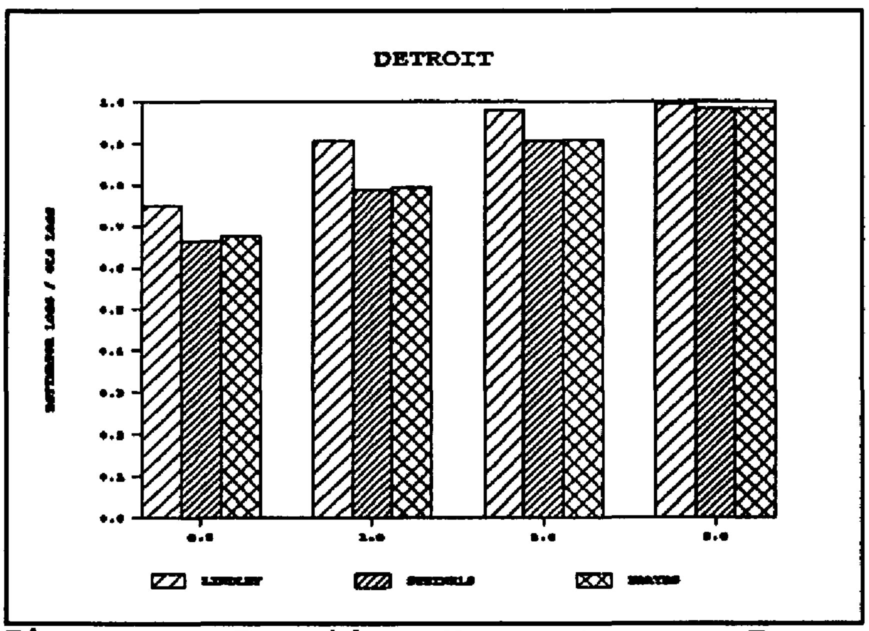

Figure 16 Ensemble Mean Square Error. Detroit. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

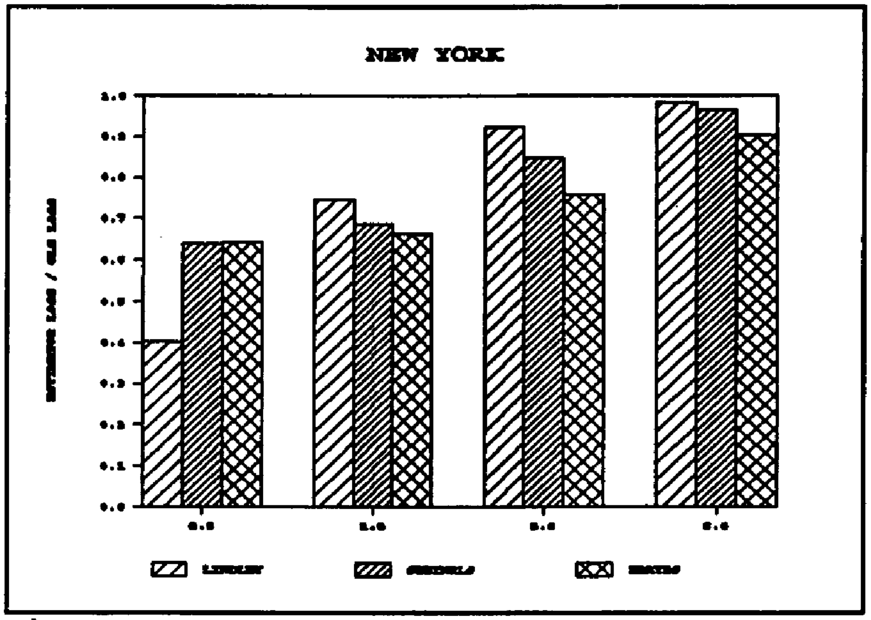

Figure 17 Ensemble Mean Square Error. New York. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

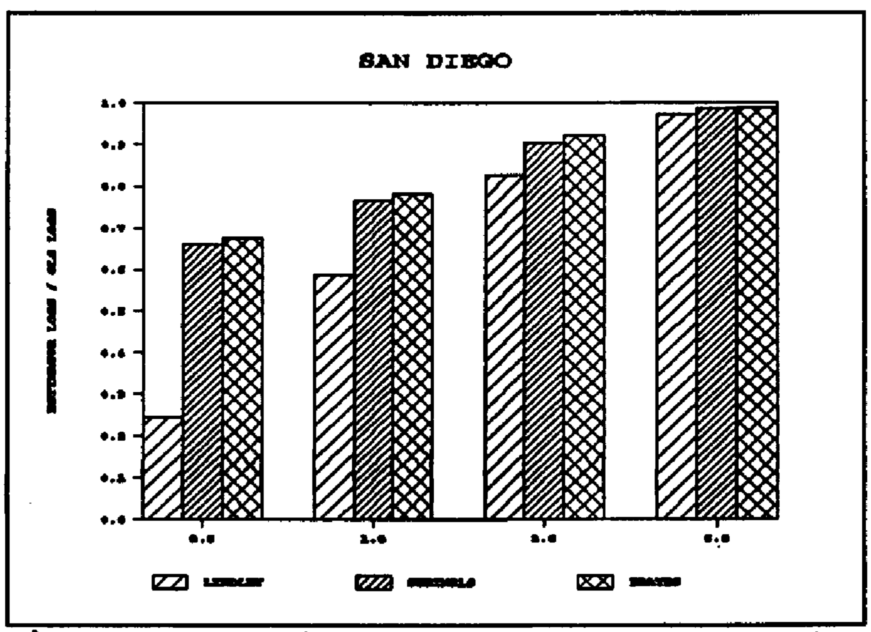

Figure 18 Ensemble Mean Square Error. San Diego. X-axis measures multiple of "true" parameter vector length. Y-axis measures ratio of estimator loss to OLS loss.

CHAPTER FOUR

SUMMARY AND CONCLUSIONS

Although Stein-like estimation has been available from econometric theory for some 30 years, and although this technique makes OLS estimates inadmissable (estimation using the Stein rule is never worse and is better than OLS in some circumstances), still the application of this technique to prediction problems has been infrequent. This dissertation explores the benefit from applying Stein-like estimation to the prediction of housing prices.

To evaluate the worth of the estimation technique in the context of predicting residential real estate prices, several Stein-like rules are considered alongside those econometric estimators that have been previously proposed for this application. The nine estimators considered are summarized in table 1. Specifically, the Stein-like rules of this study (STEIN, LINDLEY, STEINRLS, and EBAYES) are compared with Ordinary Least Squares (OLS), three ridge estimators (RIDGE, ITHKB, and STRAW), and a principal components method (PCSTEIN). The estimators are examined in the real estate settings in which they are most frequently encountered: the prediction of house price as in the appraisal or assessment process; and, the use of housing

price predictions in constructing residential price indexes, cross-sectionally or over time.

Comparisons are made using two different types of data. The first comparison is designed to rank estimator performance using the quality of data and the type of hedonic model available to appraisers and assessors attempting to predict housing prices. Multiple Listing Service data for Baton Rouge, Louisiana over five time periods are used, and the relative performance of the estimators is examined for two functional forms, linear and semilog.

With this data, the hedonic pricing model is highly descriptive, explaining between 85% and 93% of the total variation in housing prices. In this setting, the relative performance of the estimators is mixed. The iterative ridge rule offers the best (period 4, semilog, +15%) and the worst (period 3, semilog, -53%) performance as compared with OLS. Otherwise, the estimator performance is quite similar, showing modest benefits and costs over the five time periods of the study.

The second context in which the estimators are evaluated differs from the first in two ways. First, the data is considerably less descriptive, being drawn from the 1983 American Housing Survey for six cities. As a consequence, the resulting hedonic pricing model, while containing a greater number of variables, has less

explanatory power, with R-squares ranging from 42% to 70%. Second, the American Housing Survey data tests the relative effectiveness of the alternative estimators in a cross-sectional rather than in a time-series environment. It uses data from six major metropolitan areas, all taken from one nine-month time period.

In this second test setting for the estimators, differences in performances become discernable. While most of the estimators appear to offer little benefit over OLS, a few stand out as being either especially appealing or especially risky. The iterative ridge rule (ITHKB), offered no gains in two of the cities, and losses up to 30% as compared with OLS in the other four cities. On the other end of the spectrum, partitioned Stein-like estimation (STEINRLS and EBAYES) offers gains of up to 15% in four of the cities, and loses modest ground to OLS in the other two. In the cities where STEINRLS and EBAYES underperform OLS, so do all of the alternatives.

Recognizing the limitation on inferences that can be drawn from the limited number of estimator comparisons in the foregoing tests, a Monte Carlo experiment is designed as an alternative evaluation framework. The numerical simulations utilize the American Housing Survey data, and provide an opportunity to observe comparative estimator performance in a repeated sampling context.

Full sample data is used to arrive at the

specifications of the "true" parameter vector for each city. Then each city's data is divided into four subsets and estimator performances predicting housing prices in these four subsets are compared. Comparisons are conducted for four assumed loss functions: three quadratic variations of the mean square error criteria; and one popular non-quadratic loss measure, mean absolute percentage error. Also, model R-square is artificially adjusted in this experiment, providing performance observations at four separate signal-to-noise levels.

The design of the experiment thus provides the opportunity to compare estimator performances from three perspectives: 1) on a city to city basis; 2) under varying assumptions concerning the appropriate loss criteria; and, 3) at different levels of the model's assumed signal-to-noise ratio.

In these environments, three of the estimators (LINDLEY, STEINRLS, and EBAYES) stand out as clearly superior to the others evaluated. For all three mean square error criteria examined, these three estimators share the best performance relative to OLS for every city at every level of model R-square. Over the entire spectrum of evaluation, STEINRLS and EBAYES perform almost identically and slightly better than LINDLEY.

The effectiveness of these three estimators, as compared with OLS, ridge estimators, and the basic Stein-

like rules, is linked to the characteristic which is common to all three. That characteristic is the partitioning of the model and the imparting of non-sample information through this partitioning.

The dramatic reduction in risk (lower mean square error loss) using partitioned Stein-like techniques emphasizes the potential benefit from imparting non-sample information into the estimation process. In this experiment, the added information is the straightforward conjecture that many of the explanatory attributes in the hedonic model of housing carry little predictive content.

The substantial prediction gains achieved under MSE criteria by those estimators making use of this information suggest three aspects of the estimators probably influence this performance: 1) Variance in estimating the model is reduced as a result of the added information. This is true regardless of the quality of the information. 2) The bias caused by imposing the information is not severe. Clearly, the tradeoff between bias and variance in estimation for this model is favorable. 3) The Stein-like rules provide a superior method for adding non-sample information. These rules allow the extent to which added information is used in estimation to depend on the extent to which the data support the "truth" of the added information.

The markedly different performance between OLS and the partitioned Stein estimators provides encouragement for

fruitful extensions of this work on two grounds. First, the favorable bias-variance tradeoff from imposing restrictions on variables deemed to have less predictive content indicates that there is a certain degree of truth to the restrictions. This holds significant implications for improving the accuracy of hedonic price indexes. Second, the superior performance of the partitioned Stein estimators confirms the value of this technique for imparting non-sample information. The risk reduction is directly proportional to the quality of the information added to the estimation process.

The first and most natural extension of this work is to compare estimator performance with the full model used in index construction. There are two insights from the present research that make such an extension economically feasible. First, four of the compared estimators (RIDGE, ITHKB, STRAW, and STEIN) are not likely to perform very differently with the more elaborate model and can be excluded from consideration. Second, the knowledge that the estimators perform quite similarly across cities allows for a further reduction in scope.

The second possible extension is experimentation with different types of information that might be used to improve prediction. In this study, the benefit from using the Stein-like rules drops as the explanatory power of the model rises, implying limited usefulness in environments where

good data and good models are available. However, if the quality of the non-sample information is especially good (i.e. true or nearly so), then its addition to the estimation process can provide substantial reductions in risk. An example of such information might be the use of expert opinion, such as that of appraisers or brokers, to increase the precision of parameter estimation. Other possible sources of non-sample information are parameter coefficient estimates from previous research, knowledge of construction costs for physical attributes, and information about rental rates for comparable properties.

Finally, the Stein-like rules, introduced here to a residential real estate application, can be used beneficially in many other areas. There are many problems in economics and finance where estimation is adversely affected by the quality of the sample data, and where non-sample information of high quality might be available. Stein-like rules are ideally suited for these situations, as this dissertation has shown in the context of the hedonic model for housing.

REFERENCES

Abelson, Peter N. 1979. Property prices and the value of urban amenities. Journal of Environmental Economics and Management 6: 11-28.

Anderson, John E. 1981. Ridge estimation of house value determinants. Journal of Urban Economics 9: 286-297.

Atkinson, Scott E. and Thomas D. Crocker. 1987. A Bayesian approach to assessing the robustness of hedonic property value studies. Journal Of Applied Econometrics 2: 27-45.

Barish, Norman N. and Frederick H. Siff. 1969. Operational gaming simulation with application to a stock market. Management Science 15 (June): B530-541.

Bartik, Timothy J. 1987. The estimation of demand parameters in hedonic price models. Journal of Political Economy 95: 81-88.

Belsley, D., E. Kuh and R. Welsch. 1980. Regression diagnostics. New York: John Wiley.

Biddle, Jeff E. and Gary A. Zarkin. 1988. Worker preferences and market compensation for job risk. Review of Economics and Statistics 70 (November): 660-667.

Blackley, Dixie M., James R. Follain, and Haeduck Lee. 1986. An evaluation of hedonic price indexes for thirty-four large SMSAs. AREUEA Journal 14: 179-205.

Boyle, Phelim P. 1977. Options: A Monte Carlo approach. Journal of Financial Economics 4 (May): 323-388.

Brennan, Michael J., and Eduardo S. Schwartz. 1977. The valuation of American put options. Journal of Finance 32 (May): 449-462.

Brown, Gardner, Jr., and Robert Mendelsohn. 1984. The hedonic travel cost method. Review of Economics and Statistics 66 (August): 427-433.

Brown, Gardner and Henry O. Pollakowski. 1977. Economic valuation of shoreline. Review of Economics and Statistics: 272-278.

Brown, James N. and Harvey S. Rosen. 1982. On the estimation of structural hedonic price models. Econometrica 50 (May): 765-768.

Brown, R. J. 1974. On the selection of the best predictive model in multiple regression analysis." The Appraisal Journal (October): 572-578.

Case, Bradford and John M. Quigley. 1989. Statistical analysis of sales data to verify appraisal information. Property Tax Journal 8 (March): 15-25.

Church, Albert M. 1975. An econometric model for appraising. AREUEA Journal 3.

Clarkson, Geoffrey P. and Allan H. Meltzer. 1960. Portfolio selection: A heuristic approach. Journal of Finance 15 (December): 465-480.

Dale-Johnson, D. 1982. An alternative approach to housing market segmentation using hedonic price data. Journal of Urban Economics 11: 311-332.

Dinopoulos, Elias and Mordechai E. Kreinin. 1988. Effects of the U.S.-Japan auto VER on European prices and on U.S. welfare. Review of Economics and Statistics 70 (August): 484-491.

Epple, Dennis. 1987. Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. Journal of Political Economy 95: 59-80.

Farrar, D. and R. Glauber. 1967. Multicollinearity in regression analysis: The problem revisited. Review of Economics and Statistics 49: 92-107.

Ferreira, Eurico J. and G. Stacy Sirmans. 1988. Ridge regression in real estate analysis. The Appraisal Journal (July): 311-319.

Follain, James R., and Emmanuel Jimenez. 1985. Estimating the demand for housing characteristics: A survey and critique. Regional Science and Urban Economics 15: 77-107.

Follain, J. R., and S. Malpezzi. 1980. Dissecting housing value and rent: Estimates of hedonic indexes for thirty-nine large SMSA's. The Urban Institute Press.

Fomby, Thomas B., R. Carter Hill, and S. R. Johnson. 1984. Advanced econometric methods. New York: Springer-Verlag.

Gau, George W. and Daniel B. Kohlhepp. 1978. Multicollinearity and reduced form price equations for residential markets. AREUEA Journal (Spring): 50-69.

Gruenstein, John M. L. and Harindra de Silva. 1989. Hedonic index estimation for commercial buildings: Assessors and economists and the parallel search for the optimal functional form. Property Tax Journal 8 (March): 71-83.

Harrison, David, Jr. and Daniel L. Rubinfeld. 1978. Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management: 81-102.

Hartman, Raymond S. 1987. Product quality and market efficiency: The effect of product recalls on resale prices and firm valuation. Review of Economics and Statistics 69 (May): 367-372.

Hendry, D. F. 1984. Monte Carlo experimentation in econometrics. Chapter 16 of Handbook of econometrics (Vol. II) ed. Griliches, Z. and M. D. Intriligator. Amsterdam: North Holland.

Hill, R. Carter, Phillip A. Cartwright, and Julia F. Arbaugh. 1989. The use of biased predictors in marketing research. Working Paper. Louisiana State University.

Hill, R. Carter and Thomas B. Fomby. 1986. The effects of extrapolation on minimax Stein-rule prediction. Presented at the American Statistical Association Meetings. Chicago, Illinois.

Hoerl, Arthur E. and Robert W. Kennard. 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12 (February): 55-67.

________. 1970. Ridge regression: Applications to nonorthogonal problems. Technometrics 12 (February): 69-82.

Hoerl, A. E., R. W. Kennard and K. F. Baldwin. 1975. Ridge regression, some simulations. Communications in Statistics: A 4: 105-123.

Ibbotson, Roger G., Jeffery J. Diermeier and Laurence B. Siegel. 1984. The demand for capital market returns: A new equilibrium theory. Financial Analyst's Journal 40: 22-33.

Intriligator, Michael D. 1978. Econometric models, techniques and applications. Englewood Cliffs, NJ: Prentice-Hall.

Isakson, Hans R. 1986. The accuracy of arbitrage pricing versus hedonic pricing valuation methodologies in computer assisted mass appraisal systems. Property Tax Journal 5 (June): 97-109.

Jarrow, Robert, and Andrew Rudd, November 1982, Approximate Option Valuation for Arbitrary Stochastic Processes, Journal of Financial Economics 10, 347-369.

Jud, G. Donald. 1980. The effects of zoning on single family residential property values in Charlotte, North Carolina. Land Economics: 142-154.

Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. 1985. The theory and practice of econometrics. Second edition. New York: John Wiley and Sons.

Kalos, Malvin H. and Paula A. Whitlock. 1986. Monte Carlo methods. Volume I: Basics. New York: John Wiley and Sons.

Kennedy, Peter. 1985. A guide to econometrics. Second edition. Cambridge: MIT Press.

Kniesner, Thomas J. and John D. Leeth. 1988. Simulating hedonic labor market models: Computational issues and policy applications. International Economic Review 29 (November): 755-789.

Kowalski, Joseph G. and Peter F. Colwell. 1986. Market versus assessed values of industrial land. AREUEA Journal 14 (Summer): 361-373.

Krumm, Ronald J. 1980. Neighborhood amenities: An economic analysis. Journal of Urban Economics: 208-224.

Lancaster, Kelvin J. 1966. A new approach to consumer theory. Journal of Political Economics (April): 132-156.

Leamer, E. E. 1983. Model choice and specification analysis. Chapter 5 of Handbook of econometrics (Vol. I) ed. Griliches, Z. and M. D. Intriligator. Amsterdam: North Holland.

Linneman, Peter. 1981. The demand for residence site characteristics. Journal of Urban Economics: 129-148.

Malpezzi, S., L. Ozanne and T. Thibodeau. 1980. Characteristic prices of housing in fifty-nine metropolitan areas. Urban Institute Working Paper 1367-1. The Urban Institute Press.

Mark, Jonathan and Michael A. Goldberg. 1988. Multiple regression analysis and mass assessment: A review of the issues. The Appraisal Journal (January): 89-109.

McDonald, Gary and Diane I. Galarneau. 1975. A Monte Carlo evaluation of some ridge type estimators. Journal of the American Statistical Association 70 (June): 407-416.

McMillan, Melville L., Bradford G. Reid and David W. Dilen. 1980. An extension of the hedonic approach for estimating the value of quiet. Land Economics: 315-328.

Miezskowski, P., and A. Saper. 1978. An estimate of the effects of airport noise on property values. Journal of Urban Economics 5: 425-440.

Mikhail, W. M. 1975. A comparative Monte Carlo study of the properties of econometric estimators. Journal of the American Statistical Association 70 (March): 94-104.

Miller, N. G. 1982. Residential property hedonic pricing models: A review. In Research in Real Estate 2 ed. C. F. Sirmans.

Moore, James S., Alan K. Reichert, and Chien-Ching Cho. 1984. Analyzing the temporal stability of appraisal model coefficients: An application of ridge regression techniques. AREUEA Journal 12: 50-71.

Morton, T. Gregory. 1977. Factor analysis, multicollinearity, and regression appraisal models. The Appraisal Journal (October): 578-588.

Muth, Richard F. 1966. Household production and consumer demand functions. Econometrica 34 (July): 699-708.

Nagar, A. L. 1960. A Monte Carlo study of alternative simultaneous equation estimators. Econometrica 28 (July): 573-590.

Nelson, Jon P. 1978. Residential choice, hedonic prices, and the demand for urban air quality. Journal of Urban Economics: 357-369.

Ohsfeldt, Robert L. 1988. Implicit markets and the demand for housing characteristics. Regional Science and Urban Economics 18: 321-343.

Ohsfeldt, Robert L. and Barton A. Smith. 1985. Estimating the demand for heterogeneous goods. Review of Economics and Statistics 61: 165-171.

________. 1988. Assessing the accuracy of structural parameter estimates in analyses of implicit markets. Land Economics 64 (May): 135-146.

Pace, R. Kelley, and Otis W. Gilley. 1990. Estimation employing a priori information within mass appraisal and hedonic pricing models. Journal of Real Estate Finance and Economics 3: 55-72.

Pardew, Jolie B. 1988. Estimating how quality characteristics and marketing services affect alfalfa hay prices. Agribusiness 4 (March): 167-175.

Parkinson, Michael. 1977. Option pricing: The American put. Journal of Business 50 (January): 21-36.

Perry, Larry G., Cronan, Timothy P., and Epley, Donald R. 1986. Ranking comparable properties prior to their use in regression on a large or small sample. The Appraisal Journal: 57-65.

Reichert, Alan K., and James S. Moore. 1986. Using latent root regression to identify nonpredictive collinearity in statistical appraisal models. AREUEA Journal 14: 136-152.

Ridker, R. G., and John A. Henning. 1967. The determinants of residential property value with special references to air pollution. Review of Economics and Statistics: 246-255.

Rosen, Sherwin. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy 82: 34-55.

Smith, V. Kerry and Yoshiaki Kaoru. 1987. The hedonic travel cost model: A view from the trenches. Land Economics 63 (May): 179-192.

Sowey, E. R. 1973. A classified bibliography of Monte Carlo studies in econometrics. Journal of Econometrics 1: 377-395.

Strawderman, W. E. 1978. Minimax adaptive generalized ridge regression estimators. J.A.S.A. 73: 623-627.

Sweetland, Doug and William Colclough. 1986. Ridge regression: A word of caution. The Appraisal Journal: 294-300.

Thaler, Richard. 1978. A note on the value of crime control: Evidence from the property market. Journal of Urban Economics 5: 137-145.

Thibodeau, Thomas G. 1989. Housing price indexes from the 1974-1983 SIMSA Annual Housing Surveys. AREUEA Journal 1: 100-117.

Thibodeau, Thomas G. and Kerry D. Vandell. 1985. Using multiple regression analysis to determine the accuracy of mass appraisals. Property Tax Journal: 119-141.

Trajtenberg, Manuel. 1989. The welfare analysis of product innovations, with an application to computed tomography scanners. Journal of Political Economy 97 (April): 444-479.

Varian, Hal R. 1974. A Bayesian approach to real estate assessment. Studies in Bayesian Econometrics and Statistics: 195-208.

Wagner, H. 1958. A Monte Carlo study of estimates of simultaneous linear structural equations. Econometrica 26 (January): 117-133.

APPENDIX A

PREDICTION WITH THE HEDONIC PRICING MODEL: THE EFFECT OF SIMULTANEITY

In applications of the hedonic pricing model, the problem of simultaneity is almost always discussed. Rosen [1974], in his pioneering theoretical work, recognizes the simultaneity and identification problems in studies which use a simple hedonic approach. He suggests the use of a two-step procedure or the use of data from multiple markets to overcome the difficulty. Since his seminal research, several authors have considered the issue of simultaneity and have proposed methods of dealing with the problem in the context of the hedonic model.[1] Ironically, many studies applying the hedonic pricing model take steps aimed at correcting simultaneity problems when no such problems exist.[2] This appendix addresses the pertinence of the simultaneity issue

---

[1] For example, Brown and Rosen [1982] point out that this method does not solve the problem unless the functional form of the equation is specified, and Bartik [1987] notes that in addition to the demand-supply simultaneity problem there also exists a price-quantity demanded simultaneity problem in most hedonic regressions. Dale-Johnson [1982] proposes segmenting the data by location, structural characteristics or income as an alternative to Rosen's [1974] approach to the problem of multiple equilibria. Epple [1987], proposes a stochastic structure for a hedonic equilibrium model.

[2] See Follain and Jimenez [1985] for examples of this.

to the current study.

Imbedded within the hedonic pricing model are three distinct facets of simultaneity.[3] First is the fact that the demand for and the supply of the characteristics are being simultaneously determined. Therefore, without further information, it is impossible to determine if the price-quantity locus observed is representative of demand, supply, or the interaction of the two.

Second, when measuring a single market of finite size, an aggregation bias is introduced by the fact that the supply of attributes in any one product is determined at the same time as the total supply of those attributes in the market as a whole. Finally, price and quantity of each of the measured attributes are being simultaneously determined in estimating the model. This leads to non-linearity in the price parameter if diminishing marginal utility for the characteristics is assumed.

In this study, supply/demand simultaneity is not considered problematic. When the time period of estimation is of short duration, the supply of residential real estate can be considered fixed, or nearly so. All potential sets of suppliers of residential real estate are impeded from expeditious response to price changes. The home builders' reaction is slowed by the time needed for new construction;

---

[3] These are discussed at length by Follain and Jimenez [1985] in their survey of the literature on estimating housing demand characteristics.

current home owners who might be induced to enter the market are restricted by the high transactions cost of selling and moving; and landlords are unable to respond quickly because of binding lease agreements. No estimation period in this study exceeds one year. Thus, it is deemed reasonable to assume that supply is fixed.

Simultaneity attributable to aggregation is also considered inconsequential in the current application. All of the samples chosen for estimation purposes comprise only a small portion of the total geographic market available to home buyers. Thus these samples also contribute only a small percentage of the total attributes available.

The bias caused by the simultaneous determination of price and attribute quantity for any single observation is germane to the predictive purpose of this research. For example, if number of bedrooms enters the hedonic regression, linearity of the parameter requires that the marginal price of an additional bedroom be constant. It does not seem reasonable to assume that the representative consumer of housing services would pay equal amounts for the third bedroom and for the eighth bedroom. Fortunately, many of the proxies for characteristics enter the regression as binary choice variables and are not affected by this simultaneity. Elsewhere, the problem is recognized and the model is transformed to diminish the adverse impact of non-linearity.

It should be noted that the emphasis in this study is not on predicting the value of a particular attribute, not on predicting the elasticity of demand for a given characteristic, and not on the interpretation of the hedonic prices as the marginal prices of the specified attributes. The emphasis here is on predicting the equilibrium price of the composite product. To achieve this, it is not necessary to control the interpretation of the components of this equilibrium price to be either bid prices or offer prices. Such a restriction in fact may be counterproductive.

APPENDIX B

MULTICOLLINEARITY

Multicollinearity is recognized as a problem frequently encountered when applying the hedonic model to housing. The issue of collinearity in this context has been the focus of a number of studies.[1] It is, in fact, the motivation for the first application of biased estimation in a residential real estate context (Anderson [1981]).

Several methods are available for detecting and assessing the severity of collinearity. One of these, the condition index technique proposed by Belsley, Kuh, and Welsch [1980], is used for the purposes of this study and is explained in chapter two. This appendix defines multicollinearity, and describes other popular methods of its identification.

The classical linear regression model assumes that the matrix of independent variables is of full column rank. This condition is only rarely fully met even in the experimental sciences, much less in the social sciences where the investigator has little or no control over the value of the explanatory variables. On the other hand, the condition is rarely completely violated by the ability to express one

---

[1] See, for example, Morton [1977], Gau and Kohlhepp [1978], Perry, Cronan and Epley [1986], Reichert and Moore [1986], Atkinson and Crocker [1987], and Mark and Goldberg [1988].

variable as an exact scalar multiple of one of the others. Multicollinearity refers to the covariation among explanatory variables, and its presence can mask the true relationship between the dependent and independent variables.

The most popular method of detecting a collinearity problem is a simple inspection of the correlation matrix for the explanatory variables. If there are more than two independent variables, however, this method is inadequate since it does not reveal the more complex linear relations among three or more variables. High simple correlations are a sufficient but not a necessary condition for the problem of multicollinearity.

Inverting the correlation matrix and examining the diagonal elements provides one method of interpreting the interrelationships among variables. The diagonal elements of this matrix are

$$\frac{1}{(1 - R^2_i)} \qquad \text{B.1}$$

where the R^2_i represents the coefficient of determination resulting from regressing the ith independent variable on all of the other independent variables. The higher the r-squared, the smaller the denominator. Thus large numbers for the diagonal elements, called variance inflation factors, imply the existence of multicollinearity. Kennedy [1985] suggests that values larger than 10, as a rule of thumb, might be considered as indicative of a problematic relationship among

independent variables.

Farrar and Glauber [1967] propose that the severity of collinearity in data can best be measured in terms of the degree to which the data departs from orthogonality. If X'X is in correlation matrix form, the determinant of this matrix approaches one as orthogonality is approached and one as singularity is approached. Thus, inspection of the determinant of this matrix provides intuition into the collinearity problem.

APPENDIX C

SUPPLEMENTAL TABLES

This appendix supplements the results and tables of chapters two and three. Tables included here provide parameter estimates for four of the five time periods and five of the six cities studied in chapter two; comparative estimator performance under mean absolute percentage error and partitioned Stein-rule estimator performance for three of the four subsamples of the Monte Carlo experiment in chapter three.

Tables 9 and 15 in chapter two provide the parameter estimates of the nine compared estimators in time series and cross-sectional settings respectively. Since individual parameter vector element estimates are not a principal concern of that chapter, the results only for period 5 (Table 9) and Boston (Table 15) are provided as part of the main text. Tables 20 through 23 display the parameter estimate results for periods 1 through 4, and tables 24 through 28 provide the same information for Chicago, Dallas, Detroit, New York, and San Diego.

In chapter three, it is found by Monte Carlo experiment that none of the compared estimators outperforms Ordinary Least Squares under Mean Absolute Percentage Error (MAPE).

The specific results under this loss criterion for the four subsamples appear in tables 29 through 32 of this appendix.

Tables 17 through 19 in chapter three report results for the partitioned Stein-rule estimators under three quadratic loss criteria; Mean Square Error of In-Sample Prediction, Mean Square Error, and Ensemble Mean Square Error. These tables show only the results for the first subsample, because the results in other subsamples are quite similar. Results for the other three subsamples under these three loss criteria are included here as tables 33 through 41.

TABLE 20

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
BATON ROUGE DATA
RESULTS FOR TIME PERIOD 1

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.31520	10.36150	10.71340	10.32840	10.32880	10.31520	10.31510	10.35450	10.35100
LIVAREA	0.00053	0.00050	0.00025	0.00052	0.00052	0.00053	0.00053	0.00054	0.00054
AGE	-0.00458	-0.00429	-0.00207	-0.00452	-0.00452	-0.00458	-0.00458	-0.00466	-0.00465
DPOOL	0.04896	0.04995	0.04472	0.04841	0.04839	0.04896	0.04896	0.04781	0.04791
DISTLSU	-0.00978	-0.01039	-0.01245	-0.00967	-0.00966	-0.00978	-0.00978	-0.01019	-0.01015
CASH	-0.02171	-0.01734	0.01406	-0.02146	-0.02146	-0.02171	-0.02171	-0.01878	-0.01904
BR	-0.03840	-0.02290	0.07677	-0.03797	-0.03796	-0.03840	-0.03840	-0.02943	-0.03023
FULLBATH	0.00616	0.01480	0.07555	0.00609	0.00609	0.00616	0.00616	0.00472	0.00485
NETAREA	0.00007	0.00010	0.00021	0.00007	0.00007	0.00007	0.00007	0.00005	0.00005
LOTSIZE	0.13889	0.17245	0.39390	0.13732	0.13728	0.13889	0.13889	0.10643	0.10932
GARAGE	0.17870	0.16089	0.08566	0.17668	0.17663	0.17870	0.17870	0.13693	0.14065
COVTPRCH	0.00017	0.00015	0.00018	0.00017	0.00017	0.00017	0.00017	0.00013	0.00014
CARPORT	0.07442	0.06328	0.01290	0.07358	0.07356	0.07442	0.07442	0.05703	0.05857
DBRICK	-0.01640	-0.01845	-0.02328	-0.01621	-0.01621	-0.01640	-0.01640	-0.01257	-0.01291
DFIREPL	0.01223	0.01786	0.05443	0.01209	0.01209	0.01223	0.01223	0.00937	0.00962
SPECFIN	0.01036	0.01076	-0.00354	0.01025	0.01024	0.01036	0.01036	0.00794	0.00816
DCORP	0.01760	0.01721	0.02665	0.01741	0.01740	0.01760	0.01760	0.01349	0.01386
DSBP	-0.00964	-0.00711	0.01774	-0.00953	-0.00953	-0.00964	-0.00964	-0.00739	-0.00759

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	-32795.40	-26207.20	23365.30	-28500.90	-28512.80	-21662.90	-32701.70	-28215.90	-28584.20
LIVAREA	74.17	69.23	31.57	71.89	71.90	70.65	73.96	73.27	73.34
AGE	-762.20	-710.96	-298.11	-738.79	-738.86	-720.30	-760.03	-705.47	-710.03
DPOOL	11440.40	11522.50	8936.34	11089.00	11089.90	7544.93	11407.70	10707.00	10766.00
DISTLSU	-881.47	-979.65	-1384.84	-854.39	-854.47	-1260.72	-878.95	-1002.14	-992.44
CASH	925.42	1441.36	4515.23	896.99	897.07	613.99	922.77	1467.70	1424.09
BR	-12132.70	-9787.28	6910.14	-11760.00	-11761.10	-7981.32	-12098.10	-8270.06	-8580.68
FULLBATH	-289.30	1013.86	10287.60	-280.41	-280.44	-177.63	-288.47	-197.20	-204.60
NETAREA	-17.84	-10.02	17.66	-17.29	-17.29	-12.93	-17.79	-12.16	-12.62
LOTSIZE	23347.30	27187.10	51491.30	22630.10	22632.10	15400.70	23280.60	15914.30	16512.00
GARAGE	21837.30	18595.80	8191.01	21166.50	21168.30	14410.90	21774.90	14885.00	15444.10
COVTPRCH	46.39	40.43	29.08	44.96	44.97	42.31	46.25	31.62	32.81
CARPORT	8816.66	6861.92	-525.48	8545.83	8546.58	5778.89	8791.46	6009.72	6235.44
DBRICK	-6675.56	-6979.65	-6393.80	-6470.50	-6471.07	-4431.04	-6656.49	-4550.27	-4721.18
DFIREPL	-6600.56	-5576.72	2571.96	-6397.80	-6398.36	-4362.06	-6581.70	-4499.15	-4668.14
SPECFIN	2664.88	2654.18	-216.38	2583.02	2583.25	1742.94	2657.27	1816.47	1884.69
DCORP	-8630.68	-8354.07	-3341.51	-8365.56	-8366.30	-5691.81	-8606.02	-5882.95	-6103.91
DSBP	-1809.76	-1503.27	2774.50	-1754.16	-1754.32	-1193.58	-1804.59	-1233.59	-1279.92

TABLE 21

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
BATON ROUGE DATA
RESULTS FOR TIME PERIOD 2

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.43390	10.46720	10.55190	10.56110	10.58120	10.43390	10.41540	10.46290	10.46030
LIVAREA	0.00044	0.00042	0.00034	0.00039	0.00038	0.00044	0.00044	0.00049	0.00049
AGE	-0.00693	-0.00671	-0.00611	-0.00610	-0.00597	-0.00693	-0.00692	-0.00821	-0.00810
DPOOL	0.06701	0.07211	0.08483	0.05901	0.05775	0.06701	0.06689	0.05086	0.05229
DISTLSU	-0.00993	-0.01022	-0.01081	-0.00874	-0.00855	-0.00993	-0.00991	-0.01048	-0.01043
CASH	0.02433	0.02700	0.03267	0.02143	0.02097	0.02433	0.02429	0.01947	0.01990
BR	-0.05785	-0.04320	-0.00893	-0.05095	-0.04986	-0.05785	-0.05775	-0.03609	-0.03802
FULLBATH	0.04879	0.05853	0.08020	0.04297	0.04205	0.04879	0.04871	0.03044	0.03206
NETAREA	0.00029	0.00028	0.00026	0.00025	0.00025	0.00029	0.00029	0.00018	0.00019
LOTSIZE	0.33831	0.37999	0.46481	0.29796	0.29157	0.33831	0.33771	0.21106	0.22231
GARAGE	0.00238	0.01311	0.03369	0.00209	0.00205	0.00238	0.00237	0.00148	0.00156
COVTPRCH	-0.00005	-0.00002	0.00005	-0.00004	-0.00004	-0.00005	-0.00005	-0.00003	-0.00003
CARPORT	-0.01581	-0.01345	-0.00955	-0.01392	-0.01362	-0.01581	-0.01578	-0.00986	-0.01039
DBRICK	-0.00662	-0.00786	-0.01123	-0.00583	-0.00570	-0.00662	-0.00660	-0.00413	-0.00435
DFIREPL	0.06209	0.06802	0.07972	0.05468	0.05351	0.06209	0.06198	0.03874	0.04080
SPECFIN	-0.06040	-0.06423	-0.07225	-0.05320	-0.05206	-0.06040	-0.06029	-0.03768	-0.03969
DCORP	-0.14971	-0.14378	-0.13014	-0.13185	-0.12903	-0.14971	-0.14945	-0.09340	-0.09838
DSBP	-0.00949	-0.00983	-0.01097	-0.00835	-0.00818	-0.00949	-0.00947	-0.00592	-0.00623

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	-11345.70	-5924.21	8254.37	7829.45	5736.96	-87.17	-10911.40	-9239.74	-9239.74
LIVAREA	53.94	48.99	36.68	45.35	46.28	59.73	51.88	68.31	68.31
AGE	-1127.23	-1078.10	-919.62	-947.57	-967.18	-1171.06	-1084.08	-1292.05	-1292.05
DPOOL	12662.70	14485.40	17739.20	10644.50	10864.80	8.21	12177.90	11368.70	11368.70
DISTLSU	-1053.98	-1124.62	-1241.65	-886.00	-904.33	-1765.16	-1013.63	-1322.26	-1322.26
CASH	2949.98	3358.62	4260.94	2479.82	2531.12	-1.26	2837.04	1899.46	1899.46
BR	-4508.09	-2677.80	2139.70	-3789.60	-3868.00	70.63	-4335.51	0.00	0.00
FULLBATH	8700.17	10004.10	12953.90	7313.55	7464.87	73.30	8367.10	0.00	0.00
NETAREA	6.59	11.45	19.86	5.54	5.66	11.01	6.34	0.00	0.00
LOTSIZE	36747.00	42033.10	53306.30	30890.40	31529.50	-7.15	35340.30	0.00	0.00
GARAGE	19381.70	18340.00	16346.90	16292.70	16629.80	13.20	18639.80	0.00	0.00
COVTPRCH	25.31	23.31	22.83	21.27	21.71	20.01	24.34	0.00	0.00
CARPORT	1928.81	1114.28	-388.06	1621.40	1654.95	-113.66	1854.97	0.00	0.00
DBRICK	387.92	199.00	-432.09	326.09	332.84	-89.76	373.06	0.00	0.00
DFIREPL	676.39	1529.04	3722.30	568.59	580.35	-46.68	650.50	0.00	0.00
SPECFIN	-4026.01	-4431.92	-5525.97	-3384.35	-3454.38	-46.41	-3871.89	0.00	0.00
DCORP	-8898.12	-8646.86	-7969.86	-7479.95	-7634.71	-5.40	-8557.47	0.00	0.00
DSBP	-7350.39	-7156.07	-6637.46	-6178.90	-6306.74	0.05	-7069.00	0.00	0.00

TABLE 22

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
BATON ROUGE DATA
RESULTS FOR TIME PERIOD 3
A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.34410	10.38740	10.88960	10.35560	10.35590	10.34410	10.34390	10.55180	10.55180
LIVAREA	0.00051	0.00048	0.00017	0.00050	0.00050	0.00051	0.00051	0.00054	0.00054
AGE	-0.00091	-0.00087	-0.00105	-0.00090	-0.00090	-0.00091	-0.00091	-0.00440	-0.00440
DPOOL	0.07143	0.07185	0.07539	0.07073	0.07070	0.07143	0.07143	0.05844	0.05844
DISTLSU	-0.01264	-0.01325	-0.01268	-0.01251	-0.01251	-0.01264	-0.01264	-0.01544	-0.01544
CASH	-0.03283	-0.03682	-0.04702	-0.03251	-0.03250	-0.03283	-0.03283	-0.03124	-0.03124
BR	-0.02298	-0.00908	0.10030	-0.02276	-0.02275	-0.02298	-0.02298	0.00000	0.00000
FULLBATH	-0.05874	-0.03989	0.09992	-0.05816	-0.05815	-0.05874	-0.05874	0.00000	0.00000
NETAREA	0.00020	0.00020	0.00019	0.00020	0.00020	0.00020	0.00020	0.00000	0.00000
LOTSIZE	0.05388	0.07657	0.27380	0.05335	0.05334	0.05388	0.05388	0.00000	0.00000
GARAGE	0.02570	0.02264	0.02364	0.02545	0.02544	0.02570	0.02570	0.00000	0.00000
COVTPRCH	0.00005	0.00008	0.00023	0.00005	0.00005	0.00005	0.00005	0.00000	0.00000
CARPORT	0.01774	0.01505	0.00279	0.01756	0.01756	0.01774	0.01774	0.00000	0.00000
DBRICK	-0.03345	-0.03182	-0.00539	-0.03312	-0.03311	-0.03345	-0.03345	0.00000	0.00000
DFIREPL	0.06349	0.06676	0.06337	0.06287	0.06285	0.06349	0.06349	0.00000	0.00000
SPECFIN	-0.03329	-0.03470	-0.05735	-0.03296	-0.03295	-0.03329	-0.03329	0.00000	0.00000
DCORP	-0.01551	-0.01785	-0.04916	-0.01536	-0.01535	-0.01551	-0.01551	0.00000	0.00000
DSBP	0.08300	0.08153	0.04179	0.08219	0.08216	0.08300	0.08300	0.00000	0.00000

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	-24868.60	-18465.80	32783.40	-23158.10	-23120.20	-85.98	-24808.40	-12811.20	-12811.20
LIVAREA	65.50	61.25	25.74	64.68	64.66	58.38	65.34	66.12	66.12
AGE	-200.38	-191.44	-156.94	-197.86	-197.81	-475.21	-199.89	-460.19	-460.19
DPOOL	15286.10	15318.30	14380.70	15094.10	15089.90	9.30	15249.10	18003.30	18003.30
DISTLSU	-1141.95	-1240.89	-1468.88	-1127.61	-1127.29	-1976.01	-1139.18	-1229.09	-1229.09
CASH	-6206.78	-6858.71	-8142.08	-6128.84	-6127.11	-2.39	-6191.75	-6797.82	-6797.82
BR	-7203.44	-5027.01	10321.30	-7112.99	-7110.98	56.23	-7186.00	0.00	0.00
FULLBATH	-4137.47	-1524.70	14441.30	-4085.51	-4084.36	52.32	-4127.45	0.00	0.00
NETAREA	16.89	18.69	23.22	16.68	16.67	7.89	16.85	0.00	0.00
LOTSIZE	-20892.20	-17218.50	18042.90	-20629.90	-20624.10	-7.10	-20841.60	0.00	0.00
GARAGE	9316.92	8187.55	6033.51	9199.93	9197.34	-0.71	9294.36	0.00	0.00
COVTPRCH	22.59	23.91	32.45	22.31	22.30	27.26	22.53	0.00	0.00
CARPORT	680.87	-14.81	-1506.77	672.32	672.13	-80.54	679.22	0.00	0.00
DBRICK	-1478.39	-1284.25	266.82	-1459.83	-1459.42	-49.12	-1474.81	0.00	0.00
DFIREPL	-334.07	280.98	3896.38	-329.87	-329.78	-45.68	-333.26	0.00	0.00
SPECFIN	-1173.98	-1399.59	-4813.92	-1159.24	-1158.91	-26.68	-1171.14	0.00	0.00
DCORP	995.78	609.87	-4655.65	983.27	983.00	-13.31	993.37	0.00	0.00
DSBP	13614.60	13333.50	8282.50	13443.60	13439.80	-4.62	13581.60	0.00	0.00

TABLE 23

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
BATON ROUGE DATA
RESULTS FOR TIME PERIOD 4

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.43450	10.46210	10.70470	10.60150	10.58590	10.43450	10.41060	10.45810	10.45610
LIVAREA	0.00055	0.00053	0.00030	0.00047	0.00048	0.00055	0.00055	0.00056	0.00056
AGE	-0.00515	-0.00510	-0.00408	-0.00437	-0.00444	-0.00515	-0.00513	-0.00542	-0.00539
DPOOL	-0.00367	0.00581	0.07713	-0.00312	-0.00317	-0.00367	-0.00366	0.00444	0.00374
DISTLSU	-0.01841	-0.01855	-0.01716	-0.01563	-0.01589	-0.01841	-0.01836	-0.01903	-0.01898
CASH	0.00277	0.00676	0.03341	0.00235	0.00239	0.00277	0.00276	0.01023	0.00959
BR	-0.02616	-0.01513	0.06898	-0.02222	-0.02259	-0.02616	-0.02610	-0.01978	-0.02033
FULLBATH	-0.02518	-0.01312	0.08048	-0.02139	-0.02174	-0.02518	-0.02512	-0.01904	-0.01957
NETAREA	0.00013	0.00014	0.00019	0.00011	0.00011	0.00013	0.00013	0.00010	0.00010
LOTSIZE	0.14399	0.16758	0.33352	0.12229	0.12432	0.14399	0.14366	0.10886	0.11190
GARAGE	-0.09316	-0.08047	0.00491	-0.07912	-0.08043	-0.09316	-0.09295	-0.07043	-0.07240
COVTPRCH	-0.00004	-0.00003	0.00011	-0.00003	-0.00003	-0.00004	-0.00004	-0.00003	-0.00003
CARPORT	-0.03457	-0.03154	-0.01336	-0.02936	-0.02985	-0.03457	-0.03449	-0.02614	-0.02687
DBRICK	-0.02012	-0.01925	-0.01463	-0.01709	-0.01737	-0.02012	-0.02008	-0.01521	-0.01564
DFIREPL	0.10912	0.10657	0.08629	0.09268	0.09421	0.10912	0.10887	0.08250	0.08480
SPECFIN	-0.01709	-0.02151	-0.05498	-0.01451	-0.01475	-0.01709	-0.01705	-0.01292	-0.01328
DSBP	-0.19269	-0.18856	-0.14067	-0.16366	-0.16637	-0.19269	-0.19225	-0.14568	-0.14975

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	-5356.58	-1269.07	36488.20	-2597.42	2707.57	-109.23	-5133.00	-7020.67	-7020.67
LIVAREA	72.75	68.66	28.47	71.11	67.96	65.37	69.72	68.23	68.23
AGE	-726.54	-719.10	-466.07	-710.17	-678.70	-799.31	-696.21	-766.02	-766.02
DPOOL	3513.61	4863.65	14494.20	3434.45	3282.25	5.44	3366.95	7669.02	7669.02
DISTLSU	-1930.16	-1958.91	-1577.82	-1886.67	-1803.06	-2380.12	-1849.59	-2154.41	-2154.41
CASH	-2076.65	-1491.15	2688.49	-2029.86	-1939.91	-7.81	-1989.97	670.47	670.47
BR	-7022.22	-5282.73	9291.31	-6864.01	-6559.83	67.59	-6729.11	0.00	0.00
FULLBATH	-3646.13	-1837.52	12736.90	-3563.99	-3406.05	83.00	-3493.94	0.00	0.00
NETAREA	9.50	11.28	22.48	9.28	8.87	-1.49	9.10	0.00	0.00
LOTSIZE	-26851.90	-22638.60	18597.50	-26247.00	-25083.80	-10.60	-25731.10	0.00	0.00
GARAGE	-10578.40	-8976.78	3641.95	-10340.10	-9881.89	-3.05	-10136.90	0.00	0.00
COVTPRCH	7.08	7.81	23.00	6.93	6.62	15.65	6.79	0.00	0.00
CARPORT	-4971.29	-4673.49	-2032.01	-4859.28	-4643.94	-124.31	-4763.78	0.00	0.00
DBRICK	-4516.91	-4302.85	-2366.62	-4415.15	-4219.49	-78.99	-4328.38	0.00	0.00
DFIREPL	3999.31	3762.69	3965.40	3909.20	3735.97	-46.14	3832.37	0.00	0.00
SPECFIN	-1486.83	-2147.60	-7404.89	-1453.33	-1388.93	-37.06	-1424.77	0.00	0.00
DSBP	-21705.00	-21179.60	-13190.30	21216.00	-20275.80	-7.94	-20799.80	0.00	0.00

TABLE 24

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR CHICAGO

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.5933	10.8376	11.1884	11.2396	10.6824	10.5933	10.5625	10.6853	10.7172
CENCITY	-0.2018	-0.1953	-0.0881	0.0000	-0.1739	-0.2018	-0.2012	-0.2241	-0.2319
BATHMORE	0.2671	0.2985	0.1359	0.0000	0.2303	0.2671	0.2663	0.3017	0.3137
ROOMMORE	0.1212	0.0859	0.0104	0.0000	0.1045	0.1212	0.1208	0.0979	0.0899
AGE	0.0220	-0.0052	-0.0002	0.0000	0.0190	0.0220	0.0220	0.0166	0.0147
STRUCTUR	0.0059	0.0067	-0.0054	0.0000	0.0051	0.0059	0.0059	0.0030	0.0019
EXCNBHD	-0.1485	-0.0828	0.0701	0.0000	-0.1280	-0.1485	-0.1481	-0.0468	-0.0116
BATH1	0.0401	0.0529	0.0053	0.0000	0.0345	0.0401	0.0399	0.0294	0.0257
BATH2	0.0885	0.0867	0.0068	0.0000	0.0763	0.0885	0.0883	0.0649	0.0568
ROOM4	0.5363	0.2962	-0.0788	0.0000	0.4624	0.5363	0.5348	0.3934	0.3440
ROOM5	0.6630	0.4255	-0.0365	0.0000	0.5716	0.6630	0.6610	0.4863	0.4252
BED1	0.0766	0.0899	0.0024	0.0000	0.0661	0.0766	0.0764	0.0562	0.0491
BED2	0.0205	0.0260	-0.0399	0.0000	0.0177	0.0205	0.0204	0.0150	0.0131
BED4	-0.0619	-0.0207	0.0554	0.0000	-0.0534	-0.0619	-0.0617	-0.0454	-0.0397
BEDMORE	-0.0565	-0.0363	0.0041	0.0000	-0.0488	-0.0565	-0.0564	-0.0415	-0.0363
ATTCHD	-0.1165	-0.1015	-0.0218	0.0000	-0.1005	-0.1165	-0.1162	-0.0855	-0.0747
GARAGE	0.0628	0.0627	0.0503	0.0000	0.0541	0.0628	0.0626	0.0460	0.0403
BASEMENT	0.1173	0.1120	0.0303	0.0000	0.1011	0.1173	0.1169	0.0860	0.0752
AGE_SQ	-0.0017	-0.0001	0.0000	0.0000	-0.0015	-0.0017	-0.0017	-0.0012	-0.0011
AGE_CUBE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRIOR40	-0.1066	-0.2028	-0.0584	0.0000	-0.0919	-0.1066	-0.1063	-0.0782	-0.0684
ROOMHEAT	-0.3260	-0.3167	-0.2013	0.0000	-0.2811	-0.3260	-0.3251	-0.2391	-0.2091
STMHEAT	0.1248	0.1150	0.0251	0.0000	0.1076	0.1248	0.1244	0.0915	0.0800
ROOMAIR	0.0436	0.0389	-0.0348	0.0000	0.0376	0.0436	0.0435	0.0320	0.0279
CENAIR	0.1224	0.1246	0.0680	0.0000	0.1055	0.1224	0.1220	0.0898	0.0785
RMWOHT	-0.0803	-0.0835	-0.0484	0.0000	-0.0693	-0.0803	-0.0801	-0.0589	-0.0515
NOPRIV	-0.0228	-0.0001	0.0044	0.0000	-0.0196	-0.0228	-0.0227	-0.0167	-0.0146
RMWOELEC	0.0356	0.0342	0.0037	0.0000	0.0307	0.0356	0.0355	0.0261	0.0228
TEN	-0.0125	-0.0059	-0.0006	0.0000	-0.0108	-0.0125	-0.0125	-0.0092	-0.0080
TEN_SQ	0.0004	0.0002	0.0000	0.0000	0.0004	0.0004	0.0004	0.0003	0.0003
OLDTEN	-0.1940	-0.1422	-0.0454	0.0000	-0.1672	-0.1940	-0.1934	-0.1423	-0.1244
PPERROOM	-0.0778	-0.0850	-0.0651	0.0000	-0.0671	-0.0778	-0.0776	-0.0571	-0.0499
BLACKHD	-0.1651	-0.1640	-0.0778	0.0000	-0.1423	-0.1651	-0.1646	-0.1211	-0.1059
SPANISHD	-0.0902	-0.1023	-0.1080	0.0000	-0.0778	-0.0902	-0.0900	-0.0662	-0.0579
GOODNBHD	-0.3356	-0.2716	-0.0566	0.0000	-0.2893	-0.3356	-0.3346	-0.2462	-0.2152
POORNBHD	-0.3892	-0.3077	-0.0806	0.0000	-0.3356	-0.3892	-0.3881	-0.2855	-0.2496
ABANDON	-0.0484	-0.0558	-0.0639	0.0000	-0.0417	-0.0484	-0.0482	-0.0355	-0.0310

TABLE 24
(Continued)

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR CHICAGO

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
ONE	3418.7	28991.0	76771.1	84632.8	27952.3	3418.7	3058.0	29532.9	35362.0
CENCITY	-16730.4	-16165.9	-8903.6	0.0	-11676.4	-16730.4	-14965.0	-18131.7	-18444.5
BATHMORE	29344.7	32029.6	19814.9	0.0	20480.1	29344.7	26248.3	35801.6	37242.9
ROOMMORE	14183.6	11051.3	1180.7	0.0	9899.0	14183.6	12687.0	8778.9	7572.5
AGE	2224.0	-527.2	-39.2	0.0	1552.2	2224.0	1989.3	1213.1	987.5
STRUCTUR	306.4	395.4	-568.9	0.0	213.8	306.4	274.0	-180.5	-289.2
EXCNBHD	-13489.5	-9764.6	6855.9	0.0	-9414.5	-13489.5	-12066.1	1508.9	4856.9
BATH1	-1847.0	-573.0	-2133.3	0.0	-1289.0	-1847.0	-1652.1	-973.1	-778.1
BATH2	9024.1	8872.2	1832.5	0.0	6298.1	9024.1	8071.9	4754.6	3801.5
ROOM4	79954.4	58113.3	-4931.2	0.0	55801.3	79954.4	71517.8	42125.8	33681.8
ROOM5	84884.7	63748.2	-2425.5	0.0	59242.3	84884.7	75927.9	44723.5	35758.7
BED1	-2833.1	-1526.0	-2866.0	0.0	-1977.2	-2833.1	-2534.1	-1492.7	-1193.5
BED2	1087.9	1677.6	-2431.9	0.0	759.3	1087.9	973.1	573.2	458.3
BED4	-9559.7	-5829.7	6843.9	0.0	-6671.9	-9559.7	-8551.0	-5036.8	-4027.2
BEDMORE	-8399.9	-6520.5	-115.5	0.0	-5862.4	-8399.9	-7513.5	-4425.7	-3538.6
ATTCHD	-13561.6	-12483.9	-5276.5	0.0	-9464.8	-13561.6	-12130.6	-7145.2	-5713.0
GARAGE	3710.8	3925.4	4468.4	0.0	2589.8	3710.8	3319.2	1955.1	1563.2
BASEMENT	11205.8	10651.3	5365.8	0.0	7820.7	11205.8	10023.4	5904.0	4720.6
AGE_SQ	-171.5	-14.4	-1.7	0.0	-119.7	-171.5	-153.4	-90.3	-72.2
AGE_CUBE	3.1	0.6	0.0	0.0	2.2	3.1	2.8	1.6	1.3
PRIOR40	-6489.0	-17038.9	-5787.4	0.0	-4528.8	-6489.0	-5804.3	-3418.9	-2733.6
ROOMHEAT	-9195.0	-8538.6	-11475.6	0.0	-6417.3	-9195.0	-8224.7	-4844.6	-3873.5
STMHEAT	17577.2	16821.0	6501.9	0.0	12267.4	17577.2	15722.5	9261.0	7404.6
ROOMAIR	727.0	423.5	-3633.1	0.0	507.4	727.0	650.3	383.0	306.3
CENAIR	10627.6	10885.6	7338.1	0.0	7417.2	10627.6	9506.2	5599.4	4477.0
RMWOHT	-2223.9	-2543.9	-3378.8	0.0	-1552.1	-2223.9	-1989.2	-1171.7	-936.8
NOPRIV	-5619.9	-3548.8	-466.2	0.0	-3922.2	-5619.9	-5026.9	-2961.0	-2367.5
RMWOELEC	-3516.6	-3458.4	-179.9	0.0	-2454.3	-3516.6	-3145.5	-1852.8	-1481.4
TEN	-796.8	-282.1	-17.3	0.0	-556.1	-796.8	-712.7	-419.8	-335.6
TEN_SQ	31.8	13.1	0.0	0.0	22.2	31.8	28.4	16.7	13.4
OLDTEN	-16542.7	-12238.0	-5236.7	0.0	-11545.4	-16542.7	-14797.2	-8715.9	-6968.8
PPERROOM	-13197.8	-13619.3	-10132.4	0.0	-9210.9	-13197.8	-11805.2	-6953.6	-5559.7
BLACKHD	-12073.1	-12044.9	-8262.6	0.0	-8426.0	-12073.1	-10799.2	-6361.0	-5085.9
SPANISHD	-1516.9	-2671.0	-9403.5	0.0	-1058.7	-1516.9	-1356.8	-799.2	-639.0
GOODNBHD	-29559.9	-25910.8	-6375.9	0.0	-20630.3	-29559.9	-26440.9	-15574.3	-12452.5
POORNBHD	-32077.1	-26796.5	-7268.1	0.0	-22387.1	-32077.1	-28692.4	-16900.6	-13512.9
ABANDON	-5773.2	-6203.6	-6747.8	0.0	-4029.2	-5773.2	-5164.1	-3041.8	-2432.0

TABLE 25

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS RESULTS FOR DALLAS

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.5899	10.9718	11.1449	11.1065	10.6342	10.5899	10.3851	10.6727	10.6794
CENCITY	0.4379	0.3997	0.0482	0.0000	0.4004	0.4379	0.4295	0.3772	0.3723
BATHMORE	0.7441	0.6349	0.1393	0.0000	0.6803	0.7441	0.7298	0.7035	0.7002
ROOMMORE	0.0461	0.0238	0.0095	0.0000	0.0422	0.0461	0.0452	0.0447	0.0446
AGE	0.0242	-0.0013	-0.0022	0.0000	0.0221	0.0242	0.0237	0.0167	0.0161
STRUCTUR	0.0267	0.0194	-0.0346	0.0000	0.0244	0.0267	0.0262	-0.0176	-0.0212
EXCNBHD	0.3643	0.1346	0.0531	0.0000	0.3330	0.3643	0.3572	0.3206	0.3171
BATH1	0.0249	-0.0378	-0.0674	0.0000	0.0228	0.0249	0.0245	0.0199	0.0195
BATH2	0.3533	0.2583	0.0395	0.0000	0.3230	0.3533	0.3464	0.2816	0.2758
ROOM4	0.1341	-0.0623	-0.1169	0.0000	0.1226	0.1341	0.1315	0.1069	0.1047
ROOM5	0.3566	0.1515	-0.0238	0.0000	0.3260	0.3566	0.3497	0.2843	0.2784
BED2	-0.0641	-0.0514	-0.0431	0.0000	-0.0586	-0.0641	-0.0628	-0.0511	-0.0500
BED4	-0.1203	-0.0852	0.0321	0.0000	-0.1100	-0.1203	-0.1180	-0.0959	-0.0939
BEDMORE	0.0132	0.0268	0.0167	0.0000	0.0121	0.0132	0.0130	0.0105	0.0103
ATTCHD	0.0571	0.0493	-0.0205	0.0000	0.0522	0.0571	0.0560	0.0455	0.0446
GARAGE	-0.0137	0.0081	0.0676	0.0000	-0.0125	-0.0137	-0.0134	-0.0109	-0.0107
AGE_SQ	-0.0013	-0.0001	-0.0001	0.0000	-0.0012	-0.0013	-0.0013	-0.0011	-0.0010
AGE_CUBE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRIOR40	0.0634	-0.0231	-0.0097	0.0000	0.0579	0.0634	0.0622	0.0505	0.0495
ROOMHEAT	-0.3251	-0.2485	-0.1011	0.0000	-0.2972	-0.3251	-0.3188	-0.2591	-0.2538
ROOMAIR	-0.2231	-0.1814	-0.0655	0.0000	-0.2040	-0.2231	-0.2188	-0.1779	-0.1742
CENAIR	-0.1193	0.0002	0.0849	0.0000	-0.1091	-0.1193	-0.1170	-0.0951	-0.0932
RMWDHT	0.0917	0.0710	0.0053	0.0000	0.0838	0.0917	0.0899	0.0731	0.0716
NOPRIV	-0.1449	-0.2097	-0.1229	0.0000	-0.1325	-0.1449	-0.1421	-0.1155	-0.1131
RMWDELEC	-0.0125	-0.1079	-0.0916	0.0000	-0.0114	-0.0125	-0.0122	-0.0099	-0.0097
TEN	-0.0453	-0.0178	-0.0021	0.0000	-0.0414	-0.0453	-0.0444	-0.0361	-0.0353
TEN_SQ	0.0014	0.0004	-0.0001	0.0000	0.0012	0.0014	0.0013	0.0011	0.0011
OLDTEN	-0.2394	-0.0987	-0.0609	0.0000	-0.2188	-0.2394	-0.2347	-0.1908	-0.1869
PPERROOM	0.1304	0.0590	-0.0452	0.0000	0.1192	0.1304	0.1279	0.1039	0.1018
BLACKHD	-0.2910	-0.2904	-0.0867	0.0000	-0.2660	-0.2910	-0.2853	-0.2319	-0.2272
SPANISHD	-0.2706	-0.2411	-0.0675	0.0000	-0.2474	-0.2706	-0.2654	-0.2157	-0.2113
GOODNBHD	0.2187	0.0023	-0.0122	0.0000	0.1999	0.2187	0.2145	0.1743	0.1707
POORNBHD	0.1855	-0.0633	-0.0796	0.0000	0.1696	0.1855	0.1819	0.1478	0.1448
ABANDON	-0.5075	-0.4861	-0.1861	0.0000	-0.4639	-0.5075	-0.4976	-0.4045	-0.3962

TABLE 25
(Continued)

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR DALLAS

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	60967.0	66617.7	86306.8	82548.1	64674.8	32437.8	50267.9	64058.4	63971.4
CENCITY	51500.1	48136.1	6285.7	0.0	42652.0	40522.4	42462.3	32405.4	32943.3
BATHMORE	65448.4	61581.1	15295.9	0.0	54203.9	44376.8	53962.8	61055.9	61179.6
ROOMMORE	5777.8	3985.7	872.9	0.0	4785.2	5285.1	4763.9	3747.9	3805.1
AGE	2893.1	-539.1	-182.9	0.0	2396.0	1502.4	2385.4	284.3	357.8
STRUCTUR	12361.8	11375.5	-370.0	0.0	10238.0	11747.0	10192.4	4165.5	4396.4
EXCNBHD	-17322.2	-1978.9	4292.4	0.0	-14346.1	2798.5	-14282.3	2342.6	1788.6
BATH1	-4258.0	-5363.9	-7225.2	0.0	-3526.5	-9950.2	-3510.8	-1432.8	-1512.4
BATH2	19891.9	16504.3	29.3	0.0	16474.3	9547.0	16401.0	6693.4	7065.2
ROOM4	35114.7	19689.0	-5739.8	0.0	29081.8	10137.7	28952.4	11815.7	12472.1
ROOM5	42918.6	26803.3	-3273.1	0.0	35544.8	31881.0	35386.7	14441.6	15243.8
BED2	-14772.1	-13033.0	-3446.0	0.0	-12234.2	-76.0	-12179.7	-4970.7	-5246.8
BED4	-7814.5	-5036.9	3841.2	0.0	-6471.9	2707.1	-6443.1	-2629.5	-2775.6
BEDMORE	600.5	1727.4	1215.4	0.0	497.4	2156.4	495.2	202.1	213.3
ATTCHD	686.3	1870.6	-4271.4	0.0	568.4	-1929.0	565.9	230.9	243.8
GARAGE	-11385.5	-9865.0	2835.1	0.0	-9429.4	12404.1	-9387.5	-3831.1	-4043.9
AGE_SQ	-215.5	-10.8	-3.8	0.0	-178.5	-129.7	-177.7	-72.5	-76.5
AGE_CUBE	3.8	0.5	-0.1	0.0	3.1	2.3	3.1	1.3	1.3
PRIOR40	28887.1	17557.2	4414.7	0.0	23924.1	12482.8	23817.7	9720.2	10260.1
ROOMHEAT	-16699.3	-13222.1	-5808.6	0.0	-13830.2	-12061.2	-13768.7	-5619.1	-5931.3
ROOMAIR	-14757.7	-15523.5	-5320.5	0.0	-12222.2	-2964.5	-12167.9	-4965.8	-5241.7
CENAIR	5836.1	6685.0	5656.0	0.0	4833.4	10491.7	4811.9	1963.8	2072.9
RMWOHT	1480.3	814.8	-119.7	0.0	1226.0	-368.2	1220.5	498.1	525.8
NOPRIV	-387.0	-5169.9	-6738.8	0.0	-320.5	-11043.1	-319.1	-130.2	-137.5
RMWOELEC	-7059.2	-12675.7	-2815.8	0.0	-5846.4	-2905.5	-5820.4	-2375.3	-2507.3
TEN	-3822.8	-2253.0	-276.7	0.0	-3166.0	-2387.9	-3152.0	-1286.3	-1357.8
TEN_SQ	109.9	52.7	-7.1	0.0	91.0	52.3	90.6	37.0	39.0
OLDTEN	-22747.2	-12074.3	349.9	0.0	-18839.0	-7044.3	-18755.3	-7654.2	-8079.4
PPERROOM	14751.8	11356.9	-4075.8	0.0	12217.4	5314.6	12163.0	4963.8	5239.6
BLACKHD	-36149.7	-35467.6	-6263.2	0.0	-29938.9	-21546.8	-29805.8	-12164.0	-12839.7
SPANISHD	-28816.9	-23728.1	-5655.1	0.0	-23865.9	-9889.8	-23759.8	-9696.6	-10235.2
GOODNBHD	-29295.0	-12829.6	-1176.8	0.0	-24261.9	-9542.2	-24154.0	-9857.5	-10405.0
POORNBHD	-33779.8	-19255.0	-6767.0	0.0	-27976.2	-16170.3	-27851.7	-11366.5	-11997.9
ABANDON	-17385.7	-15368.7	-9510.7	0.0	-14398.7	-19714.9	-14334.6	-5850.1	-6175.1

TABLE 26

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR DETROIT

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	10.0738	10.4409	10.8084	10.7976	10.2652	10.0738	10.0722	10.2078	10.2365
CENCITY	-0.3730	-0.3785	-0.2135	0.0000	-0.2744	-0.3730	-0.3730	-0.4399	-0.4543
BATHMORE	0.4418	0.4513	0.2083	0.0000	0.3250	0.4418	0.4418	0.4421	0.4422
ROOMMORE	0.0775	0.0305	0.0112	0.0000	0.0570	0.0775	0.0774	0.0690	0.0671
AGE	-0.0198	-0.0045	0.0001	0.0000	-0.0146	-0.0198	-0.0198	-0.0156	-0.0147
STRUCTUR	-0.0720	-0.0679	-0.0468	0.0000	-0.0530	-0.0720	-0.0720	-0.0804	-0.0822
EXCNBHD	0.5110	0.3753	0.1158	0.0000	0.3758	0.5110	0.5109	0.4454	0.4313
BATH1	0.0840	0.0671	-0.0013	0.0000	0.0618	0.0840	0.0840	0.0659	0.0620
BATH2	0.2780	0.2687	0.0842	0.0000	0.2045	0.2780	0.2779	0.2178	0.2050
ROOM4	0.2907	-0.0193	-0.1339	0.0000	0.2138	0.2907	0.2906	0.2278	0.2143
ROOM5	0.4423	0.1335	-0.0222	0.0000	0.3253	0.4423	0.4422	0.3466	0.3261
BED1	0.5344	0.2323	0.0025	0.0000	0.3931	0.5344	0.5343	0.4188	0.3940
BED2	-0.1141	-0.1038	-0.0859	0.0000	-0.0839	-0.1141	-0.1141	-0.0894	-0.0841
BED4	-0.0399	0.0280	0.0659	0.0000	-0.0293	-0.0399	-0.0399	-0.0313	-0.0294
BEDMORE	0.0020	0.0294	0.0165	0.0000	0.0015	0.0020	0.0020	0.0016	0.0015
ATTCHD	-0.0952	-0.0705	0.0052	0.0000	-0.0700	-0.0952	-0.0952	-0.0746	-0.0702
GARAGE	0.1898	0.1894	0.1267	0.0000	0.1396	0.1898	0.1897	0.1487	0.1399
BASEMENT	0.0267	0.0322	0.0052	0.0000	0.0197	0.0267	0.0267	0.0210	0.0197
AGE_SQ	0.0004	-0.0001	0.0000	0.0000	0.0003	0.0004	0.0004	0.0003	0.0003
AGE_CUBE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRIOR40	-0.4497	-0.3119	-0.1508	0.0000	-0.3308	-0.4497	-0.4497	-0.3524	-0.3316
ROOMHEAT	0.1255	0.1243	0.0359	0.0000	0.0923	0.1255	0.1254	0.0983	0.0925
STMHEAT	-0.0322	-0.0265	-0.0277	0.0000	-0.0237	-0.0322	-0.0322	-0.0252	-0.0238
ELECHEAT	-0.1691	-0.1258	0.0020	0.0000	-0.1244	-0.1691	-0.1691	-0.1325	-0.1247
ROOMAIR	-0.0066	-0.0094	-0.0316	0.0000	-0.0048	-0.0066	-0.0066	-0.0052	-0.0049
CENAIR	0.0751	0.0829	0.0891	0.0000	0.0553	0.0751	0.0751	0.0589	0.0554
RMWOHT	0.0748	0.0674	0.0254	0.0000	0.0550	0.0748	0.0748	0.0586	0.0551
NOPRIV	0.2493	0.2230	0.0654	0.0000	0.1834	0.2493	0.2493	0.1954	0.1838
RMWOELEC	0.2251	0.2384	0.0712	0.0000	0.1656	0.2251	0.2251	0.1764	0.1660
TEN	0.0105	0.0009	-0.0020	0.0000	0.0077	0.0105	0.0105	0.0082	0.0077
TEN_SQ	-0.0005	-0.0003	-0.0001	0.0000	-0.0004	-0.0005	-0.0005	-0.0004	-0.0004
OLDTEN	-0.0752	-0.1067	-0.0719	0.0000	-0.0553	-0.0752	-0.0751	-0.0589	-0.0554
PPERROOM	-0.0609	-0.0666	-0.0488	0.0000	-0.0448	-0.0609	-0.0608	-0.0477	-0.0449
BLACKHD	-0.1080	-0.1109	-0.1226	0.0000	-0.0794	-0.1080	-0.1079	-0.0846	-0.0796
SPANISHD	-0.1938	-0.1491	-0.0826	0.0000	-0.1426	-0.1938	-0.1938	-0.1519	-0.1429
GOODNBHD	0.4060	0.2639	0.0048	0.0000	0.2986	0.4060	0.4059	0.3181	0.2993
POORNBHD	0.3057	0.1587	-0.0822	0.0000	0.2249	0.3057	0.3057	0.2396	0.2254

TABLE 26
(Continued)

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR DETROIT

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	14749.8	39216.5	59378.9	58172.3	20081.4	12179.9	14502.6	24347.1	26740.4
CENCITY	-7986.0	-9007.7	-7721.8	0.0	-7005.3	-8084.3	-7852.4	-12991.3	-14239.4
BATHMORE	38144.4	39186.9	15220.1	0.0	33460.4	37108.1	37506.5	38803.8	38968.3
ROOMMORE	7182.5	2869.5	731.6	0.0	6300.6	6223.9	7062.4	5779.5	5429.7
AGE	-1703.8	-606.8	-46.4	0.0	-1494.6	-99.1	-1675.3	-1273.1	-1165.7
STRUCTUR	-3055.1	-3027.4	-1889.5	0.0	-2680.0	-2993.9	-3004.0	-3643.4	-3790.1
EXCNBHD	20978.5	17430.1	6303.0	0.0	18402.5	18697.5	20627.7	18878.6	18354.9
BATH1	5116.3	3767.8	-1150.7	0.0	4488.0	3587.0	5030.7	3676.6	3317.6
BATH2	12803.3	13019.6	2735.5	0.0	11231.1	12010.8	12589.2	9200.5	8302.1
ROOM4	38079.8	10103.3	-5544.7	0.0	33403.8	25523.2	37443.0	27364.3	24692.3
ROOM5	43726.4	15407.5	-1877.9	0.0	38357.0	36227.9	42995.2	31422.0	28353.8
BED1	43012.5	16146.2	-1459.6	0.0	37730.7	31026.1	42293.2	30909.0	27890.8
BED2	-7673.7	-6915.6	-4248.1	0.0	-6731.4	-4384.4	-7545.4	-5514.4	-4975.9
BED4	-2822.3	3241.3	5624.9	0.0	-2475.7	-833.2	-2775.1	-2028.1	-1830.1
BEDMORE	-2155.3	297.1	392.6	0.0	-1890.7	-1584.1	-2119.3	-1548.8	-1397.6
ATTCHD	-12325.6	-9563.1	-581.9	0.0	-10812.0	-12012.4	-12119.4	-8857.2	-7992.3
GARAGE	5864.0	6033.6	5122.9	0.0	5144.0	6739.0	5766.0	4213.9	3802.5
BASEMENT	2981.5	3664.1	1384.0	0.0	2615.4	4439.3	2931.7	2142.6	1933.3
AGE_SQ	24.1	-4.2	-2.3	0.0	21.1	-51.8	23.7	17.3	15.6
AGE_CUBE	0.1	0.2	-0.1	0.0	0.1	1.1	0.1	0.1	0.1
PRIOR40	-29144.2	-18922.4	-6326.3	0.0	-25565.4	-20518.3	-28656.8	-20943.1	-18898.1
ROOMHEAT	8458.7	9244.2	2281.6	0.0	7420.0	6930.4	8317.3	6078.5	5484.9
STMHEAT	2331.3	2377.1	-459.7	0.0	2045.0	3130.2	2292.3	1675.3	1511.7
ELECHEAT	-16847.7	-13347.7	-1957.4	0.0	-14778.9	-12613.7	-16566.0	-12106.8	-10924.7
ROOMAIR	-172.9	-698.9	-2162.2	0.0	-151.7	-441.2	-170.0	-124.2	-112.1
CENAIR	7433.7	7262.3	5578.8	0.0	6520.9	7282.4	7309.4	5341.9	4820.3
RMWOHT	5872.2	5554.1	1959.2	0.0	5151.1	6038.8	5774.0	4219.8	3807.8
NOPRIV	11645.7	10423.9	2510.2	0.0	10215.6	8329.7	11450.9	8368.6	7551.5
RMWOELEC	20.4	2294.8	-403.6	0.0	17.9	-190.0	20.1	14.7	13.2
TEN_SQ	-19.3	-10.2	-4.1	0.0	-16.9	-13.7	-19.0	-13.9	-12.5
OLDTEN	-6417.8	-6967.0	-3719.8	0.0	-5629.7	-4234.7	-6310.5	-4611.9	-4161.5
PPERROOM	-5347.1	-6697.1	-4058.5	0.0	-4690.5	-1139.1	-5257.6	-3842.4	-3467.2
BLACKHD	-9552.5	-9141.2	-5228.0	0.0	-8379.5	-9918.1	-9392.8	-6864.5	-6194.2
SPANISHD	-21100.0	-17654.5	-5485.9	0.0	-18509.0	-14934.9	-20747.1	-15162.5	-13682.0
GOODNBHD	12347.0	8492.3	-1189.1	0.0	10830.8	9247.6	12140.5	8872.6	8006.2
POORNBHD	7879.1	3803.6	-4386.8	0.0	6911.6	4495.7	7747.3	5661.9	5109.1
ABANDON	-10811.9	-10999.1	-6104.6	0.0	-9484.3	-11003.5	-10631.1	-7769.5	-7010.8

TABLE 27

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR NEW YORK

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
ONE	10.9504	10.8649	10.9504	11.3369	11.0267	10.9504	10.8787	10.9837	10.9895
CENCITY	-0.0007	-0.0009	-0.0007	0.0000	-0.0006	-0.0007	-0.0007	0.0134	0.0158
BATHMORE	0.3172	0.3114	0.3172	0.0000	0.2546	0.3172	0.3151	0.2964	0.2928
ROOMMORE	0.0698	0.0410	0.0698	0.0000	0.0560	0.0698	0.0693	0.0532	0.0504
AGE	-0.0162	-0.0062	-0.0162	0.0000	-0.0130	-0.0162	-0.0161	-0.0099	-0.0088
STRUCTUR	0.1072	0.0985	0.1072	0.0000	0.0861	0.1072	0.1065	0.0955	0.0935
EXCNBHD	-0.0445	0.0705	-0.0445	0.0000	-0.0357	-0.0445	-0.0442	0.0723	0.0925
BATH1	0.0977	0.0910	0.0977	0.0000	0.0784	0.0977	0.0970	0.0513	0.0432
BATH2	0.2649	0.2421	0.2649	0.0000	0.2126	0.2649	0.2632	0.1391	0.1172
ROOM4	0.0635	-0.1113	0.0635	0.0000	0.0510	0.0635	0.0631	0.0333	0.0281
ROOM5	0.3802	0.1641	0.3802	0.0000	0.3052	0.3802	0.3777	0.1996	0.1682
BED1	0.1878	0.1606	0.1878	0.0000	0.1507	0.1878	0.1866	0.0986	0.0831
BED2	0.0156	0.0235	0.0156	0.0000	0.0125	0.0156	0.0155	0.0082	0.0069
BED4	-0.0418	-0.0113	-0.0418	0.0000	-0.0336	-0.0418	-0.0415	-0.0219	-0.0185
BEDMORE	-0.0058	0.0097	-0.0058	0.0000	-0.0047	-0.0058	-0.0058	-0.0030	-0.0026
ATTCHD	0.0867	0.0695	0.0867	0.0000	0.0696	0.0867	0.0861	0.0455	0.0384
GARAGE	0.0681	0.0733	0.0681	0.0000	0.0546	0.0681	0.0676	0.0357	0.0301
BASEMENT	0.0851	0.0799	0.0851	0.0000	0.0683	0.0851	0.0845	0.0447	0.0376
AGE_SQ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AGE_CUBE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRIOR40	-0.1541	0.0031	-0.1541	0.0000	-0.1237	-0.1541	-0.1531	-0.0809	-0.0682
ROOMHEAT	-0.2284	-0.1977	-0.2284	0.0000	-0.1833	-0.2284	-0.2269	-0.1199	-0.1011
STMHEAT	-0.0273	-0.0204	-0.0273	0.0000	-0.0219	-0.0273	-0.0271	-0.0143	-0.0121
ELECHEAT	0.0211	-0.0097	0.0211	0.0000	0.0169	0.0211	0.0210	0.0111	0.0093
ROOMAIR	0.0715	0.0697	0.0715	0.0000	0.0574	0.0715	0.0710	0.0375	0.0316
CENAIR	0.2732	0.2524	0.2732	0.0000	0.2193	0.2732	0.2714	0.1434	0.1209
RMWOHT	-0.0177	-0.0057	-0.0177	0.0000	-0.0142	-0.0177	-0.0176	-0.0093	-0.0078
NOPRIV	-0.0295	-0.0330	-0.0295	0.0000	-0.0237	-0.0295	-0.0293	-0.0155	-0.0131
RMWOELEC	0.0303	0.0096	0.0303	0.0000	0.0243	0.0303	0.0301	0.0159	0.0134
TEN	-0.0233	-0.0092	-0.0233	0.0000	-0.0187	-0.0233	-0.0231	-0.0122	-0.0103
TEN_SQ	0.0007	0.0002	0.0007	0.0000	0.0006	0.0007	0.0007	0.0004	0.0003
OLDTEN	-0.1394	-0.0965	-0.1394	0.0000	-0.1119	-0.1394	-0.1385	-0.0732	-0.0617
PPERROOM	0.2162	0.1537	0.2162	0.0000	0.1735	0.2162	0.2148	0.1135	0.0957
BLACKHD	-0.3049	-0.2907	-0.3049	0.0000	-0.2447	-0.3049	-0.3029	-0.1601	-0.1349
SPANISHD	-0.0500	-0.0468	-0.0500	0.0000	-0.0401	-0.0500	-0.0497	-0.0263	-0.0221
GOODNBHD	-0.2043	-0.0969	-0.2043	0.0000	-0.1640	-0.2043	-0.2029	-0.1072	-0.0904
POORNBHD	-0.2916	-0.1804	-0.2916	0.0000	-0.2341	-0.2916	-0.2897	-0.1531	-0.1290

TABLE 27
(Continued)

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR NEW YORK

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	53222.9	47979.5	53222.9	92018.2	57916.0	30876.0	51576.6	61522.2	57994.5
CENCITY	-2475.8	-2359.1	-2475.8	0.0	-2176.3	2958.4	-2399.2	147.1	-967.8
BATHMORE	32516.8	32161.0	32516.8	0.0	28583.3	22756.2	31511.0	31067.9	31683.8
ROOMMORE	6590.3	3924.6	6590.3	0.0	5793.1	5415.4	6386.4	4332.5	5292.2
AGE	-1323.1	-702.6	-1323.1	0.0	-1163.0	1371.5	-1282.1	-753.7	-995.7
STRUCTUR	8259.7	7665.0	8259.7	0.0	7260.5	8718.8	8004.2	7253.7	7681.3
EXCNBHD	918.8	7885.4	918.8	0.0	807.6	12697.3	890.3	10821.5	6612.2
BATH1	8899.1	8357.1	8899.1	0.0	7822.6	-1511.1	8623.8	3654.2	5883.6
BATH2	24876.0	22752.0	24876.0	0.0	21866.8	15685.4	24106.5	10214.8	16446.8
ROOM4	16650.8	-811.4	16650.8	0.0	14636.6	8814.0	16135.8	6837.3	11008.7
ROOM5	40570.3	19658.1	40570.3	0.0	35662.6	19865.3	39315.4	16659.4	26823.1
BED1	10308.9	9528.9	10308.9	0.0	9061.9	6671.3	9990.0	4233.1	6815.8
BED2	811.9	2309.3	811.9	0.0	713.7	3773.5	786.8	333.4	536.8
BED4	-3870.1	-1119.5	-3870.1	0.0	-3401.9	-1336.7	-3750.4	-1589.2	-2558.7
BEDMORE	744.5	2056.4	744.5	0.0	654.5	1936.7	721.5	305.7	492.3
ATTCHD	10435.6	8705.1	10435.6	0.0	9173.2	1408.0	10112.8	4285.2	6899.5
GARAGE	2615.5	3625.6	2615.5	0.0	2299.1	4282.9	2534.6	1074.0	1729.2
BASEMENT	9655.2	8996.1	9655.2	0.0	8487.2	6415.6	9356.6	3964.7	6383.6
AGE_SQ	-23.9	-0.9	-23.9	0.0	-21.0	-132.1	-23.1	-9.8	-15.8
AGE_CUBE	1.2	0.4	1.2	0.0	1.1	2.5	1.2	0.5	0.8
PRIOR40	-13247.2	726.4	-13247.2	0.0	-11644.7	4060.5	-12837.4	-5439.7	-8758.4
ROOMHEAT	-20178.2	-17473.8	-20178.2	0.0	-17737.3	-6169.7	-19554.0	-8285.7	-13340.8
STMHEAT	-1575.7	-1022.8	-1575.7	0.0	-1385.1	-624.1	-1526.9	-647.0	-1041.8
ELECHEAT	9831.8	5787.2	9831.8	0.0	8642.4	4538.7	9527.7	4037.2	6500.3
ROOMAIR	6022.3	5745.8	6022.3	0.0	5293.8	7286.9	5836.0	2472.9	3981.7
CENAIR	25093.5	22995.9	25093.5	0.0	22058.0	17427.5	24317.3	10304.1	16590.6
RMWOHT	-3669.5	-2084.8	-3669.5	0.0	-3225.6	-3200.8	-3556.0	-1506.8	-2426.1
NOPRIV	-1845.4	-1896.6	-1845.4	0.0	-1622.2	-3530.2	-1788.3	-757.8	-1220.1
RMWOELEC	-1282.3	-3347.3	-1282.3	0.0	-1127.2	-172.9	-1242.6	-526.5	-847.8
TEN	-2802.1	-1157.4	-2802.1	0.0	-2463.2	-2017.0	-2715.4	-1150.6	-1852.6
TEN_SQ	84.7	26.1	84.7	0.0	74.5	62.8	82.1	34.8	56.0
OLDTEN	-17751.8	-12311.8	-17751.8	0.0	-15604.4	-2584.3	-17202.7	-7289.4	-11736.6
PPERROOM	22863.1	17153.5	22863.1	0.0	20097.3	27368.0	22155.9	9388.2	15115.9
BLACKHD	-25860.9	-24636.7	-25860.9	0.0	-22732.5	-28053.5	-25061.0	-10619.2	-17097.9
SPANISHD	-7704.9	-7137.8	-7704.9	0.0	-6772.8	-4680.9	-7466.5	-3163.8	-5094.1
GOODNBHD	-13895.4	-7805.0	-13895.4	0.0	-12214.5	-3328.3	-13465.5	-5705.8	-9186.9
POORNBHD	-18557.6	-11996.3	-18557.6	0.0	-16312.7	-10213.3	-17983.6	-7620.3	-12269.4

TABLE 28

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR SAN DIEGO

A. SEMILOG MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
ONE	11.6596	11.8540	11.6596	11.6311	11.6521	11.6596	11.5336	11.5303	11.5303
CENCITY	-0.0228	-0.0263	-0.0228	0.0000	-0.0167	-0.0228	-0.0225	-0.0509	-0.0509
BATHMORE	0.1319	0.1531	0.1319	0.0000	0.0970	0.1319	0.1304	0.2542	0.2542
ROOMMORE	0.0528	0.0386	0.0528	0.0000	0.0388	0.0528	0.0523	0.0413	0.0413
AGE	0.0258	-0.0051	0.0258	0.0000	0.0189	0.0258	0.0255	-0.0057	-0.0057
STRUCTUR	0.0296	0.0219	0.0296	0.0000	0.0218	0.0296	0.0293	0.0320	0.0320
EXCNBHD	0.0979	0.0457	0.0979	0.0000	0.0720	0.0979	0.0969	0.0482	0.0482
BATH1	-0.1870	-0.1862	-0.1870	0.0000	-0.1375	-0.1870	-0.1850	0.0000	0.0000
BATH2	-0.0491	-0.0443	-0.0491	0.0000	-0.0361	-0.0491	-0.0485	0.0000	0.0000
ROOM4	0.0315	-0.0695	0.0315	0.0000	0.0232	0.0315	0.0312	0.0000	0.0000
ROOM5	0.2231	0.1341	0.2231	0.0000	0.1640	0.2231	0.2207	0.0000	0.0000
BED2	-0.0466	-0.0284	-0.0466	0.0000	-0.0343	-0.0466	-0.0461	0.0000	0.0000
BED4	-0.0499	-0.0330	-0.0499	0.0000	-0.0367	-0.0499	-0.0493	0.0000	0.0000
ATTCHD	0.0796	0.0967	0.0796	0.0000	0.0585	0.0796	0.0787	0.0000	0.0000
GARAGE	-0.0795	-0.0393	-0.0795	0.0000	-0.0585	-0.0795	-0.0787	0.0000	0.0000
BASEMENT	0.2542	0.2954	0.2542	0.0000	0.1869	0.2542	0.2515	0.0000	0.0000
AGE_SQ	-0.0019	-0.0001	-0.0019	0.0000	-0.0014	-0.0019	-0.0019	0.0000	0.0000
AGE_CUBE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PRIOR40	0.0145	-0.0868	0.0145	0.0000	0.0107	0.0145	0.0144	0.0000	0.0000
ROOMHEAT	-0.1575	-0.1382	-0.1575	0.0000	-0.1158	-0.1575	-0.1558	0.0000	0.0000
ROOMAIR	0.0354	0.0288	0.0354	0.0000	0.0261	0.0354	0.0351	0.0000	0.0000
CENAIR	0.2393	0.2296	0.2393	0.0000	0.1760	0.2393	0.2367	0.0000	0.0000
RMWOHT	0.0455	0.0331	0.0455	0.0000	0.0335	0.0455	0.0450	0.0000	0.0000
NOPRIV	-0.0680	-0.0805	-0.0680	0.0000	-0.0500	-0.0680	-0.0673	0.0000	0.0000
RMWOELEC	0.3642	0.3730	0.3642	0.0000	0.2677	0.3642	0.3602	0.0000	0.0000
TEN	-0.0141	-0.0081	-0.0141	0.0000	-0.0103	-0.0141	-0.0139	0.0000	0.0000
TEN_SQ	0.0003	0.0001	0.0003	0.0000	0.0002	0.0003	0.0003	0.0000	0.0000
PPERROOM	-0.3863	-0.3828	-0.3863	0.0000	-0.2841	-0.3863	-0.3822	0.0000	0.0000
BLACKHD	-0.2276	-0.2344	-0.2276	0.0000	-0.1673	-0.2276	-0.2251	0.0000	0.0000
SPANISHD	0.0835	0.0778	0.0835	0.0000	0.0614	0.0835	0.0826	0.0000	0.0000
GOODNBHD	0.0273	-0.0192	0.0273	0.0000	0.0201	0.0273	0.0270	0.0000	0.0000
POORNBHD	0.0867	0.0442	0.0867	0.0000	0.0637	0.0867	0.0857	0.0000	0.0000

TABLE 2B
(Continued)

PARAMETER ESTIMATES OF THE COMPARED ESTIMATORS
RESULTS FOR SAN DIEGO

B. LINEAR MODEL

PARAMETER	OLS	RIDGE	ITHKB	LINDLEY	STRAW	PCSTEIN	STEIN	STEINRLS	EBAYES
INTERCEPT	122263.0	142860.0	122263.0	122025.0	122224.0	85951.6	103516.0	110225.0	110604.0
CENCITY	-5650.4	-5516.6	-5650.4	0.0	-4725.8	-438.5	-4784.0	-5988.4	-5977.8
BATHMORE	16590.5	22457.1	16590.5	0.0	13875.6	14635.0	14046.6	35748.0	35145.0
ROOMMORE	8406.1	5077.8	8406.1	0.0	7030.5	8493.0	7117.1	5248.3	5347.6
AGE	2449.0	-481.1	2449.0	0.0	2048.2	5151.6	2073.5	142.0	214.6
STRUCTUR	4041.8	2099.6	4041.8	0.0	3380.4	6873.0	3422.1	3859.2	3864.9
EXCNBHD	6907.8	6138.2	6907.8	0.0	5777.4	7124.3	5848.6	8716.5	8659.6
BATH1	-20664.1	-18407.4	-20664.1	0.0	-17282.6	-11752.8	-17495.6	-4771.5	-5271.7
BATH2	-17194.8	-13933.8	-17194.8	0.0	-14381.1	-12860.6	-14558.3	-3970.4	-4386.6
ROOM4	20813.2	-486.0	20813.2	0.0	17407.4	10748.5	17621.9	4805.9	5309.7
ROOM5	40105.5	18142.1	40105.5	0.0	33542.7	33493.0	33956.0	9260.6	10231.5
BED2	-3928.0	-1675.6	-3928.0	0.0	-3285.3	-6295.7	-3325.7	-907.0	-1002.1
BED4	-11627.0	-8750.9	-11627.0	0.0	-9724.4	-6664.0	-9844.2	-2684.8	-2966.2
ATTCHD	-3238.5	-1298.9	-3238.5	0.0	-2708.5	-2729.4	-2741.9	-747.8	-826.2
GARAGE	-8870.1	-2234.4	-8870.1	0.0	-7418.6	-1197.6	-7510.0	-2048.2	-2262.9
BASEMENT	29657.4	35458.5	29657.4	0.0	24804.3	20834.6	25109.9	6848.1	7566.0
AGE_SQ	-190.6	-7.9	-190.6	0.0	-159.4	-355.6	-161.4	-44.0	-48.6
AGE_CUBE	3.2	0.2	3.2	0.0	2.7	6.0	2.7	0.7	0.8
PRIOR40	2411.6	-6337.6	2411.6	0.0	2017.0	2926.1	2041.9	556.9	615.2
ROOMHEAT	-26191.6	-22710.4	-26191.6	0.0	-21905.6	-19306.7	-22175.5	-6047.8	-6681.8
ROOMAIR	-3937.4	-4810.2	-3937.4	0.0	-3293.1	-2661.1	-3333.7	-909.2	-1004.5
CENAIR	33201.6	30811.2	33201.6	0.0	27768.5	24085.4	28110.7	7666.5	8470.2
RMWOHT	8364.5	6588.4	8364.5	0.0	6995.7	7708.8	7081.9	1931.4	2133.9
NOPRIV	-17195.4	-16576.8	-17195.4	0.0	-14381.6	-8807.7	-14558.8	-3970.6	-4386.8
RMWOELEC	35429.8	35078.0	35429.8	0.0	29632.1	23904.0	29997.2	8181.0	9038.6
TEN	-1058.1	-563.8	-1058.1	0.0	-884.9	-456.2	-895.8	-244.3	-269.9
TEN_SQ	18.3	-0.2	18.3	0.0	15.3	9.5	15.5	4.2	4.7
PPERROOM	-42788.4	-41771.6	-42788.4	0.0	-35786.6	-26693.5	-36227.5	-9880.2	-10915.9
BLACKHD	-15306.9	-18378.2	-15306.9	0.0	-12802.1	-10580.1	-12959.8	-3534.5	-3905.0
SPANISHD	7313.4	6790.7	7313.4	0.0	6116.7	5280.4	6192.0	1688.7	1865.8
GOODNBHD	-4310.0	-3813.5	-4310.0	0.0	-3604.7	-1833.6	-3649.1	-995.2	-1099.5
POORNBHD	-10299.3	-8225.6	-10299.3	0.0	-8613.9	-7178.8	-8720.0	-2378.2	-2627.5

TABLE 29

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS
UNDER MEAN ABSOLUTE PERCENTAGE ERROR
LOSS CRITERION*

FIRST SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON					
	OLS	5.520	3.609	2.648	2.076
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.002	1.000	1.000	1.000
	LINDLEY	1.112	1.048	1.009	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.041	1.020	1.004	1.000
	EBAYES	1.039	1.025	1.006	1.001
CHICAGO					
	OLS	5.416	3.189	2.087	1.430
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.051	1.015	1.003	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.023	1.012	1.002	1.000
	EBAYES	1.024	1.017	1.005	1.001
DALLAS					
	OLS	7.845	4.859	3.378	2.508
	RIDGE	1.008	1.002	1.000	1.000
	ITHKB	1.011	1.002	1.000	1.000
	LINDLEY	1.116	1.036	1.007	1.001
	STRAW	0.998	1.000	1.000	1.000
	STEIN	0.998	1.000	1.000	1.000
	STEINRLS	1.060	1.029	1.006	1.001
	EBAYES	1.051	1.028	1.006	1.001

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 29
(Continued)

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS UNDER MEAN ABSOLUTE PERCENTAGE ERROR LOSS CRITERION*

FIRST SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
DETROIT	OLS	6.359	3.728	2.431	1.653
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.047	1.013	1.002	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.032	1.019	1.004	1.000
	EBAYES	1.029	1.020	1.005	1.001
NEW YORK	OLS	5.989	3.493	2.266	1.531
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.072	1.043	1.011	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.027	1.017	1.004	1.000
	EBAYES	1.027	1.024	1.017	1.003
SAN DIEGO	OLS	5.968	4.135	3.225	2.687
	RIDGE	1.005	1.001	1.000	1.000
	ITHKB	1.005	1.001	1.000	1.000
	LINDLEY	1.162	1.116	1.027	1.002
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.060	1.044	1.017	1.002
	EBAYES	1.052	1.036	1.014	1.001

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 30

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS
UNDER MEAN ABSOLUTE PERCENTAGE ERROR
LOSS CRITERION*

SECOND SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON					
	OLS	5.531	3.603	2.646	2.076
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.109	1.042	1.009	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.040	1.022	1.005	1.000
	EBAYES	1.038	1.027	1.008	1.001
CHICAGO					
	OLS	5.400	3.183	2.084	1.428
	RIDGE	1.000	1.000	1.000	1.000
	ITHKB	1.000	1.000	1.000	1.000
	LINDLEY	1.051	1.015	1.003	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.025	1.013	1.003	1.000
	EBAYES	1.025	1.019	1.006	1.001
DALLAS					
	OLS	7.826	4.859	3.372	2.507
	RIDGE	1.007	1.002	1.000	1.000
	ITHKB	1.010	1.002	1.000	1.000
	LINDLEY	1.126	1.046	1.009	1.001
	STRAW	0.998	1.000	1.000	1.000
	STEIN	0.998	1.000	1.000	1.000
	STEINRLS	1.063	1.036	1.008	1.001
	EBAYES	1.049	1.023	1.004	1.000

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 30
(Continued)

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS
UNDER MEAN ABSOLUTE PERCENTAGE ERROR
LOSS CRITERION*

SECOND SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
DETROIT	OLS	6.338	3.718	2.426	1.651
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.002	1.000	1.000	1.000
	LINDLEY	1.054	1.013	1.003	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.037	1.022	1.006	1.001
	EBAYES	1.034	1.023	1.007	1.001
NEW YORK	OLS	6.007	3.510	2.274	1.533
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.070	1.039	1.009	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.025	1.016	1.004	1.000
	EBAYES	1.024	1.022	1.016	1.003
SAN DIEGO	OLS	5.968	4.141	3.235	2.685
	RIDGE	1.003	1.001	1.000	1.000
	ITHKB	1.004	1.001	1.000	1.000
	LINDLEY	1.164	1.118	1.028	1.003
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.058	1.046	1.020	1.002
	EBAYES	1.050	1.035	1.011	1.001

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 31

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS UNDER MEAN ABSOLUTE PERCENTAGE ERROR LOSS CRITERION*

THIRD SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	5.523	3.610	2.648	2.077
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.107	1.043	1.008	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.041	1.029	1.007	1.001
	EBAYES	1.037	1.027	1.008	1.001
CHICAGO	OLS	5.420	3.191	2.088	1.429
	RIDGE	1.000	1.000	1.000	1.000
	ITHKB	1.000	1.000	1.000	1.000
	LINDLEY	1.047	1.013	1.003	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.024	1.010	1.002	1.000
	EBAYES	1.025	1.017	1.005	1.001
DALLAS	OLS	7.838	4.865	3.377	2.506
	RIDGE	1.007	1.001	1.000	1.000
	ITHKB	1.009	1.002	1.000	1.000
	LINDLEY	1.124	1.043	1.008	1.001
	STRAW	0.998	1.000	1.000	1.000
	STEIN	0.998	1.000	1.000	1.000
	STEINRLS	1.060	1.027	1.005	1.001
	EBAYES	1.050	1.022	1.004	1.000

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 31
(Continued)

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS
UNDER MEAN ABSOLUTE PERCENTAGE ERROR
LOSS CRITERION*

THIRD SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
DETROIT					
	OLS	6.367	3.732	2.438	1.653
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.057	1.015	1.003	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.032	1.019	1.004	1.000
	EBAYES	1.029	1.019	1.005	1.001
NEW YORK					
	OLS	6.017	3.514	2.275	1.534
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.070	1.046	1.012	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.026	1.016	1.003	1.000
	EBAYES	1.025	1.022	1.017	1.004
SAN DIEGO					
	OLS	5.996	4.142	3.238	2.687
	RIDGE	1.004	1.001	1.000	1.000
	ITHKB	1.005	1.001	1.000	1.000
	LINDLEY	1.163	1.139	1.040	1.004
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.057	1.046	1.026	1.003
	EBAYES	1.051	1.038	1.017	1.002

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 32

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS UNDER MEAN ABSOLUTE PERCENTAGE ERROR LOSS CRITERION*

FOURTH SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	5.522	3.611	2.644	2.077
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.113	1.048	1.009	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.045	1.029	1.006	1.001
	EBAYES	1.040	1.026	1.006	1.000
CHICAGO	OLS	5.411	3.188	2.085	1.429
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.057	1.017	1.004	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.025	1.013	1.003	1.000
	EBAYES	1.024	1.018	1.006	1.001
DALLAS	OLS	7.798	4.839	3.366	2.502
	RIDGE	1.007	1.002	1.000	1.000
	ITHKB	1.010	1.002	1.000	1.000
	LINDLEY	1.130	1.054	1.011	1.001
	STRAW	0.998	1.000	1.000	1.000
	STEIN	0.998	1.000	1.000	1.000
	STEINRLS	1.061	1.039	1.009	1.001
	EBAYES	1.050	1.029	1.006	1.001

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 32
(Continued)

RELATIVE PERFORMANCE OF COMPARED ESTIMATORS UNDER MEAN ABSOLUTE PERCENTAGE ERROR LOSS CRITERION*

FOURTH SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
DETROIT	OLS	6.408	3.751	2.444	1.657
	RIDGE	1.001	1.000	1.000	1.000
	ITHKB	1.001	1.000	1.000	1.000
	LINDLEY	1.059	1.016	1.003	1.000
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.033	1.018	1.005	1.000
	EBAYES	1.029	1.017	1.004	1.000
NEW YORK	OLS	6.000	3.497	2.269	1.531
	RIDGE	1.000	1.000	1.000	1.000
	ITHKB	1.000	1.000	1.000	1.000
	LINDLEY	1.069	1.037	1.008	1.001
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.026	1.017	1.004	1.000
	EBAYES	1.024	1.021	1.016	1.003
SAN DIEGO	OLS	5.975	4.129	3.226	2.684
	RIDGE	1.009	1.002	1.000	1.000
	ITHKB	1.012	1.002	1.000	1.000
	LINDLEY	1.166	1.111	1.026	1.002
	STRAW	0.999	1.000	1.000	1.000
	STEIN	0.999	1.000	1.000	1.000
	STEINRLS	1.057	1.046	1.029	1.005
	EBAYES	1.051	1.039	1.021	1.003

* Table shows actual calculation of Mean Absolute Percentage Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 33

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS UNDER
MEAN SQUARE ERROR OF IN-SAMPLE PREDICTION
LOSS CRITERION*

SECOND SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	1.261	1.282	1.249	1.266
	LINDLEY	1.062	1.188	1.055	1.006
	STEINRLS	0.672	0.835	0.939	0.984
	EBAYES	0.680	0.876	0.989	0.992
CHICAGO	OLS	1.454	1.450	1.473	1.457
	LINDLEY	1.026	1.025	1.018	1.001
	STEINRLS	0.679	0.853	0.959	0.989
	EBAYES	0.685	0.900	1.024	0.998
DALLAS	OLS	3.022	3.043	3.058	3.070
	LINDLEY	0.956	1.070	1.017	1.007
	STEINRLS	0.846	1.069	1.024	1.004
	EBAYES	0.836	1.012	1.002	1.000
DETROIT	OLS	1.920	1.925	1.953	1.888
	LINDLEY	1.043	0.997	1.003	1.002
	STEINRLS	0.696	0.899	0.969	0.999
	EBAYES	0.705	0.924	0.999	1.006
NEW YORK	OLS	1.847	1.828	1.801	1.865
	LINDLEY	0.794	1.137	1.047	1.008
	STEINRLS	0.582	0.671	0.820	0.963
	EBAYES	0.588	0.663	0.848	1.000
SAN DIEGO	OLS	1.397	1.383	1.370	1.426
	LINDLEY	0.972	1.714	1.286	1.046
	STEINRLS	0.648	0.866	1.052	1.024
	EBAYES	0.660	0.852	0.981	1.004

* Table shows actual calculation of Mean Square Error of In-Sample Prediction for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 34

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS UNDER
MEAN SQUARE ERROR OF IN-SAMPLE PREDICTION
LOSS CRITERION*

THIRD SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	1.285	1.272	1.258	1.253
	LINDLEY	1.053	1.194	1.043	1.010
	STEINRLS	0.677	0.875	0.980	0.993
	EBAYES	0.686	0.882	1.006	0.999
CHICAGO	OLS	1.409	1.436	1.441	1.418
	LINDLEY	1.071	1.016	1.018	0.998
	STEINRLS	0.691	0.844	0.951	0.990
	EBAYES	0.700	0.904	1.005	0.999
DALLAS	OLS	2.990	3.013	3.016	3.046
	LINDLEY	0.981	1.072	1.018	1.002
	STEINRLS	0.832	1.007	1.000	1.002
	EBAYES	0.830	0.998	0.998	1.001
DETROIT	OLS	1.926	1.913	1.869	1.930
	LINDLEY	1.039	1.002	1.003	1.000
	STEINRLS	0.713	0.912	0.972	0.993
	EBAYES	0.719	0.934	1.000	0.997
NEW YORK	OLS	1.840	1.818	1.804	1.832
	LINDLEY	0.724	1.189	1.073	1.013
	STEINRLS	0.578	0.662	0.822	0.961
	EBAYES	0.583	0.650	0.835	1.000
SAN DIEGO	OLS	1.378	1.396	1.371	1.407
	LINDLEY	0.922	1.809	1.460	1.067
	STEINRLS	0.618	0.780	1.042	1.022
	EBAYES	0.629	0.779	0.984	1.000

* Table shows actual calculation of Mean Square Error of In-Sample Prediction for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 35

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS UNDER
MEAN SQUARE ERROR OF IN-SAMPLE PREDICTION
LOSS CRITERION*

FOURTH SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	1.266	1.263	1.273	1.252
	LINDLEY	1.032	1.241	1.057	1.014
	STEINRLS	0.727	0.942	0.996	1.001
	EBAYES	0.732	0.944	1.007	1.003
CHICAGO	OLS	1.442	1.441	1.450	1.436
	LINDLEY	1.037	1.036	1.011	1.001
	STEINRLS	0.686	0.859	0.951	0.989
	EBAYES	0.691	0.905	1.000	0.997
DALLAS	OLS	3.018	3.035	3.042	3.086
	LINDLEY	0.917	1.087	1.023	1.003
	STEINRLS	0.773	0.994	1.014	0.999
	EBAYES	0.774	0.971	1.004	0.998
DETROIT	OLS	1.915	1.956	1.926	1.960
	LINDLEY	1.052	1.025	1.010	1.000
	STEINRLS	0.752	0.946	0.981	0.999
	EBAYES	0.757	0.963	0.993	1.001
NEW YORK	OLS	1.830	1.867	1.824	1.869
	LINDLEY	0.818	1.127	1.033	1.008
	STEINRLS	0.582	0.680	0.837	0.966
	EBAYES	0.590	0.676	0.882	1.005
SAN DIEGO	OLS	1.367	1.391	1.401	1.413
	LINDLEY	1.036	1.603	1.224	1.036
	STEINRLS	0.596	0.698	0.949	1.031
	EBAYES	0.609	0.707	0.918	1.001

* Table shows actual calculation of Mean Square Error of In-Sample Prediction for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 36

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER MEAN SQUARE ERROR
LOSS CRITERION*

SECOND SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.296	0.308	0.300	0.302
	LINDLEY	0.318	0.651	0.888	0.978
	STEINRLS	0.406	0.598	0.824	0.965
	EBAYES	0.419	0.609	0.823	0.961
CHICAGO	OLS	0.186	0.185	0.183	0.183
	LINDLEY	0.390	0.739	0.929	0.988
	STEINRLS	0.335	0.562	0.838	0.968
	EBAYES	0.345	0.552	0.823	0.961
DALLAS	OLS	0.816	0.819	0.798	0.826
	LINDLEY	0.321	0.660	0.891	0.983
	STEINRLS	0.472	0.748	0.906	0.983
	EBAYES	0.506	0.764	0.920	0.986
DETROIT	OLS	0.364	0.358	0.372	0.359
	LINDLEY	0.454	0.786	0.940	0.990
	STEINRLS	0.287	0.504	0.792	0.963
	EBAYES	0.314	0.520	0.789	0.963
NEW YORK	OLS	0.230	0.224	0.221	0.225
	LINDLEY	0.203	0.522	0.831	0.971
	STEINRLS	0.262	0.384	0.663	0.929
	EBAYES	0.271	0.357	0.608	0.908
SAN DIEGO	OLS	0.410	0.407	0.397	0.403
	LINDLEY	0.260	0.602	0.828	0.971
	STEINRLS	0.386	0.637	0.908	0.993
	EBAYES	0.411	0.643	0.861	0.981

* Table shows actual calculation of Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 37

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER MEAN SQUARE ERROR
LOSS CRITERION*

THIRD SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.245	0.254	0.245	0.243
	LINDLEY	0.371	0.678	0.898	0.985
	STEINRLS	0.449	0.634	0.851	0.967
	EBAYES	0.469	0.646	0.868	0.970
CHICAGO	OLS	0.159	0.165	0.161	0.155
	LINDLEY	0.419	0.762	0.936	0.988
	STEINRLS	0.341	0.574	0.848	0.972
	EBAYES	0.352	0.554	0.817	0.963
DALLAS	OLS	0.781	0.788	0.782	0.762
	LINDLEY	0.334	0.664	0.896	0.981
	STEINRLS	0.447	0.724	0.900	0.982
	EBAYES	0.476	0.745	0.911	0.984
DETROIT	OLS	0.267	0.272	0.260	0.284
	LINDLEY	0.487	0.780	0.939	0.989
	STEINRLS	0.401	0.631	0.849	0.969
	EBAYES	0.421	0.642	0.857	0.969
NEW YORK	OLS	0.238	0.239	0.239	0.242
	LINDLEY	0.193	0.488	0.807	0.964
	STEINRLS	0.285	0.397	0.677	0.929
	EBAYES	0.291	0.368	0.607	0.906
SAN DIEGO	OLS	0.464	0.471	0.463	0.482
	LINDLEY	0.231	0.558	0.826	0.955
	STEINRLS	0.349	0.634	1.132	1.040
	EBAYES	0.367	0.632	1.020	1.001

* Table shows actual calculation of Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 38

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER MEAN SQUARE ERROR
LOSS CRITERION*

FOURTH SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.304	0.311	0.306	0.304
	LINDLEY	0.309	0.636	0.880	0.981
	STEINRLS	0.450	0.641	0.845	0.971
	EBAYES	0.469	0.656	0.858	0.973
CHICAGO	OLS	0.178	0.176	0.179	0.175
	LINDLEY	0.382	0.722	0.916	0.985
	STEINRLS	0.330	0.551	0.820	0.965
	EBAYES	0.342	0.546	0.799	0.959
DALLAS	OLS	0.812	0.828	0.811	0.823
	LINDLEY	0.317	0.636	0.886	0.978
	STEINRLS	0.448	0.730	0.913	0.979
	EBAYES	0.471	0.732	0.919	0.981
DETROIT	OLS	0.216	0.224	0.225	0.226
	LINDLEY	0.524	0.810	0.946	0.990
	STEINRLS	0.483	0.687	0.879	0.979
	EBAYES	0.499	0.703	0.885	0.980
NEW YORK	OLS	0.224	0.231	0.220	0.219
	LINDLEY	0.228	0.540	0.835	0.972
	STEINRLS	0.268	0.384	0.672	0.930
	EBAYES	0.283	0.356	0.596	0.890
SAN DIEGO	OLS	0.720	0.740	0.729	0.759
	LINDLEY	0.169	0.455	0.779	0.962
	STEINRLS	0.252	0.482	1.043	1.117
	EBAYES	0.279	0.496	0.960	1.042

* Table shows actual calculation of Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 39

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER
ENSEMBLE MEAN SQUARE ERROR LOSS CRITERION*

SECOND SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.123	0.128	0.120	0.122
	LINDLEY	0.498	0.794	0.939	0.986
	STEINRLS	0.744	0.821	0.916	0.985
	EBAYES	0.751	0.821	0.911	0.983
CHICAGO	OLS	0.065	0.064	0.065	0.066
	LINDLEY	0.596	0.837	0.957	0.994
	STEINRLS	0.651	0.770	0.914	0.984
	EBAYES	0.656	0.762	0.902	0.980
DALLAS	OLS	0.377	0.370	0.357	0.363
	LINDLEY	0.290	0.638	0.881	0.980
	STEINRLS	0.608	0.778	0.915	0.982
	EBAYES	0.643	0.811	0.933	0.986
DETROIT	OLS	0.088	0.093	0.094	0.090
	LINDLEY	0.781	0.902	0.972	0.995
	STEINRLS	0.713	0.791	0.913	0.986
	EBAYES	0.730	0.810	0.923	0.989
NEW YORK	OLS	0.077	0.072	0.070	0.074
	LINDLEY	0.422	0.783	0.935	0.990
	STEINRLS	0.624	0.698	0.826	0.967
	EBAYES	0.628	0.677	0.769	0.940
SAN DIEGO	OLS	0.182	0.176	0.173	0.173
	LINDLEY	0.312	0.688	0.870	0.975
	STEINRLS	0.597	0.658	0.781	0.949
	EBAYES	0.619	0.686	0.821	0.963

* Table shows actual calculation of Ensemble Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 40

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER
ENSEMBLE MEAN SQUARE ERROR LOSS CRITERION*

THIRD SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.106	0.109	0.105	0.108
	LINDLEY	0.557	0.813	0.942	0.997
	STEINRLS	0.777	0.827	0.932	0.980
	EBAYES	0.787	0.829	0.933	0.979
CHICAGO	OLS	0.059	0.061	0.060	0.057
	LINDLEY	0.605	0.840	0.961	0.992
	STEINRLS	0.587	0.708	0.899	0.983
	EBAYES	0.600	0.680	0.863	0.976
DALLAS	OLS	0.297	0.308	0.301	0.288
	LINDLEY	0.330	0.649	0.899	0.982
	STEINRLS	0.629	0.794	0.930	0.990
	EBAYES	0.652	0.813	0.939	0.992
DETROIT	OLS	0.109	0.118	0.109	0.118
	LINDLEY	0.742	0.864	0.970	0.994
	STEINRLS	0.703	0.807	0.917	0.981
	EBAYES	0.710	0.812	0.920	0.980
NEW YORK	OLS	0.092	0.091	0.093	0.095
	LINDLEY	0.356	0.712	0.902	0.980
	STEINRLS	0.600	0.633	0.806	0.955
	EBAYES	0.601	0.611	0.690	0.869
SAN DIEGO	OLS	0.164	0.171	0.168	0.171
	LINDLEY	0.339	0.744	0.920	0.956
	STEINRLS	0.669	0.839	1.169	1.019
	EBAYES	0.679	0.839	1.089	0.997

* Table shows actual calculation of Ensemble Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

TABLE 41

RELATIVE PERFORMANCE OF
LINDLEY, STEINRLS, AND EBAYES ESTIMATORS
UNDER
ENSEMBLE MEAN SQUARE ERROR LOSS CRITERION*

FOURTH SUBSAMPLE

CITY	ESTIMATOR	SPECIFICATION OF "TRUE" BETA 0.5	1.0	2.0	5.0
BOSTON	OLS	0.140	0.144	0.132	0.138
	LINDLEY	0.444	0.755	0.923	0.988
	STEINRLS	0.773	0.852	0.945	0.986
	EBAYES	0.780	0.855	0.949	0.986
CHICAGO	OLS	0.067	0.066	0.065	0.065
	LINDLEY	0.604	0.842	0.947	0.992
	STEINRLS	0.583	0.697	0.881	0.979
	EBAYES	0.591	0.678	0.850	0.972
DALLAS	OLS	0.337	0.342	0.331	0.346
	LINDLEY	0.308	0.630	0.878	0.980
	STEINRLS	0.617	0.781	0.924	0.986
	EBAYES	0.635	0.787	0.930	0.988
DETROIT	OLS	0.105	0.108	0.110	0.109
	LINDLEY	0.729	0.900	0.973	0.993
	STEINRLS	0.734	0.820	0.928	0.987
	EBAYES	0.740	0.823	0.928	0.987
NEW YORK	OLS	0.069	0.073	0.069	0.068
	LINDLEY	0.503	0.825	0.930	0.992
	STEINRLS	0.681	0.700	0.842	0.968
	EBAYES	0.690	0.668	0.746	0.905
SAN DIEGO	OLS	0.286	0.285	0.276	0.294
	LINDLEY	0.202	0.497	0.790	0.965
	STEINRLS	0.430	0.657	1.198	1.160
	EBAYES	0.451	0.666	1.107	1.081

* Table shows actual calculation of Ensemble Mean Square Error for OLS. Values in the rows of other estimators are ratios of the loss of those estimators to that of OLS.

APPENDIX D

SAS PROGRAM FOR THE MONTE CARLO EXPERIMENT

This appendix provides the program written in PROC MATRIX language of SAS. Data for the program consist of variables of the American Housing Survey hedonic model. The program: 1) Divides the data for each city into four subsets; 2) Specifies the "true" parameter vector;

3) Performs estimation with OLS and seven biased estimators; and, 4) Compares estimator performance under four separate loss criteria.

```
PROC MATRIX;
*************************************************************;
***    DEFINE THE VARIABLES AND MATRICES FOR USE        ***;
***               IN THE PROGRAM                        ***;
*************************************************************;
ONE=1;  SEED = 45674 ;
NESTS = 8 ;  NREPS = 1000 ;  NVARS = 13;
 VARRN=
  'ONE' 'CENCITY' 'BATHMORE' 'ROOMMORE' 'AGE' 'STRUCTUR'
  'EXCNBHD' 'BATH1' 'BED2' 'GARAGE' 'CENAIR' 'NOPRIV'
  'PPERROOM' ;
 CNRSUM='RISK TB1' 'SE TB1' 'RISK TB2' 'SE TB2'
          'RISK TB3' 'SE TB3' 'RISK TB4' 'SE TB4';
 CNGSUM='TB1' 'TB2' 'TB3' 'TB4';
 RN= 'OLS' 'RIDGE' 'ITHKB' 'LINDLEY' 'STRAW' 'STEIN'
      'STEINRLS' 'EBAYES';
 CN= 'RMSE-XTX' 'STD ERR' 'RMSE-I' 'STD ERR'
      'MAPE' 'STD ERR';
 CN2= 'MSE-XTX' 'MSE-I' 'MAPE' 'EMSE' ;
 BCOLS= 'INTERCEPT' 'STD ERR' 'CENCITY' 'STD ERR'
        'BATHMORE' 'STD ERR' 'ROOMMORE' 'STD ERR'
        'AGE' 'STD ERR' 'STRUCTUR' 'STD ERR'
        'EXCNBHD' 'STD ERR' 'ENSEMBLE' 'STD ERR';
 BCOLS2='INTERCEPT' 'CENCITY' 'BATHMORE' 'ROOMMORE' 'AGE'
        'STRUCTUR' 'EXCNBHD' 'ENSEMBLE';
```

```
LMA1=J(NREPS,NESTS,0); LMA2=LMA1; LMA3=LMA1;
SUMBIAS=J(NVARS,NESTS,0);
B1MA=J(NREPS,NESTS,0); B2MA=B1MA; B3MA=B1MA;
B4MA=B1MA; B5MA=B1MA; B6MA=B1MA; B7MA=B1MA;
EMSEMAT=B1MA;
GMATA=J(NESTS,3,0); BGMATA=GMATA;
R1SUM=J(8,8,0); R2SUM=R1SUM; R3SUM=R1SUM;
B1SUM=J(8,8,0); B2SUM=B1SUM; B3SUM=B1SUM;
B4SUM=J(8,8,0);   B5SUM=B1SUM;    B6SUM=B1SUM;    B7SUM=B1SUM;
EMSUM=B1SUM;
G1SUM=J(8,4,0); G2SUM=G1SUM; G3SUM=G1SUM;
BG1SUM=J(8,4,0); BG2SUM=BG1SUM; BG3SUM=BG1SUM;
BG4SUM=J(8,4,0); BG5SUM=BG1SUM; BG6SUM=BG1SUM; BG7SUM=BG1SUM;
EMGSUM=J(8,4,0);
*************************************************************;
***          CALL IN THE SAMPLE OF CITY DATA             ***;
*************************************************************;
 %LET INDVAR=
  ONE CENCITY BATHMORE ROOMMORE AGE STRUCTUR EXCNBHD
  BATH1 BED2  GARAGE CENAIR NOPRIV PPERROOM    ;
FETCH Y  DATA =CHICA(KEEP=LHV);
FETCH XX  DATA=CHICA(KEEP=&INDVAR);
  K=NCOL(XX);
  T=NROW(XX);
     X1=XX(,1:7)  ;
     X2=XX(,8:K);
     X = X1 || X2  ;
*************************************************************;
***    PERFORM OLS ON THE FULL SAMPLE OF CITY DATA       ***;
***       AND SPECIFY THE TRUE PARAMETER VECTOR          ***;
*************************************************************;
LINK OLS;
  TRUEBETA=BHAT;
  IF BHAT(3,1) < 0 THEN TRUEBETA(3,1) = .3813;
  IF BHAT(4,1) < 0 THEN TRUEBETA(4,1) = .0736;
  IF BHAT(5,1) > 0 THEN TRUEBETA(5,1) = -.0029;
  IF BHAT(6,1) > 0 THEN TRUEBETA(6,1) = 0 ;
  IF BHAT(7,1) < 0 THEN TRUEBETA(7,1) = .1696;
  IF BHAT(10,1) < 0 THEN TRUEBETA(10,1) = .0696;
  IF BHAT(11,1) < 0 THEN TRUEBETA(11,1) = .1018;
  IF BHAT(12,1) > 0 THEN TRUEBETA(12,1) = -.0207;
  IF BHAT(13,1) > 0 THEN TRUEBETA(13,1) = -.0545;
  SDEV=SQRT(SIGHAT2);
```

```
*************************************************************;
***     DIVIDE THE CITY SAMPLE INTO FOUR SUBSAMPLES      ***;
***   AND SELECT A SUBSAMPLE FOR ESTIMATOR COMPARISON    ***;
*************************************************************;
  SUBSIZE=INT(T#/4);
  XA=J(SUBSIZE,K,0); XB=XA; XCC=XA; XDD=XA;
  YA=J(SUBSIZE,1,0); YB=YA; YC=YA; YD=YA;
  I=4;
SUBSETS:IF I GT T-3 THEN GOTO GETOUT;
  XA(I#/4,)=XX(I,);
  YA(I#/4,)=Y(I,);
     I=I+1;
  XB((I-1)#/4,)=XX(I,);
  YB((I-1)#/4,)=Y(I,);
     I=I+1;
  XCC((I-2)#/4,)=XX(I,);
  YC((I-2)#/4,)=Y(I,);
     I=I+1;
  XDD((I-3)#/4,)=XX(I,);
  YD((I-3)#/4,)=Y(I,);
     I=I+1;
GOTO SUBSETS;
GETOUT:
     X1=XA(,1:7)  ;
     X2=XA(,8:K);
     X = X1 || X2  ;
     Y = YA  ;
     T = NROW(X);
*************************************************************;
***   SPECIFY THE LENGTH OF THE PARAMETER VECTOR AND     ***;
***  LINK THE SUBROUTINES THAT PERFORM THE ESTIMATION    ***;
***                  AND PRINT THE RESULTS               ***;
*************************************************************;
     TB=TRUEBETA#.5  ;
     TBIT=1; TBS=1; TBE=2;  SUMBIAS=J(NVARS,NESTS,0);
LINK GENERR;
   NOTE SUMMARY FOR TRUE BETA MULTIPLE = .5 ;
   NOTE TRUE BETA SPECIFICATION;
   PRINT TB ROWNAME=VARRN  ;
LINK PRINTRES;
LINK TBSUM;
     TB=TRUEBETA;
     TBIT=2; TBS=3; TBE=4;  SUMBIAS=J(NVARS,NESTS,0);
LINK GENERR;
   NOTE PAGE SUMMARY OF LOSSES FOR TRUE BETA MULTIPLE =1.0;
   NOTE TRUE BETA SPECIFICATION;
   PRINT TB ROWNAME=VARRN;
LINK PRINTRES;
LINK TBSUM;
```

```
      TB=TRUEBETA#2.0;
      TBIT=3; TBS=5; TBE=6;   SUMBIAS=J(NVARS,NESTS,0);
LINK GENERR;
   NOTE PAGE SUMMARY OF LOSSES FOR TRUE BETA MULTIPLE = 2.0;
   NOTE TRUE BETA SPECIFICATION;
   PRINT TB ROWNAME=VARRN;
LINK PRINTRES;
LINK TBSUM;
      TB=TRUEBETA#5.0;
      TBIT=4; TBS=7; TBE=8;  SUMBIAS=J(NVARS,NESTS,0);
LINK GENERR;
   NOTE PAGE SUMMARY FOR TRUE BETA MULTIPLE = 5.0;
   NOTE TRUE BETA SPECIFICATION;
   PRINT TB ROWNAME=VARRN;
LINK PRINTRES;
LINK TBSUM;
RETURN;
**********************************************************;
***                PRINT SUMMARY TABLES               ***;
**********************************************************;
**********************************************************;
***********   RELATIVE PERFORMANCE SUMMARIES *************;
********************************************* ************;
NOTE PAGE RELATIVE PERFORMANCE SUMMARIES FOR THE FOUR
          SPECIFICATIONS;
NOTE  MSE -- Q=XTX;
  PRINT G1SUM ROWNAME=RN COLNAME=CNGSUM;
NOTE  MSE -- Q=I  ;
  PRINT G2SUM ROWNAME=RN COLNAME=CNGSUM;
NOTE MAPE;
  PRINT G3SUM ROWNAME=RN COLNAME=CNGSUM;
NOTE PAGE RELATIVE PERFORMANCE SUMMARIES FOR INDIVIDUAL
          PARAMETERS;
NOTE BETA 1 MSE;
  PRINT BG1SUM ROWNAME=RN  COLNAME=CNGSUM;
NOTE BETA 2 MSE;
  PRINT BG2SUM ROWNAME=RN  COLNAME=CNGSUM;
NOTE BETA 3 MSE;
  PRINT BG3SUM ROWNAME=RN  COLNAME=CNGSUM;
NOTE BETA 4 MSE;
  PRINT BG4SUM ROWNAME=RN  COLNAME=CNGSUM;
NOTE PAGE BETA 5 MSE;
  PRINT BG5SUM ROWNAME=RN  COLNAME=CNGSUM;
NOTE BETA 6 MSE;
  PRINT BG6SUM ROWNAME=RN  COLNAME=CNGSUM;
NOTE BETA 7 MSE;
  PRINT BG7SUM ROWNAME=RN  COLNAME=CNGSUM;
NOTE ENSEMBLE MSE;
  PRINT EMGSUM ROWNAME=RN  COLNAME=CNGSUM;
```

```
*****************************************************************;
************            RISK SUMMARIES                 *********;
*****************************************************************;
NOTE PAGE ESTIMATOR COMPARISONS FOR THE FOUR SPECIFICATIONS;
NOTE OF TRUE BETA;
NOTE    ;
NOTE MSE -- Q=XTX;
  PRINT R1SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE MSE -- Q=I  ;
  PRINT R2SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE MAPE;
  PRINT R3SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE PAGE INDIVIDUAL PARAMETER SUMMARIES FOR THE FOUR
          SPECIFICATIONS;
NOTE BETA 1 MSE;
  PRINT B1SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE BETA 2 MSE;
  PRINT B2SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE BETA 3 MSE;
  PRINT B3SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE BETA 4 MSE;
  PRINT B4SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE PAGE BETA 5 MSE;
  PRINT B5SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE BETA 6 MSE;
  PRINT B6SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE BETA 7 MSE;
  PRINT B7SUM ROWNAME=RN COLNAME=CNRSUM;
NOTE ENSEMBLE MSE;
  PRINT EMSUM ROWNAME=RN COLNAME=CNRSUM;
   STOP;
****************************************************************;
***                     SUBROUTINES                         ***;
****************************************************************;

GENERR:
*This subroutine generates a vector of normally distributed*
* error terms and the "true" house price against which the*
* estimates are compared.  Then it calls the subroutine*
* "DOITALL" that performs the estimation*;
ERRVEC=J(T,1,0);
REP=0;
DO REPS = 1 TO NREPS;
REP = REP + 1;
   DO ELEMENT= 1 TO T;
     ERRVEC(ELEMENT,)=SDEV*RANNOR(SEED);
   END;
Y=X*TB + ERRVEC;
    LINK DOITALL;
END;
RETURN;
```

```
DOITALL:
*This subroutine links the calculation subroutines for*
*the various estimators that are compared.  Each estimator*
*subroutine calls the subroutine that calculates losses*;
DOITALL:
T=NROW(X);
K=NCOL(X);  K2=K+1; K1=K-1; YBAR=Y(.,); XBAR=X(.,);
LINK OLS;
  LINK OSUMMARY;
    LINK RIDGE;
     LINK LINDLEY;
      LINK STRAW;
        LINK STEIN;
         LINK STEINRLS;
          LINK EBAYES ;
RETURN;
**********************************************************;
******************      OLS       ***********************;
**********************************************************;
OLS:
EST=1;
XTX =XMULT(X', X);
EIGEN D V XTX;
SINV=SOLVE(XTX,I(K));
BHAT=SINV*X'*Y;
LBHAT=SQRT(SSQ(BHAT));
S=SSQ(Y-X*BHAT);
SIGHAT2=S#/(T-K);
COVB=SIGHAT2#SINV;
SD=SQRT(VECDIAG(COVB));
TVAL=BHAT#/SD;
ALL=BHAT||SD||TVAL;
CN = 'BHAT' 'SD' 'TVAL';
R2=1-(S#/(SSQ(Y-Y(.,))));
ADJ_R2=1-(1-R2)#(T-1)#/(T-K);
U1=(R2#(T-K))#/((1-R2)#(K-1));
PVAL1=1-PROBF(U1,K,T-K);
YHAT=X*BHAT;
BETAEST=BHAT;

**********************************************************;
*********************      RIDGE      ********************;
**********************************************************;
RIDGE:
EST = 2 ;
SIGHAT2=S#/(T-K);
B=INV(XTX)*X'*Y;
KHAT=(K)#SIGHAT2#/SSQ(B);
BRIDGE=INV(XTX+KHAT#I(K))*X'*Y;
COVRIDGE=SIGHAT2#INV(XTX+KHAT#I(K))*XTX*INV(XTX+KHAT#I(K));
STDDEV=SQRT(VECDIAG(COVRIDGE));
*NOTE PAGE HKB ESTIMATOR---WITHOUT ITERATION;
```

```
B_HKB=BRIDGE;
LHKB=SQRT(SSQ(B_HKB));
BETAEST=B_HKB;
YHAT=X*BETAEST;
***  COMPUTATION OF LOSSES;
LINK OSUMMARY;
******     ITERATE ON HKB ESTIMATOR    ******;
CRIT=1;
KOLD=KHAT;
DO IND=1 TO 20 WHILE (CRIT GT .0001);
KHAT=(K)#SIGHAT2 #/ SSQ(BRIDGE);
CRIT=KHAT-KOLD;
BRIDGE=INV(XTX+KHAT#I(K))*X'*Y;
KOLD=KHAT;
END;
EST = 3 ;
COVRIDGE=SIGHAT2#INV(XTX+KHAT#I(K))*XTX*INV(XTX+KHAT#I(K));
STDDEV=SQRT(VECDIAG(COVRIDGE));
*NOTE PAGE HKB ESTIMATOR---WITH ITERATION;
ITB_HKB=BRIDGE;
BETAEST=ITB_HKB;
YHAT=X*BETAEST;
***  COMPUTATION OF LOSSES *** ;
LINK OSUMMARY;
RETURN;
*************************************************************;
*********************     LINDLEY     *********************;
*************************************************************;
LINDLEY:
EST = 4 ;
ESTIMATE=1;
Q=XTX;
A=I(K);
R=A(2:K,);
IRX=INV(R*SINV*R');
EIGEN L PMAT IRX;
AHALF=PMAT*DIAG(SQRT(L))*PMAT';
MU =AHALF*R*SINV*Q*SINV*R'*AHALF;
EIGEN L PMAT MU;
LMAX=MAX(L);
AMAX=(2#/(T-K+2))#((TRACE(MU)#/LMAX)-2);
IF AMAX LE 0 THEN A=0;
ELSE A=AMAX;
BRLS=YBAR//J(K1,1,0);
LINK LINDSUB;
LINK OSUMMARY;
RETURN;

LINDSUB:
CONST=1-(A#S#/(BHAT'*R'*IRX*R*BHAT));
ZZ=CONST//0;
CONST=MAX(ZZ);
```

```
BLINDLEY=CONST#(BHAT-BRLS)+BRLS; VECLEN=SQRT(SSQ(BLINDLEY));
BETAEST=BLINDLEY;
YHAT=X*BETAEST;
RETURN;
***********************************************************;
********************       STRAW        ****************;
***********************************************************;
STRAW:
EST = 5 ;
ESTIMATE=2;
Q1 = XTX ;
C1 = INV(Q1)*XTX ;
EIGEN LM1 LMV1 C1 ;
LMAX1 = LM1(1,1) ;
  AMAX = 2#(K-2)#/((T-K+2)#LMAX1);
  A = AMAX ;
  C=C1;
LINK STRAWSUB;
LINK OSUMMARY;
RETURN;

STRAWSUB:
  G=(K#2)#/(T-K+2) ;
  LITTLEB=INV(XTX)*X'*Y;
  SHRINK= A#S#/((LITTLEB'*XTX*LITTLEB)+(G#S)) ;
  CONST=(I(K)+SHRINK*C) ;
  S1=INV(CONST)*LITTLEB;
  BETAEST=S1;
  YHAT=X*BETAEST;
  RETURN;
********************************************************;
******************   STEIN     *************************;
********************************************************;
STEIN:
EST = 6 ;
ESTIMATE=4;
AMAX=2#(K-2)#/(T-K+2);
A=AMAX;
LINK STEINSUB;
LINK OSUMMARY;
RETURN;

STEINSUB:
CONST=(1-A#S#/(BHAT'*XTX*BHAT));
ZZ=CONST//0; CONST=MAX(ZZ);
BETAEST=CONST#BHAT;
BSTEIN=BETAEST;
YHAT=X*BETAEST;
RETURN;
```

```
****************************************************************;
***********************    STEINRLS    **********************;
****************************************************************;
STEINRLS:
EST = 7 ;
ESTIMATE = 5;
K1=NCOL(X1); K2=NCOL(X2); K11=K1+1;
BTOP=INV(X1'*X1)*X1'*Y;
BBOT=J(K2,1,0);
BRLS=BTOP//BBOT;
RS=SSQ(Y-X*BRLS);
RR2=1-(RS#/(SSQ(Y-Y(.,))));
U2=(RR2#(T-K1))#/((1-RR2)#(K1-1));
PVAL2=1-PROBF(U2,K1,T-K1);
U3=(SSQ(Y-X*BRLS)-S)#/(K2#SIGHAT2);
PVAL3=(1-PROBF(U3,K2,T-K,K2#/2));
Q=XTX ;
A=I(K);
R=A(8:K,);
IRX=INV(R*SINV*R');
EIGEN L PMAT IRX;
AHALF=PMAT*DIAG(SQRT(L))*PMAT';
MU =AHALF*R*SINV*Q*SINV*R'*AHALF;
EIGEN L PMAT MU;
LMAX=MAX(L);
AMAX=(2#/(T-K+2))#((TRACE(MU)#/LMAX)-2);
IF AMAX LE 0 THEN A=0;
ELSE A=AMAX;
LINK RLSSUB;
LINK OSUMMARY;
RETURN;

RLSSUB:
CONST=1-(A#S#/(BHAT'*R'*IRX*R*BHAT));
ZZ=CONST//0;
CONST=MAX(ZZ);
BETAEST=CONST#(BHAT-BRLS)+BRLS;
BSTRLS=BETAEST;
YHAT=X*BETAEST;
RETURN;
*************************************************************;
*****************    EBAYES    *****************************;
*************************************************************;
EBAYES:
EST = 8;
ESTIMATE = 6        ;
BHAT = SINV*X'*Y;
K1=NCOL(X1); K2=NCOL(X2); K11=K1+1;
B1=BHAT(1:K1,);
B2=BHAT(K11:K,);
B1RLS=INV(X1'*X1)*X1'*Y;
AMAX=(2#(K2-2))#/(T-K+2);
```

```
IF AMAX LE 0 THEN A=0;
ELSE A=AMAX;
LINK EBSUB;
LINK OSUMMARY;
RETURN;

EBSUB:
CONST=1-(A#S#/(B2'*(X2'*X2-X2'*X1*INV(X1'*X1)
        *X1'*X2)*B2));
Z=0//CONST;
CONST=MAX(Z);
B1EB=B1RLS+CONST#(B1-B1RLS);
B2EB=CONST#B2;
BETAEST=B1EB//B2EB;
BBAYES=BETAEST;
YHAT=X*BETAEST;
RETURN;

OSUMMARY:
*These subroutines calculate the out-of-sample losses
* at the individual observation level and fill temporary
* matrices with these values*;
LINK LOSSES;
LMA1(REP,EST)=LOSS1; LMA2(REP,EST)=LOSS2; LMA3(REP,EST)=LOSS3;
SUMBIAS(,EST)=SUMBIAS(,EST) + LOSS4;
  LINK BLOSS;
    B1MA(REP,EST)=B1L;
    B2MA(REP,EST)=B2L;
    B3MA(REP,EST)=B3L;
    B4MA(REP,EST)=B4L;
    B5MA(REP,EST)=B5L;
    B6MA(REP,EST)=B6L;
    B7MA(REP,EST)=B7L;
    EMSEMAT(REP,EST)=EMSE;
RETURN;
LOSSES:
LOSS1=(BETAEST-TB)'*XTX*(BETAEST-TB);
LOSS2=(BETAEST-TB)'*I(K)*(BETAEST-TB);
PE=ABS((Y-YHAT)#/Y);
LOSS3=(PE(+,)#/T)*100;
LOSS4=(BETAEST)-(TB);
RETURN;
BLOSS:
B1L=(BETAEST(1,1)-TB(1,1))##2;
B2L=(BETAEST(2,1)-TB(2,1))##2;
B3L=(BETAEST(3,1)-TB(3,1))##2;
B4L=(BETAEST(4,1)-TB(4,1))##2;
B5L=(BETAEST(5,1)-TB(5,1))##2;
B6L=(BETAEST(6,1)-TB(6,1))##2;
B7L=(BETAEST(7,1)-TB(7,1))##2;
EMSE=B1L+B2L+B3L+B4L+B5L+B6L+B7L;
RETURN;
```

```
PRINTRES:
*This subroutine calculates prediction risk estimates,
* estimates of the variance of prediction risk, and gains
* relative to OLS.  It also creates tables for displaying
these estimates*;
*************************************************************;
*  TABLES OF MEAN SQUARE ERROR OF IN-SAMPLE PREDICTION, **;
*  MEAN SQUARE ERROR, AND MEAN ABSOLUTE PERCENTAGE ERROR *;
*************************************************************;
   RHAT1=LMA1(.,); RHAT2=LMA2(.,); RHAT3=LMA3(.,);
   RVAR1=(LMA1-J(NREPS,1,1)*RHAT1)##2;  RVAR1A=RVAR1(.,);
    STERR1=SQRT(RVAR1A#/NREPS);    ST1=STERR1';
   RVAR2=(LMA2-J(NREPS,1,1)*RHAT2)##2;  RVAR2A=RVAR2(.,);
    STERR2=SQRT(RVAR2A#/NREPS);    ST2=STERR2';
   RVAR3=(LMA3-J(NREPS,1,1)*RHAT3)##2;  RVAR3A=RVAR3(.,);
    STERR3=SQRT(RVAR3A#/NREPS);     ST3=STERR3';
   RMAT=RHAT1'||ST1||RHAT2'||ST2||RHAT3'||ST3;
 ***CREATE TABLE FOR LOSSES AS A PERCENT OF OLS LOSSES****;
 RMATA=RHAT1'||RHAT2'||RHAT3';
 OLSROW=RMATA(1,);
 GMATA=RMATA#/(J(NESTS,1,1)*OLSROW);
 GMATA(1,)=OLSROW;
*************************************************************;
*  TABLES OF MEAN SQUARE ERROR FOR INDIVIDUAL PARAMETERS *;
*           AND ENSEMBLE MEAN SQUARE ERROR              *;
*************************************************************;
   BRHAT1=B1MA(.,);  BRHAT2=B2MA(.,);  BRHAT3=B3MA(.,);
   BRHAT4=B4MA(.,);  BRHAT5=B5MA(.,);  BRHAT6=B6MA(.,);
   BRHAT7=B7MA(.,);  EMHAT=EMSEMAT(.,);
   BRVAR1=(B1MA-J(NREPS,1,1)*BRHAT1)##2;
     BRVAR1A=BRVAR1(+,)#/NREPS;
     BSTERR1=SQRT(BRVAR1A#/NREPS);  BST1=BSTERR1';
   BRVAR2=(B2MA-J(NREPS,1,1)*BRHAT2)##2;
     BRVAR2A=BRVAR2(+,)#/NREPS;
     BSTERR2=SQRT(BRVAR2A#/NREPS);  BST2=BSTERR2';
   BRVAR3=(B3MA-J(NREPS,1,1)*BRHAT3)##2;
     BRVAR3A=BRVAR3(+,)#/NREPS;
     BSTERR3=SQRT(BRVAR3A#/NREPS);  BST3=BSTERR3';
   BRVAR4=(B4MA-J(NREPS,1,1)*BRHAT4)##2;
     BRVAR4A=BRVAR4(+,)#/NREPS;
     BSTERR4=SQRT(BRVAR4A#/NREPS);   BST4=BSTERR4';
   BRVAR5=(B5MA-J(NREPS,1,1)*BRHAT5)##2;
     BRVAR5A=BRVAR5(+,)#/NREPS;
     BSTERR5=SQRT(BRVAR5A#/NREPS);   BST5=BSTERR5';
   BRVAR6=(B6MA-J(NREPS,1,1)*BRHAT6)##2;
     BRVAR6A=BRVAR6(+,)#/NREPS;
     BSTERR6=SQRT(BRVAR6A#/NREPS);   BST6=BSTERR6';
   BRVAR7=(B7MA-J(NREPS,1,1)*BRHAT7)##2;
     BRVAR7A=BRVAR7(+,)#/NREPS;
     BSTERR7=SQRT(BRVAR7A#/NREPS);   BST7=BSTERR7';
```

```
   EMVAR=(EMSEMAT-J(NREPS,1,1)*EMHAT)##2;
     EMVARA=EMVAR(+,)#/NREPS;
     EMSTERR=SQRT(EMVARA#/NREPS);  EMST=EMSTERR';
   BRMAT=BRHAT1'||BST1||BRHAT2'||BST2||BRHAT3'||BST3||
         BRHAT4'||BST4||BRHAT5'||BST5||BRHAT6'||BST6||
         BRHAT7'||BST7||EMHAT'||EMST;
*****CREATE TABLE OF LOSSES AS A PERCENT OF OLS LOSSES****;
BRMATA=BRHAT1'||BRHAT2'||BRHAT3'||BRHAT4'||
       BRHAT5'||BRHAT6'||BRHAT7'||EMHAT';
 BOLSROW=BRMATA(1,);
 BGMATA=BRMATA#/(J(NESTS,1,1)*BOLSROW);
 BGMATA(1,)=BOLSROW;
 EMSEVEC=BGMATA(,8);
    NOTE FULL MODEL AND ENSEMBLE GAINS;
    GMATA=GMATA||EMSEVEC;
    PRINT GMATA ROWNAME=RN COLNAME=CN2 ;
    NOTE;
    NOTE RELATIVE PERFORMANCE FOR INDIVIDUAL PARAMETERS;
    NOTE;
    PRINT BGMATA ROWNAME=RN COLNAME=BCOLS2;
    NOTE PAGE BIAS;
    BIAS=SUMBIAS#/NREPS;
    PRINT BIAS ROWNAME=VARRN COLNAME=RN ;
    BIAS_SQ=BIAS##2;
    NOTE BIAS SQUARED;
    PRINT BIAS_SQ ROWNAME=VARRN  COLNAME=RN;
    INDMSE=BRHAT1//BRHAT2//BRHAT3//BRHAT4//
           BRHAT5//BRHAT6//BRHAT7;
    NOTE VARIANCE;
    VARIANCE=INDMSE-BIAS_SQ(1:7,);
    PRINT VARIANCE ROWNAME=VARRN COLNAME=RN;
    NOTE PAGE MSE MATRIX FOR INDIVIDUAL PARAMETERS;
    PRINT INDMSE ROWNAME=VARRN COLNAME=RN;
RETURN;

TBSUM:
*This subroutine creates a summary of prediction risks
*  and gains relative to OLS for all four parameter
*  length specifications *;
   R1SUM(,TBS:TBE)=RMAT(,1:2); **RISK SUMMARY FOR Q=XTX**;
   R2SUM(,TBS:TBE)=RMAT(,3:4); **RISK SUMMARY FOR Q=I  **;
   R3SUM(,TBS:TBE)=RMAT(,5:6); **RISK SUMMARY FOR MAPE **;
   G1SUM(,TBIT)=GMATA(,1); **GAIN SUMMARY FOR Q=XTX**;
   G2SUM(,TBIT)=GMATA(,2); **GAIN SUMMARY FOR Q=I  **;
   G3SUM(,TBIT)=GMATA(,3); **GAIN SUMMARY FOR MAPE **;
   B1SUM(,TBS:TBE)=BRMAT(,1:2); **RISK SUMMARY FOR B1**;
   B2SUM(,TBS:TBE)=BRMAT(,3:4); **RISK SUMMARY FOR B2**;
   B3SUM(,TBS:TBE)=BRMAT(,5:6); **RISK SUMMARY FOR B3**;
   B4SUM(,TBS:TBE)=BRMAT(,7:8); **RISK SUMMARY FOR B4**;
   B5SUM(,TBS:TBE)=BRMAT(,9:10); **RISK SUMMARY FOR B5**;
   B6SUM(,TBS:TBE)=BRMAT(,11:12); **RISK SUMMARY FOR B6**;
   B7SUM(,TBS:TBE)=BRMAT(,13:14); **RISK SUMMARY FOR B7**;
```

```
   EMSUM(,TBS:TBE)=BRMAT(,15:16); **RISK SUMMARY FOR EMSE**;
   BG1SUM(,TBIT)=BGMATA(,1); **GAIN SUMMARY FOR B1**;
   BG2SUM(,TBIT)=BGMATA(,2); **GAIN SUMMARY FOR B2**;
   BG3SUM(,TBIT)=BGMATA(,3); **GAIN SUMMARY FOR B3**;
   BG4SUM(,TBIT)=BGMATA(,4); **GAIN SUMMARY FOR B4**;
   BG5SUM(,TBIT)=BGMATA(,5); **GAIN SUMMARY FOR B5**;
   BG6SUM(,TBIT)=BGMATA(,6); **GAIN SUMMARY FOR B6**;
   BG7SUM(,TBIT)=BGMATA(,7); **GAIN SUMMARY FOR B7**;
   EMGSUM(,TBIT)=BGMATA(,8); **GAIN SUMMARY FOR EMSE;
RETURN;
/*
//
```

VITA

JOHN R. KNIGHT

AREAS OF INTEREST

Teaching: Real Estate, Investments, Portfolios.

Research: Real Estate Finance and Investments.

VISA STATUS

U. S. Citizen

EDUCATION

Louisiana State University	Finance	1987-	Ph.D(1990)
Louisiana State University	Business	1975-78	MBA
Tulane University	Economics	1965-69	BA

HONORS

Alumni Fellow	Louisiana State University	1987-90
Scholars & Fellows	Tulane University	1965-68
Beta Gamma Sigma	Louisiana State University	1978

TEACHING EXPERIENCE

January 1984 - August 1987: Adjunct Faculty Member, College of the Desert, Palm Desert, California. Taught courses in Business Statistics, Macroeconomics, Financial Accounting, and Personal Finance.

WORK EXPERIENCE

August 1981-January 1987: Various positions in the fields of accounting and finance. Experience includes staff accounting with a CPA firm, controllership of a real estate development group, a retailing vice-presidency, and two years as Chief Accountant of Coachella Valley Water District.

January 1970-July 1981: U. S. Naval Officer.

WORKING PAPERS

Frankfurter, George M. and Knight, John R., "Increasing Portfolio Efficiency Through the Prediction of Management Buyouts," Working Paper, Louisiana State University.

Topic: "Biased Estimation in the Context of the Hedonic Pricing Model for Housing."

Abstract: This dissertation compares eight biased estimators as alternatives to Ordinary Least Squares estimation in the context of predicting residential real estate prices. It considers ridge rule estimation and principal components regressions, and it introduces the use of Stein-like rules for predicting housing prices. The study examines relative performance of these estimators in three data settings and under four separate assumptions regarding loss criteria.

The partitioned Stein-like estimators do well in all three data environments. Two of them provide especially impressive performance. Under quadratic loss, in the Monte Carlo experiment, these estimators outperform all compared alternatives across the entire range of generated samples.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: John R. Knight

Major Field: Business Administration (Finance)

Title of Dissertation: BIASED ESTIMATION IN THE CONTEXT OF THE HEDONIC PRICING MODEL FOR HOUSING

Approved:

[illegible]

Major Professor and Chairman

[illegible]

Dean of the Graduate School

EXAMINING COMMITTEE:

R. Carter Hill

James D. Shilling

John L. Glascock

John S. Howe

[illegible]

Date of Examination:

June 6, 1990