ESSAYS IN APPLIED ECONOMETRICS

By

Mark Harrison Showalter

B.A., Economics, Brigham Young University (1986)

Submitted to the School of Management

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

May 1991

© Massachusetts Institute of Technology, 1991. All rights reserved.

Signature of Author ______________________
School of Management
May 1991

Certified by ______________________
Ernst R. Berndt
Professor of Applied Economics
Thesis Supervisor

Accepted by ______________________
James B. Orlin
Chairman, Ph.D. Program

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUN 21 1991

LIBRARIES

ARCHIVES

ESSAYS IN APPLIED ECONOMETRICS

By

Mark Harrison Showalter

Submitted to the School of Management
in May, 1991 in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
at the
Massachusetts Institute of Technology

Abstract

This thesis is composed of three essays on various topics in applied econometrics. The first essay (coauthored with Ernst R. Berndt and Jeffrey M. Wooldridge) compares two estimators: the well-known Box-Cox (BC) estimator and the relatively new nonlinear least squares (NLS) alternative developed by Wooldridge. The estimators are each applied to three classical data sets to determine if significant differences are apparent between the two estimators both regarding to estimates of parameters and to goodness-of-fit statistics.

The second essay is a Monte Carlo study of the BC and NLS estimators. This is the first Monte Carlo study of the NLS estimator and it also extends previous Monte Carlo work for the BC estimator by examining the effects of autocorrelation and heteroskedasticity. The study is constructed to allow comparison between the BC and the NLS estimators.

The third essay examines monopoly behavior in the presence of intertemporal demands. The defining characteristic of intertemporal demands is that demand in one period is affected by demand in previous and possibly future periods. Results for the measurement of monopoly power and the time path of consumption are examined in detail. An empirical framework is developed to test for forward-looking behavior on the part of consumers and the monopolist. This framework is then applied to a panel data set of cigarette prices and consumption.

Thesis Committee:

Ernst R. Berndt, *Professor of Applied Economics, MIT (Chair)*
Jeffrey M. Wooldridge, *Assistant Professor of Economics, MIT*
Ann F. Friedlaender, *Professor of Economics, MIT*

Acknowledgements

I am deeply indebted to the members of my thesis committee, Ernst R. Berndt, Jeffrey M. Wooldridge, and Ann F. Friedlaender. I could not have asked for more support, more encouragement or wiser guidance than was freely given by each of these people. Ernie has been much more of a father to me than an advisor. Although severely inadequate, I express my deep appreciation to him for his help and guidance both in matters of academics and of life. Nan has been extremely supportive of all my efforts and I count it as my privilege to have worked with her. My awe for Jeff has only increased as I have gone from being his student to being a coauthor and friend. This thesis would not have been possible without these three individuals.

I am also grateful for my fellow students at MIT. Special thanks go to the four hardy souls who began this adventure with me: Chyi-Mei Chen, Lawrence Loh, Raghuram Rajan, and Miquel Villas-Boas. I feel it a privilege to have associated with these men of integrity and honor and look forward to following their paths of success in the future. Likewise the association with Ann Beatty, Vivian Ellis, Yegin Chen, Hua He, Mark McCabe, Arun Muralidhar, Kazu Ohasi, and Sung-Hwan Shin among numerous others has been a great pleasure and my sincere thanks go to each of them.

My deepest appreciation and love is reserved for my long-suffering wife and faithful companion, Marlene. Without her help and encouragement I would not have completed my first year of graduate school, let alone this thesis. My children: Brent, Tonya, Heidi, and Trina all deserve special thanks for giving me something to live for other than work and study.

Finally I express my gratitude to my parents who taught me the principles of being a good man. Someday I hope to attain them.

Contents

Essay I:

A Theoretical and Empirical Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative

by

Ernst R. Berndt
Mark H. Showalter
Jeffrey M. Wooldridge

Abstract:

The Box-Cox (BC) estimation procedure has been widely used in empirical work, often to aid in choosing a functional form. There are a number of problems with the BC estimator, however, including the lack of invariance of Wald test statistics involving slope parameter estimates to arbitrary scaling of the data, and difficulties in computing "fitted values" when the BC power transformation parameter differs from zero or one.

In this paper we employ an alternative generalized functional form estimator, recently developed by Jeffrey Wooldridge, and then compare it to the BC estimator both theoretically and empirically. We begin by briefly reviewing the drawbacks of the BC estimator, and then we propose a non-linear least squares (NLS) alternative which retains the desirable qualities of the BC estimator without involving the associated problems. We compare estimation results--point estimates, inferences and fitted values--using data from several classic hedonic regression studies. These include a wage rate equation, and two computer hedonic regression equations, one from Gregory Chow and the other an IBM data set that formed the basis of the new official BLS computer price index.

We discuss at length the lack-of-invariance property of Wald statistics inherent in nonlinear estimators, and illustrate the potential seriousness of the problem with empirical examples; then we propose a possible solution for the NLS estimator. We conclude by summarizing our findings and suggesting possible avenues for future research.

We have benefitted from the comments of participants at the MIT-Harvard joint econometrics seminar, at Boston College and the University of Montréal.

1. Introduction

The choice of an appropriate functional form is a very important issue in applied econometrics, for in many cases the underlying economic theory provides only limited guidance, e.g., use a functional form in which theoretically-consistent homogeneity and/or symmetry restrictions can be imposed. In practice, the Box-Cox and Box-Tidwell procedures have often been employed to choose among alternative functional forms. Spitzer (1984), however, has noted a very serious lack-of-scale invariance for t-statistics that emerges when one uses the Box-Cox or Box-Tidwell method. Moreover, unless the Box-Cox transformation parameter λ on the dependent variable y equals 0 or 1, one cannot solve for or compute $\hat{y}$, the fitted value of y, in closed form.

Recently, Wooldridge (1990) has developed an alternative to the Box-Cox procedure, one based on direct nonlinear least squares methods. With the nonlinear least squares (NLS) procedure, one can easily solve the fitted value problem inherent to the Box-Cox (BC) method, and although t-statistics on slope coefficients for the NLS estimator also lack invariance to arbitrary scaling of the dependent variable, Wooldridge has outlined how scale-invariant test statistics for exclusionary hypotheses can be conducted using the Lagrange multiplier test procedure. Moreover, since one important characteristic of the Box-Cox procedure is that it transforms the distribution of the dependent variable, Wooldridge has also derived computational formulae for obtaining heteroskedasticity-robust standard error estimates.

To understand better how important these theoretical issues might be in practice, it is necessary to implement the various procedures empirically and then to compare them. In this paper, therefore, we undertake an empirical comparison of the Box-Cox, nonlinear least squares, and weighted nonlinear least squares estimation procedures. Although choice of functional form is an important issue in almost all areas of applied econometrics, this issue is of special interest in labor economics and in hedonic pricing studies. We have therefore chosen one data set comparable to those commonly used in labor economics and two data sets previously employed in hedonic applications for our empirical comparisons.

The first data set, called CPS78, is similar to that used by many labor economists in estimating wage rate (or statistical earnings) equations and returns to education; this data set consists of 550 observations, randomly drawn from the May 1978 U.S. Current Population Survey. The second data set, called COLE, is that used by Cole et al. (1986) in their hedonic pricing study of mainframe computers in the U.S. from 1972 to 1984. This study is of special interest since in part it formed the basis of the official quality-adjusted price indexes for mainframe computers recently published by the U.S. Bureau of Economic Analysis.[1] The third data set, called CHOW, is that underlying the classic study by Chow (1967) of prices and the price elasticity of demand for mainframe computers in the U.S. from 1960 to 1965.

The outline of this paper is as follows. We begin with a theoretical overview, drawn in part from Wooldridge (1990). After providing a brief summary discussion of data sets and sources in section 3, in section 4 we present empirical evidence on the extent of scale invariance (or lack thereof), a comparison of parameter estimates and inference for the BC, NLS and WNLS estimators, and a comparison of residuals and goodness-of-fit measures. In section 5 we summarize, conclude, and outline future research issues.

---

[1]See Cole et al. (1986) and Cartwright (1986).

2. Approaches to Generalizing Functional Form

The primary purpose of this paper is to compare competing methodologies for generalizing functional form in econometrics. Although there are several approaches to generalizing functional form, most can be put into one of two broad classes: (i) transformation methods, and (ii) methods that directly specify flexible functional forms for $E(y|x)$. The former category is dominated by the approach popularized by Box and Cox (1964), and its numerous extensions; for a survey, see Spitzer (1982). An approach which works directly with $E(y|x)$, which will simply be referred to as "nonlinear regression methods", has recently been suggested by Wooldridge (1990). In this section we discuss each of these approaches in turn, focusing on the issues of robustness, efficiency, and scale invariance. We also consider goodness-of-fit measures for choosing among the approaches in actual empirical applications.

2.1 Transformation Methods

Let $y > 0$ be the variable of interest, and let $x = (1, x_2, \dots, x_K)$ be the vector of explanatory variables.[2] Throughout this paper, $\mathbf{x}$ can represent nonlinear transformations of an underlying set of variables; this is in contrast to y, which should be the economic variable of interest. Note that we assume $\mathbf{x}$ contains a constant.

Rather than postulating a model for $E(y|x)$ directly, the simplest transformation methods seek to find a transformation of y which has a linear conditional expectation. Although other classes could be considered, this paper focuses on the well-known Box-Cox transformation. For strictly positive y, define

[2]The Box-Cox procedure is ill defined for the case $y \leq 0$.

$$y(\lambda) \equiv \frac{y^{\lambda}-1}{\lambda}, \qquad \lambda \neq 0 \tag{2.1}$$

$$\equiv \log(y), \qquad \lambda = 0. \tag{2.2}$$

Given this definition, the weakest assumption employed in transformation methods is that, for some $\lambda \in \Re$ and some K x 1 vector $\beta \in \Re^K$,

$$E[y(\lambda) \mid x] = x\beta \tag{2.3}$$

(throughout, the "true" values of lambda and beta are denoted λ and β). By itself, of course, assumption (2.3) is not enough to recover $E(y|x)$ as a function of x, β , and λ. This is not to say that λ and β cannot be estimated under (2.3); see, for example, Amemiya and Powell (1981).[3]

Because we take $E(y|x)$ to be of substantial interest, (2.3) in isolation is almost useless. While β_j measures the effect of x_j on $E(y(\lambda)|x)$, it generally tells us nothing about the effect of x_j on $E(y|x)$. As discussed in Wooldridge (1990), a consistent framework for defining elasticities, semi-elasticities, and other economic quantities generally requires these quantities to be defined in terms of $E(y|x)$. To move from $E(y(\lambda)|x)$ to $E(y|x)$, at least one additional distributional assumption on $D(y|x)$ is needed. One natural assumption is that log(y) is normally distributed with constant variance. Although plausible, this is not the preferred assumption in the literature, as the transformation $y(\lambda)$ is presumed to "regularize" the distribution in addition to yielding a linear conditional expectation. Thus, the original Box-Cox model assumes that, for some $\lambda \in \Re$, there exists $\beta \in \Re^K$ and $\sigma^2 > 0$ such that

---

[3]However, Khazzoom (1989) has recently revealed some shortcomings of Amemiya and Powell's nonlinear 2SLS estimator in the Box-Cox context.

$$y(\lambda)|x \sim N(x\beta,\sigma^2) \tag{2.4}$$

(see also Spitzer (1982) and Hinkley and Runger (1984)). As has been observed by many statisticians and econometricians, (2.4) cannot strictly be true unless $\lambda = 0$. Thus, the distribution D(y | x) -- and in particular the expectation E(y | x) -- implied by (2.4) is not well-defined, nor do the traditional consistency properties of maximum likelihood estimation carry over (see, for example, Draper and Cox (1969) and Amemiya and Powell (1981)).

From our perspective, what is important is how well the approximation (2.4) allows one to estimate E(y | x). This raises the rather important issue of how one obtains predictions of y in transformation models. Under (2.4), one might use the naive approach as suggested by equations (2.5) and (2.6):

$$E(y|x)=[1+\lambda x\beta]^{1/\lambda},\lambda\neq 0 \tag{2.5}$$

$$=\exp(x\beta),\lambda=0. \tag{2.6}$$

Of course these are incorrect given (2.4), but neither is there a well-defined conditional expectation function. However, one might want to use the normality assumption more intensively. If one believes that u | x is distributed (approximately) as $N(0,\sigma^2)$, then the natural expectations are

$$E(y|x)=\int_{u\geq -\frac{1}{\lambda}(1+\lambda x\beta)}^{+\infty}[1+\lambda x\beta+\lambda u]^{1/\lambda}\frac{(1/\sigma)\phi(u/\sigma)}{\Phi(\frac{1}{\lambda\sigma}(1+\lambda x\beta))}du, \qquad \lambda>0 \tag{2.7}$$

$$= \int_{-\infty}^{u \le -\frac{1}{\lambda}(1+\lambda x\beta)} [1+\lambda x\beta+\lambda u]^{1/\lambda} \frac{(1/\sigma)\phi(u/\sigma)}{\Phi(-\frac{1}{\lambda\sigma}(1+\lambda x\beta))} du, \quad \lambda<0 \qquad (2.8)$$

$$= \exp(\sigma^2/2+x\beta), \lambda=0, \qquad (2.9)$$

where $\phi(z)$ denotes the standard normal density and $\Phi(z)$ denotes the standard normal distribution function. The integration in (2.7) and (2.8) must be restricted to the regions $u \geq -\lambda^{-1}(1 + \lambda x\beta)$ and $u \leq -\lambda^{-1}(1 + \lambda x\beta)$, respectively, to ensure that the expectation is well-defined. Estimates of $E(y|x)$ are obtained from (2.7)-(2.9) once $\hat{\lambda}$, $\hat{\beta}$, and $\hat{\sigma}^2$ have been computed, usually by quasi-maximum likelihood methods.

Given observations $\{(x_t,y_t):t=1,2,...,N\}$, the QMLE's of λ, β, and σ^2 solve the problem

$$\underset{\lambda,\beta,\sigma^2}{Max} \sum_{t=1}^{N} \ell_t(\lambda,\beta,\sigma^2) \qquad (2.10)$$

where $\ell_t(\lambda,\beta,\sigma^2)$ is the conditional log-likelihood of y_t given x_t for observation t:

$$\ell_t(\lambda,\beta,\sigma^2) \equiv k_0 - (\frac{1}{2})\log(\sigma^2) - (\frac{1}{2\sigma^2})(y_t(\lambda) - x_t\beta)^2 + (\lambda-1)\log(y_t) \qquad (2.11)$$

(in a time series context, it is assumed that $D(y_t|x_t) = D(y_t|x_t, \; y_{t-1},x_{t-1},...)$, i.e. there is no dynamic misspecification). Let $\theta \equiv (\lambda,\beta,\sigma^2)$ be the vector of parameters, and let

$$s_t(\theta) \equiv \nabla_\theta \ell_t(\theta) \qquad (2.12)$$

be the 1 x (K+2) gradient of the conditional log-likelihood with respect to θ. If the distributional assumption (2.4) (approximately) holds then the QMLE $\hat{\theta}$ is asymptotically normally distributed about θ with asymptotic variance estimated most easily by

$$\hat{\mathfrak{I}}^{-1} \equiv \left(\sum_{t=1}^{N} s_t(\hat{\theta})' s_t(\hat{\theta}) \right)^{-1}. \tag{2.13}$$

The asymptotic standard errors are obtained as the square roots of the diagonal elements of $\hat{\mathfrak{I}}^{-1}$. [4] If (2.4) fails to hold then $\hat{\theta}$ is generally inconsistent for θ. Nevertheless, Draper and Cox (1969) argue that $\hat{\theta}$ is approximately consistent provided the distribution of $y(\lambda)$ is symmetric.[5] When $u|x$ depends on x, e.g. $\mathrm{Var}(y(\lambda)|x)$ is nonconstant, the QMLE based on the normality assumption (2.3) can be poorly behaved (e.g. Amemiya and Powell (1981); Seakes and Layson (1983)). In addition, Poirier (1978) finds that even if u is independent of x, the QMLE can exhibit severe asymptotic bias if $y(\lambda)$ is asymmetric. Generally speaking, the Box-Cox MLE can be very sensitive to the assumptions of homoskedasticity and normality of $y(\lambda)$. This is not a very desirable property of a method if it is intended primarily to generalize functional form.

In addition to it being nonrobust, the Box-Cox approach has another undesirable feature. This has to do with the lack of scale invariance of the t-statistics on $\hat{\beta}_j$, $j=1,...,K$. As was pointed out by Spitzer (1984) and others, the t-statistics on the coefficients $\hat{\beta}_j$, $j=1,...,K$, can be altered simply by multiplying y_t by a nonzero constant. This is unfortunate because the units of measurement of y is frequently arbitrary in economics (e.g. whether price is recorded in hundreds or thousands of dollars should be irrelevant). Dagenais and Dufour (1986) have noted that certain other nonlinear models

---

[4]On this, however, see Amemiya and Powell (1981).

[5]Note that Draper and Cox also assume that u is independent of x, so that theirs is a very limited finding.

have this feature (below, we show that the nonlinear regression approach also suffers from this problem).

One solution to this problem is, rather than to report t-statistics for coefficients, to use scale invariant likelihood ratio or Lagrange multiplier statistics for testing exclusion of each variable. For example, let $\ddot{s}_t \equiv s_t(\ddot{\theta}_{(j)})$ denote the 1 x (K+2) score evaluated at $\ddot{\theta}_{(j)}$, where $\ddot{\theta}_{(j)}$ is the QMLE computed under the restriction $\beta_j = 0$. Dagenais and Dufour (1986) show that the outer product LM statistic, obtained as $N - SSR = NR_u^2$ from the regression

$$1 \; on \; \ddot{s}_t, \quad t=1,\ldots,N, \tag{2.14}$$

is invariant to the scale of y. Under H_0: $\beta_j = 0$, NR_u^2 is distributed approximately as χ_1^2, provided (2.4) holds. While this is an attractive alternative, it is computationally expensive because it requires estimation of K distinct Box-Cox regression models (one for each β_j). Also, outer product forms of LM statistics are notorious for their poor finite sample properties; see Bollershev and Wooldridge (1988).

Rather than use an LM statistic, Dagenais and Dufour suggest using a particular version of Neyman's (1959) $C(\alpha)$ statistic. For the Box-Cox model, this results in significant computational advantages because λ need only be estimated once, from the unrestricted model. The reader is referred to the Dagenais and Dufour (1986) paper for further details.

Pedagogically, the LM and $C(\alpha)$ approaches are unsatisfying because they do not allow construction of confidence intervals; hypotheses of the form H_0: $\beta_j = b_j$ using an LM or $C(\alpha)$ test require a separate computation whenever b_j changes. Researchers typically prefer to see standard errors attached to parameter estimates, so that a confidence interval can be constructed, and t-statistics (Wald statistics) can be computed for testing any hypothesis of the form H_0: $\beta_j = b_j$. Thus, it seems useful to attempt to salvage the usual t-statistic.

The problem with the usual t-statistic is that, as the analyst searches over different scalings of y, the t-statistics of the $\hat{\beta}_j$ can be changed, sometimes (as we shall show) dramatically. This is, of course, a form of data mining, and is not attractive given the usual assumptions underlying statistical inference. One solution to this problem is to force the analyst to estimate a scale parameter using the sample data. This is what Spitzer (1984) recommends in order to obtain scale-invariant t-statistics. He suggests dividing each observation y_t by the sample geometric mean of $\{y_t:t=1,...,N\}$, say $\hat{v}$. The new variable $y_t/\hat{v}$ is trivially scale invariant, so one might think that this solves the problem.[6]

However, there are two potential problems with this approach. The first is that the coefficients $\hat{\beta}_j$ might become more difficult to interpret. Fortunately, this turns out not to be much of a problem because the estimate of marginal effects from (2.5) or (2.7) - (2.9) are affected exactly as one would expect: they are simply scaled down by $\hat{v}$. Moreover, point estimates of elasticities and semi-elasticities are invariant to the scaling of y.

The second problem is potentially more serious: one needs to account for the randomness of $\hat{v}$ when computing standard errors of $\hat{\beta}_j$ and $\hat{\lambda}$. The simplest way to address this problem is to view the estimator $\hat{\theta}$ as a two-step estimator. Thus consider the extended model

---

[6]One additional advantage of employing this geometric mean transformation is that with the transformed data, maximizing the log-likelihood function is equivalent to minimizing the sum of squared residuals. For a discussion of this computational nuance, see Zarembka (1968).

$$y(\nu,\lambda) \sim N(x\beta,\sigma^2) \tag{2.15}$$

$$\nu \equiv \exp[E(\log(y))] \tag{2.16}$$

where $y(\nu,\lambda) = ((y/\nu)^{\lambda} - 1)/\lambda$, $\lambda \neq 0$, $y(\nu,\lambda) = \log(y/\nu)$, $\lambda = 0$. The parameter ν is the population geometric mean of y. Admittedly, ν could be defined to be one of a variety of other scale parameters, e.g. $\nu = E(y)$. A researcher is free to choose any moment of y, provided the moment exists. The important point is that, no matter how ν is defined, because it is estimated using sample data, the variance of the estimator $\hat{\nu}$ should be accounted for in any inference procedures.

The QMLE $\hat{\theta}$ now solves[7]

$$\max_{\theta} \sum_{t=1}^{N} \ell_t(\theta;\hat{\nu}), \qquad \text{where} \quad \hat{\nu} \equiv \exp\left(\frac{1}{N}\sum_{t=1}^{N} \log(y_t)\right) \quad \text{and} \tag{2.17}$$

$$\ell_t(\theta,\nu) \equiv k_0 - (1/2)\log(\sigma^2) - (1/2)(y_t(\nu,\lambda) - x_t\beta)^2/\sigma^2 + (\lambda-1)\log(y_t/\nu). \tag{2.18}$$

A standard mean value expansion can be used to derive an estimate of the asymptotic variance of $\hat{\theta}$. Redefining s_t to account for the dependence upon η gives $s_t(\theta,\nu) \equiv \nabla_\theta \ell_t(\theta,\nu)$, the 1x(K+2) score of the log-likelihood for observation t. Let $\hat{s}_t \equiv s_t(\theta,\hat{\nu})$ equal the score evaluated at the estimated values $\hat{\theta},\hat{\nu}$. Now define $\hat{g}_t \equiv \hat{s}_t + C_N \cdot \hat{\nu} \cdot \log(y_t/\hat{\nu})$ where $\hat{C}_N \equiv \frac{1}{N} \cdot \sum_{t=1}^{N} \nabla_\nu \hat{s}_t'$. Then, as can be shown using

[7]Alternatively, one might simply want arbitrarily to choose one observation as "numeraire". While such a procedure might be "natural" in a time series context (say, take the first observation), in a cross-section context the choice of numeraire observation would seem to be totally arbitrary, yielding somewhat capricious t-statistics.

methods similar to those employed in the appendix to Wooldridge (1990), a consistent estimator of the asymptotic variance of $\hat{\theta}$ is

$$\left(\sum_{t=1}^{N}\hat{s}_t'\hat{s}_t\right)^{-1}\left(\sum_{t=1}^{N}\hat{g}_t'\hat{g}_t\right)\left(\sum_{t=1}^{N}\hat{s}_t'\hat{s}_t\right)^{-1}. \tag{2.19}$$

Note that if the estimation of $\hat{\nu}$ were ignored, this reduces to the usual outer product of the score estimator. Surprisingly, the standard error of λ, $se(\hat{\lambda})$, obtained from (2.19), differs from that obtained from (2.13), even though $se(\hat{\lambda})$ is scale invariant for any <u>fixed</u> scaling of y.

2.2 The Nonlinear Least Squares Approach

Let y and $x = (1,x_2,...,x_k)$ be defined as in the previous subsection. Without any assumptions on the conditional distribution of y given x (except that its support is contained in $[0,\infty)$), consider the following model for $E(y|x)$:

$$E(y|x)=[1+\lambda x\beta]^{1/\lambda},\lambda\neq 0 \tag{2.20}$$

$$=\exp(x\beta),\lambda=0. \tag{2.21}$$

When $\lambda = 1$ (2.20) reduces to a linear model for $E(y|x)$.[8] The exponential regression model (2.21) is particularly appealing for a strictly positive y because it ensures that the predicted values are well-defined and positive for all x and any value of β, whereas this is not necessarily the case for (2.20). Note that the semi-elasticity for this model is

---

[8]Interestingly, the conditional mean functions (2.20) and (2.21) can be derived from a modified version of the Box-Cox model if $P(y > 0) = 1$. If the conditional mean assumption (2.3) is supplemented with the assumption that $\log(y)|x$ is normally distributed with constant variance, then (2.20) and (2.21) can be shown to hold; see Wooldridge (1990) for details.

$$\frac{\partial E(y|x)}{\partial x_j}\cdot\frac{1}{E(y|x)}=[1+\lambda x\beta]^{-1}\beta_j, \tag{2.22}$$

while the elasticity is

$$\frac{\partial E(y|x)}{\partial x_j}\cdot\frac{x_j}{E(y|x)}=[1+\lambda x\beta]^{-1}\beta_j x_j, \tag{2.23}$$

To estimate β and λ by nonlinear least squares (NLS) or weighted NLS (WNLS), the derivatives of the regression function are needed. Define the (K+1)x1 parameter vector $\delta = (\beta,\lambda)'$ and express the parameterized regression function for E(y | x) as

$$\begin{aligned} m(x;\delta) &= [1+\lambda x\beta]^{1/\lambda}, \qquad \lambda \neq 0 \\ &= \exp(x\beta), \qquad \lambda = 0. \end{aligned} \tag{2.24}$$

For $\lambda \neq 0$ the gradient of m(x;δ) with respect to β is the 1xK vector

$$\nabla_\beta m(x;\delta)=[1+\lambda x\beta]^{(1/\lambda)-1}x. \tag{2.25}$$

For $\lambda = 0$,

$$\nabla_\beta m(x;\beta,0)= \exp(x\beta)x. \tag{2.26}$$

The derivative of m(x;δ) with respect to λ, when $\lambda \neq 0$, is derived in Wooldridge (1990) as

$$\nabla_\lambda m(x;\beta,\lambda)=\frac{1}{\lambda^2}[1+\lambda x\beta]^{\frac{(1-\lambda)}{\lambda}}[\lambda x\beta-(1+\lambda x\beta)\log(1+\lambda x\beta)]. \tag{2.27}$$

For $\lambda = 0$ it equals

$$\nabla_\lambda m(x;\beta,0)= -\frac{\exp(x\beta)(x\beta)^2}{2} \tag{2.28}$$

Equation (2.27) is the basis for the LM statistic for the hypothesis H_0: $\lambda = \lambda_0$, while (2.28) is the basis for the LM test of H_0: $\lambda = 0$; see Wooldridge (1990) for further details.

Under the assumption that (2.24) holds, δ can be consistently estimated by NLS. In addition, if V(y | x) is constant and equal to, say η^2, then standard formulae are available for estimating the asymptotic variance of the NLS estimator. Let $\hat{\delta}$ be the NLS estimator, let $\hat{e}_t \equiv y_t - m(x_t, \hat{\delta})$ denote the NLS residuals, and estimate η^2 by the degrees-of-freedom adjusted estimator

$$\hat{\eta}^2 \equiv \frac{1}{N-P}\sum_{t=1}^{N} \hat{e}_t^2, \tag{2.29}$$

where P = K+1 is the number of parameters. A standard estimate of the asymptotic variance of $\hat{\delta}$ is

$$\hat{\eta}^2\left(\sum_{t=1}^{N} \nabla_\delta \hat{m}_t' \nabla_\delta \hat{m}_t\right)^{-1} \tag{2.30}$$

which is valid provided that $V(y_t | x_t) = \eta^2$ and, in a time series context, $E(y_t | x_t) = E(y_t | x_t, y_{t-1}, x_{t-1}, \ldots)$. (This latter condition ensures that the errors $e_t = y_t - E(y_t | x_t)$ are conditionally serially uncorrelated.)

The heteroskedasticity-robust asymptotic variance estimator of β and λ can be obtained by using the approach of White (1980). That estimator in this context is

$$\frac{N}{(N-P)}\left(\sum_{t=1}^{N} \nabla_\delta \hat{m}_t' \nabla_\delta \hat{m}_t\right)^{-1}\left(\sum_{t=1}^{N} \hat{e}_t^2 \nabla_\delta \hat{m}_t' \nabla_\delta \hat{m}_t\right)\left(\sum_{t=1}^{N} \nabla_\delta \hat{m}_t' \nabla_\delta \hat{m}_t\right)^{-1}, \tag{2.31}$$

which uses a degrees-of-freedom adjustment to enhance finite sample performance.

Although the NLS estimator is robust to heteroskedasticity and requires no distributional assumption, the estimate in (2.31) might be large if $V(y_t | x_t)$ is highly variable. Improvements in efficiency might be realized by using a weighted NLS approach. Let $\hat{\omega}_t \equiv \omega(x_t, \hat{\gamma})$ be a set of weights that can depend on x_t and a vector of estimates $\hat{\gamma}$. Then, if

$$V(y_t|x_t) = \eta^2 \omega(x_t, \gamma) \qquad (2.32)$$

and $\hat{\gamma}$ is $\sqrt{N}$-consistent for γ, η^2 is estimated as in (2.29) where $\hat{e}_t$ is now weighted by $1/\sqrt{\hat{\omega}_t}$. An estimator of the asymptotic variance of the WNLS estimator $\hat{\delta}$ is given by (2.30), except that $\hat{e}_t$ and $\nabla_\delta \hat{m}_t$ are weighted by $1/\sqrt{\hat{\omega}_t}$. In the empirical work, we use the NLS estimator and a WNLS estimator with

$$\omega(x, v) = [m(x, \delta)]^2; \qquad (2.33)$$

the estimated $\hat{\omega}_t$ weights are obtained as the squared fitted values from the initial NLS regression.

The Lagrange multiplier tests for the linear ($\lambda = 1$) and exponential($\lambda = 0$) models are simple to compute. In the general WNLS case, the test for H_0: $\lambda = 1$ is obtained as NR_u^2 from the regression

$$\frac{\hat{e}_t}{\sqrt{\hat{\omega}_t}} \quad on \quad \frac{x_t}{\sqrt{\hat{\omega}_t}}, \frac{\hat{y}_t \log(\hat{y}_t)}{\sqrt{\hat{\omega}_t}} \quad t = 1, \ldots, N, \qquad (2.34)$$

where $\hat{y}_t$ and $\hat{e}_t$ are the fitted values and residuals from the WNLS estimation. LM is distributed as χ_1^2 under H_0 if the variance assumption (2.33) holds. To obtain a robust form of the statistic, first compute the residuals $\tilde{r}_t$ from the regression

$$\frac{\hat{y}_t}{\sqrt{\hat{\omega}_t}} \log(\hat{y}_t) \quad on \quad \frac{x_t}{\sqrt{\hat{\omega}_t}}, \qquad (2.35)$$

and then form LM = N - SSR from the regression

$$1 \; on \; \tilde{e}_t \tilde{r}_t, t=1,\ldots,N, \tag{2.36}$$

where $\tilde{e}_t \equiv \hat{e}_t / \sqrt{\hat{\omega}_t}$. Testing H_0: $\lambda = 0$ requires WNLS estimation of an exponential regression model. Let $\hat{y}_t$ and $\hat{e}_t$ be the fitted values, and let $\tilde{y}_t$ and $\tilde{e}_t$ be the weighted quantities. Then, from Wooldridge (1990), compute $LM = NR_u^2$ from the OLS regression

$$\tilde{e}_t \; on \; \tilde{y}_t \cdot x_t, \; \tilde{y}_t \cdot [\log(\hat{y}_t)]^2; \tag{2.37}$$

again, $LM \overset{a}{\sim} \chi_1^2$ under H_0 and (2.33). The robust test is based on the residuals, $\tilde{r}_t$, from the OLS regression

$$\tilde{y}_t [\log(\hat{y}_t)]^2 \; on \; \tilde{y}_t \cdot x_t, \tag{2.38}$$

and the LM test statistic is then computed as LM = N - SSR, as in (2.36).

One problem that the NLS and WNLS approaches share with transformation methods is that the t-statistics of $\hat{\beta}_1,\ldots,\hat{\beta}_K$ lack invariance to the scaling of y whenever λ is estimated along with the β's. Wooldridge (1990) shows that the estimate and standard error for λ (both the usual and robust form) are invariant to rescaling of y, and moreover, that the LM statistic for testing any exclusion restrictions on the β_k slope coefficients is also invariant to the scaling of y; hence LM tests can be used as alternatives to t-statistics. Unfortunately, if there are many x's, this can be computationally expensive.

As one method of obtaining scale invariant t-statistics, consider adding a scale parameter to (2.24):

$$m(x_t,\beta,\lambda,\nu) = \nu[1+\lambda x_t\beta]^{\frac{1}{\lambda}}, \quad \lambda \neq 0,$$
$$m(x_t,\beta,\lambda,\nu) = \nu\exp(x_t\beta), \quad \lambda = 0, \tag{2.39}$$
$$\nu \equiv \exp[E(\log(y_t))];$$

again, this definition of η is arbitrary, but it is estimable from the data. Let $\hat{\nu}$ be the sample geometric mean of $\{y_t\}$. Then substituting $\hat{\nu}$ into (2.39) and using NLS or WNLS is the same as dividing each y_t by $\hat{\nu}$ and estimating the model as before. As in the Box-Cox case, the variation in $\hat{\nu}$ must be taken into account. Wooldridge (1990) has derived a consistent estimator of the asymptotic variance of $\hat{\delta}$ where $\hat{\delta}$ now refers to the set of parameters in the (sample geometric mean scaled) model. Let

$$\tilde{C}_N \equiv \frac{1}{N}\sum_{t=1}^{N}\left(\frac{\nabla_\delta \hat{m}_t}{\sqrt{\hat{\omega}_t}}\right)'\left(\frac{\nabla_\nu \hat{m}_t}{\sqrt{\hat{\omega}_t}}\right), \tag{2.40}$$

where $\nabla_\delta \hat{m}_t$ is the same as before except that it is not multiplied by $\hat{\nu}$; also, note that $\nabla_\nu \hat{m}_t = [1+\hat{\lambda}x\hat{\beta}]^{1/\hat{\lambda}}$ is simply the fitted value for the <u>scaled</u> regressand $y_t/\hat{\nu}$. A consistent estimate of the asymptotic variance of $\hat{\theta}$ is

$$\left(\sum_{t=1}^{N}\nabla_\delta\tilde{m}_t'\nabla_\delta\tilde{m}_t\right)^{-1}[I_P|-\tilde{C}_N]\left(\sum_{t=1}^{N}\tilde{g}_t'\tilde{g}_t\right)[I_P|-\tilde{C}_N]'\left(\sum_{t=1}^{N}\nabla_\delta\tilde{m}_t'\nabla_\delta\tilde{m}_t\right)^{-1} \tag{2.41}$$

where $P \equiv K+1$, $\tilde{g}_t$ is the 1x(P+1) vector

$$\tilde{g}_t \equiv \left(\tilde{e}_t\nabla_\delta\tilde{m}_t, \ \hat{\nu}\log\left(\frac{y_t}{\hat{\nu}}\right)\right) \tag{2.42}$$

and all "~" variables are weighted by $1/\sqrt{\hat{\omega}_t}$. This expression is robust to variance misspecification and also accounts for the randomness of $\hat{\nu}$.

2.3 Goodness-of-Fit Measures

From our perspective, the ultimate goal of any exercise to generalize functional form is to obtain reliable estimates of E(y | x). Therefore, it is important to have goodness of fit measures that allow discrimination among alternative methods. One natural measure is simply an R-squared defined in terms of the untransformed variable y. Given fitted values $\hat{y}_t$, t=1,...,N, the R-squared is simply

$$R^2 \equiv 1 - \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \bar{y})^2}. \tag{2.43}$$

This R-squared measures the percentage of the variation in y "explained by" the x's, regardless of how the fitted values $\hat{y}_t$ are obtained. In the context of NLS or WNLS, $\hat{y}_t$ is obtained directly as $\hat{y}_t \equiv m(x_t,\hat{\delta})$. The fitted values from a transformation model can be obtained only once an expression for E(y | x) is available. In the Box-Cox context, this expression is given by (2.7)-(2.9).

The R-squared defined by (2.43) is not without its problems. First, for a given functional form for E(y | x), R^2 is always maximized by the NLS estimator. Consequently, (2.43) cannot be used to choose between weighted and unweighted least squares estimators. This is less of a problem that it might seem because, provided E(y | x) is correctly specified, the fitted values from these procedures should be similar.

In this paper, the primary use of R^2 is to compare the direct procedures, NLS and WNLS, to transformation methods (specifically Box-Cox). Because the models for E(y | x) are nonnested for these two approaches, R^2 can legitimately be used as a goodness-of-fit measure. However, since transformation methods do not directly minimize the sum of squared residuals in y_t, R^2 criteria will tend to favor NLS.

An alternative to a standard R-squared measure would be a metric of how well the models estimated the conditional variance as well as the conditional mean. One possibility is simply to evaluate a normal log-likelihood function at the implied conditional mean and conditional variance. It is well-known that the true conditional mean and the true conditional variance maximize the expected log-likelihood whether or not y conditional on x is normally distributed. We do not pursue this approach for two reasons. First, estimation of $V(y|x)$ in the Box-Cox context is computationally cumbersome (it is more difficult than estimating $E(y|x)$). Second, a primary motivation for using NLS and WNLS is that the estimates are robust to heteroskedasticity in the former case and variance misspecification in the latter case. A goodness-of-fit measure based on the conditional mean and conditional variance essentially ignores the robustness considerations.

In addition to aggregate R^2 statistics which summarize the sample information in a single statistic, one might be interested in how well individual order statistics of the various empirical distributions of the fitted values match with those of the dependent variable. While this does not provide a metric for comparison across samples, it should provide clues to possible anomolies in the estimation procedures.

In a similar vein we can calculate the correlation matrices of the various fitted values and the dependent variables. This gives a direct measure of how well the various procedures match the dependent variable and also what similarities exist among the estimation procedures.

In the empirical section of the paper we shall use several of these variations to assess goodness-of-fit for each of the sample data sets.

3. Data and Empirical Implementation

For the BC procedure, we compute standard error estimates employing the variance-covariance matrix of analytic first derivatives as outlined by Berndt, Hall, Hall and Hausman (1974), and the estimate corrected for the sample geometric mean (see (2.19)). For NLS and WNLS, we compute three sets of standard errors using: first, the Gaussian quadratic form of analytic first derivatives; second, a heteroskedasticity-robust estimator due to White (1980); and third, a heteroskedasticity-robust estimator which also accounts for the estimated geometric mean of y (equation (2.31)).

The data sets used for comparing the alternative estimators have been chosen to generate additional interest in classic findings and to facilitate replication, and are all taken from the data diskette accompanying Berndt (1990). Specifically, we employ three data sets. The first, called CPS78, is a random sample of 550 observations drawn from the May 1978 U.S. Current Population Survey, originally constructed by Henry S. Farber. This type of data set is frequently employed by labor economists to estimate statistical earnings functions, where the dependent variable is some transformation (often logarithmic) of the hourly wage rate in dollars (WAGE), and the set of regressors includes a constant term, potential experience (EXP) (measured as age minus years of education minus schooling minus six), and its square (EXP2), a race dummy variable (RACE) taking on the value one only if the individual is non-white and non-Hispanic, a gender dummy variable (FEMALE) taking on the value one only if the individual is female, and an education variable (EDN) measuring the years of schooling.

A second data set is that underlying the classic study of quality-adjusted mainframe computer prices and the demand for computers by Gregory Chow (1967). Chow related the monthly rental price of mainframe computers (PRICE) to multiplication speed (MULT), memory capacity (MEM), access time (ACCESS), and a set of annual dummy variables; we employ Chow's 1961-1965 data, whose details are discussed further in Berndt (1990,ch.4).

The third and final data set we employ is that underlying the more recent mainframe computer hard disk drive price index study by Rosanne Cole et al. (1986). This study is of interest since it has played a critical role in the U.S. Bureau of Economic Analysis decision to employ hedonic regression methods to adjust mainframe computer prices for quality change over time in its official computer price index. The Cole et al. data set encompasses 91 models over the 1972-84 time period, and relates the list price of hard disk drives (PRICE) to a constant, 1973-84 annual time dummy variables, and two performance variables, SPEED and CAP (capacity); further details on these data are given in Cole et al. (1986), Triplett (1989) and Berndt (1990,Ch. 4).

Computations for this empirical research were carried out on an IBM 4381 mainframe and an AT&T 6386 personal computer, using the statistical programs TSP and GAUSS.

4. Empirical Results

Our discussion of empirical results focuses on three issues: (a) we begin by addressing the lack of scale invariance issue, assessing its numerical significance on estimated slope coefficients, and then we implement scale invariant t-statistics using the LM test procedure for both the BC and NLS estimators; (b) we then go on to discuss similarities and differences among the BC, NLS and WNLS parameter estimates in our three data sets, as well as the estimated standard errors; and finally (c) we compare the alternative estimation procedures using a variety of goodness-of-fit criteria.

4.1 Resolving the Lack of Invariance to Scaling Issue

We begin by demonstrating empirically, in a rather persuasive manner, the lack of invariance of t-statistics on slope coefficients to arbitrary re-scaling of the dependent variable in the BC and NLS estimation procedures, and then we present LM-based scale invariant t-statistics. Using the CPS78 data we estimated a model in which WAGE was measured in dollars per hour, and then multiplied this by 100, resulting in a measure in units of cents per hour. Results from the various estimations are presented in Table 1.

With both the BC and NLS procedures, as expected, estimates of λ and its standard error are invariant to scaling; the BC parameter estimate (standard error) is 0.072 (0.037), while that for NLS is -0.275(0.300). Note that $\lambda=0$ cannot be rejected at usual significance levels, thereby lending support to the common procedure in labor economics of employing log (WAGE) as the dependent variable in statistical earnings functions.

Matters are rather different, however, when we examine estimated slope coefficients and their associated t-statistics. For the BC estimator, although the RACE coefficient and t-statistic are relatively robust under scaling, other coefficients and t-statistic vary considerably, with some t-statistics changing by a factor of more than two. For the NLS estimates, this lack of robustness also is present; the coefficients and t-statistics on the EDN and EXP variables, for example, change by a factor greater than three after arbitrary re-scaling. Notice that no sign changes occur under re-scaling for either the BC or NLS estimators. Finally, for each coefficient in the line labeled "LM t-Stat", we present NLS scale-invariant t-statistics based on the LM test statistic, computed as the (positive) square root of the χ^2_1 test statistic based on (2.35) with weights all equal to one. We conclude that while the issue of scaling is empirically significant for these models in particular, and for nonlinear models in general, scale-invariant t-statistics can be obtained using the LM test procedure.

4.2 Comparison of Parameters and Standard Errors for BC, NLS and WNLS

Having disposed of the scaling issue for t-statistics, we now compare parameter estimates and standard errors for the BC, NLS and WNLS estimators, where in each case we follow Spitzer's (1984) suggestion and transform the dependent variable by the sample geometric mean, as discussed underneath (2.36). We begin with a comparison based on the CPS78 data discussed briefly in the previous sub-section.

As seen in Table 2, although estimates of λ vary in sign based on the BC, NLS and WNLS methods, standard errors are relatively large (especially when corrected for both heteroskedasticity and the random geometric sample mean), and it is not clear these estimates differ significantly. Coefficient estimates on the EDN, RACE, FEMALE ,EXP and EXP2 variables are also very similar across the BC, NLS and WNLS estimation procedures with this data set, but both usual, robust and corrected (for the geometric mean) NLS standard errors are larger than those for the BC and WNLS. The NLS t-statistics are not always largest, however, owing to variation in parameter estimates among estimations; scale-invariant LM t-statistics bear no systematic inequality relationship to the various values based on traditional computations.

The similarity in parameter estimates and inference obtained using the BC, NLS and WNLS procedures does not occur, however, for the COLE data set. As seen in Table 3, estimates of λ based on NLS (2.60) and WNLS (4.46) differ dramatically from that based on BC (0.87). The coefficients for SPEED also differ widely across models with (13.26) for BC, (10.27) for NLS and (-0.85) for WNLS. Only the WNLS estimate would be considered insignificantly different from zero. The estimates fo CAP are all positive and small. With respect to the annual time dummy variable coefficients estimates (used as a basis for forming quality-adjusted computer price indexes in the hedonic price literature), sign differences among BC, NLS and WNLS occur for four of the twelve coefficients--1973, 1974, 1977 and 1978. Interestingly, towards the end of the sample in

1982-84, LM t-statistics on the WNLS time dummies become smaller (in absolute value) than those based on the other procedures.[9]

Finally, another interesting finding in Table 3 concerns alternative estimates of the standard errors. For the BC procedure, traditional and "correct" standard error estimates are quite similar, with no discernible inequality relationship between them occurring. For the NLS and WNLS procedures, we see that correcting for the sample geometric mean has a somewhat larger effect than in the BC case, but most of this is due to the heteroskedasticity adjustment implicit in the "correct" standard errors.

The final data set we use in comparing the BC, NLS and WNLS procedures is that underlying the classic study by Gregory Chow on estimating the prices and price elasticity of demand for computers. Our results from the CHOW data set are presented in Table 4. A number of results are worth noting.

First, estimates of λ vary dramatically across estimation procedures. While the 0.129 BC estimate of λ is positive, small and statistically insignificant, the NLS (-2.03) and especially the WNLS (-3.05) estimates are negative, large and statistically significant. A negative estimate of λ in the computer market is not entirely unexpected (see Jack E. Triplett (1989) for a conjecture that this might occur), and demonstrates the importance of allowing λ to vary outside the (0,1) interval when implementing such models empirically. For the estimated slope coefficients on the MULT, MEM and ACC variables, no sign differences occur among the three estimation procedures, but one sign variation is found among the annual time dummy variables--that for 1962.

In terms of standard error estimation procedures, for BC the traditional and "correct" method yield roughly similar results, with the "correct" estimates being slightly larger than the traditional in all cases but one (the 1961 time dummy). Interestingly, for NLS the

[9]Implications of these results for price index computation are discussed in detail by Berndt, Showalter and Wooldridge (1990).

corrections (both heteroskedasticity and sample geometric mean) actually lower the standard errors for all the parameters. For WNLS we have mixed results. Generally the robust standard errors are smaller than the traditional ones while the "correct" standard errors are, with one exception, the highest of the three estimates (the exception being the MEM coefficient where the "correct" standard error is higher than the traditional but lower than the robust).

On the basis of parameter estimates and inference, therefore, we conclude that substantial differences are found among the BC, NLS and WNLS estimates, particularly witn the COLE and CHOW data sets. We see no systematic effect of adjusting for the sample geometric mean, although there is some evidence that using White's robust standard errors might be an adequate approximation for the "correct" standard errors.

We now move on to a comparison of estimation methods using goodness-of-fit criteria. Recall from our earlier discussion that if one defines the residual as $\hat{e}_t \equiv y_t - f(x_t,\hat{\beta},\hat{\lambda})$, then by construction the NLS estimator will always produce a lower sum of squared residuals than the WNLS method. However, we cannot say that NLS will result in a lower sum of squared residuals than BC when the residuals are calculated using equations (2.7)-(2.9) (although we suspect that this will generally be the case). As a result, it is important to use criteria other than sums of squared residuals when comparing the BC, NLS and WNLS procedures.

In Table 5 we present summary statistics (min, max, 25%, 75%, std. dev. and mean) for the observed (scaled by the sample geometric mean) dependent variable, y, and fitted values, where the latter are computed in four ways, as discussed in Section 2: IBC (Incorrect Box-Cox, Box-Cox using equation (2.5)), BC (Box-Cox using equation (2.7)-(2.9)), NLS (nonlinear least squares fitted value) and WNLS (weighted nonlinear least squares fitted values), all for the CPS78 data. Corresponding summary statistics for the COLE and CHOW data sets are given in Tables 6 and 7.

As seen in Table 5, for the CPS78 data the distribution of the fitted values is roughly similar for all four procedures, although the mean of the fitted value for IBC is about 10% less than the sample mean of y, implying that for the IBC, the mean "residual" is non-zero. For the COLE data (see Table 6), while mean fitted values are all approximately equal and close to the sample observed mean, the IBC and BC minimum fitted values (0.573 and 0.583) are substantially larger than those for NLS (0.296) and WNLS (0.312), and for the sample observed min (0.294). However, maximum fitted values of the IBC (2.755) and BC (2.757) are close to the observed sample max (2.766), but these maximum values are larger than the maximum fitted values based on the NLS (2.439) and WNLS (2.128) procedures.

Finally, for the CHOW data (see Table 7), greater diversity appears. The mean fitted values for IBC (2.170), NLS (1.999) and WNLS (2.029) are quite close to the sample mean of the observed y (2.034), but the mean fitted value from BC (2.574) is about 20% larger. Although the min (0.086) and max (21.221) fitted values from WNLS are virtually identical to those observed (0.090 and 21.218), the min fitted values for NLS (0.021) and IBC (0.055) are smaller, and the max fitted values for IBC (74.892) and BC (81.636) are much larger than for observed y (21.218). Since the max values for IBC and BC are so much larger than for the observed y, several IBC and BC residuals will be correspondingly large, one might expect that the sum of squared residuals will be correspondingly large, and therefore that the sum of squared residuals based on the BC and IBC methods and the CHOW data will be much larger than for the NLS and WNLS methods. This is in fact what occurs; as seen in Table 7, with the CHOW data the sums of squared residuals for IBC (3089.09) and BC (3829.93) are much larger than for WNLS (91.47) or NLS (87.08). For the CPS78 data (Table 5), differences in the sums of squared residuals are very small, while for the COLE data (Table 6), the differences are only slightly larger.

In the last panel of Tables 5,6 and 7, we present simple correlations between the observed and four sets of fitted values (IBC, BC, NLS and WNLS), both centered about their sample means and uncentered. For the CPS78 data (Table 5), the uncentered correlations are all very large (above 0.9), and for the centered correlations with observed y, the fitted value correlations are all very similar (about 0.59). For the COLE data (Table 6), simple correlations display a bit more diversity, but differences are not dramatic. With the CHOW data, however (Table 7), two groups of correlations differ. While the IBC and BC fitted values are very highly correlated (the centered and uncentered correlation are each 1.000) with each other, and while the NLS and WNLS reveal similarly high correlations (0.997 centered, 0.998 uncentered), simple correlations between the IBC, BC and NLS-WNLS fitted values are lower, around 0.8 for both the centered and uncentered data.

These correlations among fitted values and between fitted and actual values of y imply correlation structures among residuals. Simple (uncentered) residual correlations for the CPS78, COLE and CHOW data sets are presented in the bottom panel of Tables 5, 6 and 7, respectively. As seen in Table 5, the inter-correlations among the IBC, BC, NLS and WNLS residual for the CPS78 data set are all very high--greater than 0.99. For the COLE data set (Table 6), we find that the WNLS residuals have a relatively low correlation with the BC-IBC residuals (0.641), with the remaining correlations 0.86 or above.set (Table 7), however, three clusters of correlations become evident. While correlations between IBC and BC residuals (0.999) and between NLS and WNLS residuals (0.977) remain very high, simple correlations between one of IBC-BC and one of NLS-WNLS are very low--between about 0.11 and 0.13. With the CHOW data, therefore, two very distinct groups of residuals emerge--one set based on Box-Cox variants, and the other on nonlinear least squares variants. For this data set in particular, the transformation and nonlinear least squares methods yield very different results.

5. Concluding Remarks

Our purpose in this paper has been to compare empirically two distinct approaches to choosing a functional form--the Box-Cox and nonlinear least squares procedures--based on three publicly available data sets.

We can summarize our findings as follows. First, we provided a rather persuasive empirical example demonstrating that with both the Box-Cox (BC) and nonlinear least squares (NLS) procedures, while t-statistics on the transformation parameter λ are invariant to arbitrary scaling of the dependent variable, the t-statistics on slope coefficients, intercepts and dummy variables can be changed dramatically simply by arbitrarily re-scaling the data. We also noted that since the t-statistic is a Wald test statistic, this lack of invariance is not surprising, and we eliminated it by employing a computationally more cumbersome Lagrange multiplier test statistic, systematically excluding one variable at a time and re-estimating. We conclude, therefore, that while in practice in these nonlinear models scaling issues are very important, they can be resolved though use of the LM test statistic procedure. Future research that focuses on necessary and sufficient conditions for such lack of scaling invariance, as well as on more computationally efficient ways of doing scale-invariant testing of exclusionary restrictions, would appear to be most useful.

Second, we have found that differences among the BC, NLS and weighted nonlinear least squares (WNLS) parameter estimates vary by data set, and that little in general can be stated concerning what a researcher should expect with a particular data set. Specifically, in one data set (CPS78) parameter estimates differed very little among alternative estimators, in a second data set (COLE) the differences were substantially larger--sometimes even resulting in different signs for estimated coefficients, and in our third data set (CHOW) the differences were very large, with the estimated transformation parameter λ having a different sign depending on the estimation procedure employed. Since in some cases we find substantial differences among estimators, we are now faced with issues assessing which estimator is "best" in terms of yielding estimates closest to the "true" parameters. Our results therefore imply that empirical assessment of these alternative estimators based on a well-designed Monte-Carlo approach is warranted.[10]

Third, we have computed standard error estimates using traditional, heteroskedasticity-robust and simultaneous heteroskedasticity-robust and sample geometric mean-adjusted computational procedures. Our results suggest that when there are differences among these alternative standard error estimates, most of the difference can be attributed to adjusting for heteroskedasticity; the marginal change induced by adjusting for the random sample geometric mean of the dependent variable is relatively minor.

Fourth, in terms of fitted values and residuals, we have found that in some cases the common but incorrect Box-Cox (IBC) and correct Box-Cox(BC) procedures yield fitted values much greater than (less than) the sample maximum (minimum) values of the observed y, and that in such cases the resulting extremely large residuals for IBC and BC yield very large sums of squared residuals, much larger than that for NLS and WNLS. In these cases, while the correlations between IBC and BC residuals, and between NLS and WNLS residuals, are very high, the IBC-BC and NLS-WNLS residuals tend to cluster in two distinct groups, with simple correlations between any one of IBC-BC and one of NLS-WNLS being very small (less than 0.15). Which of these residuals are more

---

[10]Research on this topic is currently underway. See Showalter (1990).

"correct" depends of course on the true parameters and model, and in this study those are still unknown. Further research on this topic using Monte Carlo approaches would be useful

Finally, in this paper we have reported results using only the Box-Cox transformation on the dependent variable, and have employed "natural" (i.e., untransformed) values for the explanatory variables. In particular, we have not reported results when some explanatory variables are transformed into logarithms (as was done in the original COLE and CHOW studies) or are transformed using the Box-Tidwell procedures. We have done some research on these issues, however, and can briefly report that when one employs logarithmic transformations of explanatory variables as was done in the original studies by Cole et al. and Chow, differences among the BC, NLS and WNLS estimates of λ become rather small, and typically our λ estimates were insignificantly different from zero, thereby lending support to the log-log functional form specification used by Cole et al. and Chow.[11] However, when Box-Tidwell-type procedures are employed, differences among the various estimation procedures re-emerge.

[11]For further discussion, see Berndt, Showalter and Wooldridge (1990).

--- Table 1 ---

Resolution of the Lack of Invariance to Scaling Issue

CPS 1978 Data

	Box-Cox		NLS	
	Unscaled ($ per hour)	Scaled (*100)	Unscaled ($ per hour)	Scaled (*100)
Lambda	0.072	0.072	-0.275	-0.275
Std Err	0.037	0.037	0.300	0.300
t-Stat	1.917	1.917	-0.917	-0.917
LM t-Stat	0.816	0.816	0.563	0.563
Constant	0.540	6.204	0.702	2.807
Std Err	0.120	0.596	0.099	1.685
t-Stat	4.515	10.412	7.087	1.666
LM t-Stat	9.734	9.734	4.844	4.844
FEMALE	-0.371	-0.515	-0.200	-0.056
Std Err	0.046	0.134	0.115	0.109
t-Stat	-8.054	-3.849	-1.741	-0.514
LM t-Stat	9.092	9.092	6.890	6.890
RACE	-0.131	-0.182	-0.094	-0.026
Std Err	0.050	0.075	0.063	0.052
t-Stat	-2.602	-2.420	-1.490	-0.507
LM t-Stat	2.485	2.485	2.236	2.236
EDN	0.080	0.111	0.042	0.012
Std Err	0.010	0.029	0.026	0.024
t-Stat	8.308	3.822	1.614	0.501
LM t-Stat	9.152	9.152	5.596	5.596
EXP	0.034	0.047	0.020	0.006
Std Err	0.007	0.015	0.012	0.011
t-Stat	4.734	3.078	1.690	0.511
LM t-Stat	6.412	6.412	5.695	5.695
EXP2	-4.10E-04	-5.69E-04	-2.35E-04	-6.62E-05
Std Err	1.40E-04	2.40E-04	1.47E-04	1.29E-04
t-Stat	-2.928	-2.377	-1.599	-0.513
LM t-Stat	3.777	3.777	3.173	3.173

Note: The LM t-Stats are based on the LM test statistic (computed as the square root of the chi-squared test statistic), allowing for heteroskedasticity.

--- Table 2 ---

Alternative Estimates of Standard Errors and t-Statistics

CPS 1978 Data

		BC			NLS	
		Std.Err.	t-stat		Std.Err.	t-stat
Lambda	0.07151			-0.27532		
Usual		0.03730	1.917		0.30020	-0.917
Robust					0.45707	-0.602
Correct		0.03730	1.917		0.45707	-0.602
Constant	-1.10529			-1.02175		
Usual		0.10211	-10.825		0.14443	-7.074
Robust					0.11650	-8.771
Correct		0.10633	-10.395		0.11869	-[illegible].60[illegible]
EDN	0.07109			0.06712		
Usual		0.00692	10.269		0.00950	7.062
Robust					0.00805	8.336
Correct		0.00692	10.268		0.00807	8.317
RACE	-0.11591			-0.14922		
Usual		0.04548	-2.548		0.05757	-2.592
Robust					0.05113	-2.919
Correct		0.04549	-2.548		0.05111	-2.920
FEMALE	-0.32857			-0.31723		
Usual		0.03461	-9.492		0.04589	-6.913
Robust					0.04456	-7.119
Correct		0.03463	-9.487		0.04433	-7.156
EXP	0.02980			0.03115		
Usual		0.00574	5.194		0.00583	5.348
Robust					0.00574	5.426
Correct		0.00574	5.192		0.00575	5.419
EXP2	-0.00036			-0.00037		
Usual		0.00012	-3.024		0.00012	-3.204
Robust					0.00013	-2.957
Correct		0.00012	-3.024		0.00013	-2.957

--- Table 2 (Continued) ---

Alternative Estimates of Standard Errors and t-Statistics

CPS 1978 Data

		WNLS	
		Std.Err.	t-stat
Lambda	-0.25248		
Usual		0.30592	-0.825
Robust		0.30563	-0.826
Correct		0.30563	-0.826
Constant	-1.08120		
Usual		0.10850	-9.965
Robust		0.10329	-10.467
Correct		0.10593	-10.206
EDN	0.07051		
Usual		0.00766	9.200
Robust		0.00676	10.425
Correct		0.00679	10.389
RACE	-0.11212		
Usual		0.04684	-2.394
Robust		0.04856	-2.309
Correct		0.04857	-2.308
FEMALE	-0.30226		
Usual		0.03675	-8.225
Robust		0.03637	-8.311
Correct		0.03621	-8.347
EXP	0.03224		
Usual		0.00501	6.430
Robust		0.00427	7.559
Correct		0.00429	7.523
EXP2	-0.00040		
Usual		0.00011	-3.717
Robust		0.00009	-4.215
Correct		0.00009	-4.209

--- Table 3 ---
Alternative Estimates of Standard Errors and t-Statistics
Cole Data

		BC			NLS	
		Std.Err.	t-stat		Std.Err.	t-stat
Lambda	0.86257			2.58990		
Usual		0.21118	4.085		0.38164	6.786
Robust					0.39420	6.570
Correct		0.25534	3.378		0.39420	6.570
Constant	-0.64311			-0.55997		
Usual		0.12193	-5.274		0.07618	-7.351
Robust					0.08288	-6.756
Correct		0.11730	-5.483		0.08418	-6.652
SPEED	13.41629			10.54324		
Usual		2.66228	5.039		3.40578	3.096
Robust					4.03067	2.616
Correct		2.37244	5.655		4.00958	2.630
CAPACITY	0.00014			0.00127		
Usual		0.00017	0.808		0.00046	2.733
Robust					0.00043	2.938
Correct		0.00016	0.847		0.00048	2.667
1973	0.23921			0.07760		
Usual		0.11940	2.003		0.11749	0.661
Robust					0.11883	0.653
Correct		0.11882	2.013		0.11880	0.653
1974	0.03888			-0.07403		
Usual		0.10020	0.388		0.08205	-0.902
Robust					0.08402	-0.881
Correct		0.09982	0.389		0.08520	-0.869
1975	-0.06402			-0.12453		
Usual		0.14118	-0.453		0.06563	-1.897
Robust					0.04985	-2.498
Correct		0.14160	-0.452		0.05051	-2.465
1976	-0.06146			-0.14651		
Usual		0.10056	-0.611		0.05743	-2.551
Robust					0.04446	-3.295
Correct		0.10052	-0.611		0.04540	-3.227

--- **Table 3 (Continued)** ---

Alternative Estimates of Standard Errors and t-Statistics

Cole Data

		BC			NLS	
		Std.Err.	t-stat		Std.Err.	t-stat
1977	0.02961			-0.11924		
Usual		0.11167	0.265		0.09137	-1.305
Robust					0.07297	-1.634
Correct		0.11099	0.267		0.07103	-1.679
1978	0.21434			-0.18883		
Usual		0.19080	1.123		0.27742	-0.681
Robust					0.22777	-0.829
Correct		0.18700	1.146		0.22597	-0.836
1979	-0.03698			-0.60416		
Usual		0.32848	-0.113		0.26236	-2.303
Robust					0.20254	-2.983
Correct		0.32614	-0.113		0.20737	-2.914
1980	-0.25360			-0.82602		
Usual		0.22352	-1.135		0.23597	-3.500
Robust					0.20057	-4.118
Correct		0.22263	-1.139		0.21764	-3.795
1981	-0.25941			-0.82932		
Usual		0.24131	-1.075		0.23729	-3.495
Robust					0.19870	-4.174
Correct		0.23882	-1.086		0.21587	-3.842
1982	-0.23589			-0.85830		
Usual		0.40696	-0.580		0.26206	-3.275
Robust					0.20860	-4.115
Correct		0.40553	-0.582		0.22578	-3.802
1983	-0.52679			-1.02977		
Usual		0.28303	-1.861		0.25643	-4.016
Robust					0.20253	-5.084
Correct		0.28643	-1.839		0.22645	-4.547
1984	-0.62440			-1.06818		
Usual		0.28959	-2.156		0.24510	-4.358
Robust					0.21192	-5.041
Correct		0.29171	-2.140		0.23639	-4.519

--- Table 3 (Continued) ---
Alternative Estimates of Standard Errors and t-Statistics
Cole Data

		WNLS	
		Std.Err.	t-stat
Lambda	2.57957		
Usual		0.50334	5.125
Robust		0.42342	6.092
Correct		0.42342	6.092
Constant	-0.51235		
Usual		0.08453	-6.061
Robust		0.07367	-6.955
Correct		0.07452	-6.875
SPEED	7.49824		
Usual		3.37800	2.220
Robust		3.39850	2.206
Correct		3.30321	2.270
CAPACITY	0.00150		
Usual		0.00054	2.762
Robust		0.00045	3.323
Correct		0.00050	2.966
1973	0.04042		
Usual		0.10073	0.401
Robust		0.09236	0.438
Correct		0.09273	0.436
1974	-0.08976		
Usual		0.05832	-1.539
Robust		0.07314	-1.227
Correct		0.07435	-1.207
1975	-0.12700		
Usual		0.04356	-2.915
Robust		0.04503	-2.820
Correct		0.04618	-2.750
1976	-0.13947		
Usual		0.03813	-3.658
Robust		0.04194	-3.325
Correct		0.04378	-3.186

--- Table 3 (Continued) ---
Alternative Estimates of Standard Errors and t-Statistics
Cole Data

		WNLS	
		Std.Err.	t-stat
1977	-0.06888		
Usual		0.07378	-0.934
Robust		0.05468	-1.260
Correct		0.05245	-1.313
1978	-0.17382		
Usual		0.36390	-0.478
Robust		0.22203	-0.783
Correct		0.22026	-0.789
1979	-0.59740		
Usual		0.30069	-1.987
Robust		0.21982	-2.718
Correct		0.22217	-2.689
1980	-0.80706		
Usual		0.26348	-3.063
Robust		0.21954	-3.676
Correct		0.23102	-3.494
1981	-0.80916		
Usual		0.26508	-3.052
Robust		0.21680	-3.732
Correct		0.22800	-3.549
1982	-0.86386		
Usual		0.29019	-2.977
Robust		0.22379	-3.860
Correct		0.23557	-3.667
1983	-0.99009		
Usual		0.26393	-3.751
Robust		0.20154	-4.913
Correct		0.22070	-4.486
1984	-0.99120		
Usual		0.25817	-3.839
Robust		0.20059	-4.941
Correct		0.21965	-4.513

--- **Table 4** ---

Alternative Estimates of Standard Errors and t-Statistics

Chow Data

		BC			NLS	
		Std.Err.	t-stat		Std.Err.	t-stat
Lambda	0.12913			-2.02749		
Usual		0.14056	0.919		0.43165	-4.697
Robust					0.40761	-4.974
Correct		0.14167	0.911		0.40761	-4.974
Constant	0.04120			0.63441		
Usual		0.24733	0.167		0.19839	3.450
Robust					0.17167	3.987
Correct		0.30742	0.134		0.16964	4.034
MULT	-0.00001			-0.00094		
Usual		9.82E-06	-0.947		0.00052	-1.804
Robust					0.00043	-2.171
Correct		1.00E-05	-0.870		0.00045	-2.084
MEM	0.00042			0.00001		
Usual		7.10E-05	5.894		0.00001	2.106
Robust					0.00001	2.446
Correct		9.00E-05	4.824		0.00001	2.570
ACCESS	-0.00008			-0.04812		
Usual		4.74E-05	-1.694		0.02594	-1.855
Robust					0.02155	-2.233
Correct		5.00E-05	-1.679		0.02150	-2.238
1961	-0.15710			-0.29943		
Usual		0.38903	-0.404		0.14187	-2.111
Robust					0.12095	-2.476
Correct		0.38590	-0.407		0.11645	-2.571
1962	0.09273			-0.15963		
Usual		0.29238	0.317		0.08247	-1.936
Robust					0.06872	-2.323
Correct		0.30276	0.306		0.06821	-2.340
1963	-0.25717			-0.21562		
Usual		0.30548	-0.842		0.10775	-2.001
Robust					0.09052	-2.382
Correct		0.30768	-0.836		0.08889	-2.426
1964	-0.14770			-0.22731		
Usual		0.35770	-0.413		0.11520	-1.973
Robust					0.09655	-2.354
Correct		0.35788	-0.413		0.09526	-2.386
1965	-0.73291			-0.23831		
Usual		0.31279	-2.343		0.12272	-1.942
Robust					0.10301	-2.314
Correct		0.31540	-2.324		0.10206	-2.335

--- Table 4 (Continued) ---
Alternative Estimates of Standard Errors and t-Statistics
Chow Data

		WNLS	
		Std.Err.	t-stat
Lambda	-3.05296		
Usual		0.23361	-13.069
Robust		0.24741	-12.340
Correct		0.24741	-12.340
Constant	0.36960		
Usual		0.02142	17.253
Robust		0.02788	13.256
Correct		0.03127	11.820
MULT	-0.00125		
Usual		0.00044	-2.813
Robust		0.00048	-2.604
Correct		0.00049	-2.535
MEM	0.00001		
Usual		0.00000	2.668
Robust		0.00000	3.071
Correct		0.00000	2.744
ACCESS	-0.00247		
Usual		0.00154	-1.606
Robust		0.00168	-1.472
Correct		0.00209	-1.181
1961	-0.16786		
Usual		0.05856	-2.867
Robust		0.05497	-3.054
Correct		0.06123	-2.742
1962	-0.04633		
Usual		0.01608	-2.881
Robust		0.01452	-3.190
Correct		0.01703	-2.720
1963	-0.09324		
Usual		0.03105	-3.003
Robust		0.02990	-3.118
Correct		0.03403	-2.740
1964	-0.08507		
Usual		0.02779	-3.061
Robust		0.02647	-3.214
Correct		0.03098	-2.746
1965	-0.08294		
Usual		0.02742	-3.025
Robust		0.02653	-3.126
Correct		0.03021	-2.745

Table 5 - CPS 1978 Data

Summary Statistics

	Y	IBC	BC	NLS	WNLS
Min	0.116	0.477	0.516	0.532	0.542
Max	5.314	2.129	2.268	2.546	2.566
25%	0.698	0.816	0.878	0.859	0.857
75%	1.396	1.239	1.327	1.335	1.330
Std Dev	0.607	0.314	0.334	0.362	0.361
Mean	1.129	1.050	1.126	1.129	1.129

Sum of Squared Residuals

IBC	135.641
BC	131.487
NLS	130.351
WNLS	130.402

Correlation Matrices

	--Centered--					--Uncentered--				
Y	1.000					1.000				
IBC	0.592	1.000				0.924	1.000			
BC	0.592	1.000	1.000			0.924	1.000	1.000		
NLS	0.596	0.995	0.995	1.000		0.925	0.999	0.999	1.000	
WNLS	0.595	0.995	0.995	1.000	1.000	0.925	0.999	0.999	1.000	1.000

	--Residuals--			
IBC	1.000			
BC	0.999	1.000		
NLS	0.993	0.996	1.000	
WNLS	0.993	0.996	1.000	1.000

Table 6 - Cole Data

Summary Statistics

	Y	IBC	BC	NLS	WNLS
Min	0.294	0.571	0.581	0.294	0.319
Max	2.766	2.757	2.759	2.439	2.448
25%	0.742	0.760	0.766	0.773	0.791
75%	1.291	1.178	1.182	1.283	1.294
Std Dev	0.565	0.502	0.501	0.511	0.513
Mean	1.132	1.126	1.131	1.132	1.132

Sum of Squared Residuals

IBC	6.659
BC	6.668
NLS	5.175
WNLS	5.214

Correlation Matrices

	--Centered--					--Uncentered--				
Y	1.000					1.000				
IBC	0.877	1.000				0.977	1.000			
BC	0.876	1.000	1.000			0.977	1.000	1.000		
NLS	0.905	0.968	0.967	1.000		0.982	0.995	0.995	1.000	
WNLS	0.905	0.967	0.966	0.999	1.000	0.982	0.994	0.994	1.000	1.000

	--Residuals--			
IBC	1.000			
BC	1.000	1.000		
NLS	0.880	0.880	1.000	
WNLS	0.875	0.875	0.997	1.000

Table 7 - Chow Data

Summary Statistics

	Y	IBC	BC	NLS	WNLS
Min	0.090	0.055	0.091	0.021	0.086
Max	21.218	74.892	81.636	21.218	21.221
25%	0.398	0.810	1.057	0.610	0.853
75%	2.254	1.299	1.646	2.019	1.889
Std Dev	3.003	8.215	8.944	2.844	2.780
Mean	2.034	2.170	2.574	1.999	2.029

Sum of Squared Residuals

IBC --	3089.093
BC --	3829.932
NLS --	87.081
WNLS --	91.472

Correlation Matrices

	--Centered--					--Uncentered--				
Y	1.000					1.000				
IBC	0.778	1.000				0.766	1.000			
BC	0.782	1.000	1.000			0.778	1.000	1.000		
NLS	0.939	0.813	0.816	1.000		0.958	0.790	0.801	1.000	
WNLS	0.935	0.827	0.831	0.997	1.000	0.956	0.797	0.808	0.998	1.000

	--Residuals--			
IBC	1.000			
BC	0.999	1.000		
NLS	0.125	0.107	1.000	
WNLS	0.139	0.117	0.977	1.000

References

Amemiya, T. and J.L. Powell (1981), "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-Linear Two Stage Least Squares Estimator," Journal of Econometrics, 17, 351-381.

Berndt, E.R. (1990), The Practice of Econometrics: Classic and Contemporary, Reading, MA: Addison-Wesley Publishers, Forthcoming.

Berndt, E.R., B.H. Hall, R.E. Hall, and J.A. Hausman (1974), "Estimation and inference in nonlinear structural models," Annals of Economic and Social Measurement, 3, 653-665.

Berndt, E.R., M.H. Showalter and J.M. Wooldridge (1990), "Hedonic Price Indexes for Computers and Peripherals: Box-Cox Formulations and a Nonlinear Least Squares Alternative," Massachusetts Institute of Technology, mimeo, July.

Bollershev, T. and J.M. Wooldridge (1988), "Quasi-maximum Likelihood Estimation of Dynamic Models with Time Varying Covariances," Working Paper 505, Department of Economics, Massachusetts Institute of Technology.

Box, G.E.P. and D.R. Cox (1964), "An Analysis of Transformations," Journal of the Royal Statistical Society, Series B, 26, 211-252.

Chow, G.C. (1967), "Technological Change and the Demand for Computers," American Economic Review, 57, 1117-1130.

Cole, R., Y.C. Chen, J.A. Barquin-Stolleman, E. Dulberger, N. Helvacian and J.H. Hodge (1986), "Quality-Adjusted Price Indexes form Computer Processors and Selected Peripheral Equipment," Survey of Current Business, 66, 41-50.

Dagenais, M.G. and J.-M. Dufour (1986), "Invariance, Nonlinear Models and Asymptotic Tests," Working Paper 3287, Department of Economics, University of Montreal, December.

Draper, N.R. and D.R. Cox (1969), "On Distributions and Their Transformation to Normality,"Journal of the Royal Statistical Society, Series B, 31, 472-476.

Hinkley D.V. and G. Runger (1984), "The Analysis of Transformed Data," Journal of the American Statistical Association, 79, 302-309.

Khazzoom, J.D. (1989), "A Note on the Application of the Nonlinear Two-Stage Least-Squares Estimator to a Box-Cox-Transformed Model," Journal of Econometrics, 42, 377-379.

Neyman, J. (1959), "Optimal Asymptotic Tests of Composite Statistical Hypotheses," in U. Grenander, ed., Probability and Statistics, Almqvist and Wiksell: Stockholm, 213-234.

Poirier, D.J. and A. Melino (1978), "A Note on the Interpretation of Regression Coefficients Within a Class of Truncated Distributions," Econometrica, 46, 1207-1209.

Seakes, T.G. and S.K. Layson (1983), "Box-Cox Estimation with Standard Econometric Problems," Review of Economics and Statistics, 65, 160-164.

Showalter, M.H. (1990), "A Monte Carlo Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative", MIT mimeo.

Spitzer, J.J. (1982), "A Primer on Box-Cox Estimation," Review of Economics and Statistics, 64, 307-313.

Spitzer, J.J. (1984), "Variance Estimates in Models with Box-Cox Transformations: Implications for Estimation and Hypothesis Testing," Review of Economics and Statistics, 66, 645-652.

Triplett J.E. (1989), "Price and Technological Change in a Capital Good: A Survey of Research on Computers," Chapter 4 in Dale W. Jorgenson and Ralph Landau, eds., Technology and Capital Formation, Cambridge, MA: MIT Press, 127-213.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, 48, 817-838.

Wooldridge, J.M. (1990), "Some Alternatives to the Box-Cox Regression Model," mimeo, MIT Department of Economics, January.

Zarembka, P. (1968), "Functional Form in the Demand for Money," Journal of the American Statistical Association, 63, 502-511.

Essay II:

A Monte Carlo Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative

by

Mark H. Showalter

Abstract:

This paper is a Monte Carlo study of the Box-Cox and the nonlinear least squares (NLS) estimators. This is the first Monte Carlo study of the NLS estimator and it also extends previous Monte Carlo work for the BC estimator by examining the effects of autocorrelation and heteroskedasticity. The study is constructed to allow comparison between the Box-Cox and the NLS estimators.

The results suggest: 1) The transformation parameter in the Box-Cox model appears to be inconsistently estimated in the presence of conditional heteroskedasticity. 2) The constant term in both the Box-Cox and the NLS models is poorly estimated in small samples. 3) The ratio of the variance of the error term to the total variance of the dependent variable overshadows the effects of misspecification (autocorrelation or heteroskedasticity). 4) The effect of sample size is more evident in the NLS model than in the Box-Cox counterpart. 5) Asymptotic size is generally overestimated in the NLS model when Gaussian variance estimators are used. 6) The true value of the transformation parameter appears to be important in determining the fit of the data in the Box-Cox model while it is unimportant in the NLS model.

1. Introduction

Since its introduction, the Box-Cox model has been widely used as a framework for generalizing functional form in econometric research. This has occurred despite the fact that the distributional assumptions underlying Box-Cox (BC) cannot be true except in the special case of the transformation parameter, usually denoted λ, equalling zero. Additionally, the BC procedure estimates a transformation of the dependent variable when, in fact, researchers are most frequently interested in the original untransformed variable. This use of a (not necessarily consistent) estimate of the transformed dependent variable poses difficulties in correctly determining the marginal effects of changes in the independent variables on the untransformed dependent variable, and also in computing fitted values for the untransformed dependent variable.

Nonetheless, the importance of the BC transformation for distinguishing among alternative specifications has transcended its theoretical drawbacks. The BC transformation provides some statistical rigor to the otherwise ad hoc choice between a model which uses an untransformed dependent variable and a model which measures the dependent variable in logs. It additionally allows for choices other than these two common ones. The Box-Cox procedure also has the convenient property that parameter estimates can be (laboriously) obtained using ordinary least squares techniques.

This tension between the desire for a statistical methodology to determine the most appropriate functional form and the obvious theoretical flaws of the BC model provides fertile ground for a Monte Carlo study of the properties of the BC procedure. Such a study would provide researchers with a better idea of the limitations involved when using the BC model. Since the estimation of a BC model involves either maximum likelihood estimation or iterated OLS, Monte Carlo simulation has been rather limited. Spitzer (1978) is probably the best of the work done thus far and he acknowledges the limitations of his study, due in large part to the cost of doing nonlinear simulations.

However, as computing costs have gone down, use of the BC procedure appears to be increasing in the empirical literature. This would therefore seem an appropriate time for a more detailed Monte Carlo study of the BC procedure.

A nonlinear least squares (NLS) alternative to the BC model has recently been proposed by Jeff Wooldridge (1990) which, like the BC model, allows the data to determine the most appropriate functional form--be it linear, logarithmic or some other form--but avoids the theoretical problems of the BC model in that no comprehensive distributional assumptions are required. The model is specified such that fitted values of the dependent variable are easy to calculate and the marginal effects of the explanatory variables are straightforward to compute, requiring only differentiation of the estimated conditional mean function. This flexibility comes at a cost, however. Unlike the BC model, consistent parameter estimates generally cannot be obtained from an OLS regression. Also, one might wonder if perhaps BC might be more efficient in the sense of smaller standard errors if in fact the normality assumption of the BC model were close to the true distribution.

This paper investigates both the BC model and the NLS alternative using simple Monte Carlo methods. The organization of the paper is as follows: First we present a brief review of the BC model. Then we outline more extensively the NLS alternative proposed by Wooldridge and provide details on the simulation methodology. We then present the simulation results and analysis, and conclude with a brief summary.

2. The Box-Cox Model

The BC model typically assumes the following form:

$$\frac{y_t^\lambda - 1}{\lambda} = x_t\beta + \epsilon_t$$

x_t - *1xk vector of explanatory variables; first element equal to 1*
β - *kx1 vector of coefficients* (1)
λ - *scalar transformation parameter*
$y_t > 0$
ϵ_t IID $N(0,\sigma^2)$
ϵ_t, x_t *independent*

By performing a change of variables from ϵ_t to y_t we find the log likelihood function of the sample of y_t's to be

$$Logl = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\frac{y_t^\lambda - 1}{\lambda} - x_t\beta\right)^2 + (\lambda - 1)\sum_{t=1}^{T}\ln y_t \tag{2}$$

The BC procedure consists of maximizing this log likelihood function over (λ,σ^2,β). Conceptually this allows the data to decide what is the most appropriate transformation of the dependent variable y.

However it is well known that the normality assumption for ϵ_t cannot be true except in the special case, $\lambda=0$, due to the positivity constraint on y_t. This implies that the likelihood function is misspecified and therefore we cannot apply the usual maximum likelihood properties of consistency or efficiency. Instead, at best, we must use quasi-maximum likelihood results and evaluate the objective function with the true distribution.

Various proposals for fixing the BC model have been advanced. Amemiya and Powell (1981) offer a two-stage nonlinear least squares estimator while Poirier (1978) suggests using a truncated normal distributional assumption. Even with the misspecification, work by Draper and Cox (1969) suggest that the results from BC might not be too wrong if in fact the true distribution is close to normal.

3. The NLS Model

An alternative to the BC model has been proposed by Jeff Wooldridge (1990), representing a distinct departure from these adjustments. Wooldridge motivates his model by noting that in the great majority of economic research the interest in estimation focuses on the mean of the untransformed dependent variable, conditional on the explanatory variables. For example, in a hedonic specification of the price of computers, the interest might lie in predicting the price of a new model for use in a price index--not in predicting a transformation of the price as the BC model would do. Or a researcher might want to find the marginal effect of a change in the price with respect to a change in the computing speed. Such questions obviously center around the untransformed dependent variable.

Observing the quandary surrounding BC estimation, Wooldridge proposes instead a conditional mean function

$$\mu(x_t) \equiv E(y_t|x_t) = (1+\lambda x_t\beta)^{\frac{1}{\lambda}}. \tag{3}$$

Such a model can be consistently estimated using NLS and has some very attractive properties. First it generalizes the functional form in much the same way as does the BC model: For $\lambda=1$ we have a linear model

$$\mu(x_t) = 1+x_t\beta \tag{4}$$

and for $\lambda = 0$ we have an exponential model

$$\mu(x_t) = Exp(x_t\beta), \tag{5}$$

which is analogous to the logarithmic case for BC (In fact, (5) is implied by the BC model when $\lambda=0$). As such it generalizes the choice of functional form in much the same way as BC, but does not have the theoretical flaws.

Second, it relaxes the independence assumption of the BC model (ϵ_t and x_t independently distributed) and can be expressed having either an additive error (see equation 8a below) or a multiplicative error (see equation 8b).

Having introduced the NLS alternative, we make two additional points. First, since the Wooldridge approach is a nonlinear model, standard nonlinear estimation problems arise (e.g. selection of starting values). These type of problems can often be avoided with a BC model since a BC model can usually be estimated with software having OLS and matrix algebra capabilities. Second, both the BC and NLS models can be viewed as approximations to the following more general framework

$$y_t = f(x_t, \theta, \epsilon_t) \tag{6}$$

where $\theta' = (\beta', \lambda)$.

In the BC case

$$f(x_t, \epsilon_t, \theta) = (1 + \lambda(x_t\beta + \epsilon_t))^{\frac{1}{\lambda}}. \tag{7}$$

In the NLS case

$$f(x_t, \epsilon_t, \theta) = (1 + \lambda x_t\beta)^{\frac{1}{\lambda}} + \epsilon_t \tag{8a}$$

or

$$f(x_t,\epsilon_t,\theta)=(1+\lambda x_t\beta)^{\frac{1}{\lambda}}\epsilon_t \tag{8b}$$

It is an open question which one better approximates the true function, $f(x_t,\epsilon_t,\theta)$.

A problem common to both the BC and the NLS models is the lack of invariance of standard Wald-type statistics (t-tests) to multiplicative scaling of the dependent variable. In the basic OLS model, multiplicative scaling of the dependent variable simply scales the estimated slope parameters and their associated standard errors by the same factor leaving t-statistics--hence inference results--unchanged.

This is not necessarily true with a nonlinear model. In general, there is no guarantee that inference results remain unchanged after scaling in a nonlinear model. This has unsettling implications. This lack of invariance means that when estimating, for example, a nonlinear wage equation, how you measure wage--in cents, or dollars, for example--could affect the statistical significance of the estimated coefficients. Such an effect has been noted by Spitzer (1984) for the BC model, by Wooldridge (1990) for the NLS model and its possible effects have been highlighted for both models by Berndt, Showalter, and Wooldridge (1990a). Each of these papers suggest as a possible solution scaling the original dependent variable by the sample geometric mean which makes the dependent variable trivially scale invariant, while Wooldridge and Berndt et al suggest as an alternative a Lagrange multiplier test which is invariant to multiplicative scaling.

The purpose of this paper is to use Monte Carlo methods to simulate both the BC model and the NLS model under alternative error covariance structures. Since no hypothesis framework is available to decide statistically which model performs the best, the research strategy sets the simulations such that the models are as similar as possible. The decision to focus attention on alternative error covariance structures is motivated largely by the observation that this will be the likely stumbling block of the BC model, and it would be useful to evaluate the sensitivity of the BC model to various assumptions about the error term. Although several studies have suggested that misspecification of the BC model will likely lead to values of λ biased toward zero (Amemiya and Powell (1981); Seaks and Layson (1983)), relatively little is known about the effects of misspecification on the estimated slope coefficients.[1]

The dependent variables in both models are normalized by the sample geometric mean. This is done for two reasons. First it makes models with differing λ's comparable. A previous Monte Carlo simulation of the BC model by Spitzer (1978) found that not normalizing can lead to the wrong conclusions. Second, this seems the more logical approach if in fact researchers use the models as Spitzer, Wooldridge and Berndt et al have suggested.

However, this normalization brings with it some problems. As Wooldridge points out, the sample geometric mean should be considered an estimate of the true geometric mean. Since this use of the sample geometric mean defines a new "true" parameter vector, each model with a different value for λ will have a different true parameter vector. Perhaps an example is in order.

Suppose we have the following BC specification

$$y_t=(1+\lambda x_t\beta+\lambda e_t)^{\frac{1}{\lambda}} \tag{9}$$

Let the sample geometric mean be

---

[1]Amemiya and Powell (1981) assume a two-parameter gamma distribution in which the magnitude of λ affects both the sign and the size of the bias on the coefficient estimates (coefficients include a constant term and one slope parameter).

$$\hat{g}=Exp(\frac{1}{T}\sum_{t=1}^{T}\ln y_t) \tag{10}$$

Now we redefine our model to be normalized by $\hat{g}$

$$y_t^* \equiv \frac{y_t}{\hat{g}}=(1+\lambda x_t\delta+\lambda e_t^*)^{\frac{1}{\lambda}}$$

where

$$\begin{aligned} 1+\lambda\delta_1 &= \hat{g}^{-\frac{1}{\lambda}}(1+\lambda\beta_1) \\ \delta_i &= \hat{g}^{-\frac{1}{\lambda}}\beta_i \qquad i=2,\ldots,k \\ e_t^* &= \hat{g}^{-\frac{1}{\lambda}}e_t \end{aligned} \tag{11}$$

Equation (11) is the model we would estimate. Conceptually, $\hat{g}$ is an estimate of the population geometric mean $g = \exp(E(\log(y_t)))$. As the value of λ changes the value of the population geometric mean will change also. But this implies that our true parameter vector δ will be different for different values of λ. Also, we note that for the (additive) NLS model,

$$y_t=(1+\lambda x_t\delta)^{\frac{1}{\lambda}}+e_t^*. \tag{12}$$

Here, even if β, x_t, ϵ_t and λ are the same as in equation (11), the geometric mean--both sample and population--will be different from the BC model. This implies that direct comparison between the two models will not be possible.

The approach taken here is one of cautious pragmatism: It is of some interest to determine the properties of the BC model given its theoretical flaws. It is also of interest to compare the performance of the BC and NLS models. But differing objective functions and unobservable "true" parameter vectors make direct comparisons between models and even within a model across different λ's somewhat arbitrary. Since it is of interest to compare models when λ varies it will be necessary to normalize the models. We will proceed by making a large sample estimate of the true geometric mean for each permutation of λ, the estimation method (BC or NLS) and the covariance structure. We will make this our reference point in determining the performance of the estimators. All comparisons will be done in percentage terms to make interpretation easier.

The simulation methodology focuses on the effects of both λ and the error distribution on the estimation of the slope coefficients. Following Hendry's (1984) suggestion to estimate a response function over the parameter space, we select several values for λ and also for the error distribution. The values chosen for λ-- -1.0, -0.5, 0, 0.5 and 1.0 --are similar to those used in Amemiya and Powell (1981) and in Spitzer (1978). They also encompass the range of λ estimated in several empirical studies which employ the BC model (e.g. Heckman and Polachek (1974), and Dinan and Miranowski (1989)).

The error structures are chosen to include interesting cases, three of which fit the classical BC assumptions on the error term (IID normal errors), and two error structures which are commonly found in empirical work but which violate the BC paradigm-- namely, heteroskedastic errors and autocorrelated errors.

The three classical error structures were chosen to account for approximately 10 percent (Case 1), 30 percent (Case 2) and 50 percent (Case 3) of the total variation of y (hereafter these percentages will be referred to as the variance ratio).[2] These values

[2]By comparison, Spitzer (1978) looked only at the case of errors accounting for 5 percent of the total variation.

correspond loosely to standard R-squared measures of approximately 0.9, 0.7 and 0.5, respectively.[3]

For the heteroskedastic model (Case 4) we choose an error structure similar to those discussed in Greene (1990, pp. 408-409) having a conditional variance of the form $V(\epsilon_t | x_t) = \sigma^2(1 + x_t\beta)$ with a variance ratio approximately equal to 10 percent.

For the autocorrelation model (Case 5) we choose a model with simple first order autocorrelation of the form $\epsilon_t = 0.7^*\epsilon_{t-1} + v_t$, v_t a white noise disturbance and the variance ratio again approximately equal to 10 percent.

With the primary focus of the simulation being on the effect of λ and the error structure, the selection of the β's and the x's is done to facilitate those ends. Again following Spitzer's lead, we choose a model with a constant term and two continuous regressors. Like Spitzer we have opposing signs on the two slope coefficients, but we simplify the values to be $\beta_1=\beta_2=-\beta_3=1$ with β_2 and β_3 being the coefficients on the continuous regressors. The value of $x_t\beta$ is chosen to equal zero in expectation to keep the variance ratio approximately the same across differing values of λ and across models (BC and NLS). We generalize Spitzer's specification somewhat by drawing a new set of x's for each replication rather than keeping the x's fixed. This is done to make the simulation results dependent upon a distribution (in our case, a bivariate normal) rather than on a particular draw from a distribution, as in Spitzer. A more detailed description of the estimation procedure is given in the appendix.

---

[3]The approximation is due to differences between the BC and the NLS models and across different values of λ.

4. Analysis

The presentation of the simulation results highlight the similarities and differences between the BC and the NLS models. First we examine the parameter estimation results, and then we investigate the fit of the two models to the data.

We are first interested in the rates of convergence of the various parameter estimates in the two models. To gain an intuitive appreciation for the rates of convergence, we present a series of graphs for each parameter in each of the two models. Each graph displays five line plots, one for each of the covariance structures. Since our results suggest a great similarity across different values of λ, we present results only for the case of $\lambda=1$. All terms are represented as percentage deviations from the true parameters.

Figures 1a and 1b give the mean squared percentage error (MSPE) for the constant term where MSPE is defined in equation (13).

$$MSPE=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\theta}_i-\theta}{\theta}\right)^2 \tag{13}$$

where:
$\hat{\theta}_i$ is the parameter estimate
θ is the true parameter
N is the number of simulations

It is readily seen from the graphs that the MSPE decreases in all cases as sample size grows, even in the case of heteroskedasticity and autocorrelation for the BC model. The graphs also suggest that the largest determinant of MSPE is the variance ratio. Recall that the simulation was set up to give the autocorrelation and heteroskedasticity cases approximately the same variance ratio (10 percent) as model 1. A comparison of the BC and NLS models reveal a striking similarity with trends matching in magnitudes across the two models. Figures 2a, 2b, 3a and 3b reveal similar results for the two slope parameters. The major difference between the slope parameters and the constant term is the magnitude of the MSPE. For T=30, model 3 (variance ratio = 50 percent), the MSPE for the constant is 0.26 while for the first slope coefficient it is 0.06. This suggests that while the estimate of the constant may be consistent, in all cases it may be poorly estimated in small samples when the variance ratio is large.

Figures 4a and 4b give the MSPE results for λ. This is a case where BC differs substantially from NLS. The first major difference is that NLS generally exhibits much larger MSPE than does the BC model. Much of the discrepancy comes from a few very large outliers in the NLS model, the largest being an estimate of 25 when the true value is 1. Secondly, the differences between the various error structures do not seem as pronounced as in the case of the other parameters. Figure 5 shows a graph of the mean bias for the transformation parameter in the BC model. The graph indicates that in the case of heteroskedasticity, the BC estimate of λ is in fact inconsistent.

A more formal analysis of the above observations involves the calculation of a response function for each of the parameters. Interestingly, this is a potentially excellent use of the BC-NLS generalization framework. A large number of Monte Carlo simulations have the format of a positive variable (e.g. MSPE) explained by a set of explanatory variables (e.g. sample size).[4] The exact functional form is unknown and therefore various approximations are used. Both the BC and NLS models allow for a considerable amount of generality in functional form including interaction between explanatory variables with a relatively parsimonious form. For this reason we use both the BC and the NLS models to calculate response functions for the MSPE of each of the

[4]For example, see Lo and MacKinlay (1989) or Donald and Donner (1990).

parameters. The response functions are shown in tables 2-9 with an explanation of the variables in Table 1.

Through this set of tables several striking patterns emerge. First is the confirmation of the results shown in the previous graphs: heteroskedasticity and autocorrelation have relatively little impact on MSPE relative to the other variables. Coefficients for heteroskedasticity and autocorrelation are almost always small and insignificant. The notable exception is for the transformation parameter in the BC model in the presence of heteroskedasticity (Table 9: columns 3 and 4) It is also interesting to note that the OLS approximations of the response function do not give as clear results in this particular case. They instead show heteroskedasticity as statistically insignificant.

The effect of λ on the estimation results also shows some variation between the BC and the NLS models. In the NLS model the value of λ generally does not matter; the coefficients are typically small and insignificant. However for the BC model a pattern of statistical significance and negative coefficients seems to predominate. For instance, looking at Table 5 (BC parameter 2) we see that three of the four estimates of the effect of λ are negative and significant.

Looking at the effect of sample size we see that the rates of convergence are higher in the NLS model (larger coefficient values in absolute value). As an example compare the dummy coefficients for time for parameter 2 between the BC and NLS models (Tables 4 and 5 for BC and NLS, respectively). We see that the coefficients for T60 in the BC model=-0.0099 while in the NLS model it is -0.0121. Such patterns emerge consistently across all the regressions. Another possibly related pattern is that NLS usually exhibits a larger constant term than BC. We also find the NLS model is typically more sensitive to the variance ratio than the BC analog.

It is also of interest to know the general shape of the small sample parameter distribution. Again using the case when $\lambda=1$ we have calculated a set of two-tailed

Kolmogorov statistics testing for normality of the distribution of the parameter estimates using as the null the sample mean and sample variance. The results in Tables 10-13 show that generally we cannot reject the null of being normally distributed, albeit with a mean and variance different from the true values.[5] Since we are using the sample rather that true population moments, we would not necessary expect a decrease in significance as the sample size increases, whose lack of trend is evidenced in the tables. Of note again is the similarity between alternative covariance structures. Increasing variance, autocorrelation or heteroskedasticity does not seem to have a major effect on the test results.

A more useful analysis for applied work involves calculating the implied probabilities for standard t-tests. The results are listed in Tables 14-17. As a general rule, we see that the estimated probabilities of correctly "accepting" the null hypothesis underestimate the asymptotic level of 0.95.[6] Some interesting patterns emerge, however. First we see that the pattern of underestimation seem to be more true for the NLS model than for the BC model. For example, in Table 14 for parameter 1, only 2 of 20 NLS estimates equal or exceed 0.95, while in the parallel case for the BC, 8 of 20 do. Similar patterns exist for the two slope parameters. Also, as would be expected, heteroskedasticity and autocorrelation appear lower relative to the standard cases, evidencing an incorrect estimate of the standard errors. Somewhat disturbing is the apparent lack of trend toward the asymptotic values of 0.95. Also we see a very distinct difference in pattern for the BC estimate of λ in Table 17. In this table we see values ranging from 0.696 to 0.992, with a decided downward trend in the case of

---

[5]The critical value for a test of (asymptotic) size 0.05 is approximately 0.0608 (From Lindgren (1976), Table VI, N=500).

[6]More precisely, the estimate of a type I error for testing the null hypothesis that the estimated parameter equals the true parameter is generally overestimated.

heteroskedasticity. This table also shows that for relatively low sample sizes (T=30,60) the average t-statistic exceeds 0.95 while for sample sizes 200 and 500 the t-statistic is less than 0.95. The pattern for NLS more closely matches that of the other parameters.

Given the results to this point, we can say several useful things. First, both BC and NLS exhibit surprising similarities given that they are estimated with different objective functions. The constant term appears to be poorly estimated in both models. Also the trends toward consistency for the constant and slope coefficients look remarkably similar. The major distinction between that two models comes in the estimation of λ. The estimated λ for NLS has similar patterns to the NLS slope coefficients, but for the BC model λ exhibits a different pattern: It appears to be less sensitive to the variance ratio and appears inconsistent for the heteroskedastic case.

Finally, NLS appears to be more sensitive to sample size than the BC model with MSPE decreasing more rapidly (relative to BC) with an increase in sample size.

Fitted Values

Beyond asking questions about each individual parameter, we are also interested in how well the models fit the data. A reasonable and simple statistic to use would be the standard R-squared measure computed by most econometric packages. By construction, however, we know that NLS maximizes R-squared (for a given functional form) and so it will dominate the simple BC fitted values.[7] It is still of interest, however, to calculate functions to gauge marginal effects (e.g. of sample size). Also, if we use as our criterion the deviation from a "true" R-squared we have no assurance that NLS will dominate BC.

[7]Where the stochastic error is set to zero and the BC model is solved for y to compute the fitted value.

A related measure, one which does not necessarily favor NLS, is an absolute deviation based R-squared measure whose essential difference from the standard R-squared is that sums are taken over absolute diffences rather that squared differences.

We use both these R-squared measures as a means for testing goodness-of-fit. As in the parameter analysis section, we calculate response functions based on squared percentage deviations from the true R-squared value where the true R-squared involves using the true residuals. We use the same set of variables as before except now we just perform simple OLS calculations measuring all the continuous positive variables in logarithms. The results are given in Tables 18 and 19 with R1 refering to the standard R-squared and R2 the absolute deviation R-squared.

Looking first at R1 we see that the dependent mean is lower for NLS than BC. This is to be expected given the setup of NLS. Somewhat surprising, though, is that for some cases BC gets closer to the true value than NLS. As we noted, NLS maximizes R-squared, but that might mean NLS overstates the true R-squared. Since BC has a smaller value for the constant term (-9.11 versus -8.976 for NLS) if we take the case $\lambda=0$, T=30 (the norm), Variance Ratio=0.1, using the OLS1 model we see that BC estimate equals -7.53 while NLS = -7.404; i.e. BC gets closer to the true value. This special case appears to erode quickly as sample size increases. The NLS model improves at a much quicker rate than BC, roughly twice as fast looking at the coefficients for T (-0.044 for BC; -0.088 for NLS). As would be expected, the dominant determinant is the variance ratio, having roughly the same effect in both models.

Interestingly, given the poor performance of λ in the BC model in the presence of heteroskedasticity, the effect of heteroskedasticity is rather small, not evident at all in OLS1 and marginally significant in OLS2. Of course, we see no effect in the NLS model to either heteroskedasticity or autocorrelation.

The other major difference between BC and NLS is the effect of the true λ. The true λ has no effect in the NLS model while its effect in the BC model is relatively strong. OLS1 for BC suggests that the larger the true λ, the better the model fits, while OLS2 suggests a convex function in λ with a minimum around -1 (depending on the other variables).

Looking at R2 we see much the same pattern. First we see that NLS has a lower mean than BC. Also, as with R1, it would appear that BC might produce better results than NLS in small samples (T less than 60). Sample size has a strong effect in NLS with rapid improvement as sample size increases (coefficients: OLS2--BC= -0.02, NLS= -0.03). Using R2 we see that heteroskedasticity has little effect in the BC model in contrast to what we found with R1. We also see that the true λ does have an effect in the BC model, much the same as we found with R1.

The major conclusions from this analysis would suggest that relative to their own "true" models, NLS generally gives better fitted values although BC might be preferable in small samples. BC is sensitive to the true λ and shows some evidence, although not overwhelming, of being quite sensitive to the presence of heteroskedasticity.

5. Conclusions

In this paper we have presented the results of a Monte Carlo study of the Box-Cox model and a nonlinear least squares alternative proposed by Wooldridge (1990). The simulation extends previous Monte Carlo work on the Box-Cox model by looking explicitly at alternative structures of the error distribution, including the case of heteroskedastic errors and autocorrelated errors. The simulation also presents new work on the NLS estimator and was structured to allow for comparision between the BC and the NLS model. Several interesting results were discovered and can be summarized as follows:

1) λ appears to be inconsistently estimated in the BC model in the presence of conditional heteroskedasticity. This, of course, does not occur in the NLS model. All other parameters appear to be consistently estimated (decreasing MSPE as sample size increases).

2) Relative to the slope parameters, the constant term in both the BC and the NLS model is poorly estimated in small samples.

3) The variance ratio (ratio of error variance to the total variance of y) is the most important factor in determining MSPE for both models, overshadowing the effects of misspecification (autocorrelated or heteroskedastic errors).

4) While the effect of sample size is more evident in the NLS model, there is some evidence that BC might perform better than NLS in small samples (relative to the correct specification), both in terms of parameter estimation and in fitting the data. Also there is some evidence that the NLS estimator is sensitive to a high variance ratio.

5) Asymptotic size is generally overestimated in the NLS model when Gaussian variance estimators are used.

6) The true value of λ appears to be important in determining the fit of the data in the BC model while it is unimportant in the NLS model.

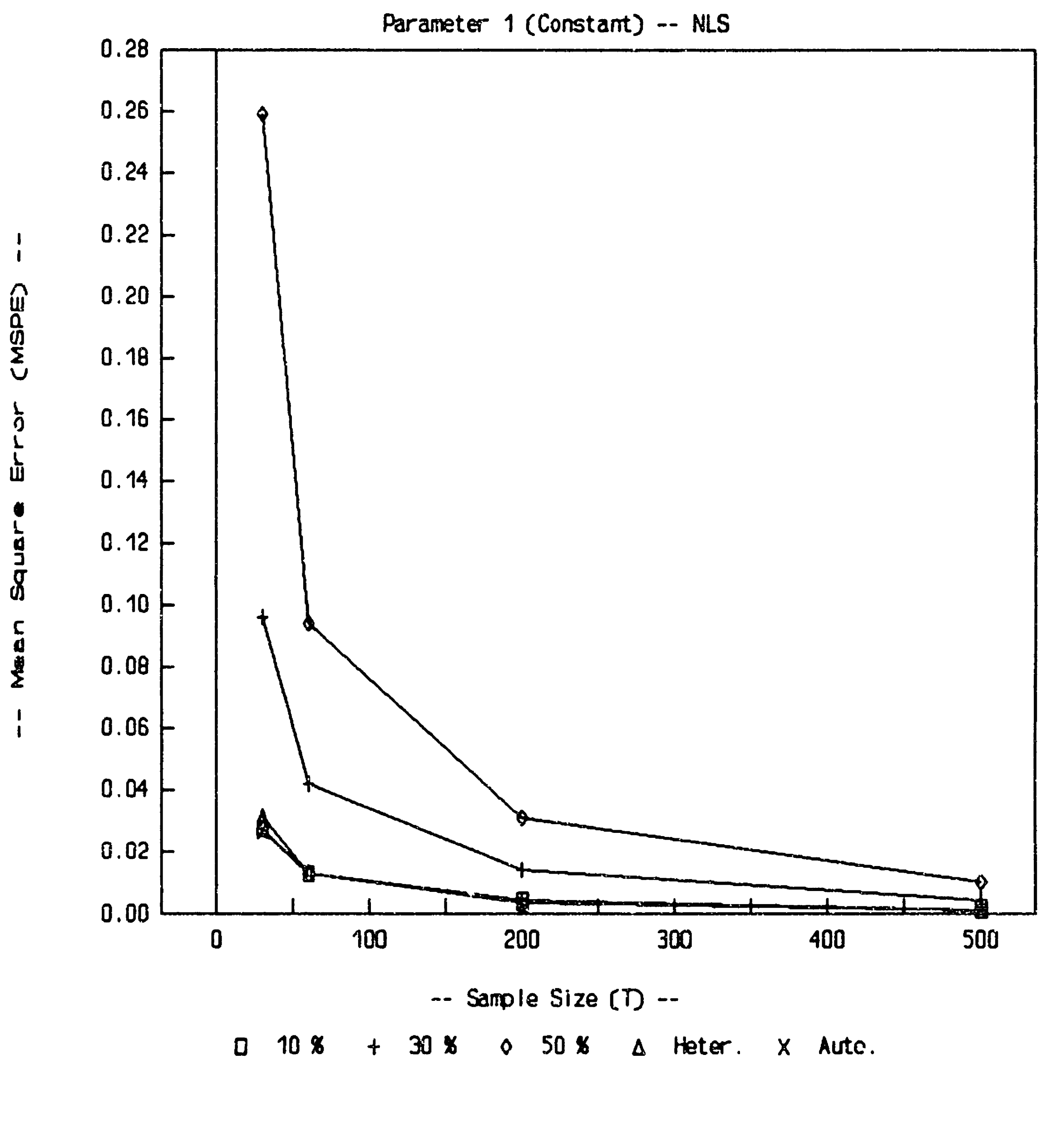
This research suggests several avenues for further exploration. Given the poor performance of both models in estimating the constant term, it is of some practical interest to study the effect of estimating a model with dummy variables. Several possible applications of the general BC-NLS framework involve hedonic studies with time dummies and in these models the effectiveness of estimating the coefficients on the dummy variables takes on significant practical importance.[8]

Also of practical significance is gauging the relative importance of transformations on the x's as well as on the y's. Initial work done by Berndt, Showalter and Wooldridge (1990b) suggests a substantial amount of substitutibility between transforming the y's and transforming the x's.

Finally, and perhaps most obviously, more refined Monte Carlo methods could be employed. These would include simulations over a larger parameter set (including variations on the distribution of the x's) in order to fit the response function better. As the price of computing power continues its expected decline, these projects should become more feasible.

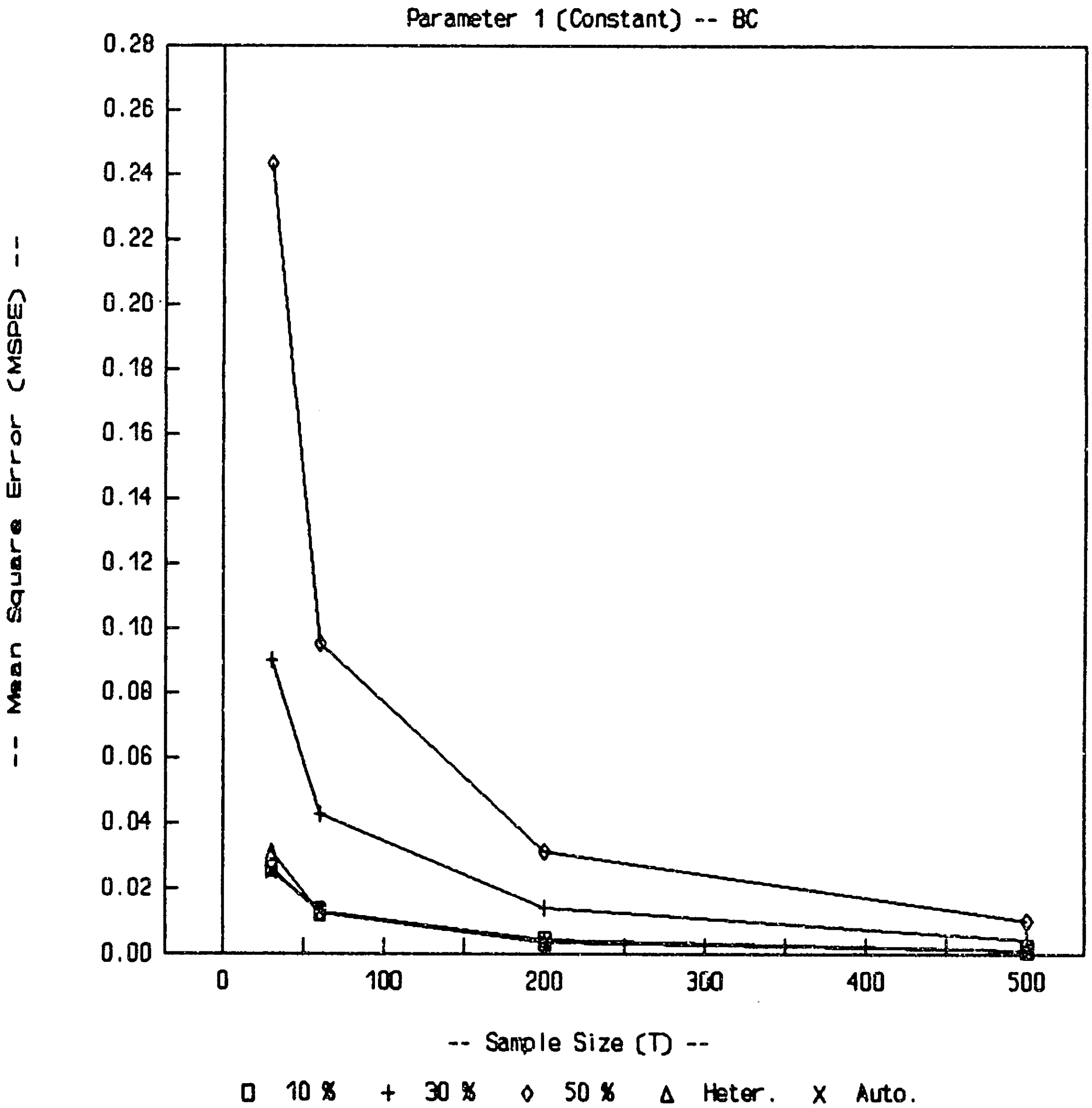
[8]For a discussion of the use of the BC-NLS framework in the context of hedonic models see Berndt, Showalter and Wooldridge (1990a) and (1990b).

Figure 1a



Parameter 1 (Constant) -- NLS
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 % + 30 % ◇ 50 % Δ Heter. X Autc.

Figure 1b



Parameter 1 (Constant) -- BC
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 % + 30 % ◊ 50 % Δ Heter. X Auto.

Figure 2a

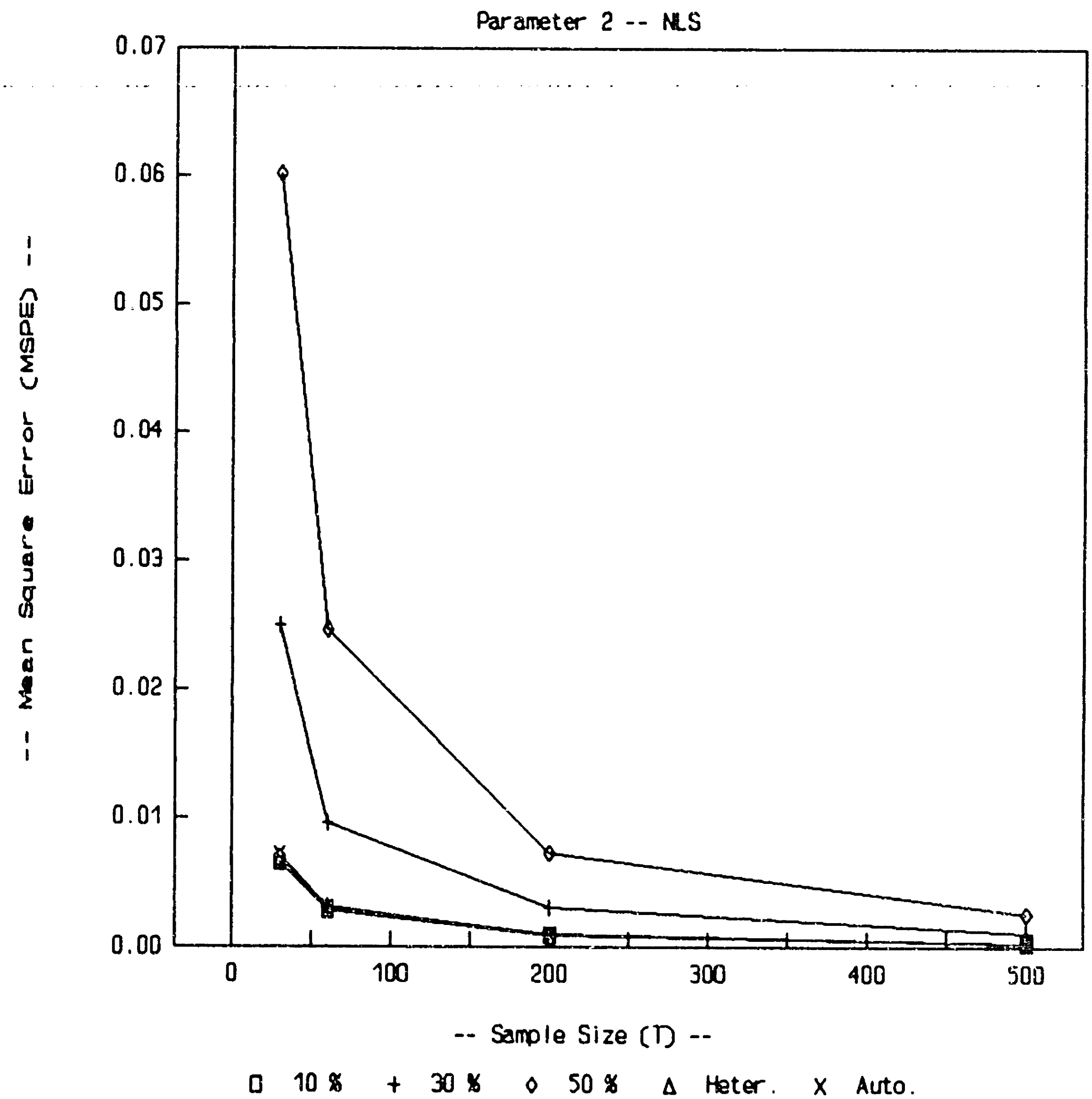

Parameter 2 -- NLS
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 % + 30 % ◊ 50 % Δ Heter. X Auto.

Figure 2b

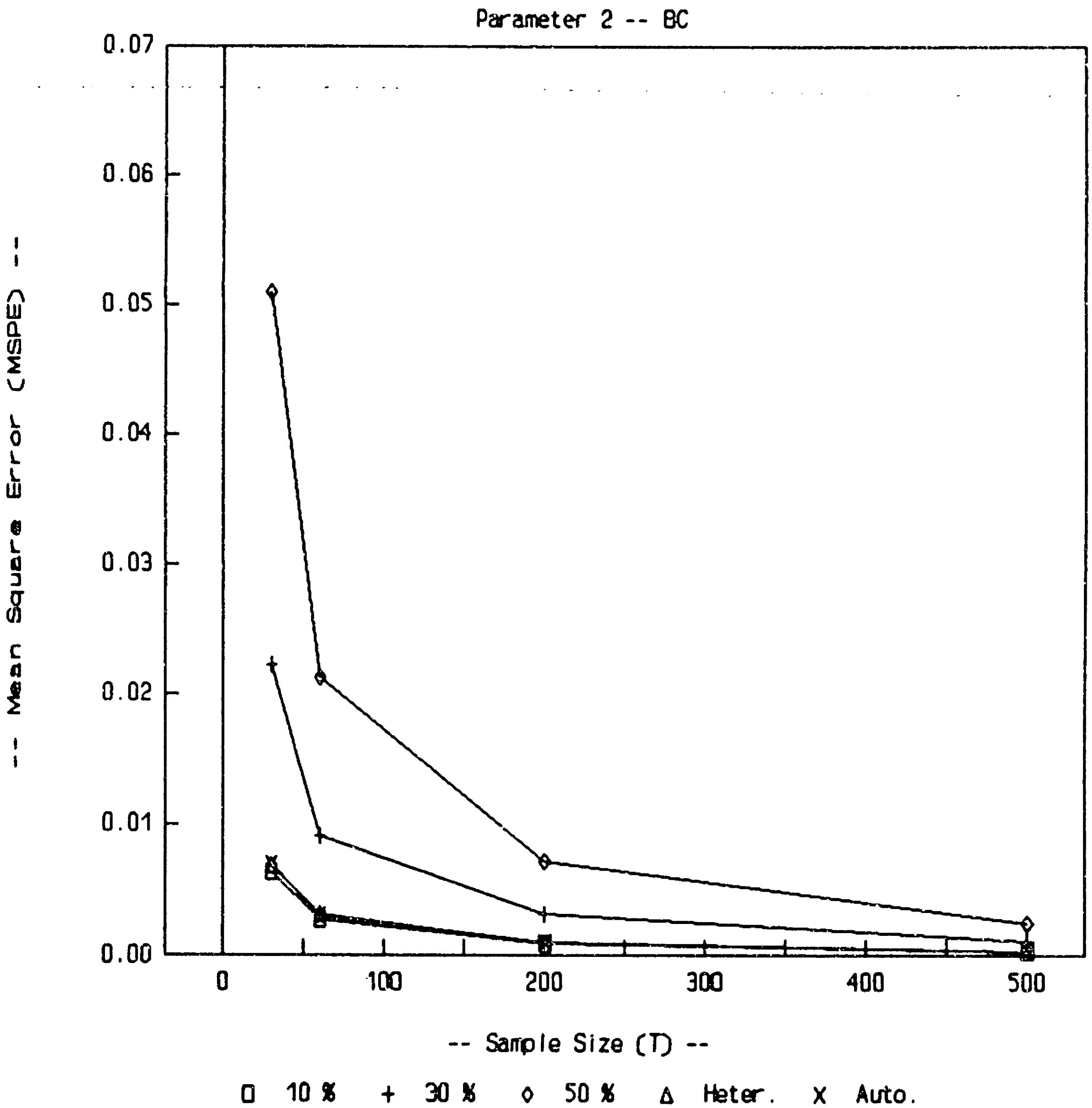
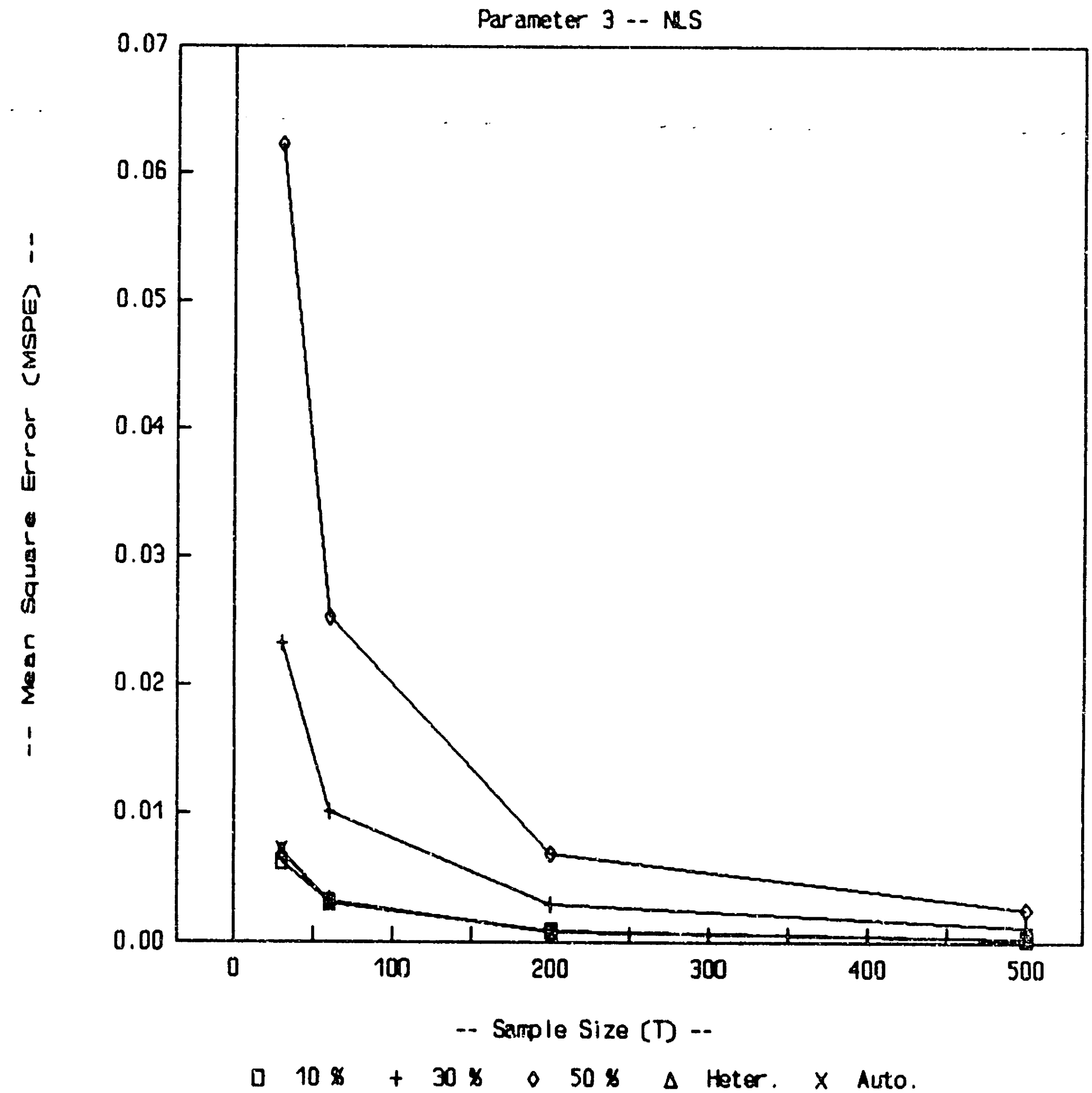
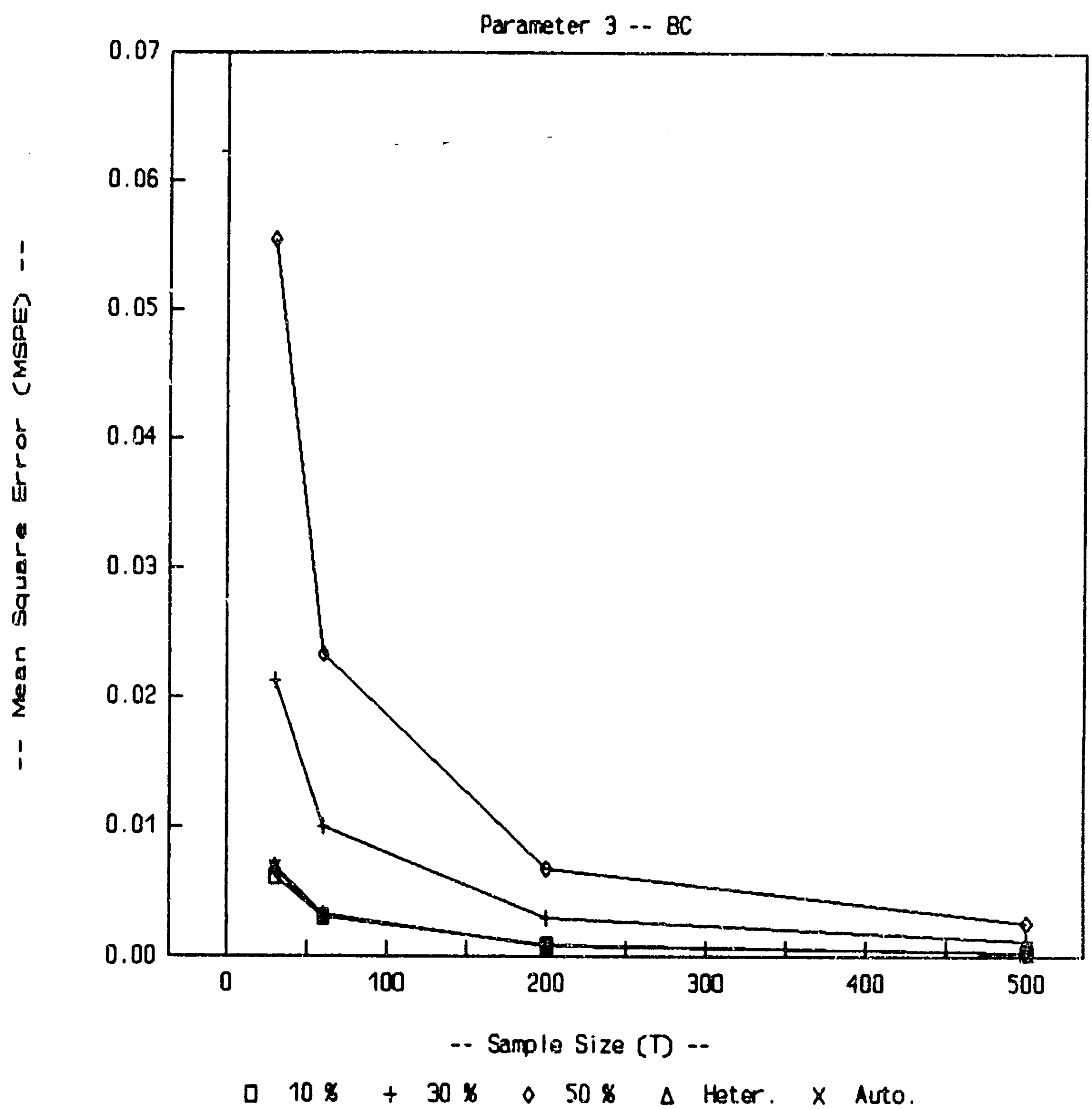

Parameter 2 -- BC
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 %   + 30 %   ◊ 50 %   Δ Heter.   X Auto.

Figure 3a



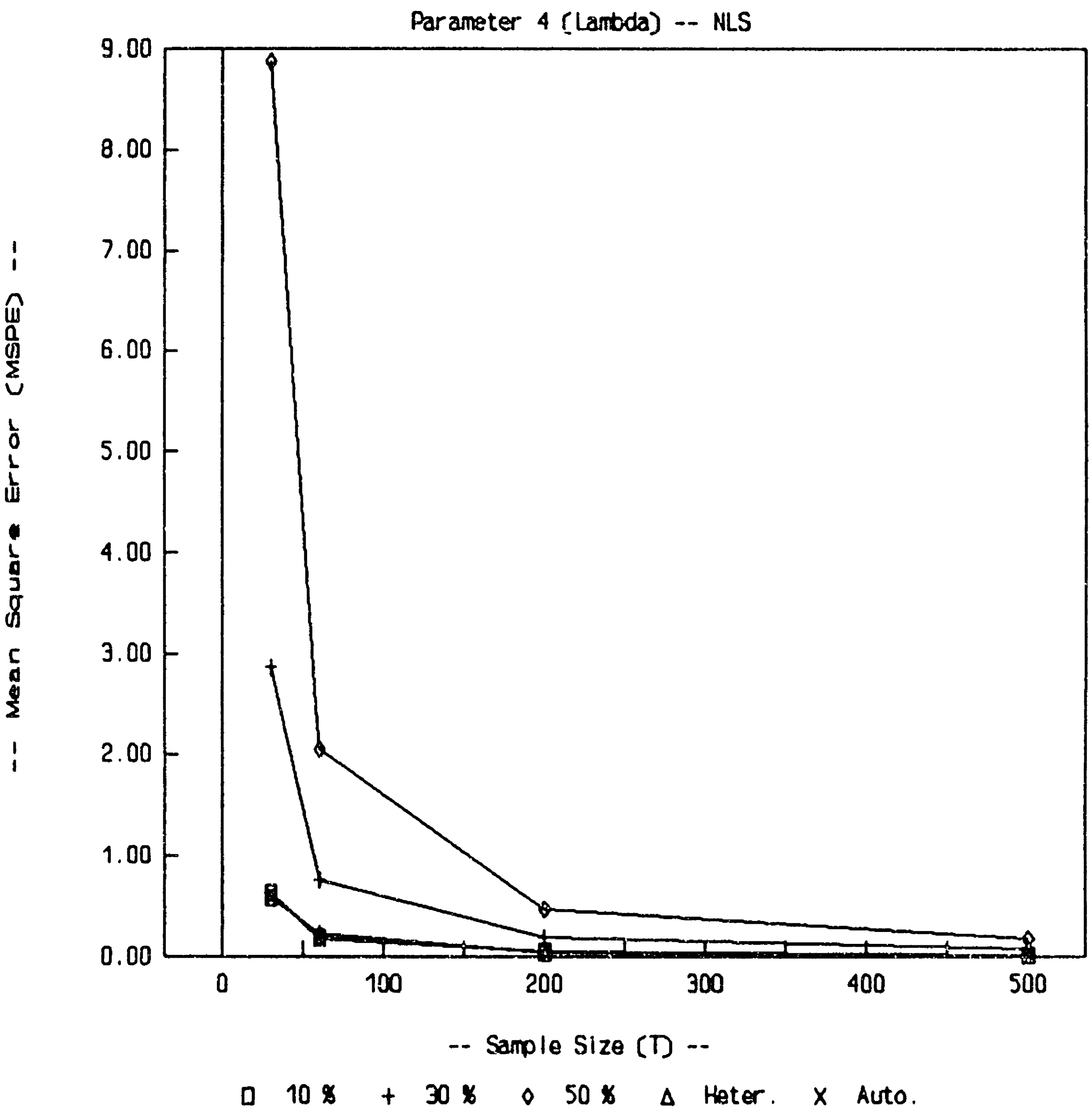
Parameter 3 -- NLS
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 % + 30 % ◊ 50 % Δ Heter. X Auto.

Figure 3b



Parameter 3 -- BC
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 % + 30 % ◇ 50 % Δ Heter. X Auto.

Figure 4a



Parameter 4 (Lambda) -- NLS
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 % + 30 % ◇ 50 % Δ Heter. X Auto.

Figure 4b

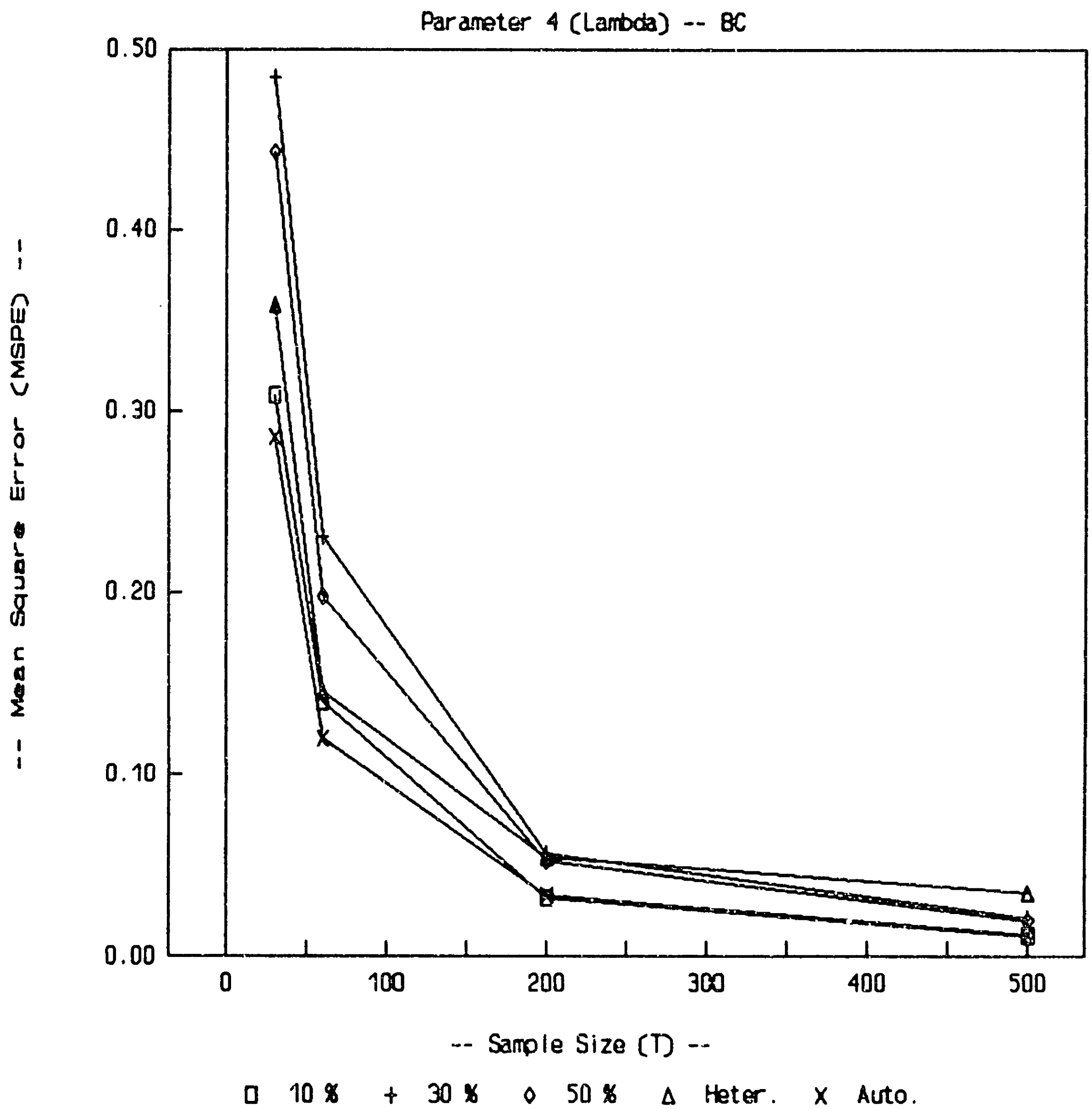

Parameter 4 (Lambda) -- BC
0.50
0.40
0.30
0.20
0.10
0.00
-- Mean Square Error (MSPE) --
0
100
200
300
400
500
-- Sample Size (T) --
□ 10 % + 30 % ◇ 50 % Δ Heter. X Auto.

Figure 5

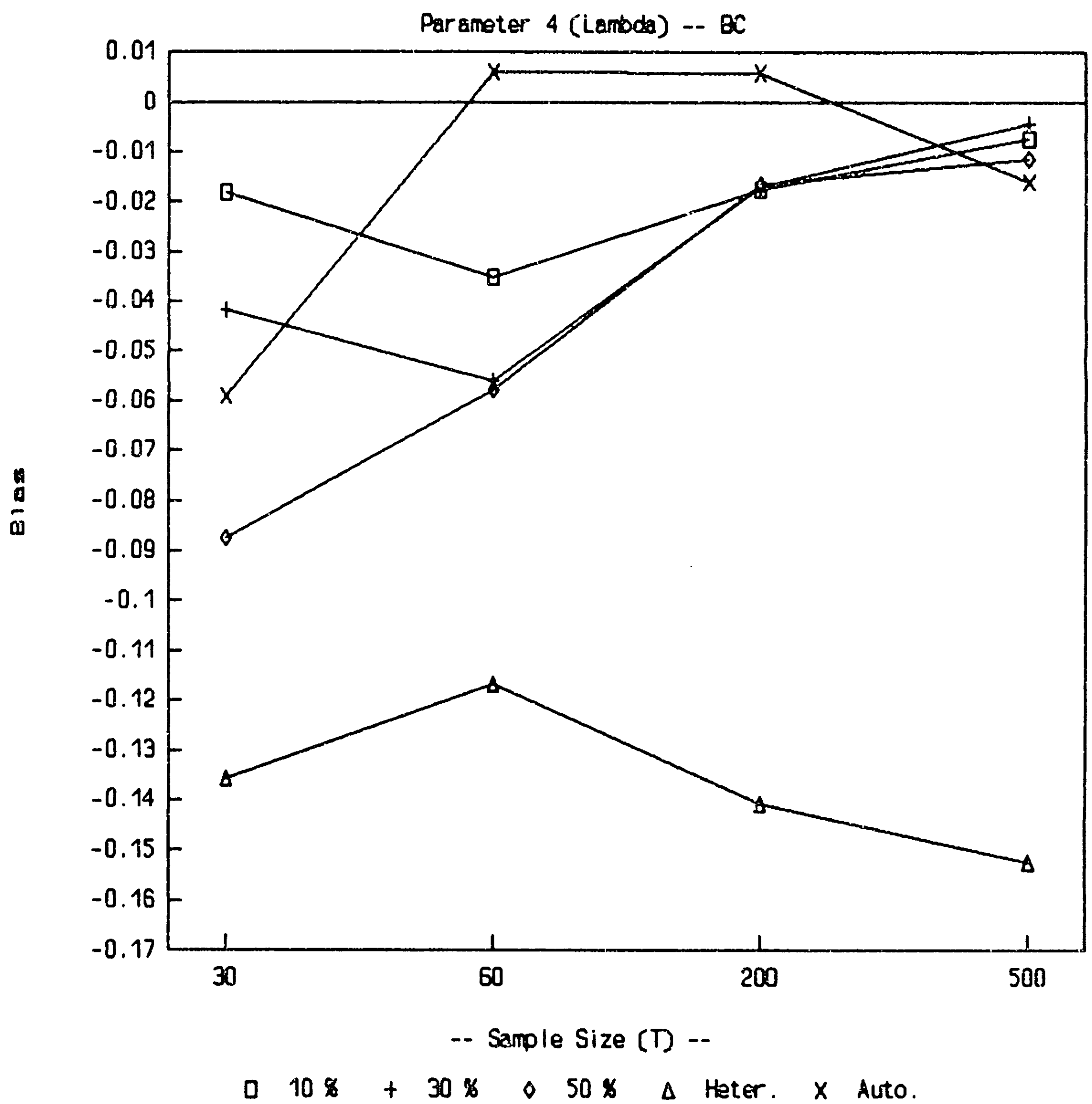

Parameter 4 (Lambda) -- BC
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
-0.07
-0.08
-0.09
-0.1
-0.11
-0.12
-0.13
-0.14
-0.15
-0.16
-0.17
Bias
30
60
200
500
-- Sample Size (T) --
□ 10 % + 30 % ◇ 50 % Δ Heter. X Auto.

Table 1

Description of variables used to calculate response functions.

LAMBDA - True Value of λ

T - Sample Size

T60,T200,T500 - Dummy variable for sample size (e.g. T60=1 if T=60, 0 otherwise)

VAR_RATIO - Estimate of true variance ratio calculated with true residuals

$$= \frac{\sum_{t=1}^{T} (\epsilon_t - \bar{\epsilon})^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

$$\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t, \qquad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

HETER - 1 if Heteroskedastic errors (Case 4), 0 otherwise.

AUTO - 1 if Autocorrelated errors (Case 5), 0 otherwise.

VAR_RATIO_SQ,LAMBDA_SQ,T_SQ,T*VAR_RATIO,T*LAMBDA, LAMBDA*VAR_RATIO - Square and Crossproduct terms.

MSPE_TP - Transformation parameter for dependent variable (MSPE).

T_TP - Transformation parameter for Time (T).

VAR_RATIO_TP - Transformation parameter for VAR_RATIO.

Table 2

Response Function Results for Parameter 1 (Constant)
--- NLS ---

	OLS1	OLS2	NLS	BC
Dep. Mean	0.0332	0.0332	0.0332	
R-Squared	0.6769	0.8184	0.9939	
CONSTANT	0.0322	0.0296	6.0328	4.0719
	(0.0107)	(0.0150)	(0.7270)	(0.2427)
LAMBDA	0.0004	0.0019	0.0248	0.0545
	(0.0047)	(0.0072)	(0.0107)	(0.0108)
T		-0.0004	-2.1146	-1.1367
		(0.0001)	(0.4681)	(0.0995)
T60	-0.0477			
	(0.0093)			
T200	-0.0711			
	(0.0093)			
T500	-0.0779			
	(0.0093)			
VAR_RATIO	0.2277	0.0880	2.9608	2.8324
	(0.0274)	(0.1136)	(0.2893)	(0.2068)
HETER	0.0057	-0.0007	-0.0038	0.0037
	(0.0099)	(0.0079)	(0.0976)	(0.0275)
AUTO	0.0053	-0.0016	-0.0816	-0.0340
	(0.0100)	(0.0081)	(0.1178)	(0.0227)
LAMBDA_SQ		-0.0004		
		(0.0060)		
T_SQ		7.40E-07		
		(1.21E-07)		
VAR_RATIO_SQ		0.5133		
		(0.1898)		
T*VAR_RATIO		-8.06E-04		
		(8.96E-05)		
LAMBDA*VAR_RATIO		-0.0029		
		(0.0238)		
T*LAMBDA		-2.23E-06		
		(1.92E-05)		
MSPE_TP			-0.0157	-0.0075
			(0.0394)	(0.0080)
T_TP			-0.1426	-0.0075
			(0.0674)	(0.0217)
VAR_RATIO_TP			0.5018	0.4769
			(0.1055)	(0.0516)

Table 3

Response Function Results for Parameter 1 (Constant)
--- BC ---

	OLS1	OLS2	NLS	BC
Dep. Mean	0.0326	0.0326	0.0326	
R-Squared	0.6935	0.8428	0.9820	
CONSTANT	0.0284	0.0209	4.1269	3.1937
	(0.0100)	(0.0132)	(0.8731)	(0.4626)
LAMBDA	-0.0048	0.0052	-0.1547	-0.0988
	(0.0043)	(0.0068)	(0.0289)	(0.0240)
T		-0.0003	-1.3652	-0.9435
		(0.0001)	(0.4966)	(0.1579)
T60	-0.0424			
	(0.0086)			
T200	-0.0657			
	(0.0086)			
T500	-0.0727			
	(0.0086)			
VAR_RATIO	0.2355	0.1100	2.0154	2.0407
	(0.0270)	(0.1036)	(0.3128)	(0.3736)
HETER	0.0053	0.0002	0.1618	-0.0115
	(0.0091)	(0.0070)	(0.1744)	(0.0565)
AUTO	0.0047	-0.0006	0.1136	-0.0306
	(0.0092)	(0.0071)	(0.2080)	(0.0604)
LAMBDA_SQ		0.0072		
		(0.0054)		
T_SQ		6.70E-07		
		(1.07E-07)		
VAR_RATIO_SQ		0.5221		
		(0.1812)		
T*VAR_RATIO		-7.98E-04		
		(8.51E-05)		
LAMBDA*VAR_RATIO		-0.0659		
		(0.0260)		
T*LAMBDA		1.07E-05		
		(1.70E-05)		
MSPE_TP			-0.0168	-0.0007
			(0.0675)	(0.0150)
T_TP			-0.0581	0.0184
			(0.1101)	(0.0340)
VAR_RATIO_TP			0.1787	0.2577
			(0.1616)	(0.1242)

Table 4

Response Function Results for Parameter 2
--- NLS ---

	OLS1	OLS2	NLS	BC
Dep. Mean	0.0084	0.0084	0.0084	
R-Squared	0.6762	0.8239	0.9984	
CONSTANT	0.0078	0.0072	4.3128	2.8639
	(0.0028)	(0.0038)	(0.3670)	(0.4323)
LAMBDA	-0.0006	0.0002	-0.0559	-0.0618
	(0.0012)	(0.0018)	(0.0063)	(0.0134)
T		-8.98E-05	-1.9785	-1.1543
		(1.77E-05)	(0.2100)	(0.1214)
T60	-0.0121			
	(0.0024)			
T200	-0.0181			
	(0.0024)			
T500	-0.0198			
	(0.0024)			
VAR_RATIO	0.0593	0.0193	2.9381	3.1487
	(0.0070)	(0.0287)	(0.1838)	(0.4524)
HETER	0.0014	-3.33E-04	-0.0623	-0.0363
	(0.0026)	(2.00E-03)	(0.0515)	(0.0273)
AUTO	0.0018	-3.48E-05	0.0874	0.0665
	(0.0026)	(2.05E-03)	(0.0559)	(0.0259)
LAMBDA_SQ		4.89E-04		
		(1.52E-03)		
T_SQ		1.88E-07		
		(3.07E-08)		
VAR_RATIO_SQ		0.1396		
		(0.0480)		
T*VAR_RATIO		-2.08E-04		
		(2.26E-05)		
LAMBDA*VAR_RATIO		-4.88E-03		
		(6.03E-03)		
T*LAMBDA		1.77E-06		
		(4.85E-06)		
MSPE_TP			0.0073	-0.0013
			(0.0202)	(0.0125)
T_TP			-0.1445	-0.0171
			(0.0343)	(0.0264)
VAR_RATIO_TP			0.5215	0.5457
			(0.0547)	(0.0810)

Table 5

Response Function Results for Parameter 2 --- BC ---				
	OLS1	OLS2	NLS	BC
Dep. Mean	0.0072	0.0072	0.0072	
R-Squared	0.6980	0.8351	0.9855	
CONSTANT	0.0070	0.0047	3.3526	1.9570
	(0.0022)	(0.0029)	(1.0158)	(0.5677)
LAMBDA	-0.0012	8.02E-04	-0.1957	-0.1227
	(0.0009)	(1.51E-03)	(0.0438)	(0.0237)
T		-7.52E-05	-1.7655	-1.0751
		(1.38E-05)	(0.5500)	(0.2023)
T60	-0.0099			
	(0.0019)			
T200	-0.0150			
	(0.0019)			
T500	-0.0164			
	(0.0019)			
VAR_RATIO	0.0500	0.0262	1.9605	2.1335
	(0.0058)	(0.0232)	(0.3641)	(0.4490)
HETER	0.0010	7.16E-05	0.0891	0.0601
	(0.0020)	(1.56E-03)	(0.1456)	(0.0564)
AUTO	0.0011	3.28E-05	0.1480	0.0364
	(0.0020)	(1.59E-03)	(0.1714)	(0.0459)
LAMBDA_SQ		2.14E-03		
		(1.20E-03)		
T_SQ		1.54E-07		
		(2.38E-08)		
VAR_RATIO_SQ		0.1070		
		(0.0405)		
T*VAR_RATIO		-1.71E-04		
		(1.90E-05)		
LAMBDA*VAR_RATIO		-1.46E-02		
		(5.81E-03)		
T*LAMBDA		3.38E-06		
		(3.79E-06)		
MSPE_TP			-0.0135	0.0111
			(0.0585)	(0.0166)
T_TP			-0.1058	-0.0203
			(0.1002)	(0.0404)
VAR_RATIO_TP			0.1845	0.3389
			(0.1366)	(0.1341)

Table 6

Response Function Results for Parameter 3 --- NLS ---				
	OLS1	OLS2	NLS	BC
Dep. Mean	0.0085	0.0085	0.0085	
R-Squared	0.6702	0.8277	0.9973	
CONSTANT	0.0076	0.0076	3.4120	3.1048
	(0.0028)	(0.0038)	(0.3790)	(0.3809)
LAMBDA	-0.0007	1.91E-04	-0.0763	-0.0595
	(0.0012)	(1.84E-03)	(0.0095)	(0.0127)
T		-9.21E-05	-1.3281	-1.2662
		(1.78E-05)	(0.1849)	(0.1220)
T60	-0.0117			
	(0.0024)			
T200	-0.0183			
	(0.0024)			
T500	-0.0197			
	(0.0024)			
VAR_RATIO	0.0602	0.0167	3.3390	3.1664
	(0.0072)	(0.0288)	(0.2747)	(0.4317)
HETER	0.0015	-2.90E-04	-0.0577	-0.0299
	(0.0026)	(2.01E-03)	(0.0712)	(0.0228)
AUTO	0.0018	-1.85E-04	0.0249	0.0076
	(0.0026)	(2.06E-03)	(0.0785)	(0.0230)
LAMBDA_SQ		0.0005		
		(0.0015)		
T_SQ		1.93E-07		
		(3.08E-08)		
VAR_RATIO_SQ		0.1470		
		(0.0482)		
T*VAR_RATIO		-2.12E-04		
		(2.27E-05)		
LAMBDA*VAR_RATIO		-0.0058		
		(0.0061)		
T*LAMBDA		2.25E-06		
		(4.87E-06)		
MSPE_TP			-0.0081	0.0039
			(0.0269)	(0.0111)
T_TP			-0.0357	-0.0400
			(0.0459)	(0.0247)
VAR_RATIO_TP			0.5682	0.5707
			(0.0728)	(0.0911)

Table 7

Response Function Results for Parameter 3
--- BC ---

	OLS1	OLS2	NLS	BC
Dep. Mean	0.0074	0.0074	0.0074	
R-Squared	0.6895	0.8389	0.9794	
CONSTANT	0.0068	0.0048	2.1932	2.1887
	(0.0023)	(0.0030)	(0.9016)	(0.5273)
LAMBDA	-0.0012	8.26E-04	-0.1983	-0.1238
	(0.0010)	(1.55E-03)	(0.0544)	(0.0222)
[illegible]		-7.76E-05	-1.1161	-1.1446
		(1.41E-05)	(0.4184)	(0.2071)
T60	-0.0096			
	(0.0019)			
	-0.0153			
	(0.0019)			
T200				
T500	-0.0166			
	(0.0019)			
VAR_RATIO	0.0523	0.0258	2.0135	2.2323
	(0.0062)	(0.0237)	(0.4492)	(0.4486)
HETER	0.0012	1.10E-04	0.1447	0.0308
	(0.0021)	(1.60E-03)	(0.1904)	(0.0514)
AUTO	0.0011	-2.74E-05	0.1612	-0.0187
	(0.0021)	(1.63E-03)	(0.2240)	(0.0455)
LAMBDA_SQ		0.0023		
		(0.0012)		
T_SQ		1.60E-07		
		(2.44E-08)		
VAR_RATIO_SQ		0.1145		
		(0.0415)		
T*VAR_RATIO		-1.79E-04		
		(1.95E-05)		
LAMBDA*VAR_RATIO		-0.0150		
		(0.0060)		
T*LAMBDA		3.40E-06		
		(3.89E-06)		
MSPE_TP			-0.0189	0.0141
			(0.0727)	(0.0151)
T_TP			0.0053	-0.0353
			(0.1225)	(0.0396)
VAR_RATIO_TP			0.1668	0.3790
			(0.1688)	(0.1340)

Table 8

Response Function Results for Parameter 4 (Lambda)
--- NLS ---

	OLS1	OLS2	NLS	BC
Dep. Mean	0.9369	0.9369	0.9369	
R-Squared	0.5401	0.6555	0.9939	
CONSTANT	1.0577	1.2354	12.8112	11.1533
	(0.5026)	(0.8101)	(4.2768)	(0.9135)
LAMBDA	-0.0578	0.0612	-0.0717	0.0991
	(0.2183)	(0.3918)	(0.0118)	(0.0240)
T		-0.0125	-3.8421	-2.7628
		(0.0038)	(2.5609)	(0.4764)
T60	-2.1825			
	(0.4361)			
T200	-2.6626			
	(0.4362)			
T500	-2.7567			
	(0.4362)			
VAR_RATIO	7.8930	-1.9911	2.7815	2.8684
	(1.2847)	(6.1541)	(0.7093)	(0.3862)
HETER	0.2834	-0.0501	-0.0390	-0.0109
	(0.4658)	(0.4294)	(0.1917)	(0.0553)
AUTO	0.3066	-0.0621	-0.0075	0.0158
	(0.4696)	(0.4392)	(0.2247)	(0.0487)
LAMBDA_SQ		-0.0536		
		(0.3266)		
T_SQ		2.77E-05		
		(6.57E-06)		
VAR_RATIO_SQ		27.4752		
		(10.2821)		
T*VAR_RATIO		-0.0306		
		(0.0049)		
LAMBDA*VAR_RATIO		-0.7552		
		(1.2916)		
T*LAMBDA		2.90E-04		
		(1.04E-03)		
MSPE_TP			-0.1901	-0.0436
			(0.0609)	(0.0104)
T_TP			-0.2174	-0.1471
			(0.1612)	(0.0379)
VAR_RATIO_TP			0.4148	0.4076
			(0.1807)	(0.0945)

Table 9

Response Function Results for Parameter 4 (Lambda) --- BC ---				
	OLS1	OLS2	NLS	BC
Dep. Mean	0.1576	0.1576	0.1576	
R-Squared	0.9389	0.8396	0.9887	
CONSTANT	0.3585	0.2663	4.6902	5.4388
	(0.0133)	(0.0394)	(0.9254)	(1.0694)
LAMBDA	-0.0042	-0.0094	-0.0381	-0.0242
	(0.0057)	(0.0203)	(0.0120)	(0.0173)
T		-0.0024	-2.0287	-2.8586
		(0.0002)	(0.6910)	(0.8241)
T60	-0.2229			
	(0.0113)			
T200	-0.3438			
	(0.0113)			
T500	-0.3709			
	(0.0113)			
VAR_RATIO	0.1838	1.0481	0.0044	0.0268
	(0.0358)	(0.3105)	(0.0035)	(0.0357)
HETER	0.0080	0.0221	0.0802	0.2209
	(0.0121)	(0.0209)	(0.0287)	(0.0707)
AUTO	-0.0252	-0.0082	0.0133	-0.0266
	(0.0122)	(0.0213)	(0.0390)	(0.0768)
LAMBDA_SQ		-0.0098		
		(0.0161)		
T_SQ		3.72E-06		
		((3.20E-07)		
VAR_RATIO_SQ		-1.2435		
		(0.5426)		
T*VAR_RATIO		-0.0009		
		(0.0003)		
LAMBDA*VAR_RATIO		0.0215		
		(0.0779)		
T*LAMBDA		2 2.10E-05		
		(5.09E-05)		
MSPE_TP			0.0325	0.1668
			(0.0929)	(0.0755)
T_TP			-0.1536	-0.2959
			(0.1155)	(0.0896)
VAR_RATIO_TP			-2.3324	-1.2126
			(0.3992)	(0.6836)

Table 10

Kolmogorov Statistics Testing for Normality
Using Sample Mean and Variance
--- Parameter 1 (Constant) ---

T	NLS Variance Structure 10 %	30 %	50 %	Heter.	Auto.
30	0.04246	0.03096	0.03357	0.02703	0.03210
60	0.02606	0.02494	0.03204	0.04376	0.03069
200	0.04184	0.04415	0.04335	0.04698	0.03049
500	0.02242	0.02402	0.03493	0.01868	0.02466

T	BC Variance Structure 10 %	30 %	50 %	Heter.	Auto.
30	0.03915	0.02992	0.04365	0.02221	0.02993
60	0.02463	0.02855	0.03128	0.04639	0.02995
200	0.03676	0.05661	0.04682	0.04023	0.03416
500	0.02264	0.01883	0.03681	0.03032	0.02974

Table 11

Kolmogorov Statistics Testing for Normality
Using Sample Mean and Variance
--- Parameter 2 ---

T	NLS Variance Structure 10 %	30 %	50 %	Heter.	Auto.
30	0.02102	0.02958	0.02228	0.03628	0.03278
60	0.02673	0.01989	0.03012	0.04696	0.04468
200	0.03835	0.03557	0.04354	0.03941	0.04992
500	0.02708	0.03011	0.03257	0.03490	0.02698

T	BC Variance Structure 10 %	30 %	50 %	Heter.	Auto.
30	0.02725	0.02221	0.01655	0.03239	0.03091
60	0.01935	0.01807	0.01744	0.03975	0.04596
200	0.03512	0.03565	0.03359	0.03687	0.05482
500	0.02476	0.03918	0.02777	0.03280	0.02865

Table 12

Kolmogorov Statistics Testing for Normality
Using Sample Mean and Variance
--- Parameter 3 ---

T	NLS Variance Structure: 10 %	30 %	50 %	Heter.	Auto.
30	0.02725	0.02221	0.01655	0.03239	0.03091
60	0.01935	0.01807	0.01744	0.03975	0.04596
200	0.03512	0.03565	0.03359	0.03687	0.05482
500	0.02476	0.03918	0.02777	0.03280	0.02865

T	BC Variance Structure: 10 %	30 %	50 %	Heter.	Auto.
30	0.02745	0.03074	0.04515	0.02874	0.03113
60	0.02094	0.01543	0.03159	0.03615	0.02541
200	0.07640	0.06839	0.03725	0.02143	0.02143
500	0.02249	0.02919	0.02087	0.03151	0.03105

Table 13

Kolmogorov Statistics Testing for Normality
Using Sample Mean and Variance
--- Parameter 4 (Lambda) ---

T	NLS Variance Structure: 10 %	30 %	50 %	Heter.	Auto.
30	0.06304	0.06937	0.86600	0.06712	0.06232
60	0.04095	0.04741	0.05735	0.02249	0.03489
200	0.03260	0.03199	0.02813	0.02437	0.02269
500	0.02825	0.04314	0.04870	0.03837	0.02428

T	BC Variance Structure: 10 %	30 %	50 %	Heter.	Auto.
30	0.04691	0.04415	0.02730	0.03637	0.05041
60	0.02964	0.02817	0.01855	0.03118	0.02831
200	0.03974	0.04375	0.05426	0.02648	0.03342
500	0.04869	0.03119	0.02249	0.02943	0.01698

Table 14

Average "Acceptance" Rates for Testing Null of Parameter 1 Equalling True Value

T	NLS				
	Variance Structure				
	10 %	30 %	50 %	Heter.	Auto.
30	0.926	0.924	0.906	0.922	0.896
60	0.916	0.938	0.942	0.914	0.906
200	0.898	0.916	0.920	0.930	0.916
500	0.914	0.956	0.968	0.934	0.946

T	BC				
	Variance Structure				
	10 %	30 %	50 %	Heter.	Auto.
30	0.958	0.962	0.950	0.932	0.936
60	0.928	0.968	0.964	0.928	0.930
200	0.910	0.910	0.916	0.934	0.924
500	0.918	0.952	0.964	0.934	0.950

Table 15

Average "Acceptance" Rates for Testing Null of Parameter 2 Equalling True Value

T	NLS				
	Variance Structure				
	10 %	30 %	50 %	Heter.	Auto.
30	0.912	0.922	0.930	0.912	0.876
60	0.922	0.938	0.938	0.934	0.892
200	0.912	0.930	0.936	0.920	0.886
500	0.920	0.938	0.950	0.914	0.924

T	BC				
	Variance Structure				
	10 %	30 %	50 %	Heter.	Auto.
30	0.944	0.952	0.956	0.928	0.906
60	0.928	0.948	0.958	0.936	0.924
200	0.902	0.924	0.948	0.916	0.894
500	0.924	0.942	0.956	0.918	0.930

Table 16

Average "Acceptance" Rates for Testing Null of Parameter 3 Equalling True Value

NLS — Variance Structure

T	10 %	30 %	50 %	Heter.	Auto.
30	0.918	0.928	0.922	0.900	0.866
60	0.898	0.922	0.916	0.890	0.898
200	0.898	0.916	0.924	0.926	0.910
500	0.908	0.928	0.938	0.918	0.956

BC — Variance Structure

T	10 %	30 %	50 %	Heter.	Auto.
30	0.948	0.956	0.942	0.918	0.912
60	0.916	0.930	0.924	0.912	0.908
200	0.910	0.928	0.930	0.928	0.914
500	0.918	0.934	0.944	0.926	0.954

Table 17

Average "Acceptance" Rates for Testing Null of Parameter 4 Equalling True Value

NLS — Variance Structure

T	10 %	30 %	50 %	Heter.	Auto.
30	0.920	0.902	0.902	0.938	0.920
60	0.940	0.942	0.898	0.918	0.950
200	0.954	0.946	0.924	0.978	0.938
500	0.950	0.926	0.924	0.954	0.954

BC — Variance Structure

T	10 %	30 %	50 %	Heter.	Auto.
30	0.976	0.986	0.992	0.966	0.976
60	0.970	0.972	0.980	0.936	0.966
200	0.928	0.888	0.916	0.866	0.936
500	0.952	0.926	0.932	0.696	0.942

Table 18

Response Function Calculations for R1

	NLS		BC	
	OLS1	OLS2	OLS1	OLS2
Dep. Mean	-8.4022	-8.4022	-7.3076	-7.3076
R-Squared	0.9360	0.9235	0.8535	0.9109
CONSTANT	-8.9756	-8.9906	-9.1099	-10.3598
	(0.2701)	(0.5500)	(0.3774)	(0.5400)
LAMBDA	-0.0491	-0.4225	-1.9820	-1.0098
	(0.1173)	(0.2660)	(0.1629)	(0.2775)
T		-0.0283		-0.0142
		(0.0026)		(0.0025)
T60	-2.2620		-1.5225	
	(0.2344)		(0.3223)	
T200	-4.0415		-2.0954	
	(0.2344)		(0.3225)	
T500	-4.9075		-2.1679	
	(0.2344)		(0.3225)	
VAR_RATIO	15.7160	14.2316	15.7238	18.8286
	(0.6903)	(4.1782)	(1.0179)	(4.2516)
HETER	0.0980	0.0279	0.4570	0.5862
	(0.2503)	(0.2915)	(0.3441)	(0.2859)
AUTO	0.2400	0.1654	-0.0454	0.0956
	(0.2523)	(0.2982)	(0.3467)	(0.2918)
LAMBDA_SQ		0.3296		1.3145
		(0.2217)		(0.2205)
T_SQ		4.06E-05		2.27E-05
		(4.46E-06)		(4.38E-06)
VAR_RATIO_SQ		6.3043		-2.0587
		(6.9808)		(7.4306)
T*VAR_RATIO		-0.0109		-0.0066
		(0.0033)		(0.0035)
LAMBDA*VAR_RATIO		1.4670		-0.9464
		(0.8769)		(1.0663)
T*LAMBDA		3.76E-04		-4.03E-03
		(7.05E-04)		(6.97E-04)

Table 19

Response Function Calculations for R2

	NLS		BC	
	OLS1	OLS2	OLS1	OLS2
Dep. Mean	-7.0645	-7.0645	-6.7345	-6.7345
R-Squared	0.8981	0.8947	0.8117	0.8851
CONSTANT	-7.0838	-6.5227	-7.5289	-8.3300
	(0.3282)	(0.6213)	(0.3874)	(0.5553)
LAMBDA	-0.1400	-0.3930	-1.2899	-0.3096
	(0.1426)	(0.3005)	(0.1672)	(0.2854)
T		-0.0297		-0.0206
		(0.0029)		(0.0026)
T60	-2.3833		-1.7689	
	(0.2848)		(0.3309)	
T200	-4.3214		-2.8961	
	(0.2849)		(0.3311)	
T500	-5.2120		-3.0615	
	(0.2849)		(0.3311)	
VAR_RATIO	13.6961	6.2502	13.1394	12.3189
	(0.8389)	(4.7203)	(1.0449)	(4.3721)
HETER	0.1969	-0.0102	0.3641	0.4169
	(0.3042)	(0.3293)	(0.3532)	(0.2940)
AUTO	0.3785	0.1448	0.0357	0.0877
	(0.3066)	(0.3369)	(0.3559)	(0.3001)
LAMBDA_SQ		0.3034		1.4503
		(0.2505)		(0.2267)
T_SQ		4.36E-05		3.17E-05
		(5.04E-06)		(4.50E-06)
VAR_RATIO_SQ		17.8932		5.5739
		(7.8865)		(7.6413)
T*VAR_RATIO		-0.0145		-0.0078
		(0.0037)		(0.0036)
LAMBDA*VAR_RATIO		0.8414		-1.3156
		(0.9907)		(1.0965)
T*LAMBDA		4.56E-04		-3.79E-03
		(7.96E-04)		(7.17E-04)

Appendix:

This appendix describes in detail the simulation methodology. All the simulations were done on a MicroVAX 3400 computer using the computer program TSP.

Simulation Steps:

1) Choose Error Structure (see below: Error Structure)
2) Estimate geometric mean with large sample approximation and --compute true parameter vectors (see below: Geometric Mean)
3) Choose sample size (T=30,60,200,500)
4) Choose true λ (λ=-1, -.5, 0, .5, 1)
5) Choose iteration number (i= 1 to 500)
6) Draw x's, errors, and compute y's (see below: --Simulation)
7) Calculate sample geometric mean; then normalize y's --(see below: Normalization)
8) Estimate BC and NLS (see below: Estimation)
9) Loop over steps 5 through 9 (i)
10) Loop over steps 4 through 10 (λ)
11) Loop over steps 3 through 11 (T)
12) Loop over steps 1 through 12 (Error structure)

Notes:

1) Error Structure: As described in the text, we choose 5 error structures, three corresponding to classical BC assumptions, two to misspecifications of the BC model (autocorrelation and conditional heteroskedasticity). Specifics follow:
   Case 1) e_t distributed N(0,0.002704)
   2) e_t distributed N(0,0.010201)
   3) e_t distributed N(0,0.024025)
   4) e_t distributed N(0,0.002704*(1+$x_t\beta$))
   x_{1t}=1; x_{2t} and x_{3t} distributed joint normal
   - $E(x_{2t})=2$, $E(x_{3t})=3$, $V(x_{2t})=0.04$, $V(x_{3t})=0.04$
   - $Cov(x_{2t},x_{3t})=0.028$
   - $\beta_1=\beta_2=-\beta_3=1$
   5) $e_t = 0.7e_{t-1} + v_t$
   - v_t distributed N(0,0.002704*(1-.49))
   - e_0 distributed N(0,0.002704).

2) Geometric Mean: The large sample approximation of the geometric mean was computed by taking the average geometric mean of 1000 samples of 900 observations each.
   1) Compute 900 y_t's from a given error structure for both the BC and the NLS models (x's drawn each time from population described in error structure 4 above).
   2) Compute geometric mean of that sample discarding any negative or missing y's.
   3) Repeat steps (1) and (2) 1000 times and take average value of geometric mean as the true population geometric mean.

4) Using the estimates of the population geometric mean compute true parameter values using equation (11) from the text.

6) Simulation: x's were drawn from a population described in case 4 of step 1 (Bivariate normal population). Error terms were drawn from the appropriate distributions (according to the error structure). Then y's were computed for the BC (equation (7)) and the NLS (equation (8a)) models using $\beta_1=\beta_2=-\beta_3=1$.

7) Normalization: Sample geometric means for both the BC and the NLS models were computed discarding any missing or nonpositive y's. Then the original y's were normalized by the appropriate geometric means.

8) Estimation: The BC model was estimated by maximum likelihood using the true parameter values as the starting values. The NLS estimation also used the true parameters as starting values. Both BC and NLS procedures were allowed a maximum of 500 iterations to converge otherwise the observations would be discarded.[9] Berndt, Hall, Hall, Hausman estimates of the standard errors were used in the BC model while Gaussian standard errors were used in the NLS model.

Simulation took approximately three weeks of computer time.

[9]This occurred 3 times out of 100,000 estimations.

References

Amemiya,T. and J.L. Powell (1981), "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-Linear Two Stage Least Squares Estimator," Journal of Econometrics, 17, 351-381.

Berndt, E.R. M.H. Showalter, and J.M. Wooldridge (1990a), "A Theoretical and Empirical Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative," MIT Sloan School of Management working paper #3187-90-EFA (August) 1990.

Berndt, E.R. M.H. Showalter, and J.M. Wooldridge (1990b), "Hedonic Price Indexes for Computers and Peripherals: Box-Cox Formulations and a Nonlinear Least Squares Alternative," mimeo, MIT, (July).

Box, G.E.P. and D.R. Cox (1964), "An Analysis of Transformations," Journal of the Royal Statistical Society, Series B, 26, 211-252.

Dinan, T.M. and J.A. Miranowski (1989), "Estimating the Implicit Price of Energy Efficiency Improvements in the Residential Housing Market: A Hedonic Approach," Journal of Urban Economics, 25, 52-67.

Donald, A. and A. Donner (1990), "A Simulation Study of the Analysis of Sets of 2x2 Contingency Tables Under Cluster Sampling: Estimation of Common Odds Ratio," Journal of the American Statistical Association, 85, 537-543.

Draper, N.R. and D.R.Cox (1969), "On Distributions and Their Transformations to Normality," Journal of the Royal Statistical Society, Series B, 30, 472-476.

Greene, W.H. (1990) Econometric Analysis, New York: Macmillan Publishing Co.

Heckman J. and S. Polachek (1974), "Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship," Journal of the American Statistical Association, 69, 350-354.

Hendry, D.F. (1984) "Monte Carlo experimentation in econometrics," Chapter 16 in Z. Griliches and M.D. Intriligator (eds.) Handbook of Econometrics, Amsterdam, North-Holland Publishing Company, Volume II, 937-976.

Lindgren, B.W. (1976) Statistical Theory, Third Edition, New York: Macmillan Publishing Co.

Lo, A.W. and A.C. MacKinlay (1989), "The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation," Journal of Econometrics, 40, 203-238.

Poirier, D.J. (1978), "The Use of the Box-Cox Transformation in Limited Dependent Variable Models," Journal of the American Statistical Association, 73, 284-287.

Seaks, T.G. and S.K. Layson (1983) "Box-Cox Estimation with Standard Econometric Problems," The Review of Economics and Statistics, 65, 160-164.

Spitzer, J.J. (1984), "Variance Estimates in Models with Box-Cox Transformations: Implications for Estimation and Hypothesis Testing," Review of Economics and Statistics, 66, 645-652.

Spitzer, J.J. (1978), "A Monte Carlo Investigation of the Box-Cox Transformation in Small Samples," Journal of the American Statistical Association, 73, 488-495.

Wooldridge, J.M. (1990), "Some Alternaltive to the Box-Cox Regression Model," mimeo, MIT Department of Economics, January.

Essay III:

Monopoly Behavior with Intertemporal Demands

by

Mark H. Showalter

Abstract:

This paper investigates optimal monopoly behavior when demand for the monopolist's product is related over time. The product is called an intertemporal good and the resulting demand function is referred to as an intertemporal demand. The theoretical section of the paper shows that it might be optimal for a monopolist to produce in the inelastic portion of demand contrary to static monopoly theory. A second result is that regardless of whether the consumer is forward-looking or not, the observed time path of consumption will be related to future variables due to the forward-looking behavior of the monopolist. The paper continues by deriving four alternative models for consumer and monopoly behavior and then proceeds to estimate the models with data on cigarette consumption and prices. The results strongly suggest that firms are forward-looking while the evidence that consumers are forward-looking is less persuasive.

1. Introduction

Recently, intertemporal utility functions and the resulting intertemporal demand functions have seen an increasing amount of use in economics. Probably the most active research has been done in the empirical and theoretical finance literature where, for example, Huang and Kreps (1987) argue for the use of intertemporal utilities (time-nonseparable) in the widely used continuous-time models of consumption since it is unreasonable to assume that consumption at one instant is unaffected by consumption in an adjacent instant. Other authors (for example, Constantinides (1990)) have suggested that allowing for intertemporal utilities might help explain the famous "Equity Premium Puzzle" of Mehra and Prescott (1985).[1]

Other fields in economics have begun using intertemporal utilities in novel and interesting ways. In a provocative article, Becker and Murphy (1988, hereafter BM) build a "rational addict" model using a simple intertemporal utility function and then derive various implications, some of which are tested in an empirical companion piece by Becker, Grossman and Murphy (1990). Their work will be discussed at some length in this paper.

One unifying element of the research to date is the assumption of a perfectly competitive market in the supply of the consumption good. This assumption has proven very useful in allowing researchers to compute a time path for the consumption good and in providing estimable equations from which various hypotheses can be tested. But an equally interesting set of issues and predictions emerge when this perfect competition

---

[1]An excellent introduction to these and related topics is given in Heaton (1989).

assumption is relaxed. Demand will still be related over time owing to the consumer's intertemporal utility, but now the question arises: How will firms optimally account for these intertemporal dependencies? How will firms set their control variables, such as price and advertising? How will the nature of interfirm competition be affected by these intertemporal linkages? The usefulness of investigating these and related questions about firm behavior in the presence of intertemporal demands is suggested from any number of demand studies which find that lagged consumption is an important predictor of current consumption (see, for example, Houthakker and Taylor (1970)) suggesting that, in fact, intertemporal dependencies do exist and might be empirically important. The strength of these lags has been rationalized by explanations ranging from habit behavior and partial adjustment to simple ignorance on the part of the researcher; but regardless of the explanation, firms supplying the good have a strong incentive to account for these intertemporal linkages.

Becker, Grossman, and Murphy (1990, hereafter BGM) give one example of how intertemporal demands can affect firm behavior. BGM present a simple two-period model where demand for a good is related across time and supply is controlled by a classic monopolist. They show--counter to textbook monopoly theory--that the monopolist might in fact produce in the inelastic portion of the demand curve. As will be shown in this paper, the insight holds true for the multi-period monopolist model as well. But this has important implications for much of the New Empirical Industrial Organization (NEIO) literature which attempts to identify and quantify market power from comparative static results based on simple, static, monopoly or oligopoly models.

Examples of this literature are Bresnahan (1989), Panzer and Rosse (1987), Ashenfelter and Sullivan (1987), Sumner (1983), and Applebaum (1982). The market power indices these type of papers attempt to measure can alternatively be interpreted as measuring the degree of forward-looking behavior on the part of a monopolistic firm instead of the intended measure of oligopolist competitive behavior.

Another possible effect of intertemporal demands is that current measures of returns to advertising (and possibly other activities) could be biased downward, possibly seriously so. For example, in models with intertemporal demands, firms will be receiving a return on current advertising for possibly several future periods. This would suggest that studies looking only at current returns to advertising (as has been the practice) would seriously misstate the actual returns the firms expect and plan for.

The cigarette industry is, not surprisingly, probably the best example of this type of market. For years economists have puzzled over the extraordinary amount of advertising that the cigarette industry uses. Measured static returns to advertising (measured as the change in current consumption due to current advertising) are substantially below "rational" levels and economists have generally assumed that the advertising is simply used in interfirm competition.[2] But if, in fact, demand is intertemporally linked, as is unquestionably the case with cigarettes, then firms could conceivably be willing to spend a large amount on advertising (in the current period) to attract a few new customers in order to receive revenues from those newly "hooked" consumers in the future. The same argument could also be applied to many pharmaceutical drugs.

---

[2]See, for example, Baltagi and Levin (1986) page 153.

In a similar vein, if we think of price as the control variable, it is easy to think of examples where firms might consider pricing below marginal cost--dumping or preying in the economic vernacular--for some period in order to receive increased revenues in the future. Note that this is not the case of predatory behavior where a firm tries in the current period to drive rivals out of the market in order to enjoy a monopoly in the future. In a model with intertemporal demands, it results from a firm attempting to expand the market by attracting new customers and then recouping the loss from the initial period by an increase in profits in later periods from the "hooked" consumers.

As the preceding discussion has indicated, the study of firm behavior in the presence of intertemporal demands has many interesting applications in several areas of economics. Depending on the application, consumers might be modeled as rational and forward-looking, as has been done in the finance literature and in the work of BM and BGM; or, alternatively, intertemporal utilities "justify" the assumption of myopic habit behavior. In the empirical section of this paper I develop a framework to test rational versus myopic consumer behavior as well as rational versus myopic monopoly behavior. The framework is then applied to data on cigarette consumption. Cigarettes is the ideal consumer good to use for this particular test both because the good obviously has intertemporal qualities to its use and because cigarette manufacturing is a highly concentrated industry. Although several previously cited economic studies find little evidence of monopoly power in the cigarette industry, one of the key points of this paper is that those studies are fundamentally flawed due to the intertemporal nature of the demand for cigarettes.

The organization of the paper is as follows: In the next section I formalize the above discussion by presenting and analyzing a simple model of monopoly behavior in the presence of intertemporal demands. Following that, in section 3, I develop a framework to test empirically for rational behavior on the part of a monopolist and consumers. In section 4 I present and discuss the estimation results and in section 5 I present a brief conclusion.

2. A Model of Monopoly Behavior

The ideas presented in the introduction can be illustrated with a straightforward example. Initially, assume the existence of an economy with one individual with a concave utility function of the form $U(Q_t,Q_{t-1},Z_t)$ where Q_t and Q_{t-1} are complements. Further, suppose that the consumer discounts the future at an infinite rate so that her optimization problem only accounts for the current and the past but not the future. The optimization problem in period t with an income Y_t and Z_t as the numeraire good then becomes

$$\max_{\{Q_t,Z_t\}} U(Q_t,Q_{t-1},Z_t) \quad \text{such that } Y_t=Z_t+P_tQ_t \tag{2.1}$$

where P_t is the price of the intertemporal good. Suppose for simplicity I take the utility function as separable in Q_t and Z_t. Then the first order conditions are calculated as:

$$Q_t: \quad \frac{\partial U}{\partial Q_t}(Q_t,Q_{t-1})-\lambda P_t=0 \tag{2.2}$$

$$Z_t:\quad \frac{\partial U}{\partial Z_t}(Z_t)-\lambda=0 \tag{2.3}$$

$$\lambda:\quad Y_t - Z_t - P_tQ_t=0 \tag{2.4}$$

Substituting (2.4) into (2.3) and then (2.3) into (2.2) I then get

$$\frac{\partial U}{\partial Q_t}(Q_t,Q_{t-1})=\frac{\partial U}{\partial Z_t}(Y_t-P_tQ_t)P_t \tag{2.5}$$

which implicitly defines a demand for Q_t in terms of P_t, Y_t and Q_{t-1}. Using the implicit function theorem, Appendix A shows that the following conditions hold:

$$\frac{\partial Q_t}{\partial P_t}<0\ ,\quad \frac{\partial Q_t}{\partial Y_t}>0\ ,\quad \frac{\partial Q_t}{\partial Q_{t-1}}>0. \tag{2.6}$$

Write this demand function as

$$Q_t=Q(P_t,Y_t,Q_{t-1}) \tag{2.7}$$

Now suppose the supply of good Q_t is controlled by a monopolist who maximizes the present value of profits with respect to price and further that the monopolist has constant marginal costs of production, c. The monopolist's problem is then

$$\underset{\{P_t\}_{t=1}^{\infty}}{Max}\quad \sum_{t=1}^{\infty}\beta^{t-1}(P_t-c)Q(P_t,Y_t,Q_{t-1}) \tag{2.8}$$

where β is the one period discount rate and Q_0 is given. The first order condition for an arbitrary P_t will then be

$$\beta^{t-1}\left[Q_t+(P_t-c)\frac{\partial Q_t}{\partial P_t}+D_t\right] = 0 \tag{2.9}$$

where

$$D_t=\sum_{i=t+1}^{\infty}\beta^{i-t}(P_i-c)\left(\prod_{j=t+1}^{i}\frac{\partial Q_j}{\partial Q_{j-1}}\right)\frac{\partial Q_t}{\partial P_t} \tag{2.10}$$

D_t is the key to this intertemporal problem; it represents the cumulative discounted effect of a change in price P_t on all future periods.

It is instructive to solve for the markup equation implied by equation (2.9):

$$\frac{P_t-c}{P_t}=\left(-\frac{1}{\epsilon_t}\right)*\left(1+\frac{D_t}{Q_t}\right) = \left(\epsilon_t*\left(\frac{Q_t}{Q_t+D_t}\right)\right)^{-1} \tag{2.11}$$

where

$$\epsilon_t\equiv\frac{\partial Q_t}{\partial P_t}\frac{P_t}{Q_t} <0 \tag{2.12}$$

If the solution to the entire system implies price above marginal cost, then from the concavity assumptions I get $D_t < 0$. This in turn implies that the monopolist will set his markup lower than he would in the time-separable case where $D_t = 0$. This sustains the result suggested by BGM that a monopolist might set prices in the inelastic portion of the demand curve due to intertemporal factors, thus contradicting the elementary textbook characterization of monopoly behavior.

Also note that equation (2.11) has a form similar to that used in many of the papers in the NEIO literature. This literature equates the markup to the inverse of the product of the demand elasticity and a parameter which is interpreted as a measure of market power. The parameter theoretically varies from 1 (a monopoly or perfect cartel) to positive infinity (perfect competition). Examples of this literature are Bresnahan (1989) and Sullivan (1985). However, as (2.11) makes clear, this estimated parameter can also be interpreted as measuring the optimal value of $Q_t/(Q_t + D_t)$ for a monopolist, not market power for possibly oligopolist firms as was the intended purpose. Without further specification, the hypothesis of market power as given in the NEIO literature and an alternative of a forward-looking monopolist facing intertemporal demands are indistinguishable.

In a similar vein, methods which attempt to assess monopoly behavior by using predictions from the static monopoly model will likewise be misleading (for examples of this type of paper see Panzer and Rosse (1987) and Ashenfelter and Sullivan (1987)). These methods start from the premise that monopolists will not operate in the inelastic portion of demand. From this various testable hypotheses can be generated. But if a monopolist operates in the inelastic region of demand for intertemporal reasons, the above premise will be false.

Another interesting implication of the model emerges from (2.9). At the optimum, equation (2.9) will be satisfied for all t. But this implies that P_t (and hence Q_t) will be related to past and <u>future</u> prices. This occurs even in the case of an "irrational" myopic consumer modeled above. The forward-looking time path of prices and consumptions occurs solely due to the optimization behavior of the monopolist. This suggests an alternative explanation for the results in BGM. They find the "rational addict" model supported by their empirical results. What the above model suggests is that the intertemporal optimization of a monopolist with myopic consumers will generate a similar and possibly indistinguishable pattern from the "rational addict" model. Given that the optimization incentives are probably stronger for the firm than for consumers, it seems probable that, in fact, the BGM results come more from the supply side rather than the demand side as hypothesized by BGM. This conjecture is lent support in the empirical section of this paper.

In this section I have discussed several interesting implications of the monopoly model with intertemporal demands. Although general properties of the solution can be characterized from the first order conditions, a closed form solution to the monopolist's maximization problem requires additional parameterization. The following section continues by developing a framework for testing rational versus myopic behavior both on the part of the consumer and on the part of a monopolist.

3. Alternative Behavioral Models

In this section of the paper I develop four alternative behavioral models for interaction between a consumer and a monopolist, assuming full information. Both agents can act either "rationally", meaning that the agent takes the future into account in determining optimal current actions, or, alternatively, each can act "myopically" where the future plays no role in current optimization decisions. The four models are then:

1) Myopic Consumer--Myopic Monopolist

2) Myopic Consumer--Rational Monopolist

3) Rational Consumer--Myopic Monopolist

4) Rational Consumer--Rational Monopolist.

The basic strategy for solving each of the four models is first to posit a consumer behavior, either rational or myopic, quantified as a demand equation and then solve a monopoly optimization problem which varies both with the monopolist's optimization horizon (current (myopic) or current and future (rational)) and with the demand equation of the consumer. The end result is a two-equation system for each of the four models which can then be estimated. After developing the models, I outline the actual estimation strategy and the associated assumptions.

3.1 Myopic Consumer--Myopic Monopolist (Model 1)

This is the simplest of the four models to be considered and readily illustrates the general methodology used to solve each of the models. The consumer is modeled as a Stackelberg follower taking prices as given while the monopolist is modeled as a Stackelberg leader who optimizes expected profits with respect to price. As will become clear in the development of Models 2 and 4, a simple specification of the demand function is essential in order to obtain the solution to the monopoly optimization problem. For that reason the myopic consumer is assumed to have a linear demand of the form,

$$Q_t = \alpha_1 P_t + \alpha_2 Y_t + \alpha_3 Q_{t-1} + e_t$$
$$\alpha_1 < 0 \ , \ \alpha_2 > 0 \ , \ 0 < \alpha_3 < 1 \tag{3.1}$$

where Q_t is quantity, P_t is price, Y_t is income, and e_t is an iid stochastic term known to the consumer at time t but unknown at time t both to the monopolist and to the econometrician. The monopolist has a time t expectation of e_t of e^m while the econometrician has time t expectation of e_t of zero. Equation (3.1) and the rational demand to be introduced in Model 3 are similar to those used in BGM.

The myopic monopolist ignores the intertemporal linkages and simply maximizes expected profits,

$$\underset{P_t}{Max} \quad (P_t - c_t) E[Q_t | I_t^m] \tag{3.2}$$

for each time t where I_t^m is the information set of the monopolist at time t and is assumed to contain c_t, Y_t and Q_{t-1}. This expected profit function is strictly concave in P_t ($\alpha_1 < 0$) so the solution to the first-order condition will guarantee a maximum. The optimal price for this problem is then

$$P_t^* = \frac{1}{2} c_t - \frac{1}{2} \frac{\alpha_2}{\alpha_1} Y_t - \frac{1}{2} \frac{\alpha_3}{\alpha_1} Q_{t-1} - \frac{1}{2\alpha_1} e^m. \tag{3.3}$$

The econometrician is assumed to have a zero expectation of e^m. The optimal price is seen to be increasing in c_t, Y_t and Q_{t-1}. Equations (3.1) and (3.3) characterize Model 1.

3.2 Myopic Consumer--Rational Monopolist (Model 2)

The myopic consumer-rational monopolist model was used extensively in section 2 to outline the implications of intertemporal demands. Here, I parameterize the general specification of section 2 in a form similar to Model 1. Demand is again modeled as linear,

$$Q_t=\alpha_1 P_t+\alpha_2 Y_t+\alpha_3 Q_{t-1}+e_t, \tag{3.4}$$

with definitions and parameter restrictions the same as in Model 1. The monopolist, however, now solves the intertemporal problem

$$\underset{\{P_{t+i}\}_{i=0}^{\infty}}{Max} \sum_{i=0}^{\infty} \beta^i (P_{t+i}-c_{t+i}) E[Q_{t+i}|I_t^m] \quad 0<\beta<1 \tag{3.5}$$

The associated first-order condition for an arbitrary P_{t+i} is

$$\beta^i\left[E[Q_{t+i}|I_t^m]+\alpha_1\sum_{j=0}^{\infty}(\beta\alpha_3)^j(P_{t+i+j}-c_{t+i+j})\right] = 0 \tag{3.6}$$

Equations (3.4) and (3.5) together form what is known in the dynamic optimization literature as a stochastic linear optimal regulator problem which is characterized by a solution where the control variable (P_t in this model) is a linear function of the state variables (c_t, Y_t and Q_{t-1}).[3] However, the linear functions in general depend upon an iterated matrix which is impossible to characterize without knowing the parameter values ($\alpha_1,\alpha_2,\alpha_3,\beta$) a priori. As an alternative to the general linear regulator problem, I propose

[3]Sargent (1987; pp. 36-40) gives an excellent exposition of this type of problem.

solving a "knife-edge" special case where, using the property that these type of problems have a unique solution, I construct the problem such that the monopolist faces the same problem for each time period. I therefore assume constant real marginal costs, $c_t=c$, constant real incomes, $Y_t=Y$, $E[e_{t+i}|I_t^m] = e^m$ for $i>0$, and, in equilibrium, the monopolist rationally expects future quantities to equal Q_{t-1}. This implies that $E[Q_{t+i}|I_t^m] = \alpha_1 P_{t+i} + \alpha_2 Y + \alpha_3 Q_{t-1} + e^m$ for all i greater than zero which in turn implies that (3.6) has the same form for all P_{t+i}. This leads to the obvious solution of constant prices,

$$\alpha_1 P+\alpha_2 Y+\alpha_3 Q_{t-1}+e^m+\alpha_1(P-c)\sum_{j=0}^{\infty}(\alpha_3\beta)^j = 0. \tag{3.7}$$

With the assumption that $0<\alpha_3\beta<1$ so that the infinite sum is convergent, this can be solved for price in terms of c, Y and Q_{t-1},

$$P^* = \left(\frac{1}{2-\alpha_3\beta}\right)c - \left(\frac{\alpha_2(1-\alpha_3\beta)}{\alpha_1(2-\alpha_3\beta)}\right)Y - \left(\frac{\alpha_3(1-\alpha_3\beta)}{\alpha_1(2-\alpha_3\beta)}\right)Q_{t-1} - \left(\frac{(1-\alpha_3\beta)}{\alpha_1(2-\alpha_3\beta)}\right)e^m. \tag{3.8}$$

Note that if $\beta=0$ (the monopolist ignores the future) equation (3.8) reduces to (3.3) making Model 1 a special case of Model 2 (when income, costs, and quantities are constant). It is also seen that since $\alpha_3\beta<1$, price is an increasing function of c, Y, Q_{t-1}. Also note that the markup over marginal costs will be lower than in the myopic case since in this intertemporal setting a monopolist has an incentive to lower prices in the current period in order to gain profits in future periods. Equations (3.4) and (3.8) then form the basis for estimating Model 2.

3.3 Rational consumer--Myopic Monopolist (Model 3)

This model is similar to Model 1 and is relatively straightforward. Demand is linear of the form

$$Q_t=\alpha_1 P_t+\alpha_2 Y_t+\alpha_3 Q_{t-1}+\alpha_4 Q^c_{t+1}+e_t \tag{3.9}$$

where Q^c_{t+1} is the consumer's rational forecast of Q_{t+1}. Both the monopolist and the econometrician are assumed to have rational, although not necessarily identical, forecasts of Q^c_{t+1}. The monopolist is assumed to have the same expectation, e^m, for all e_{t+i}. In the same manner that the myopic demand incorporated all past values of income and prices through the lagged term Q_{t-1}, (3.9) incorporates consumer's expectations of all future prices and income through the lead term, Q^c_{t+1}.

The monopolist, on the other hand, is only concerned about the present and given information at time t solves

$$\underset{P_t}{Max} \quad (P_t-c_t)E[Q_t|I_t^m] \tag{3.10}$$

where Q_t is defined as in (3.9). Again the problem is concave in P_t and the resulting price equation is

$$P_t^* = \frac{1}{2}c_t - \frac{1}{2}\frac{\alpha_2}{\alpha_1}Y_t - \frac{1}{2}\frac{\alpha_3}{\alpha_1}Q_{t-1} - \frac{1}{2}\frac{\alpha_4}{\alpha_1}Q_{t+1} + \zeta_t$$

$$\zeta_t \equiv -\frac{1}{2\alpha_1}e^m - \frac{1}{2}\frac{\alpha_4}{\alpha_1}\eta_t^m, \qquad (3.11)$$

where η_t^m is the monopolist's forecast error of Q^c_{t+1}. As would be expected, if $\alpha_4 = 0$ (no forward-looking component of demand), (3.11) reduces to the myopic monopolist FOC of Model 1 (equation (3.3)). (3.9) and (3.11) then characterize Model 3.

3.4 Rational Consumer--Rational Monopolist (Model 4)

The final case models both consumers and the monopolist as rational. Demand has the same form as in Model 3,

$$Q_t = \alpha_1 P_t + \alpha_2 Y_t + \alpha_3 Q_{t-1} + \alpha_4 Q_{t+1} + e_t, \qquad (3.12)$$

but unlike Model 3, the rational monopolist accounts for the intertemporal behavior of the consumer in setting the time path of prices. The setup of this rational monopoly model will be slightly different from the one introduced in Model 2. An examination of the FOC of the monopolist assuming demand as in (3.12) will explain why a change is necessary.

Using demand as defined in (3.12) and computing a first-order condition for an arbitrary P_{t+i} (FOC for (3.5)) results in

$$\beta\left[E[Q_{t+i}|I_t^m]+(P_{t+i}-c_{t+i})\alpha_1+\alpha_1\sum_{j=1}^{t+i-1}\left(\frac{\alpha_4}{\beta}\right)^j(P_{t+i-j}-c_{t+i-j})+\alpha_1\sum_{j=1}^{\infty}(\beta\alpha_3)^j(P_{t+i+j}-c_{t+i+j})\right] = 0. \quad (3.13)$$

Define the quantities B_{t+i} and D_{t+i} as follows:

$$B_t \equiv \alpha_1\sum_{i=1}^{t-1}\left(\frac{\alpha_4}{\beta}\right)^i(P_{t-i}-c_{t-i}) \quad (3.14)$$

$$D_t \equiv \alpha_1\sum_{i=1}^{\infty}(\beta\alpha_3)^i(P_{t+i}-c_{t+i}). \quad (3.15)$$

where D_{t+i} accounts for the monopolist's forward-looking behavior and B_{t+i} represents the "backward-looking" component where the monopolist accounts for the consumer's forward-looking behavior. If α_4 equals 0, then B_{t+i} will equal 0.

This model arbitrarily starts at date 0 and B_{t+i} will hence have a differing number of elements (indexed by i) for each t+i. If t+i=1 B_{t+i} will contain no elements. If t+i=2 there will be 1. For t+i=3 there will be 2 and so on. The convenient symmetry of the FOC in Model 2 where the FOC for each P_{t+i} looks the same is lost in this case. To regain the symmetry, for this model I assume a doubly infinite horizon, i ranging from negative to positive infinity, which again restores the symmetry.

Including this additional assumption along with the assumptions used to solve Model 2 (constant costs, incomes, and expectations) and also assuming $0<(\alpha_4/\beta)<1$, equation (3.13) can be solved for P as

$$P^* = \left(\frac{(V_1+V_2+1)}{(V_1+V_2+2)}\right)c - \left(\frac{\alpha_2}{\alpha_1(V_1+V_2+2)}\right)Y - \left(\frac{\alpha_3}{\alpha_1(V_1+V_2+2)}\right)Q_{t-1} - \left(\frac{\alpha_4}{\alpha_1(V_1+V_2+2)}\right)Q_{t+1} + \zeta_t \tag{3.16}$$

$$V_1 \equiv \frac{\alpha_4}{\beta - \alpha_4} \tag{3.17}$$

$$V_2 \equiv \frac{\beta\alpha_3}{1-\beta\alpha_3} \tag{3.18}$$

where ζ_t is the zero expectation (to the econometrician) error term subsuming forecast errors on Q_{t-1}, Q_{t+1} and e_t. Note that the assumption of a doubly infinite horizon would not change any of the conditions in Models 1, 2 or 3.

3.5 Dynamics and Elasticities

From a public policy perspective, probably the most important and interesting number from the host of cigarette studies to date is the own-price elasticity of demand. The intertemporal demand equations used in this study, in particular the forward-looking "rational" demands of Models 3 and 4, offer interesting and informative variations on the standard elasticity calculations. These variations can be motivated by examining the solutions to the difference equations embodied in the two demand equations. The myopic demand is a simple first order difference equation whose general solution is

$$Q_t=\sum_{i=0}^{\infty} \alpha_3^i(\alpha_1 P_{t-i}+\alpha_2 Y_{t-i})+c_1\alpha_3^t \tag{3.19}$$

where c_1 is an arbitrary constant term.[4] If time starts at date 1 and Q_0 is given then the solution to the problem is

$$Q_t=\sum_{i=0}^{t-1} \alpha_3^i(\alpha_1 P_{t-i}+\alpha_2 Y_{t-i})+Q_0\alpha_3^t \tag{3.20}$$

These equations imply that only past and current prices enter into current period demand; any change in future prices will leave current demand unaffected.

This contrasts sharply with the rational demands of Models 3 and 4. The rational demand is a second order difference equation in Q_t whose general solution is

$$\begin{aligned} Q_t=&(1-4\alpha_3\alpha_4)^{-1/2}\sum_{i=1}^{\infty} \lambda_1^i(\alpha_1 P_{t-i}+\alpha_2 Y_{t-i}) \\ &+(1-4\alpha_3\alpha_4)^{-1/2}\sum_{i=0}^{\infty} \lambda_2^{-i}(\alpha_1 P_{t+i}+\alpha_2 Y_{t+i})+c_1\lambda_1^t+c_2\lambda_2^{-t} \\ &t=\ldots,-1,0,1,\ldots \end{aligned} \tag{3.21}$$

where c_1 and c_2 are arbitrary constants and λ_1 and λ_2 are inverses of the roots of the characteristic equation

$$\alpha_3 z^2-z+\alpha_4=0. \tag{3.22}$$

and are equal to

[4] This section ignores the error term, e_t, used in the previous section. Inclusion would unnecessarily complicate an already burdensome notation without adding anything to the discussion since the elasticities calculated later in the text are identical to those based on the expectation of Q_t with respect to the time t information set of the econometrician.

$$\lambda_1 \equiv \frac{2\alpha_3}{1+(1-4\alpha_3\alpha_4)^{1/2}},$$
$$\lambda_2 \equiv \frac{2\alpha_3}{1-(1-4\alpha_3\alpha_4)^{1/2}}. \tag{3.23}$$

The roots will be real if and only if $(1\text{-}4\alpha_3\alpha_4) \geq 0$ and the system will be stable if $|\lambda_1|,|1/\lambda_2| < 1$ and $c_1=c_2=0$.

Also implicit in (3.21) is the fact that when a price changes, <u>all</u> quantities adjust; past, present and future. Perhaps a more reasonable specification would constrain past quantities to remain unchanged when, for instance, P_t changes, forcing any adjustments to affect only current and future consumption. This is the purpose of equation (3.24) where these type of questions can be answered by analysis at t=1, with Q_0 fixed.

$$\begin{aligned} Q_t = {} & (1-4\alpha_3\alpha_4)^{-1/2}\sum_{i=1}^{t-1} \lambda_1^i(\alpha_1 P_{t-i}+\alpha_2 Y_{t-i}) \\ & +(1-4\alpha_3\alpha_4)^{-1/2}\sum_{i=0}^{\infty} \lambda_2^{-i}(\alpha_1 P_{t+i}+\alpha_2 Y_{t+i})+c_1(\lambda_1^t-\lambda_2^{-t}) \\ & \lambda^{-t}\left(Q_0-(1-4\alpha_3\alpha_4)^{-1/2}\sum_{i=1}^{\infty} \lambda_2^{-i}(\alpha_1 P_{t+i}+\alpha_2 Y_{t+i})\right) \end{aligned} \tag{3.24}$$

The major distinction between the solutions to the rational demand ((3.21) and (3.24)) and the solutions to the myopic demand ((3.19) and (3.20)) is that Q_t in the rational demand depends upon the full sequence of prices--past, present and future price changes will change current demand; while Q_t in the myopic demand is unaffected by future price changes. It follows that for rational demand, in contrast to myopic demands, a temporary price change will have a different impact from a permanent price change.

Table 1

Various measures of the change in current quantity

	Demand	
Elements held fixed:	Myopic	Rational
Temporary Price Change:		
1) Q_{t-1}, Q_{t+1}, P_i $i \neq t$	α_1	α_1
2) P_i, Q_i $i \neq t$ (Anticipated)	α_1	$\alpha_1(1-4\alpha_3\alpha_4)^{-1/2}$
3) P_i $i \neq t$, Q_i $i < t$ (Surprise)	α_1	$\alpha_1(1-4\alpha_3\alpha_4)^{-1/2}(1-\lambda_2^{-2})$
Permanent Price Change:		
4) P_i $i < t$, Q_i $i \neq t$ (Anticipated)	α_1	$\alpha_1(1-4\alpha_3\alpha_4)^{-1/2}(\lambda_2/(\lambda_2 - 1))$
5) P_i, Q_i $i < t$ (Surprise)	α_1	$\alpha_1(1-4\alpha_3\alpha_4)^{-1/2}(\lambda_2-\lambda_2^{-1})/(\lambda_2 - 1)$

Table 1 gives the 5 different measures upon which I will base my elasticity calculations. The first measure is derived directly from the demand equations (3.1) and (3.9) as $\partial Q_t/\partial P_t$. Measure 2 is $\partial Q_t/\partial P_t$ using (3.19) and (3.21) and is interpreted as the fully anticipated effect of a change in P_t on Q_t. Measure 3 is $\partial Q_1/\partial P_1$ using (3.20) and (3.25) and is interpreted as the effect of a surprise change in P_t on Q_t. Measure 4 is the fully anticipated effect of a permanent change in price and measure 5 is the analogous effect of a "surprise" permanent change in price holding Q_0 fixed. Both are computed as dQ_t/dP where dP is the differential change in prices from time t to infinity. Specifically measure 4 uses (3.19) for the myopic and (3.21) for the rational demand. Measure 5 uses (3.20) and (3.25) evaluated at t=1. Having developed the mathematical framework I now turn to estimation issues.

3.6 Data and Estimation

The data for this study comes largely from the widely used resource book, The Tax Burden on Tobacco, published by the Tobacco Institute, an industry trade group. The data is well described in Sumner (1981) and Sullivan (1987) and I will only give a brief overview here. The annual publication provides information by state and year (from 1955 to 1988 for most states) on: 1) average per capita consumption in packs of cigarettes, 2) average price per pack, and 3) average excise tax (state and federal) per pack. Additionally, I collected per capita disposable income for each of the states from 1956 to 1988 from various issues of the Survey of Current Business. From various issues of the Economic Report of the President I constructed a fiscal year price index along the lines suggested by BGM to convert incomes and prices into real (1967) terms. A detailed explanation of the data is found in Appendix B. Following Sumner (1981) and Sullivan (1987) among others, I assume a no arbitrage condition between states; i.e., there is no cross-border smuggling of cigarettes.[5] Using the assumption of no arbitrage and thus independent markets, I can use Sumner's clever insight that excise taxes can be considered an additive component of (unobservable) marginal costs, thus identifying the parameters on marginal cost. Notationally, consider price as a function of c_t and x_t,

[5]Sumner details the legal restrictions that make this a plausible assumption although there exists evidence that for some states this might not be a reasonable approximation (see Doran (1979) and BGM). Preliminary investigation on my part showed little evidence of cross-border smuggling for my data although further study is needed.

$$P_t=\theta_1 c_t+X_t\theta_2+\epsilon_t \tag{3.25}$$

where c_t is unobservable marginal cost, X_t in my models includes income, lagged quantity and possibly lead quantity, and ϵ_t is a random disturbance. Sumner's insight is that if c_t has the following form:

$$c_t=c_t^*+ex_t \tag{3.26}$$

where c_t^* is the unobservable component of marginal costs and ex_t is the observable excise tax and c_t^* is statistically independent of ex_t and X_t, then P_t can be written as

$$P_t=\theta_1 ex_t+X_t\theta_2+v_t \tag{3.27}$$

$$v_t \equiv \epsilon_t+\theta_1 c_2^* \tag{3.28}$$

where now θ_1 can be estimated with observable variables.

The estimation technique used in this study is nonlinear two-stage least squares (N2SLS) with cross-equation constraints. This provides for consistent parameter estimates (conditional on the correct specification of the model) while not imposing a complete structure on the error covariance matrix as required, for instance, by three-stage least squares. In this two equation context, N2SLS solves

$$\underset{\theta}{\text{Min}} \quad \frac{1}{2T^*} e(\theta)' \, P \, e(\theta) \tag{3.29}$$

where

$$e(\theta)\equiv\begin{pmatrix}e_1(\theta)\\ e_2(\theta)\end{pmatrix}, \tag{3.30}$$

$e_i(\theta)$ is the vector of residuals from equation i; and

$$P\equiv\begin{pmatrix}P_z & 0\\ 0 & P_z\end{pmatrix} \tag{3.31}$$

$$P_z\equiv Z(Z'Z)^{-1}Z',$$

Z is the matrix of instruments; and

$$T^*\equiv\sum_{i=1}^{51} T_i\ , \tag{3.32}$$

where T_i is the number of observation for state i.

Models 1 and 2 (myopic demand models) use as instruments excise taxes, income, lagged quantity, and state and year dummy variables. Models 3 and 4 (rational demand models) add lead quantity to the instrument set. Under traditional regularity conditions, parameter estimates will be asymptotically distributed as

$$\sqrt{T^*}\,(\theta_{N2SLS}-\theta_0)\ \sim N(0,\text{plim } T^*A^{-1}B\Omega B'A^{-1})$$
$$A\equiv\nabla_\theta F'\ P\ \nabla_\theta F, \tag{3.33}$$
$$B\equiv\nabla_\theta F'\ P.$$

where A and B are evaluated at the true parameter value, θ_0, $\nabla_\theta F$ signifies the gradient of the conditional mean function (of the two stacked equations) with respect to θ, and Ω represents the covariance of the error vector. The classical assumption on Ω is

$\Omega = \Sigma \otimes I_T^*$ with Σ defined as:

$$\Sigma \equiv \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \tag{3.34}$$

which is easily estimated from the estimated residuals. For robustness I assume two additional structures on sigma: one (WHITE) allows for unspecified heteroskedasticity within and between the two equations and is estimated using White's (1982) estimator and the other (NW) allows for general heteroskedasticity and autocorrelation and is estimated using Newey and West's (1987) estimator. The details on constructing the various covariance matrices is contained in Appendix B. The use of these more general covariance structures allows for robust tests of the various hypotheses of interest. The models are partially nested, allowing for a test of Model 1 as a special case of both Model 2 (β=0) and Model 3 ($\alpha_4 = 0$). I can also test Model 2 (α_4=0) and Model 3 ($V_1 + V_2 = 0$) as special cases of Model 4. The next section details estimation results and various hypothesis tests.

4. Estimation Results

Table 2 (following page) reports some descriptive statistics of the data used for estimation. After accounting for missing data and dropping the first and last year's observations for each state, 1517 observations remain to be used in the estimation.

Table 3 (page 123) gives the first estimation results. Each parameter estimate is associated with three values of the asymptotic t-statistic (H_0: parameter = 0)

Table 2

N=1517

	Mean	Std.Dev.	Minimum	Maximum
Quantity	126.43	31.45	54.70	297.90
Price	28.75	3.25	18.36	39.47
Tax	12.47	3.41	3.36	22.14
Income	3303.10	852.40	1229.00	6560.00
Lagged	126.38	31.59	57.20	297.90
Lead	126.32	31.37	54.70	297.90

corresponding to the three alternative covariance estimates outlined in section 3.6. The parameter estimates from each of the four models are all of the expected sign and are generally significantly different from zero regardless of the covariance estimate used. The lone exception is the coefficient on income which is small and insignificantly different from zero in all but the fourth model where it is marginally significant.

The estimates on the price coefficient (α_1) range from -0.60 in Model 4 to -2.78 in Model 1. The two myopic demand models, 1 and 2, have relatively high estimates of the coefficient on lagged quantity (α_3) of 0.73 and 0.85, respectively; while the rational demands produce estimates of 0.42 (Model 3) and 0.48 (Model 4). The estimate on lead quantity (α_4) is about the same magnitude as that on lagged quantity, equalling 0.38 in Model 3 and 0.48 in Model 4. The results from the rational demand models are interesting when compared with the results from BGM. In the BGM model the ratio of α_4/α_3 gives the consumer's discount rate with its implied interest rate ($\alpha_4/\alpha_3 = (1+r)^{-1}$, where r is the interest rate). BGM estimate implausibly high values for r of about 0.25. In contrast, the implied (real) interest rate in these rational demand models is 0.09 in

Table 3 Parameter Estimates and Various t-statistics				
	Parameter Estimates	Asymptotic t-statistics		
Model:		Classical	WHITE	NW
1 (MY-MY)				
Price (α_1)	-2.78	-30.77	-21.02	-13.56
Income (α_2)	1.56E-4	0.19	0.14	0.25
Lagged (α_3)	0.73	69.15	32.21	51.77
2 (MY-RA)				
Price (α_1)	-1.25	-10.31	-8.05	-12.56
Income (α_2)	4.02E-4	0.52	0.36	0.63
Lagged (α_3)	0.85	71.76	30.34	66.83
β	1.20	65.44	28.85	64.50
3 (RA-MY)				
Price (α_1)	-2.65	-36.65	-23.56	-10.45
Income (α_2)	7.25E-4	1.08	0.85	1.43
Lagged (α_3)	0.42	27.11	14.85	47.54
Lead (α_4)	0.38	24.67	12.33	31.44
4 (RA-RA)				
Price (α_1)	-0.60	-6.61	-5.47	-14.50
Income (α_2)	1.19E-3	2.11	1.54	3.05
Lagged (α_3)	0.49	37.14	16.10	68.21
Lead (α_4)	0.48	35.80	13.66	40.11
β	2.08	36.70	16.05	68.75

Model 3 and 0.02 in Model 4 (α_3, α_4 evaluated to the third decimal) which are much more reasonable estimates.

However, the models estimated here do not perform as well in estimating the monopolist's discount rate, β. The estimate of β in Model 2 is 1.20 and in Model 4 is 2.08 with both estimates being quite high relative to the theoretical value ($0<\beta<1$), which is probably due to the no-growth assumption required to solve these models. The high value of β implies that the monopolist is giving great weight to future profits relative to current profits and is evidenced by a relatively low price in early periods and a high price in later periods. But if in fact prices increase over time due to other factors such as rising costs and incomes, the estimated β will capture this growth effect in addition to the true discounting effect. This growth effect might explain a large part of the estimated β in Model 2 where the estimate of 1.20 is 0.25 above a reasonable estimate of 0.95. It seems unlikely to explain the very large value of 2.08 in Model 4.

In another interesting comparison to previous literature, I can compute the "monopoly power index" of Sumner (1981) as the coefficient on excise taxes in the FOC equation for the two rational monopoly Models (2 and 4). For Model 2 the index equals 1.02 which is identical to Sumner's estimate while for Model 4 it is 1.01. Sumner interprets his results as a rejection of monopoly power in the cigarette industry. However, in my framework the index value close to 1 reflects the exercise of monopoly power through a monopolist charging a low price in early periods and a high price in later periods. It is unclear how the two opposing interpretations of essentially the same data and the same numbers can be resolved without resorting to firm level data. At the very least, this study suggests a reevaluation of past market power studies which use aggregate industry data.

Table 4 (following page) gives the estimates of the various own-price elasticities described in section 3.5. The values are based on the formulas given in Table 1 multiplied by the ratio of the mean price to the mean quantity (0.227). The estimates

Table 4

Various measures of the change in current quantity

Elements held fixed:	Myopic Demand		Rational Demand	
Model:	1 (MY-MY)	2 (MY-RA)	3 (RA-MY)	4 (RA-RA)
Temporary Price Change:				
1) Q_{t-1},Q_{t+1},P_i $i \neq t$	-0.63	-0.28	-0.60	-0.14
2) P_i, Q_i $i \neq t$ (Anticipated)	-0.63	-0.28	-0.99	-0.50
3) P_i $i \neq t$, Q_i $i<t$ (Surprise)	-0.63	-0.28	-0.77	-0.22
Permanent Price Change:				
4) P_i $i<t$, Q_i $i \neq t$ (Anticipated)	-0.63	-0.28	-1.89	-2.02
5) P_i, Q_i $i<t$ (Surprise)	-0.63	-0.28	-1.46	-0.88

vary considerably across the models. For the myopic demands (Models 1 and 2) future price changes have no impact on current consumption decisions, which accounts for the lack of variation in the various elasticity estimates in these two models. The estimate for Model 1 is -0.63 and for Model 2 it is -0.28.

In sharp contrast to the myopic demands, the rational demands have a large amount of variation among the several estimates. Focusing on the results for Models 3 and 4 we see that Model 4 gives more inelastic estimates relative to Model 3 with the exception of elasticity 4 which is the effect of a permanent increase in price when all quantities are allowed to adjust. For example, elasticity 3 (a surprise temporary price increase) is -0.77 in Model 3 while it is -0.22 in Model 4. The difference between a temporary and a permanent increase in price is also striking. Using the estimates from Model 4, a temporary increase in price results in an estimate of -0.22 while a permanent increase results in -0.88. All the elasticities estimated for Model 3 seem implausibly high especially those due to a permanent price change. The estimates from Model 4 suffer similar problems although they are generally not as severe with the exception of the estimate of -2.02 for the anticipated permanent price change (elasticity 4).

Table 5 (next page) gives some goodness-of-fit statistics. The first line labeled DET gives the determinant of the matrix

$$\Sigma = \frac{1}{1517}\begin{pmatrix} \hat{e}_1'\hat{e}_1 & \hat{e}_1'\hat{e}_2 \\ \hat{e}_2'\hat{e}_1 & \hat{e}_2'\hat{e}_2 \end{pmatrix}$$

where $\hat{e}_1$ is the 1517x1 vector of estimated residuals from the demand equation and $\hat{e}_2$ is the 1517x1 vector of estimated residuals from the FOC equation. DET acts as a generalized measure of fit for the two equation system accounting for the variances of both equations as well as the covariance between the equations. The results range from 4.50 in Model 4 up to 177.37 in Model 1. The dichotomy between the rational (2 and 4) and the myopic (1 and 3) monopoly models is most apparent. Conditioning on a myopic consumer, a change from a myopic monopolist to a rational monopolist reduces the value of the determinant from 177.37 to 8.57 while conditioning on a myopic monopolist (Models 1 and 3) the determinant changes from 177.37 in the myopic

Table 5

Goodness-of-Fit Statistics

	Model (Consumer-Monopolist)			
	1 (MY-MY)	2 (MY-RA)	3 (RA-MY)	4 (RA-RA)
DET	177.37	8.57	141.47	4.50
SSR: Demand	45770.8	39733.9	31469.6	21256.1
SSR: FOC	8932.9	496.5	10383.4	488.8
R^2: Demand	0.97	0.97	0.98	0.99
R^2: FOC	0.53	0.97	0.48	0.97

consumer case only down to 141.47 in the rational consumer case. The same pattern holds true with the sum-of-squared residual terms (SSR). The SSR for the demand equations range from a low of 21256.1 in Model 4 up to 45770.8 in Model 1. A tremendous amount of variation exists in the SSR for the FOC equation with the SSR for the two rational monopoly Models (2 and 4) about 1/20 the size of the myopic monopoly Models (1 and 3). For example, Model 2 (myopic consumer-rational monopolist) has an SSR for the FOC equation of 496.5 while the SSR for Model 1 (myopic consumer-myopic monopolist) is 8932.9.

The hypothesis test results are given in Tables 6 and 7 (next page) where the conservative WHITE estimate of the standard error has been used. Table 6 gives the results where Model 1 is a special case of Model 2 ($\beta=0$) and where Model 1 is a special case of Model 3 ($\alpha_4=0$). Both tests reject the null hypothesis of Model 1 with the asymptotic two-tailed t-statistic for $\beta=0$ in Model 2 equalling 28.85 and the equivalent statistic for $\alpha_4=0$ in Model 3 equalling 12.33. Table 7 presents the results of testing Models 2 ($\alpha_4 = 0$) and 3 ($V_1 + V_2 = 0$) as special cases of Model 4. The asymptotic

Table 6		
Test Model 1 as a special case of Model 2:		
H_0: $\beta = 0$	H_A: $\beta \neq 0$	t-stat: 28.85
Test Model 1 as a special case of Model 3:		
H_0: $\alpha_4 = 0$	H_A: $\alpha_4 \neq 0$	t-stat: 12.33

Table 7		
Test Model 2 as a special case of Model 4:		
H_0: $\alpha_4 = 0$	H_A: $\alpha_4 \neq 0$	t-stat: 13.66
Test Model 3 as a special case of Model 4:		
H_0: $V_1+V_2 = 0$	H_A: $V_1+V_2 \neq 0$	t-stat: 5.23

t-statistic for α_4=0 equals 13.66 while the equivalent statistic for $V_1 + V_2 = 0$ equals 5.23 and therefore both Model 2 and 3 are rejected as special cases of Model 4 on the basis of these test statistics.

Combined with the goodness-of-fit statistics, the hypothesis test results suggest that the most general model of both rational consumer and rational monopolist best describes the data. However the actual parameter estimates for Model 4 are somewhat less than satisfactory. As noted previously the model gives an unbelievable estimate of β as well as several implied price elasticities which are quite high compared to previous studies. Model 2, while suffering from drawbacks similar to Model 4, seems to have much more defensible estimates of both the price elasticities and the estimate of β, although β is still too high. The tradeoff is a relatively slight decrease in the precision of the fit when compared to Model 4. The myopic monopoly models seem to explain the data quite poorly relative to their rational counterparts. My conclusion is that Model 4 wins the statistical horserace but that Model 2 draws even on broader considerations. Models 1 and 3 can be rejected out of hand.

5. Conclusion

In this paper I have examined both theoretically and empirically monopoly behavior with intertemporal demands. The theoretical discussion suggests that intertemporal demands might play a significant and largely unexplored, role in influencing firm behavior. In the empirical section, I established a framework for testing various alternative hypothesis about consumer and monopoly behavior, and then I employed a widely-used cigarette database to conduct the tests. The results suggest that the myopic monopoly model can be rejected in favor of the rational monopoly model, but the results were somewhat less conclusive distinguishing between the rational and the myopic consumer. These initial results suggest that further research on firm behavior in the presence of intertemporal demands could prove very fruitful. Specifically disentangling competitive oligopolistic behavior from the strictly monopolistic setting of this paper would be very useful. Also of interest is the effect on other control variables such as research and development expenditures and advertising when demands are linked over time.

Appendix A:

This appendix proves the results given in equation (2.6) in the text concerning the implicit demand equation. The utility function is assumed to be strictly increasing in each of the arguments, strictly concave, separable in Z_t, and it is also assumed that Q_t and Q_{t-1} are complements. Denoting partial derivatives with subscripts (U_1 denotes the partial with respect to Q_t, U_2 with respect to Q_{t-1}, and U_3 with respect to Z_t) I can rewrite equation (2.5) as

$$U_1(Q_t,Q_{t-1}) - U_3(Y_t-P_t*Q_t)*P_t \equiv 0$$

at the optimum. Taking the total derivative I get

$$U_{11}dQ_t + U_{12}dQ_{t-1} - U_{33}P_tdY_t + U_{33}Q_tP_tdP_t + U_{33}P_t^2dQ_t - U_3dP_t = 0.$$

First I want to sign dQ_t/dP_t (with dQ_{t-1} and dY_t equal to zero).

$$\frac{dQ_t}{dP_t} = \frac{U_3 - U_{33}Q_tP_t}{U_{11} + U_{33}P_t^2} < 0$$

which is negative due to the assumptions on the utility function.
Next I want to determine the sign of dQ_t/dY_t (with $dQ_{t-1}=dP_t=0$).

$$\frac{dQ_t}{dY_t} = \frac{U_{33}P_t}{U_{11} + U_{33}P_t^2} > 0.$$

which is positive. Finally I want to determine the sign of dQ_t/dQ_{t-1} (where $dP_t=dY_t=0$):

$$\frac{dQ_t}{dQ_{t-1}} = \frac{-U_{12}}{U_{11} + U_{33}P_t^2} > 0.$$

Appendix B:

The cigarette data (prices, taxes and quantities) are from the 1990 edition of Tax Burden on Tobacco published by the Tobacco Council. Prices are measured as the weighted average price per pack, taxes as the average excise tax per pack (state and federal), and quantities are per capita sales in packs. Quantities are measured on June 30th of each year while associated prices and taxes are measured on either November 1 or October 1 of the prior year. I adjusted prices and taxes to be on the same fiscal year basis as quantities by taking a weighted average of current year and previous year values. For concreteness, the price and tax variables for 1970--P_{70}^{*} and t_{70}^{*}, respectively--were constructed as follows: (P_{70},t_{70}), (P_{69},t_{69}) come from the data book. First subtract taxes from the price: $P_{70}^{**} = P_{70} - t_{70}$, $P_{69}^{**} = P_{69} - t_{69}$. Then weight prices according to the fraction of the year each price (or tax) was in effect: $P_{70}^{\#} = (2/3)P_{70}^{**} + (1/3)P_{69}^{**}$, $t_{70}^{\#} = (2/3)t_{70} + (1/3)t_{69}$. Add taxes back to get P_{70}^{*}: $P_{70}^{*} = P_{70}^{\#} + t_{70}^{\#}$, $t_{70}^{*} = t_{70}^{\#}$. P_{70}^{*} and t_{70}^{*} were then normalized by a fiscal year price index for 1970 (a simple average of the 1969 and 1970 CPI taken from the survey of current business).

The income variable is the personal disposable income by state and year from various issues of the survey of current business. Income was also adjusted to a real fiscal year basis.

Covariance construction

Three alternative covariance matrices were construced, each based on a different error structure assumption. Each of the (loosely speaking) estimated asymptotic covariance matrices, e_avar, has a similar form (from equation (3.33) in the text):

$$e\_avar \equiv \hat{A}^{-1}\hat{B}\Omega\hat{B}'\hat{A}^{-1}$$

where

$$\hat{A} = \nabla_{\theta}\hat{F}' \; P \; \nabla_{\theta}\hat{F}$$
$$\hat{B} = \nabla_{\theta}\hat{F}' \; P$$

with ^ signifying evaluation at the estimated parameters, $\nabla_{\theta}F$ defining the gradient of the conditional mean, and P being the block diagonal projection matrix ((3.31) in the text). The three covariance estimates differ in the estimates of $P\Omega P$ which is the inner matrix of $B\Omega B$. Writing Ω in partitioned form I have,

$$\Omega = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where Σ_{ij} represents the $T^{*}xT^{*}$ error covariance matrix for equation i and j ($T^{*} \equiv \sum_{i=1}^{51} T_i$,

T_i is the number of time periods state i is observed). Σ_{11} is the covariance for equation 1, Σ_{22} for equation 2 and Σ_{12}, Σ_{21} are the cross-equation covariance matrices. For the simple classical covariance, $\Sigma_{ij} = \sigma_{ij} I_T{}^*$ and σ_{ij} is estimated as

$$\frac{1}{T^*}\sum_{t=1}^{T^*} e_{it} e_{jt}.$$

For the WHITE covariance Σ_{ij} is assumed to be diagonal and the quantity $Z'\Sigma_{ij} Z$ estimated as

$$\frac{1}{T^*}\sum_{t=1}^{T^*} Z_t' e_{it} e_{jt} Z_t$$

To estimate the NW estimator I used a variation of the Newey-West (1987) estimator. Suppressing the equation subscripts, $Z'\Sigma Z$ can be written as

$$\begin{bmatrix} Z(1)', & \ldots, & Z(N)' \end{bmatrix} \begin{bmatrix} \Sigma(11) & \ldots & \Sigma(1N) \\ \cdot\ \cdot & & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \Sigma(N1) & \ldots & \Sigma(NN) \end{bmatrix} \begin{bmatrix} Z(1) \\ \cdot \\ \cdot \\ \cdot \\ Z(N) \end{bmatrix}$$

where the matrices have been partitioned into the component states (Z(i) is the (T_ixK) matrix of intruments for state i for all years, Σ(ij) is the T_ixT_j covariance matrix of residuals for state i and state j). This quadratic form can be written as

$$Z'\Sigma Z = \sum_{i=1}^{N}\sum_{j=1}^{N} Z(i)'\ \Sigma(ij)\ \ Z(j).$$

I estimated each of the components $Z(i)'\Sigma(ij)\ Z(j)$ with the Newey-West estimator after truncating all states to have the same number of observations in order to create a balanced panel.[6]

[6]In Newey-West notation: G(N) = 2 (=(number of years for each state) raised to the (1/4)th power).

The following four pages give parameter estimates for the dummy variables of the four different models along with WHITE standard errors.

Model 1 (MY-MY)

	State Variables				Year Variables				
	Demand		FOC			Demand		FOC	
	Param.	Std.Er	Param.	Std.Er		Param.	Std.Er	Param.	Std.Er
AK	124.69	8.82	14.52	1.66	1957	-8.86	2.48	-4.82	0.71
AL	115.91	6.72	16.00	1.16	1958	-7.08	2.44	-4.56	0.68
AR	116.33	6.73	14.13	1.18	1959	-0.99	2.49	-4.55	0.67
AZ	113.20	7.06	12.43	1.30	1960	-2.49	2.44	-5.37	0.65
CA	114.99	7.72	12.87	1.48	1961	-0.14	2.38	-5.94	0.66
CO	108.34	7.22	9.90	1.39	1962	-2.79	2.35	-6.45	0.65
CT	125.09	8.24	13.02	1.68	1963	-1.99	2.31	-6.35	0.65
DC	125.52	8.41	3.92	2.27	1964	-5.09	2.19	-6.28	0.64
DE	126.64	7.98	8.38	1.61	1965	-0.76	2.11	-5.89	0.61
FL	126.15	7.63	12.67	1.42	1966	-0.15	2.12	-5.89	0.60
GA	116.11	6.90	13.52	1.26	1967	-0.11	1.90	-5.82	0.61
HI	105.58	7.40	18.52	1.18	1968	-0.52	1.77	-5.33	0.55
IA	115.41	7.17	13.87	1.25	1969	-0.89	1.69	-5.30	0.51
ID	109.50	6.73	13.85	1.19	1970	-1.48	1.70	-4.81	0.50
IL	119.09	7.75	11.32	1.46	1971	4.42	1.54	-4.02	0.49
IN	113.64	7.12	8.04	1.44	1972	7.19	1.54	-4.36	0.49
KS	112.08	7.17	12.77	1.28	1973	3.16	1.35	-5.68	0.51
KY	116.58	6.98	3.58	1.87	1974	0.14	1.33	-7.08	0.50
LA	120.89	7.01	13.24	1.26	1975	-3.26	1.38	-8.00	0.50
MA	124.24	7.86	13.31	1.42	1976	-1.34	1.42	-7.86	0.51
MD	112.91	7.47	10.79	1.41	1977	-5.90	1.43	-8.47	0.53
ME	121.03	7.18	10.01	1.36	1978	-4.99	1.27	-7.68	0.54
MI	119.31	7.55	11.22	1.40	1979	-9.00	1.25	-7.95	0.52
MN	117.95	7.34	14.47	1.26	1980	-12.15	1.37	-8.90	0.50
MO	114.04	7.14	9.82	1.37	1981	-16.06	1.53	-9.91	0.51
MS	113.69	6.45	15.31	1.13	1982	-17.50	1.52	-9.47	0.52
MT	115.59	6.98	13.15	1.22	1983	-14.54	1.34	-6.87	0.52
NC	115.38	7.39	-0.51	1.89	1984	-9.37	1.25	-4.42	0.52
ND	112.49	6.89	14.33	1.18	1985	-3.47	1.10	-2.78	0.51
NE	112.17	7.07	13.40	1.24	1986	-1.79	1.00	-1.51	0.50
NH	142.07	8.96	-5.11	2.26					
NJ	124.60	8.11	12.84	1.49					
NM	112.53	6.81	16.59	1.10					
NV	134.17	8.56	5.04	1.88					
NY	121.79	7.90	12.51	1.51					
OH	115.52	7.30	10.62	1.38					
OK	118.12	7.08	12.19	1.30					
OR	111.86	7.28	7.44	1.63					
PA	119.39	7.43	13.26	1.31					
RI	122.35	7.54	9.61	1.46					
SC	107.44	6.38	12.13	1.28					
SD	111.24	6.74	14.25	1.15					
TN	115.77	6.79	13.54	1.22					
TX	121.00	7.36	14.15	1.27					
UT	100.75	6.48	19.08	0.98					
VA	106.54	6.89	7.86	1.43					
VT	122.25	7.29	9.62	1.43					
WA	118.85	7.59	17.46	1.24					
WI	116.74	7.28	14.12	1.25					
WV	116.25	6.82	13.81	1.18					
WY	114.23	7.40	8.53	1.46					

Model 2 (MY-RA)

	State Variables					Year Variables			
	Demand		FOC			Demand		FOC	
	Param.	Std.Er	Param.	Std.Er		Param.	Std.Er	Param.	Std.Er
AK	55.72	9.37	27.31	0.23	1957	-1.37	2.31	-8.11	0.16
AL	52.40	7.57	24.28	0.19	1958	-0.08	2.35	-7.63	0.15
AR	53.02	7.59	22.39	0.18	1959	5.50	2.31	-7.34	0.14
AZ	50.92	7.83	23.93	0.19	1960	2.86	2.29	-7.34	0.15
CA	50.93	8.38	25.11	0.23	1961	5.18	2.24	-7.46	0.15
CO	48.61	7.89	22.56	0.23	1962	1.28	2.19	-7.37	0.14
CT	55.61	8.95	24.29	0.28	1963	1.94	2.17	-7.41	0.14
DC	57.58	8.98	25.16	0.30	1964	-1.55	2.07	-7.39	0.14
DE	57.88	8.95	24.65	0.23	1965	3.18	1.98	-7.54	0.14
FL	57.04	8.56	24.16	0.22	1966	2.77	2.07	-7.51	0.14
GA	52.97	7.75	23.81	0.18	1967	2.49	1.77	-7.35	0.14
HI	45.83	7.76	23.65	0.19	1968	1.34	1.70	-6.79	0.14
IA	51.84	7.92	23.72	0.22	1969	1.39	1.60	-6.84	0.14
ID	49.03	7.49	23.52	0.20	1970	0.26	1.59	-6.57	0.14
IL	53.51	8.53	24.43	0.22	1971	5.81	1.38	-6.10	0.15
IN	52.08	8.00	23.10	0.21	1972	8.07	1.47	-6.08	0.15
KS	50.33	7.88	23.13	0.19	1973	4.66	1.19	-6.62	0.14
KY	55.32	8.17	23.40	0.22	1974	4.04	1.17	-7.29	0.13
LA	55.04	8.00	24.21	0.20	1975	2.04	1.24	-7.18	0.14
MA	55.73	8.66	23.96	0.20	1976	3.72	1.27	-6.40	0.14
MD	50.91	8.17	23.58	0.26	1977	-0.51	1.31	-6.25	0.14
ME	55.02	8.22	22.76	0.20	1978	-0.10	1.12	-5.26	0.14
MI	53.90	8.38	23.89	0.20	1979	-2.80	1.06	-5.19	0.14
MN	52.93	8.00	23.20	0.21	1980	-2.95	1.21	-5.94	0.14
MO	51.92	8.00	23.74	0.21	1981	-4.66	1.41	-6.49	0.14
MS	51.60	7.34	23.65	0.18	1982	-6.41	1.38	-5.78	0.14
MT	51.40	7.82	23.42	0.22	1983	-6.77	1.16	-3.42	0.14
NC	53.39	8.75	23.50	0.26	1984	-5.23	1.06	-2.62	0.16
ND	50.46	7.63	23.08	0.18	1985	-0.49	0.91	-2.17	0.17
NE	50.18	7.78	23.06	0.18	1986	-0.25	0.81	-1.15	0.17
NH	68.04	10.61	23.77	0.30					
NJ	55.65	8.86	23.73	0.20					
NM	49.92	7.50	23.94	0.17					
NV	61.61	9.80	25.83	0.27					
NY	54.39	8.67	24.41	0.23					
OH	52.21	8.12	22.90	0.20					
OK	53.56	7.98	22.95	0.19					
OR	49.85	8.13	22.69	0.24					
PA	53.65	8.22	22.77	0.20					
RI	55.28	8.49	22.92	0.22					
SC	49.11	7.21	23.11	0.19					
SD	49.82	7.50	22.98	0.20					
TN	52.88	7.65	23.30	0.17					
TX	54.36	8.19	23.01	0.19					
UT	44.01	6.87	23.37	0.19					
VA	48.93	7.68	23.60	0.20					
VT	55.78	8.35	23.26	0.21					
WA	52.46	8.20	24.52	0.23					
WI	52.28	8.01	22.74	0.19					
WV	52.57	7.73	23.84	0.18					
WY	51.64	8.25	23.57	0.22					

Model 3 (RA-MY)

State Variables

	Demand		FOC	
	Param.	Std.Er	Param.	Std.Er
AK	112.15	7.23	12.23	1.54
AL	105.35	5.41	14.23	1.08
AR	105.17	5.42	12.19	1.11
AZ	102.44	5.65	10.45	1.19
CA	104.15	6.17	10.87	1.35
CO	97.43	5.78	7.80	1.27
CT	113.21	6.64	10.85	1.58
DC	110.59	6.78	0.73	2.34
DE	112.70	6.35	5.59	1.47
FL	114.04	6.10	10.45	1.29
GA	104.51	5.54	11.44	1.18
HI	97.02	6.05	17.25	1.06
IA	104.46	5.77	11.93	1.13
ID	99.61	5.41	12.11	1.09
IL	107.09	6.19	9.05	1.33
IN	101.02	5.71	5.52	1.33
KS	101.16	5.76	10.80	1.17
KY	101.61	5.54	0.40	1.88
LA	109.18	5.61	11.12	1.16
MA	112.30	6.30	11.15	1.30
MD	101.04	5.98	8.53	1.28
ME	109.00	5.72	7.71	1.24
MI	107.19	6.03	8.93	1.27
MN	107.01	5.95	12.58	1.14
MO	102.00	5.69	7.48	1.25
MS	103.31	5.20	13.55	1.07
MT	105.44	5.59	11.33	1.11
NC	102.44	5.64	-3.50	1.79
ND	102.19	5.53	12.55	1.08
NE	101.74	5.67	11.55	1.12
NH	122.68	6.88	-9.54	2.16
NJ	112.51	6.51	10.64	1.36
NM	103.23	5.52	15.08	1.01
NV	118.80	6.72	1.80	1.75
NY	110.02	6.33	10.34	1.39
OH	103.80	5.83	8.41	1.27
OK	106.60	5.67	10.09	1.19
OR	101.08	5.79	5.25	1.53
PA	108.02	5.97	11.23	1.19
RI	109.92	6.03	7.20	1.33
SC	96.53	5.13	10.13	1.23
SD	101.23	5.42	12.53	1.05
TN	104.30	5.44	11.50	1.14
TX	109.93	5.91	12.22	1.15
UT	93.28	5.28	18.05	0.88
VA	94.27	5.52	5.39	1.33
VT	109.78	5.82	7.21	1.32
WA	108.73	6.13	15.83	1.11
WI	105.94	5.85	12.24	1.13
WV	105.40	5.48	11.89	1.09
WY	102.23	5.90	6.14	1.33

Year Variables

	Demand		FOC	
	Param.	Std.Er	Param.	Std.Er
1957	-10.39	2.06	-5.15	0.73
1958	-10.97	2.06	-5.34	0.69
1959	-5.49	2.11	-5.45	0.69
1960	-6.88	2.07	-6.29	0.67
1961	-3.63	1.98	-6.71	0.68
1962	-4.86	1.94	-6.98	0.67
1963	-2.83	1.88	-6.64	0.65
1964	-6.60	1.80	-6.69	0.64
1965	-3.66	1.75	-6.54	0.63
1966	-2.64	1.62	-6.47	0.64
1967	-1.88	1.58	-6.27	0.60
1968	-1.96	1.43	-5.70	0.54
1969	-1.97	1.43	-5.60	0.52
1970	-4.11	1.42	-5.38	0.52
1971	-0.70	1.45	-5.03	0.54
1972	2.10	1.35	-5.39	0.54
1973	-1.58	1.21	-6.69	0.55
1974	-4.60	1.24	-8.15	0.54
1975	-8.15	1.23	-9.13	0.55
1976	-5.36	1.23	-8.84	0.58
1977	-8.62	1.18	-9.25	0.56
1978	-7.23	1.07	-8.35	0.57
1979	-11.42	1.09	-8.67	0.55
1980	-15.06	1.27	-9.74	0.55
1981	-17.99	1.34	-10.60	0.56
1982	-17.56	1.31	-9.80	0.55
1983	-13.25	1.13	-6.88	0.54
1984	-9.37	1.02	-4.57	0.56
1985	-4.60	0.91	-3.08	0.56
1986	-2.40	0.84	-1.67	0.55

Model 4 (RA-RA)

	State Variables				Year Variables				
	Demand		FOC			Demand		FOC	
	Param.	Std.Er	Param.	Std.Er		Param.	Std.Er	Param.	Std.Er
AK	19.28	6.83	27.65	0.23	1957	-1.01	1.69	-8.04	0.15
AL	19.99	5.39	24.58	0.18	1958	-2.84	1.77	-7.52	0.14
AR	19.93	5.41	22.71	0.18	1959	1.85	1.75	-7.21	0.14
AZ	18.73	5.56	24.23	0.19	1960	-0.97	1.72	-7.21	0.14
CA	18.13	5.93	25.41	0.23	1961	2.46	1.65	-7.35	0.14
CO	17.08	5.60	22.86	0.22	1962	-0.01	1.62	-7.28	0.14
CT	19.86	6.42	24.65	0.27	1963	2.16	1.57	-7.34	0.13
DC	18.78	6.27	25.58	0.27	1964	-2.30	1.52	-7.31	0.13
DE	19.91	6.40	25.05	0.22	1965	0.80	1.47	-7.42	0.14
FL	21.10	6.08	24.52	0.22	1966	0.59	1.44	-7.40	0.14
GA	19.43	5.54	24.13	0.17	1967	1.13	1.32	-7.26	0.14
HI	16.96	5.62	23.89	0.18	1968	0.15	1.23	-6.70	0.14
IA	18.99	5.65	24.03	0.22	1969	0.80	1.29	-6.76	0.14
ID	18.42	5.33	23.80	0.19	1970	-2.47	1.18	-6.46	0.14
IL	18.82	6.06	24.77	0.21	1971	-0.15	1.24	-5.95	0.15
IN	17.92	5.79	23.45	0.20	1972	2.00	1.18	-5.92	0.14
KS	18.07	5.64	23.44	0.18	1973	-0.75	0.98	-6.47	0.14
KY	18.45	5.91	23.80	0.22	1974	-0.62	0.98	-7.14	0.13
LA	20.57	5.71	24.54	0.19	1975	-2.35	0.99	-7.03	0.13
MA	20.17	6.15	24.31	0.19	1976	0.35	1.00	-6.28	0.14
MD	17.46	5.82	23.91	0.25	1977	-2.13	0.91	-6.15	0.13
ME	20.18	5.88	23.12	0.20	1978	-1.28	0.81	-5.19	0.14
MI	19.10	5.97	24.24	0.20	1979	-3.81	0.79	-5.11	0.14
MN	19.62	5.73	23.52	0.21	1980	-3.64	0.95	-5.86	0.14
MO	18.25	5.74	24.07	0.21	1981	-3.45	1.02	-6.44	0.13
MS	19.85	5.23	23.94	0.18	1982	-2.95	0.99	-5.77	0.14
MT	19.37	5.54	23.72	0.22	1983	-2.66	0.86	-3.45	0.14
NC	18.98	6.11	23.87	0.26	1984	-3.89	0.74	-2.61	0.16
ND	18.85	5.40	23.38	0.17	1985	-0.95	0.66	-2.14	0.17
NE	18.45	5.52	23.36	0.18	1986	-0.52	0.65	-1.14	0.17
NH	22.08	7.49	24.33	0.30					
NJ	19.78	6.31	24.09	0.19					
NM	19.33	5.31	24.21	0.17					
NV	20.84	7.06	26.27	0.26					
NY	19.40	6.15	24.75	0.22					
OH	18.55	5.78	23.24	0.20					
OK	19.73	5.69	23.28	0.19					
OR	17.86	5.76	23.01	0.23					
PA	19.62	5.85	23.11	0.19					
RI	19.64	6.10	23.29	0.22					
SC	17.92	5.17	23.41	0.19					
SD	18.76	5.33	23.26	0.19					
TN	19.58	5.47	23.62	0.17					
TX	20.41	5.84	23.35	0.18					
UT	17.38	4.84	23.58	0.19					
VA	16.40	5.51	23.91	0.19					
VT	20.27	6.00	23.62	0.20					
WA	19.69	5.79	24.82	0.23					
WI	19.31	5.68	23.06	0.18					
WV	19.81	5.53	24.15	0.18					
WY	17.92	5.89	23.90	0.22					

References

Appelbaum, E. (1982), "The Estimation of the Degree of Oligopoly Power," Journal of Econometrics, 19, 287-299.

Ashenfelter, O. and D. Sullivan (1987), "Nonparametric Tests of Market Structure: An Application to the Cigarette Industry," Journal of Industrial Economics, 35, 483-498.

Baltagi, B.H. and D. Levin (1986), "Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation and Advertising Reconsidered," Review of Economics and Statistics, 68, 148-155.

Becker, G.S., M. Grossman and K.M. Murphy (1990), "An Empirical Analysis of Cigarette Addiction," Working Paper #3322, National Bureau of Economic Research, April.

Becker, G.S. and K.M. Murphy (1988), "A Theory of Rational Addiction," Journal of Political Economy, 96, 675-700.

Bishop, J.A. and J.H. Yoo (1985), " "Health Scare," Excise Taxes and Advertising Ban in the Cigarette Demand and Supply," Southern Economic Journal, 52, 402-411.

Bresnahan, T.F. (1989), "Empirical Studies of Industries With Market Power," Chapter 17 in Handbook of Industrial Organization, Richard Schmalensee and Robert D. Willig, Editors, Vol II, Amsterdam: North Holland Publishers, pp. 1011-1057.

Bulow, J.I. and P. Pfleiderer (1983), "A Note on the Effect of Cost Changes on Prices," Journal of Political Economy, 91, 182-185.

Constantinides, G.M., (1990), "Habit Formation: A Resolution of the Equity Premium Puzzle," Journal of Political Economy, 98, 519-543.

Doran, G, (1979), The Smoking Paradox: Public Regulation of the Cigarette Industry, Cambridge Mass: Abt Books.

Gallant, A.R. and D.W. Jorgenson (1979), "Statistical Inference for a System of Simultaneous, Non-Linear, Implicit Equations in the Context of Instrumental Variable Estimation," Journal of Econometrics, 11, 275-302.

Huang, C., and Kreps, D. (1987), "On Intertemporal Preferences With a Continuous Time Dimension, I: The Case of Certainty," Unpublished manuscript, April.

Heaton, J. (1989), "The Interaction Between Time-Nonseparable Preferences and Time Aggregation," Unpublished manuscript, May

Houthakker, H.S. and L.D. Taylor (1970), Consumer Demand in the United States, Second Edition, Cambridge Mass: Harvard University Press.

Mehra, R. and E.C. Prescott, (1985), "The Equity Premium: A Puzzle," Journal of Monetary Economics, 15, 145-161.

Newey, W.K. and K.D. West, (1987) "A Simple Positive-Semi Definite Estimate of the Covariance Matrix," Econometrica, .

Panzer, J.C. and J.N. Rosse (1987), "Testing for "Monopoly" Equilibrium," Journal of Industrial Economics, 35, 443-456.

Roberts, M.J. and L. Samuelson (1988), "An Empirical Analysis of Dynamic, Nonprice Competition in an Oligopolistic Industry," Rand Journal of Economics, 19, 200-220.

Sargent, T.J. (1987), Dynamic Macroeconomic Theory, Cambridge Mass: Harvard University Press.

Sullivan, D. (1987), "Testing Hypotheses about Firm Behavior in the Cigarette Industry," Journal of Political Economy, 93, 586-598.

Sumner, D.A. (1981), "Measurement of Monopoly Behavior: An Application to the Cigarette Industry," Journal of Political Economy, 89, 1010-1019.

White, H. (1982), "Instrumental Variables Regression with Independent Observations," Econometrica, 50, 483-500.