NAT'L INST. OF STAND & TECH R.I.C.
A11105 487118

NIST
PUBLICATIONS

NIST United States Department of Commerce
Technology Administration
National Institute of Standards and Technology

NISTIR 5066

ECONOMIC DATA: HANDLE WITH CARE

Janet Rogers

QC
100
.U56
NO.5066
1998

NISTIR 5066

ECONOMIC DATA: HANDLE WITH CARE

Janet Rogers

Mathematical and Computational Sciences Division
Information Technology Laboratory
National Institute of Standards and Technology
Boulder, Colorado 80303-3328

March 1998

U.S. DEPARTMENT OF COMMERCE, William M. Daley, Secretary
TECHNOLOGY ADMINISTRATION, Gary R. Bachula, Acting Under Secretary for Technology
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, Raymond G. Kammer, Director

Economic Data: Handle with Care

Janet E. Rogers*
Mathematical and Computational Sciences Division
National Institue of Standards and Technology
325 Broadway
Boulder, CO 80303-3328

This paper provides compelling evidence that common practices employed in order to simplify the least squares analysis of economic data can lead to incorrect results and conclusions. Two examples are presented using state per capita income data over the period 1965 to 1991. The first example documents how parameter estimates can be changed by seemingly inconsequential mathematical transformations of the model. The second supports the assertion that the measurement errors inherent in economic data must be addressed. The results from both studies demonstrate the importance—and computational feasibility—of formulating economic models precisely.
JEL classifications: C51; C63; O47.
Keywords: errors in variables; linear least squares; nonlinear least squares; orthogonal distance regression; total least squares.

1. Introduction

Early in their book *Estimation and Inference in Econometrics* [1, page 57], Davidson and MacKinnon state

> ...[the model that] *is linear in the parameters and in the logarithms of all the variables will be very much easier to estimate than the nonlinear model... Thus, it should come as no surprise to learn that loglinear regression models, like*
>
> $$\ln[y_t] = \beta_1 + \beta_2 \ln[z_{t,2}] + \beta_3 \ln[z_{t,3}] + v_t \ , \tag{1.1}$$
>
> *are estimated very frequently in practice, while multiplicative models with additive error terms, like*
>
> $$y_t = \exp[\beta_1] z_{t,2}^{\beta_2} z_{t,3}^{\beta_3} + u_t \ , \tag{1.2}$$
>
> *are very rarely estimated.*

*Janet.Rogers@NIST.gov

Davidson and MacKinnon do *not* advocate that researchers select between linear and nonlinear models strictly on grounds of convenience. In many instances, however, the choice between two mathematically equivalent models is not well delineated. Furthermore, few texts provide numerical examples that demonstrate the consequences of choosing one model representation over another, and the numerical comparisons that are presented generally show only relatively small changes in the estimated parameters, which could lead researchers to think that these issues are "much ado about nothing."

This paper provides researchers with empirical evidence highlighting the dramatic impact that mathematical transformations and simplifying assumptions can have on the conclusions drawn from an analyses. While the results presented here do not have universal applicability, they do serve as a clear warning to researchers who do not explicitly address the possible ramifications of such transformations. The inferences from this paper are straightforward: researchers must examine the transformations they employ and the assumptions they make—and explicitly address their implications—or risk reporting incorrect results and conclusions.

Two relatively simple examples, suggested by recent publications, are considered. Both use the method of least squares—one of the most common computational procedures employed in economic research—to estimate values for the K unknown parameters $\boldsymbol{\beta}$ of an overdetermined system of equations,

$$y_t = f_t[\boldsymbol{\beta}] + \varepsilon_t \ , \quad t = 1, \ldots, T, \tag{1.3}$$

where y_t denotes observation t of the response (dependent) variable, $f_t[\boldsymbol{\beta}]$ denotes the mathematical model being investigated, and ε_t denotes the unobservable random error or disturbance in y_t. The generalized least squares estimators for this problem are the values $\mathbf{b}$ that minimizes the residual sum of squares $S[\boldsymbol{\beta}]$, that is

$$\begin{aligned} \mathbf{b} \stackrel{\text{def}}{=} & \operatorname*{argmin}_{\boldsymbol{\beta}} S[\boldsymbol{\beta}] \\ & \text{subject to:} \qquad S[\boldsymbol{\beta}] = \mathbf{e}'\boldsymbol{\Omega}^{-1}\mathbf{e}, \end{aligned} \tag{1.4}$$

where $\mathbf{e}$ denotes the column vector with elements $e_t \stackrel{\text{def}}{=} f_t[\boldsymbol{\beta}] - y_t, \ t = 1, \ldots, T$, and $\boldsymbol{\Omega}$ denotes the $T \times T$ covariance matrix $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']$. Often the disturbances can be assumed to be identically and independently distributed, so that $\boldsymbol{\Omega} \equiv \sigma^2 I$ and $S[\boldsymbol{\beta}] = \sigma^{-2}\mathbf{e}'\mathbf{e}$.

Programs for finding the least squares solution are ubiquitous.

- Linear models are those of the form $f_t[\boldsymbol{\beta}] \equiv \mathbf{x}_t\boldsymbol{\beta}$, and *linear least squares* (LLS) procedures solve for

  $$\mathbf{b} \stackrel{\text{def}}{=} (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} \ , \tag{1.5}$$

  where $\mathbf{y}$ denotes the vector with rows $y_t, \ t = 1, \ldots, T$, and $\mathbf{X}$ denotes the $T \times K$ matrix with rows $\mathbf{x}_t \equiv (x_{t,1}, x_{t,2}, \ldots, x_{t,K})$ containing the explanatory (independent) variables associated with observation t. For uncorrelated disturbances with constant variance, eq (1.5) reduces to the familiar ordinary (unweighted) least squares solution

  $$\mathbf{b} \stackrel{\text{def}}{=} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \ . \tag{1.6}$$

- *Nonlinear least squares* (NLS) must be used when $f_t[\boldsymbol{\beta}] \not\equiv \mathbf{x}_t\boldsymbol{\beta}$. Most NLS procedures have a LLS procedure at their core. However, when the model is nonlinear in the unknown parameters the $T \times K$ derivative matrix $\mathbf{X}$ composed of rows

$$\mathbf{x}_t \stackrel{\text{def}}{=} \left(\frac{\partial f_t[\boldsymbol{\beta}]}{\partial \beta_1}, \ldots, \frac{\partial f_t[\boldsymbol{\beta}]}{\partial \beta_K}\right) \tag{1.7}$$

is a function of $\boldsymbol{\beta}$, and therefore the solution defined by eq (1.5) will change as $\boldsymbol{\beta}$ changes. NLS solutions are therefore computed iteratively, progressing from an estimate $\boldsymbol{\beta}_n$ at iteration n to an estimate $\boldsymbol{\beta}_{n+1}$ at iteration $n+1$ until the parameter estimates $\boldsymbol{\beta}_{n+1}$ are "good enough." NLS procedures have greater complexity than LLS procedures: the functional form of the model $f_t[\boldsymbol{\beta}]$—and frequently the derivatives $\partial f_t[\boldsymbol{\beta}]/\partial \beta_k$—must be supplied; starting values $\boldsymbol{\beta}_0$ are required to begin the iterative process; and finally, convergence criteria must be selected.

- Errors in variables problems, also known as measurement error problems, occur when the explanatory variables include a stochastic component $\boldsymbol{v}_t$ so that $\mathbf{x}_t = \mathbf{x}_t^* + \boldsymbol{v}_t,\ t = 1, \ldots, T$, and $E[\mathbf{X}'\boldsymbol{\varepsilon}] \neq 0$. Common weighted or unweighted least squares methods, whether linear or nonlinear, are inappropriate in this case and other solution procedures should be employed. The method of *instrumental variables* (IV), discussed by Davidson and MacKinnon [1], Greene [2] and Judge et al. [3], is one frequently applied technique for estimating $\boldsymbol{\beta}$ when there are errors in the explanatory variables. Another procedure for extending least squares data fitting procedures to measurement error problems problems is *orthogonal distance regression* (ODR), which is especially appropriate when the model is nonlinear in the explanatory variables or the variances of the measurement errors are known to within a constant.

The theoretical properties of least squares estimators are described in most econometric texts; Davidson and MacKinnon [1], Greene [2] and Judge et al. [3] are three excellent references. Under fairly weak regularity conditions, LLS, NLS, IV and ODR estimators are all consistent, even when the disturbances are not independent with constant variance. Furthermore, LLS, NLS, and ODR estimators are maximum likelihood estimators (MLE) when the disturbances are normally distributed with covariance matrix known to within a constant, that is, when $\boldsymbol{\mathcal{E}} \sim N(\mathbf{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\mathcal{E}} \stackrel{\text{def}}{=} \boldsymbol{\varepsilon}$ for LLS and NLS and $\boldsymbol{\mathcal{E}} \stackrel{\text{def}}{=} (\boldsymbol{\varepsilon}', \boldsymbol{v}_1', \ldots, \boldsymbol{v}_T')'$ for ODR [4].

Table 1. Least squares techniques.

Model w.r.t. $\boldsymbol{\beta}$	Errors in variables	Technique	MLE when $\boldsymbol{\mathcal{E}} \sim N(\mathbf{0}, \boldsymbol{\Omega})$
Linear	No	Linear least squares (LLS)	Yes
Nonlinear	No	Nonlinear least squares (NLS)	Yes
Linear or nonlinear	Yes	Instrumental variables estimation (IV)	No
Linear or nonlinear	Yes	Orthogonal distance regression (ODR)	Yes

Researchers are sometimes tempted to transform nonlinear models into linear models so that LLS can be employed rather than NLS. They are also frequently tempted to ignore measurement errors in the explanatory variables so that LLS or NLS can be used in place of IV or ODR. The consequences of inappropriately applying such transformations and making such assumptions are often not fully appreciated, however. The next section provides empirical evidence showing how the results of an analysis can change enormously as the model is transformed to have different mathematical representations. Further, the section following documents significant differences between the results obtained using NLS and those obtained using ODR, emphasizing the importance of explicitly modeling the measurement errors inherent in economic data.

2. Linear and Nonlinear Least Squares

Researchers often transform nonlinear models into linear models as a simplification. These mathematical transformations should be applied only after their effects on the error terms $\boldsymbol{\varepsilon}$ have been investigated, however. Consider, for example, the model

$$y_t = \beta_1 \exp[\beta_2 z_t] + \varepsilon_t \ , \quad t = 1, \ldots, T \ , \tag{2.1}$$

and its logarithmic transformation

$$\ln[y_t] = \ln[\beta_1] + \beta_2 z_t + \ln\left[1 + \frac{\varepsilon_t}{\beta_1 \exp[\beta_2 z_t]}\right] = \ln[\beta_1] + \beta_2 z_t + v_t \ , \tag{2.2}$$

where $v_t \stackrel{\text{def}}{=} \ln\left[1 + \frac{\varepsilon_t}{\beta_1 \exp[\beta_2 z_t]}\right]$. When it is correct to assume that $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)' \sim N(0, \boldsymbol{\Omega}_\varepsilon)$, then it is also correct to assume that $\boldsymbol{v} = (v_1, \ldots, v_T)' \sim N(0, \boldsymbol{\Omega}_v)$ although $\boldsymbol{\Omega}_\varepsilon \neq \boldsymbol{\Omega}_v$. Furthermore, if the disturbances $\boldsymbol{\varepsilon}$ are homoscedastic and $\boldsymbol{\Omega}_\varepsilon \equiv \sigma_\varepsilon^2 I$, then, unless $E\left[\beta_1 \exp[\beta_2 z_t]\right]$ is constant for all t, it must be the case that the disturbances $\boldsymbol{v}$ are heteroscedastic and $\boldsymbol{\Omega}_v \neq \sigma_v^2 I$ for some constant σ_v^2. The converse is also true.

The statistical implications of mathematical transformations of models and data are generally acknowledged, but frequently ignored. In many instances, transformations appear to be applied for no reason other than to allow the least squares solution to be found using LLS rather than NLS. Although these simplifying transformations are often innocuous, they can sometimes lead to significantly different parameter estimates. This section demonstrates this phenomenon using linear and nonlinear forms of a logistic growth model.

Sigmoidal or "S-shaped" curves like that specified by the logistic function can be used to characterize data that display a monotonic increase in growth rate up to an inflection point, after which the growth rate decreases to 0. These functional forms have been used by Griliches [5] and Mansfield [6] to analyze the diffusion rates of new technologies, and by Clark et al. [7] to model the growth of the manufacturing sector in developing economies.

Berry and Kaserman [8] also use the logistic function to model the determinants of state-level economic development in the United States. They postulate that the true per capita income $y^*_{s,t}$ for state s in year t has a sigmoidal shape—increasing rapidly at first as new technologies are introduced and then leveling off as the equilibrium value is approached—and thus can be represented by the logistic function

$$y^*_{s,t} = \frac{\alpha_s}{1 + \phi_s \exp[-\gamma_s (t - t_0)]} \ , \quad t = t_1, \ldots, t_T, \tag{2.3}$$

where α_s denotes the asymptotic (steady-state) limit on real per capita income, ϕ_s is a measure of how far initial per capita income y_{s,t_0} is from its asymptotic value, and γ_s denotes the growth rate.

Assumptions made regarding the unobservable disturbances $\varepsilon_{s,t}$ allow the statistical specification of model (2.3) to be completed.

A. One possibility is that the error $\varepsilon^{\mathsf{a}}_{s,t}$ in the observed value of real per capita income $y_{s,t}$ is additive, and thus that

$$y_{s,t} = \frac{\alpha_s}{1+\phi_s \exp[-\gamma_s(t-t_0)]} + \varepsilon^{\mathsf{a}}_{s,t}, \quad t = t_1, \ldots, t_T, \tag{2.4}$$

which is *nonlinear* in the unknown parameters α_s, ϕ_s, and γ_s.

B. Suppose, on the other hand, that the disturbance in observation $y_{s,t}$ is multiplicative, so that $\ln[y_{s,t}] = \ln[y^*_{s,t}\varepsilon_{s,t}] = \ln[y^*_{s,t}] + \varepsilon^{\mathsf{b}}_{s,t}$, where $\varepsilon^{\mathsf{b}}_{s,t} \stackrel{\text{def}}{=} \ln[\varepsilon_{s,t}]$. The statistical representation of model (2.3) then becomes

$$\ln[y_{s,t}] = \ln[\alpha_s] - \ln\left[1+\phi_s \exp[-\gamma_s(t-t_0)]\right] + \varepsilon^{\mathsf{b}}_{s,t}, \quad t = t_1, \ldots, t_T, \tag{2.5}$$

which is again *nonlinear* in the unknown quantities α_s, ϕ_s, and γ_s.

C. If the disturbance is additive but α_s is a known constant, say $\tilde{\alpha}_s$, the appropriate statistical specification might also be

$$y_{s,t} = \frac{\tilde{\alpha}_s}{1+\phi_s \exp[-\gamma_s(t-t_0)]} + \varepsilon^{\mathsf{c}}_{s,t}, \quad t = t_1, \ldots, t_T, \tag{2.6}$$

which is *nonlinear* in the two parameters ϕ_s and γ_s.

D. Finally, when α_s is the known constant $\tilde{\alpha}_s$, then an appropriate statistical specification of this logistic model could also be given by

$$\ln\left[\frac{\tilde{\alpha}_s - y_{s,t}}{y_{s,t}}\right] = \ln[\phi_s] - \gamma_s(t-t_0) + \varepsilon^{\mathsf{d}}_{s,t}, \quad t = t_1, \ldots, t_T, \tag{2.7}$$

which is *linear* in the two parameters $\ln[\phi_s]$ and γ_s.

Ordinary (unweighted) least squares techniques can be applied to each of these four models, and under the assumption that the specified disturbances are independent and identically normally distributed, the results will be MLE.[1] Note, however, that if $\boldsymbol{\varepsilon}^{\mathsf{a}}$ or $\boldsymbol{\varepsilon}^{\mathsf{c}}$ are assumed to be homoscedastic, then $\boldsymbol{\varepsilon}^{\mathsf{b}}$ and $\boldsymbol{\varepsilon}^{\mathsf{d}}$ must be heteroscedastic. Furthermore, because α_s is *not* known a priori for state per capita income data, the results obtained using nonlinear model (2.6) or linear model (2.7) are statistically questionable. That is, the arbitrary assignment of a fixed value to $\tilde{\alpha}_s$ creates a distortion in the estimated values of $\ln[\phi_s]$ and γ_s. In addition, the estimated variances

[1]Ideally, assumptions regarding the distribution of ε and how it enters the estimating equation should emerge as part of the derivation of $f_t[\boldsymbol{\beta}]$. Statistical techniques for evaluating these assumptions are provided, for example, by Greene [2], Ratkowsky [9], and Seber and Wild [10]. When the distribution of the error term cannot be hypothesized a priori, Godfrey and Wickens [11] and Leech [12] provide procedures for choosing the form of the error distribution based on analysis of the data.

and covariances of $\ln[\phi_s]$ and γ_s are likely to be incorrect, primarily because the uncertainty of α_s and its correlation with ϕ_s and γ_s is not captured in these two models.

Figure 1 is a graphical presentation of the least squares estimates of the growth parameters γ_s obtained using these four statistical representations of the logistic growth model. The response variables are real per capita income for the 48 contiguous United States over the period 1965 to 1991, adjusted to constant 1987 dollars [13, 14, 15]. For models (2.6) and (2.7), $\tilde{\alpha}_s$ is set to $1000 plus the largest real per capita income for state s over the observed time span, as suggested by Berry and Kaserman [8]. The values for ϕ_s and γ_s obtained using LLS and model (2.7) are used as starting values for the NLS procedure employed to solve model (2.6). The NLS estimated values for ϕ_s and γ_s along with the selected value of $\tilde{\alpha}_s$ are then used as starting values to fit models (2.4) and (2.5).

The thumbnail plots displayed in Figure 1 show that the asymptotic 95 percent confidence intervals for γ_s obtained using model (2.4) under the assumption that the real per capita income data include an *additive* Gaussian stochastic component are not in general significantly different from those obtained using model (2.5) under the assumption that the real per capita income data is log-normal. The results obtained using the linearized, two-parameter model (2.7), however, do not in general overlap the confidence intervals obtained using either model (2.4) or (2.5)—and frequently do not overlap the confidence intervals obtained using the nonlinear, two-parameter model (2.6)—thus showing sensitivity to the value chosen for $\tilde{\alpha}_s$. The estimates of economic growth obtained using the linear approximation (2.7) are therefore inconsistent with the nonlinear least squares estimates obtained using either model (2.4) or (2.5) and are often inconsistent with model (2.6) as well.

Given the accessibility of high quality nonlinear optimization software, it is arguably better to use NLS to solve for the parameters of a nonlinear model with additive error rather than to apply LLS to a linearization of the model: linearizing transformations change the statistical model, and it is unlikely that it can be determined beforehand under what circumstances the use of a linearizing transformation will make a significant difference in the results. The next section, however, shows that even NLS might not be adequate in the presence of measurement errors.

3. Measurement Errors and Orthogonal Distance Regression

Errors in variables, factor analysis, simultaneous equation models, and measurement error models are all names used to describe the least squares problem that arises when the explanatory variables $\mathbf{X}$ as well as the response variables $\mathbf{y}$ include measurement errors. The optimization procedure for the *linear* measurement error problem is sometimes called *total least squares* (TLS) [16]. The optimization procedure for the *nonlinear* measurement error problem is called *orthogonal distance regression* (ODR). For both linear and nonlinear models, the problem is that of finding parameter estimates that minimize the sum of the squares of the weighted orthogonal distances between each observed data point and the curve described by the modeling equation.

The issues surrounding measurement errors have been examined by a number of authors. Anderson [17, 18], for example, contains an excellent survey and explanation of the relationship between the various forms of measurement error models. Theoretical and computational analyses are also presented by Dagenais [19], Fuller [4], Morgan [20], and Van Huffel and Vandewalle [16]. Boggs et al. [21] describe the results of a Monte Carlo study of nonlinear measurement error problems. Aigner [22, 23], Chamberlain [24], Chamberlain and Griliches [25], DeLong [26], Greene [2], Griliches [27], Griliches and Ringstad [28], and Malinvaud [29] each discuss the use of measurement error models in econometrics.

The primary result of each of these references is that when both the response and the explanatory variables have significant measurement errors, then use of LLS or NLS is not always theoretically justified and will likely produce poor estimates. Other estimation methods should therefore be used. The method of instrumental variables (IV) is one commonly employed technique for solving economics measurement errors problems. Another procedure for solving measurement error problems is the method of orthogonal distance regression (ODR). Although not often employed by economists, ODR is an especially appropriate technique when the model is nonlinear or the variances of the measurement errors are known.

The importance of explicitly addressing measurement errors is demonstrated here using ODR and a variation of the analysis presented in the preceding section. Per capita incomes for the 48 contiguous United States are jointly modeled by

$$y^*_{s,t} \;=\; y^*_{s,t_0} \prod_{\tau=t_0}^{t-1} [1+\gamma_{s,\tau}] \;, \quad s=1,\ldots,48, \quad t=t_1,\ldots,t_T, \tag{3.1}$$

where $y^*_{s,t}$ again denotes the true per capita income for state s in year t, y^*_{s,t_0} denotes the true per capita income for state s in year t_0, and $\gamma_{s,t}$ is the rate of economic growth for state s in year t, predicted by

$$\gamma_{s,t} \;\stackrel{\text{def}}{=}\; \sum_{i=1}^{9} \alpha_i r_i + \phi_0 y_{s,t_0} + \sum_{j=1}^{8} \phi_j q_{s,t,j} \;\stackrel{\text{def}}{=}\; \mathbf{z}_{s,t}\boldsymbol{\beta} \;, \tag{3.2}$$

with r_i, denoting dummy variables that identify nine geographic regions of the United States, $q_{s,t,j}$, denoting the values of eight tax and expenditure variables for state s in year t, $\boldsymbol{\beta} \stackrel{\text{def}}{=} (\alpha_1,\ldots,\alpha_9,\phi_0,\phi_1,\ldots,\phi_8)'$ denoting the unknown parameters, and $\mathbf{z}_{s,t} \stackrel{\text{def}}{=} (r_1,\ldots,r_9,y_{s,t_0},q_{s,t,1},\ldots,q_{s,t,8})$ denoting all of the explanatory variables associated with observation t. The explanatory variables for state s at time t are:

- INITIAL INCOME—real per capita income in $t_0 = 1965$;
- CORPORATE INCOME TAX—real per capita state corporate income tax;
- INCOME TAX—real per capita individual state income tax;
- PROPERTY TAX—real per capita state property tax;
- SALES TAX—real per capita general state sales tax;
- OTHER TAXES—total real per capita state taxes less state corporate income, property, sales and individual income tax revenues;
- EDUCATION—real per capita spending on state institutions of higher education;

- HIGHWAYS—real per capita spending on highways;

- PUBLIC WELFARE—real per capita spending on public welfare;

All variables are adjusted to 1987 dollars using the national price deflator [13]. The five tax revenue variables exclude local taxes, and the variable EDUCATION does not include local government expenditures. Numerical characteristics of these data are listed in Table 2. The economic implications of the results presented here are discussed by Rogers in reference [30].

It is appropriate to use NLS to solve for the parameters of model (3.1) when the underlying statistical model is

$$\begin{aligned} y_{s,t} &= y_{s,t_0} \prod_{\tau=t_0}^{t-1} (1+\gamma_{s\tau}) \ + \ \varepsilon_{s,t} \\ \text{subject to: } \gamma_{s,t} &= \sum_{i=1}^{9} \alpha_i r_i + \beta_0 y_{s,t_0} + \sum_{j=1}^{8} \beta_j q_{s,t,j} \ , \end{aligned} \tag{3.3}$$

where $\varepsilon_{s,t}$ denotes the error in real per capita income for state s in year t. The statistical representation of model (3.1) that explicitly incorporates measurement error terms is

$$\begin{aligned} y_{s,t} &= (y_{s,t_0} + \varepsilon_{s,t_0}) \prod_{\tau=t_0}^{t-1} (1+\gamma_{s\tau}) \ + \ \varepsilon_{s,t} \\ \text{subject to: } \gamma_{s,t} &= \sum_{i=1}^{9} \alpha_i r_i + \beta_0 (y_{s,t_0} + \varepsilon_{s,t_0}) + \sum_{j=1}^{8} \beta_j (q_{s,t,j} + \delta_{s,t,j}) \ , \end{aligned} \tag{3.4}$$

where ε_{s,t_0} denotes the error in real per capita income for state s in year t_0, and $\delta_{s,t,j}$ is the error associated with explanatory variable $q_{s,t,j}$. Models (3.3) and (3.4) highlight the difference between solving this problem using NLS versus using ODR: the former assigns all of the random error to the response variable $y_{s,t}$, while the latter distributes this error among all of the stochastic variables.

The ODR estimators of the weighted minimization problem associated with model (3.1) are

$$\begin{aligned} \mathbf{b} \ &\stackrel{\text{def}}{=} \ \operatorname*{argmin}_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) \\ \text{subject to:} \quad S(\boldsymbol{\beta}) &= \sum_{s=1}^{48} \left[\omega_\varepsilon^2 \varepsilon_{s,t_0}^2 + \sum_{t=1}^{T} \left(\varepsilon_{s,t}^2 + \omega_v^2 \sum_{j=1}^{8} v_{s,t-1,j}^2 \right) \right] \ , \\ \varepsilon_{s,t} &= y_{s,t} - (y_{s,t_0} + \varepsilon_{s,t_0}) \prod_{\tau=t_0}^{t-1} (1+\gamma_{s,\tau}) \ , \quad t = t_1, \ldots, t_T, \\ \gamma_{s,t} &= (\mathbf{z}_{s,t} + \boldsymbol{v}_{s,t}) \boldsymbol{\beta} \ , \quad t = t_1, \ldots, t_T, \end{aligned} \tag{3.5}$$

where $\boldsymbol{v}_{s,t} \stackrel{\text{def}}{=} (v_{r_1}, \ldots, v_{r_9}, v_{y_{s,t_0}}, v_{q_{s,t,1}}, \ldots, v_{q_{s,t,8}})$ denotes the unobservable disturbances in the explanatory variables $\mathbf{z}_{s,t}$, with $v_{r_1} \equiv \cdots \equiv v_{r_9} \equiv 0$, $v_{y_{s,t_0}} \equiv \varepsilon_{y_{s,t_0}}$, and $v_{q_{s,t,j}} \equiv \delta_{q_{s,t,j}}$, $j = 1, \ldots, 8$, and ω_ε and ω_v are weights used to compensate for differences between σ_ε^2 and σ_v^2. When $\omega_\varepsilon \equiv \omega_v \equiv \infty$, the solution to model (3.5) is the same as that obtained using NLS for which $\boldsymbol{v}_{s,t} \equiv \mathbf{0} \ \forall \ t$.

The solution to problem (3.5) is found using ODRPACK [31], a portable collection of publicly available Fortran 77 subroutines designed specificly for the ODR problem.[2] Model (3.5) is encoded using a multiresponse format: the 26 values of real per capita income $y_{s,t}$, $t = 1966, \ldots, 1991$ form the "observation" for state $s = 1, \ldots, 48$. Using this multiresponse format allows the measurement error in y_{s,t_0}, as well as those in the explanatory variables $z_{s,t,j}$, to be handled correctly. Starting values for the NLS optimization are obtained using LLS and a linearization of model (3.5); the NLS results are then used as starting values for the ODR optimization. Derivatives are computed using ODRPACK's forward finite difference option, and default values are used for the convergence criteria.

Assuming the values for total state income and total state population are each accurate within approximately 1 percent, real per capita income values will be accurate within approximately 2 percent [33]. Similarly, although total tax revenues and expenditures are known precisely, the 1 percent accuracy of the state population values will result in 1 percent accuracy of the real per capita representations of these values. Use of the national price deflator to adjust the data to constant dollars also introduces errors in so far as relative and/or absolute purchasing power parity is not constant across states. The accuracy of the real per capita income values used in this study are therefore at best accurate within approximately $\pm\$250$ to $\pm\$350$ (2 to 3 percent), while tax and expenditure values are at best accurate within approximately $\pm\$1$.

To assess the effect of different magnitudes of the measurement errors ε_{s,t_0}, the analysis reported herein is repeated using $\omega_\varepsilon = 1, 2, \ldots, 10$. The ordinary NLS solution, which is equivalent to setting $\omega_\delta = \omega_\varepsilon = \infty$, is also reported. For each of the ODR analyses, $\omega_\delta = 50$.

The residual variance of the ODR solution is estimated by

$$\hat{\sigma}^2 = \frac{S(\mathbf{b})}{\nu}, \tag{3.6}$$

where $S(\mathbf{b})$ is the sum of the squares defined in model (3.5), $\mathbf{b}$ denotes the least squares estimators of the unknown parameters, and $\nu \stackrel{\text{def}}{=} n \times \bar{q} - p$ denotes the degrees of freedom in the fitted results. Here, $n = 48$ is the number of observations (states), $p = 18$ is the number of unknown parameters being estimated, and $\bar{q} \in [1, 26]$ is the number of independent responses per observation. For the results reported below, $\bar{q} = 13$ and $\nu = 48 \times 13 - 18 = 606$. This is a compromise between assuming that each response is effectively independent ($\bar{q} = 26$) as suggested by Bard [34, p. 195], and assuming that these multiple responses provide no additional degrees of freedom ($\bar{q} = 1$), as suggested by Bates and Watts [35, pp. 140–141].

The outcomes from the analyses of models (3.3) and (3.4) are shown graphically in Figures 2 and 3, and are listed in Table 3. The plots show how the 95 percent confidence intervals vary with ω_ε and ω_δ. The tables list the parameter estimates and their asymptotic standard errors, as well as the t-value for the null hypothesis that the estimated parameter value is 0 and the two-sided significance level at which this null hypothesis would be rejected. The standard errors are derived

[2] ODRPACK is an implementation of the ODR algorithm developed by Boggs, Byrd, and Schnabel [32]. It has been used to solve problems in many different fields on machine architectures ranging from PCs to supercomputers. ODRPACK accommodates many levels of user sophistication and problem difficulty; implicit as well as explicit models; multiresponse data; and correlation within the components of a multidimensional observation. Source code and documentation for ODRPACK are available from `http://www.netlib.org/netlib/odrpack/`.

under the assumption that there are no important omitted variables and that the covariance matrix of the response variables is diagonal. The validity of this assumption will be explored in future work.

These figures and tables document the consequences of ignoring measurement errors in the explanatory variables. The outcome of tests of the hypothesis that the parameters are equal to 0 are different at the 5 percent significance level for half of the explanatory variables depending on whether NLS or ODR is employed. Furthermore, when NLS is employed ($\omega_\varepsilon = \infty$), the coefficient on INITIAL INCOME is *negative* and significantly different from 0 at the 0.1 percent significance level; when ODR is applied and $10 \leq \omega_\varepsilon \leq 4$, this coefficient is not significantly different from 0 at the 5 percent significance level; and, when $\omega_\varepsilon \leq 3$, it is *positive* and significantly different from 0 at the 0.1 percent significance level. (Similar results are reported in DeLong [26].) The conclusion drawn from the NLS analysis of this dataset with respect to the coefficient for INITIAL INCOME, which is important for the test of income convergence across states and regions, is thus different than that drawn from an ODR analysis when $\omega_\varepsilon = 10$, and is diametrically opposite that drawn from the ODR analysis with $\omega_\varepsilon \geq 3$. Note also that these differences are *less* significant than those that would have been obtained if the Bard estimate of the number of degrees of freedom had been employed.

The measurement errors estimated using $\omega_\varepsilon = 1$ are shown in Figure 4, and those obtained when $\omega_\varepsilon = 4$ are shown in Figure 5. When $\omega_\varepsilon = 4$, the variance of the measurement errors ε_{st_0} is approximately \$300, or 2.5 percent of the mean adjusted real per capita income value, \$12 280. The mean is computed using adjusted real per capita income data from years 1965 to 1992; the median value is \$12 089. When $\omega_\varepsilon = 1$, this variance is approximately \$600, or 5 percent of the mean adjusted real per capita income. The measurement errors for the other explanatory variables show some pattern between states that needs further exploration.

Also, model (3.4) implicitly specifies that the coefficients of the independent variables are homogeneous across states. Canto and Webb [36] reject this hypothesis, conjecturing that tax levels and degrees of progressivity are likely to be different, and hence result in significantly different values for the various coefficients. While it is not possible to make a definitive test of this hypothesis for the analysis reported here, Figures 4 and 5 do show that in a number of instances the measurement errors for a given explanatory variable and state are either always positive or always negative, possibly indicating that the hypothesis of homogeneity should be rejected.

4. Conclusions

This paper describes two studies of United States per capita income data that show that common mathematical transformations performed on economic models in order to permit use of simpler least squares procedures can significantly alter the results obtained. The empirical evidence presented herein, although limited in scope, highlights the importance of specifically testing assumptions made during any given analyses. The first example shows that the results obtained by applying linear least squares to a linearization of an inherently nonlinear model are significantly different from those obtained when nonlinear least squares is applied directly to the nonlinear model. The second illustrates the potential consequences of not explicitly modeling measurement errors in economic data. Together, the results of these two studies demonstrate the importance—and computational feasibility—of formulating economic models precisely.

I am grateful for beneficial discussions with Jim Alm and Don Waldman (University of Colorado Department of Economics); and for helpful comments and suggestions from Ron Smith (Department of Economics at Birkbeck College, London, and visiting professor at University of Colorado Department of Economics); Paul Domich, Tony Kearsley, and Kevin Coakley (NIST Information Technology Laboratory); Matt Young (NIST Electronics and Electrical Engineering Laboratory); Edie DeWeese (NIST Publication Services Office); and Bruce McCullough (Federal Communications Commission).

5. References

[1] Davidson, R.; MacKinnon, J. G. Estimation and Inference in Econometrics. New York, NY: Oxford University Press; 1993.

[2] Greene, W. H. Econometric Analysis, third edition. Upper Saddle River, NJ: Prentice Hall; 1997.

[3] Judge, G. G.; Griffiths, W. E.; Hill, R. C.; Lütkepohl, H.; Lee, T. The Theory and Practice of Econometrics, second edition. New York, NY: John Wiley and Sons; 1985.

[4] Fuller, W. A. Measurement Error Models. New York, NY: John Wiley and Sons; 1987.

[5] Griliches, Z. Hybrid corn: An exploration in the economics of technological change. Econometrica 24: 501–522; 1957 October.

[6] Mansfield, E. Technical change and the rate of imitation. Econometrica 29: 741–766; 1961 October.

[7] Clark, D. P.; Kaserman, D. L.; Anantanasuwong, D. A diffusion model of industrial sector growth in developing economies. World Development 21: 421–428; 1993.

[8] Berry, D. M.; Kaserman, D. L. A diffusion model of long-run state economic development. Atlantic Economic Journal 21(4): 39–54; 1993 December.

[9] Ratkowsky, D. A. Nonlinear Regression Modeling. New York, NY: Marcel Dekker; 1983.

[10] Seber, G. A. F.; Wild, C. J. Nonlinear Regression. New York, NY: John Wiley and Sons; 1989.

[11] Godfrey, L.; Wickens, M. Testing linear and log-linear regressions for functional form. Review of Economic Studies 48: 487–496; 1981.

[12] Leech, D. Testing the error specification in nonlinear regression. Econometrica 43(4): 719–725; 1975 July.

[13] Economics and Statistics Administration, Bureau of the Census (U.S). Statistical Abstract of the United States; 1965-1991.

[14] Income Statistics Branch/HHES Division, Bureau of the Census (U.S.). Current Population Reports, Series P60; 1965-1991. ftp://ftp.census.gov/pub/income. [accessed 9/28/96].

[15] Income Statistics Branch/HHES Division, Bureau of the Census (U.S.). State Government Finances Reports; 1965-1991. ftp://ftp.census.gov/pub/govs/state. [accessed 9/28/96].

[16] Huffel, S. V.; Vandewalle, J. The Total Least Squares Problem: Computational Aspects and Analysis. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 1991.

[17] Anderson, T. W. Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics. Journal of the Royal Statistical Society (Series B) 38: 1–36; 1976.

[18] Anderson, T. W. Estimating linear statistical relationships. The Annals of Statistics 12: 1–45; 1984.

[19] Dagenais, M. G. Parameter estimation in regression models with errors in the variables and autocorrelated disturbances. Journal of Econometrics 64(1-2): 145–163; 1994 September/October.

[20] Moran, T. A. P. Estimating structural and functional relationships. Journal of Multivariate Analysis 1: 232–255; 1971.

[21] Boggs, P. T.; Donaldson, J. R.; Schnabel, R. B.; Spiegelman, C. H. A computational examination of orthogonal distance regression. Journal of Econometrics 38(1/2): 169–201; 1988.

[22] Aigner, D. J. Regression with binary independent variable subject to errors of observation. Journal of Econometrics 1: 49–60; 1973.

[23] Aigner, D. J. MSE dominance of least squares with errors of observation. Journal of Econometrics 2: 365–372; 1974.

[24] Chamberlain, G. Education, income, and ability revisited. Journal of Econometrics 5: 241–257; 1977.

[25] Chamberlain, G.; Griliches, Z. Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers. International Economic Review 16: 422–449; 1975.

[26] DeLong, J. B. Productivity growth, convergence, and welfare: Comment. The American Economic Review 78(5): 1138–1154; 1988.

[27] Griliches, Z. Errors in variables and other unobservables. Econometrica 42: 971–998; 1974.

[28] Griliches, Z.; Ringstad, V. Errors-in-the-variables bias in nonlinear contexts. Econometrica 38: 368–370; 1970.

[29] Malinvaud, E. Statistical Methods of Econometrics. New York, NY: American Elsevier Publishing Company; 1970.

[30] Rogers, J. E. Public policy and economic growth in the United States: An empirical study. National Institute of Standards and Technology NISTIR 5067; 1997.

[31] Boggs, P. T.; Byrd, R. H.; Rogers, J. E.; Schnabel, R. B. User's reference guide for ODRPACK Version 2.01: Software for weighted orthogonal distance regression. National Institute of Standards and Technology NISTIR 4834; 1992.

[32] Boggs, P. T.; Byrd, R. H.; Schnabel, R. B. A stable and efficient algorithm for nonlinear orthogonal distance regression. Society for Industrial and Applied Mathematics (SIAM) Journal of Scientific and Statistical Computing 8(6): 1052–1078; 1987.

[33] Ku, H. H., editor. Precision Measurement and Calibration: Selected Papers on Statistical Concepts and Procedures. National Bureau of Standards (U.S.) Special Publication 300 Volume 1; 1969.

[34] Bard, Y. Nonlinear Parameter Estimation. New York, NY: Academic Press; 1974.

[35] Bates, D. M.; Watts, D. G. Nonlinear Regression Analysis and Its Applications. New York, NY: John Wiley and Sons; 1988.

[36] Canto, V.; Webb, R. I. The effect of state fiscal policy on state relative economic performance. Southern Economic Journal 54: 186–202; 1987 July.

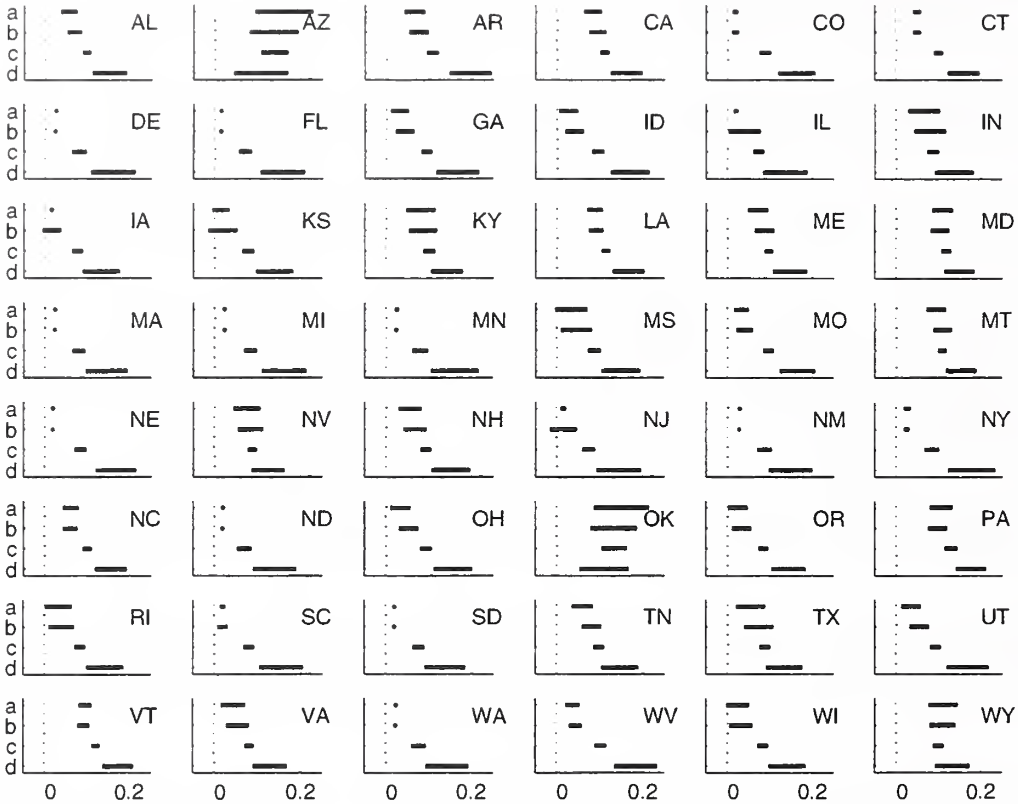

Figure 1. 95 percent asymptotic confidence intervals for γ. The x-axis limits for all states are -0.05 to 0.25, and the position of 0 is shown as the vertical dotted line. Plotted results are as follows.

Identifier	Function	Model	Error in y_{st}	Treatment of α_s
a	2.4	Nonlinear	Additive	Estimated
b	2.5	Nonlinear	Multiplicative	Estimated
c	2.6	Nonlinear	Additive	Known constant
d	2.7	Linear	—	Known constant

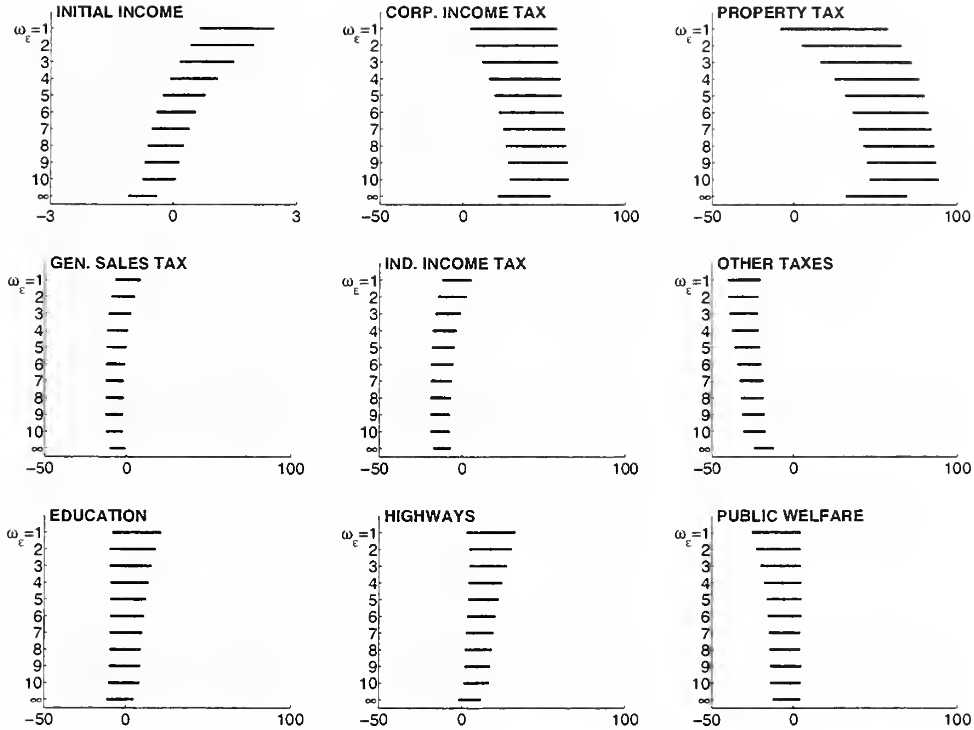

Figure 2. 95 percent confidence intervals for coefficients of explanatory variables. The weights ω_ε that were used to obtain the solutions are shown along vertical axis, and the position of 0 is shown by the vertical dotted line. All parameter values have been multiplied by 10^6.

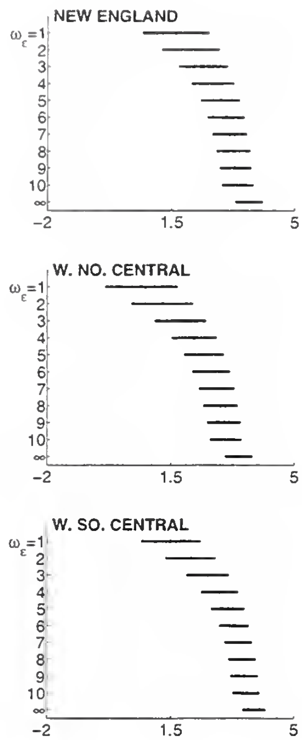

Figure 3. 95 percent confidence intervals for regional constants. The weights ω_ε that were used to obtain the solutions are shown along vertical axis, and the position of 0 is shown by the vertical dotted line. All parameter values have been multiplied by 10^2.

Figure 4. Measurement errors (y-axis) versus state (x-axis) for $\omega_\epsilon = 1$ and $\omega_\delta = 50$. All measurement errors are in constant 1987 dollars. The dotted horizontal line centered in each plot indicates the position of 0. The dashed horizontal lines indicate the position of two standard errors above and below 0, and the middle tick label on each vertical axis is the value of one standard error.

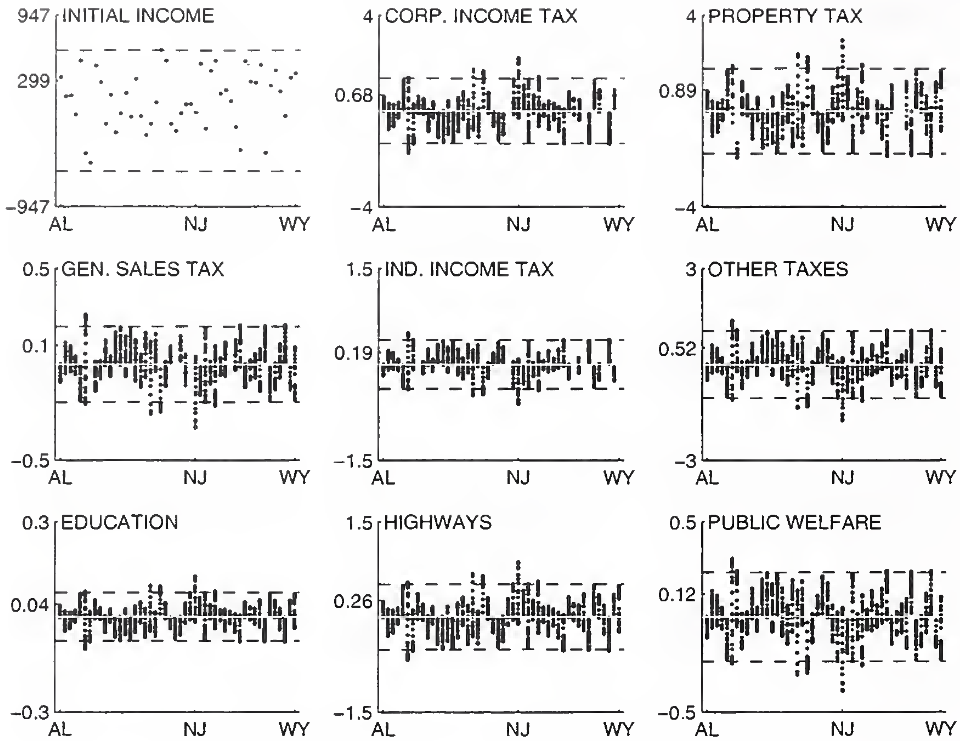

Figure 5. Measurement errors (y-axis) versus state (x-axis) for $\omega_\epsilon = 4$ and $\omega_\delta = 50$. All measurement errors are in constant 1987 dollars. The dotted horizontal line centered in each plot indicates the position of 0. The dashed horizontal lines indicate the position of two standard errors above and below 0, and the middle tick label on each vertical axis is the value of one standard error.

Table 2. Response and explanatory variables.

Variables		Mean	Median	Minimum	Maximum
Response:	PER CAPITA INCOME	12401.41	12200.56	5652.90	22618.77
Explanatory:	INITIAL INCOME	9087.61	9117.42	5652.90	11956.17
	CORPORATE INCOME TAXES	62.95	50.19	0.00	2823.85
	PROPERTY TAXES	23.09	5.01	0.00	1802.04
	SALES TAXES	231.49	232.59	0.00	811.15
	INDIVIDUAL INCOME TAXES	184.50	167.54	0.00	925.00
	OTHER TAXES	315.49	261.46	24.81	5023.03
	HIGHER EDUCATION	205.10	198.32	1.02	662.36
	HIGHWAYS	214.91	189.25	67.92	1402.01
	PUBLIC WELFARE	216.17	194.25	30.66	727.44

- INITIAL INCOME values are computed for the year 1965.
- All other values are computed using data from years 1965 to 1992.
- All data are converted to 1987 dollars using the national price deflator [13].

Table 3. Statistics for estimated parameters.

Variable	Weights ω_ϵ	ω_δ	Parameter value $\times 10^{-6}$	Standard error $\times 10^{-6}$	t-value	Two-sided significance level %
INITIAL INCOME	1	50	1.548	0.451	3.434	0.06
	2	50	1.197	0.388	3.081	0.22
	3	50	0.824	0.330	2.496	1.28
	4	50	0.513	0.288	1.784	7.49
	5	50	0.270	0.259	1.043	29.72
	6	50	0.080	0.239	0.335	73.76
	7	50	−0.065	0.226	−0.287	77.46
	8	50	−0.176	0.217	−0.810	41.84
	9	50	−0.264	0.211	−1.253	21.06
	10	50	−0.333	0.206	−1.617	10.65
	∞	∞	−0.745	0.177	−4.219	0.00
CORPORATE INCOME TAXES	1	50	31.006	13.202	2.349	1.92
	2	50	32.885	12.539	2.623	0.89
	3	50	35.285	11.667	3.024	0.26
	4	50	37.621	10.866	3.462	0.06
	5	50	39.764	10.240	3.883	0.01
	6	50	41.649	9.788	4.255	0.00
	7	50	43.310	9.469	4.574	0.00
	8	50	44.727	9.244	4.839	0.00
	9	50	45.892	9.082	5.053	0.00
	10	50	46.860	8.965	5.227	0.00
	∞	∞	37.555	8.004	4.692	0.00
PROPERTY TAXES	1	50	24.945	16.372	1.524	12.81
	2	50	35.560	15.248	2.332	2.00
	3	50	44.501	13.986	3.182	0.15
	4	50	50.966	12.924	3.944	0.01
	5	50	55.681	12.133	4.589	0.00
	6	50	59.334	11.566	5.130	0.00
	7	50	62.173	11.172	5.565	0.00
	8	50	64.402	10.897	5.910	0.00
	9	50	66.225	10.699	6.190	0.00
	10	50	67.646	10.554	6.410	0.00
	∞	∞	50.562	9.317	5.427	0.00

Table 3. Statistics for estimated parameters, continued.

Variable	Weights ω_ϵ	Weights ω_δ	Parameter value $\times 10^{-6}$	Standard error $\times 10^{-6}$	t-value	Two-sided significance level %
GENERAL SALES TAXES	1	50	0.715	3.813	0.188	85.13
	2	50	−2.193	3.629	−0.604	54.59
	3	50	−4.324	3.398	−1.272	20.37
	4	50	−5.600	3.185	−1.758	7.92
	5	50	−6.338	3.017	−2.101	3.61
	6	50	−6.769	2.891	−2.342	1.95
	7	50	−7.042	2.800	−2.515	1.22
	8	50	−7.227	2.736	−2.642	0.85
	9	50	−7.347	2.688	−2.733	0.65
	10	50	−7.430	2.653	−2.801	0.53
	∞	∞	−5.409	2.370	−2.282	2.28
INDIVIDUAL INCOME TAXES	1	50	−3.368	4.424	−0.761	44.67
	2	50	−6.235	4.160	−1.499	13.45
	3	50	−8.807	3.847	−2.289	2.24
	4	50	−10.547	3.576	−2.949	0.33
	5	50	−11.623	3.371	−3.448	0.06
	6	50	−12.242	3.221	−3.800	0.02
	7	50	−12.646	3.118	−4.056	0.01
	8	50	−12.924	3.049	−4.238	0.00
	9	50	−13.085	2.997	−4.366	0.00
	10	50	−13.197	2.958	−4.461	0.00
	∞	∞	−12.125	2.651	−4.573	0.00
OTHER TAXES	1	50	−30.454	4.964	−6.135	0.00
	2	50	−31.117	4.684	−6.643	0.00
	3	50	−30.732	4.351	−7.063	0.00
	4	50	−29.681	4.051	−7.327	0.00
	5	50	−28.441	3.815	−7.455	0.00
	6	50	−27.213	3.640	−7.476	0.00
	7	50	−26.173	3.523	−7.430	0.00
	8	50	−25.326	3.430	−7.383	0.00
	9	50	−24.603	3.368	−7.305	0.00
	10	50	−24.020	3.321	−7.233	0.00
	∞	∞	−18.059	2.953	−6.115	0.00

Table 3. Statistics for estimated parameters, continued.

Variable	Weights ω_ϵ	ω_δ	Parameter value $\times 10^{-6}$	Standard error $\times 10^{-6}$	t-value	Two-sided significance level %
HIGHER EDUCATION	1	50	6.586	7.570	0.870	38.46
	2	50	4.174	6.990	0.597	55.07
	3	50	2.946	6.363	0.463	64.36
	4	50	2.166	5.844	0.371	71.10
	5	50	1.476	5.460	0.270	78.69
	6	50	0.703	5.186	0.136	89.22
	7	50	0.114	5.004	0.023	98.18
	8	50	−0.315	4.872	−0.065	94.86
	9	50	−0.767	4.780	−0.161	87.25
	10	50	−1.128	4.712	−0.239	81.09
	∞	∞	−3.260	4.183	−0.779	43.60
HIGHWAYS	1	50	17.811	7.389	2.410	1.62
	2	50	17.571	6.496	2.705	0.70
	3	50	16.250	5.676	2.863	0.43
	4	50	14.658	5.068	2.892	0.40
	5	50	13.208	4.650	2.840	0.47
	6	50	11.967	4.355	2.748	0.62
	7	50	10.971	4.159	2.638	0.86
	8	50	10.178	4.020	2.532	1.16
	9	50	9.540	3.920	2.433	1.52
	10	50	9.038	3.847	2.349	1.91
	∞	∞	4.971	3.375	1.473	14.12
PUBLIC WELFARE	1	50	−10.590	7.364	−1.438	15.09
	2	50	−9.350	6.794	−1.376	16.92
	3	50	−7.799	6.189	−1.260	20.81
	4	50	−6.631	5.694	−1.165	24.46
	5	50	−5.889	5.329	−1.105	26.96
	6	50	−5.461	5.072	−1.077	28.21
	7	50	−5.212	4.894	−1.065	28.74
	8	50	−5.061	4.770	−1.061	28.90
	9	50	−4.985	4.680	−1.065	28.73
	10	50	−4.944	4.615	−1.071	28.45
	∞	∞	−4.159	4.121	−1.009	31.33

Table 3. Statistics for estimated parameters, continued.

Variable	Weights ω_ϵ	ω_δ	Parameter value $\times 10^{-6}$	Standard error $\times 10^{-6}$	t-value	Two-sided significance level %
NEW ENGLAND						
(CT, ME, MA, NH, RI, VT)	1	50	1.642	0.468	3.506	0.05
	2	50	2.049	0.404	5.069	0.00
	3	50	2.408	0.346	6.969	0.00
	4	50	2.682	0.303	8.855	0.00
	5	50	2.887	0.274	10.533	0.00
	6	50	3.046	0.255	11.964	0.00
	7	50	3.164	0.242	13.091	0.00
	8	50	3.253	0.233	13.975	0.00
	9	50	3.325	0.226	14.685	0.00
	10	50	3.381	0.222	15.248	0.00
	∞	∞	3.709	0.191	19.402	0.00
MIDDLE ATLANTIC						
(NJ, NY, PA)	1	50	1.243	0.483	2.573	1.03
	2	50	1.711	0.414	4.137	0.00
	3	50	2.129	0.350	6.076	0.00
	4	50	2.446	0.305	8.021	0.00
	5	50	2.679	0.274	9.761	0.00
	6	50	2.855	0.254	11.239	0.00
	7	50	2.986	0.240	12.428	0.00
	8	50	3.084	0.231	13.334	0.00
	9	50	3.161	0.225	14.068	0.00
	10	50	3.220	0.220	14.649	0.00
	∞	∞	3.567	0.188	18.935	0.00
EAST NORTH CENTRAL						
(IL, IN, MI, OH, WI)	1	50	0.252	0.531	0.475	63.52
	2	50	0.826	0.452	1.829	6.80
	3	50	1.338	0.381	3.510	0.05
	4	50	1.721	0.331	5.207	0.00
	5	50	2.002	0.297	6.742	0.00
	6	50	2.212	0.274	8.068	0.00
	7	50	2.368	0.259	9.133	0.00
	8	50	2.484	0.249	9.972	0.00
	9	50	2.576	0.242	10.652	0.00
	10	50	2.647	0.236	11.195	0.00
	∞	∞	3.045	0.203	15.013	0.00

Table 3. Statistics for estimated parameters, continued.

Variable	Weights ω_ϵ	ω_δ	Parameter value $\times 10^{-6}$	Standard error $\times 10^{-6}$	t-value	Two-sided significance level %
WEST NORTH CENTRAL						
(IA, KS, MN, MO, NE, ND, SD)	1	50	0.663	0.513	1.291	19.71
	2	50	1.255	0.435	2.887	0.40
	3	50	1.776	0.366	4.859	0.00
	4	50	2.161	0.317	6.826	0.00
	5	50	2.439	0.284	8.586	0.00
	6	50	2.646	0.262	10.091	0.00
	7	50	2.798	0.248	11.294	0.00
	8	50	2.912	0.238	12.229	0.00
	9	50	3.001	0.231	12.984	0.00
	10	50	3.070	0.226	13.588	0.00
	∞	∞	3.430	0.194	17.710	0.00
SOUTH ATLANTIC						
(DE, FL, GA, MD, NC, SC, VA, WV)	1	50	1.327	0.458	2.898	0.39
	2	50	1.875	0.387	4.843	0.00
	3	50	2.358	0.325	7.257	0.00
	4	50	2.713	0.281	9.655	0.00
	5	50	2.967	0.252	11.785	0.00
	6	50	3.155	0.232	13.588	0.00
	7	50	3.293	0.219	15.014	0.00
	8	50	3.395	0.210	16.130	0.00
	9	50	3.475	0.204	17.024	0.00
	10	50	3.537	0.199	17.733	0.00
	∞	∞	3.848	0.171	22.566	0.00
EAST SOUTH CENTRAL						
(AL, KY, MS, TN)	1	50	1.399	0.418	3.348	0.09
	2	50	1.908	0.356	5.364	0.00
	3	50	2.350	0.301	7.814	0.00
	4	50	2.672	0.262	10.204	0.00
	5	50	2.900	0.236	12.295	0.00
	6	50	3.067	0.218	14.045	0.00
	7	50	3.188	0.207	15.401	0.00
	8	50	3.277	0.199	16.478	0.00
	9	50	3.347	0.193	17.324	0.00
	10	50	3.400	0.189	17.994	0.00
	∞	∞	3.690	0.163	22.674	0.00

Table 3. Statistics for estimated parameters, continued.

Variable	Weights ω_ϵ	ω_δ	Parameter value $\times 10^{-6}$	Standard error $\times 10^{-6}$	t-value	Two-sided significance level %
WEST SOUTH CENTRAL						
(AR, LA, OK, TX)	1	50	1.521	0.420	3.618	0.03
	2	50	2.082	0.354	5.883	0.00
	3	50	2.557	0.296	8.635	0.00
	4	50	2.896	0.256	11.306	0.00
	5	50	3.132	0.230	13.624	0.00
	6	50	3.303	0.212	15.546	0.00
	7	50	3.426	0.201	17.053	0.00
	8	50	3.517	0.193	18.197	0.00
	9	50	3.587	0.188	19.109	0.00
	10	50	3.641	0.184	19.830	0.00
	∞	∞	3.867	0.158	24.479	0.00
MOUNTAIN						
(AZ, CO, ID, MT, NV, NM, UT, WY)	1	50	0.674	0.507	1.329	18.43
	2	50	1.211	0.435	2.783	0.55
	3	50	1.685	0.370	4.554	0.00
	4	50	2.038	0.323	6.308	0.00
	5	50	2.294	0.291	7.872	0.00
	6	50	2.486	0.270	9.210	0.00
	7	50	2.626	0.256	10.273	0.00
	8	50	2.730	0.246	11.105	0.00
	9	50	2.813	0.239	11.775	0.00
	10	50	2.876	0.234	12.309	0.00
	∞	∞	3.250	0.201	16.180	0.00
PACIFIC						
(CA, OR, WA)	1	50	0.384	0.550	0.698	48.56
	2	50	0.919	0.473	1.943	5.25
	3	50	1.396	0.402	3.471	0.06
	4	50	1.757	0.351	5.003	0.00
	5	50	2.022	0.317	6.380	0.00
	6	50	2.222	0.294	7.566	0.00
	7	50	2.370	0.279	8.508	0.00
	8	50	2.480	0.268	9.251	0.00
	9	50	2.568	0.261	9.850	0.00
	10	50	2.635	0.255	10.328	0.00
	∞	∞	3.103	0.219	14.138	0.00