Estimation of Short-Run and Long-Run Elasticities of Energy Demand From Panel Data Using Shrinkage Estimators

G. S. MADDALA
Department of Economics, The Ohio State University, Columbus, OH 43210-1172

Robert P. TROST
Department of Economics, The George Washington University, Washington, DC 20052

Hongyi LI
Department of Management, The Chinese University of Hong Kong, Shatin N.T., Hong Kong

Frederick JOUTZ
Department of Economics, The George Washington University, Washington, DC 20052

This article discusses the problem of obtaining short-run and long-run elasticities of energy demand for each of 49 states in the United States using data for 21 years. Estimation using the time series data by each state gave several wrong signs for the coefficients. Estimation using pooled data was not valid because the hypothesis of homogeneity of the coefficients was rejected. Shrinkage estimators gave more reasonable results. The article presents in a unified framework the classical, empirical Bayes, and Bayes approaches for deriving these estimators.

KEY WORDS: Bayesian shrinkage estimator; Random-coefficient model; Stein-rule estimator.

In the analysis of panel data, it is customary to pool the observations, with or without individual-specific dummies. These dummy variables are assumed to be fixed (fixed-effects models) or random (random-effects or variance-components models). In these models the slope coefficients are assumed to be equal. The homogeneity of the slope coefficients is often an unreasonable assumption, and one can allow for cross-sectional heterogeneity and/or heterogeneity over time. A commonly used model to allow for cross-sectional heterogeneity is the random-coefficient model, in which the parameters are assumed to come from a common distribution. A commonly used model to allow for heterogeneity over time is the time-varying parameter model, which can be estimated by state-space methods. In any case, one needs to impose some structure on the coefficients if one allows for parameter heterogeneity.

In this article we will concentrate on cross-sectional heterogeneity only. The article considers several shrinkage-type estimators and illustrates their use in the estimation of short-run and long-run price and income elasticities of residential demand for electricity and natural gas in the United States based on a panel data on 49 states over 21 years (1970–1990). Robertson and Symons (1992) and Pesaran and Smith (1995) discussed the biases that are likely to occur in the estimation of long-run elasticities if parameter heterogeneity is ignored and the data are pooled. They assumed a random-coefficient model and were interested in the mean of the coefficients. The focus of this article is on the estimation of the individual parameters for each cross-section under parameter heterogeneity of the form considered by these authors. The framework used is the random-coefficient model, and even if interest centers on the mean of the coefficients, it will be pointed out that there are some problems with the use of two-step procedures in the general least squares (GLS) estimation, a procedure that is most commonly used. Moreover, an iterative procedure will be suggested and illustrated with the empirical example.

Because a review of the estimation of energy-demand models can be found in the book by Berndt (1991), we will not elaborate on the derivation of the demand model estimated. The model we use is a standard dynamic linear regression (DLR) model derived by taking account of error-correction and partial-adjustment mechanisms. For a review of these, see Nickell (1985) and Alogoskoufis and Smith (1991).

1. ALTERNATIVE ESTIMATION METHODS FOR THE ESTIMATION OF THE INDIVIDUAL HETEROGENEOUS PARAMETERS

Consider the linear model

$$y_i \sim N(X_i\beta_i, \sigma_i^2 I), \qquad i = 1, 2, \dots, N, \tag{1}$$

where y_i is a $T \times 1$ vector, X_i is a $T \times k$ matrix of observations for the ith cross-section, and β_i is a $k \times 1$ vector of parameters. Unlike pure cross-section data, panel data allow us to take care of the dynamic structure. We will, therefore, assume that X_i include lagged values of y_i. This

Journal of Business & Economic Statistics
January 1997, Vol. 15, No. 1

creates problems when we consider the random-coefficient model later. In this case the assumption of strict exogeneity of X_i is no longer valid, and the results we discuss are valid only asymptotically under the usual regularity conditions assumed in dynamic regression models.

With panel data, however, there is the question of whether to pool the data and obtain a single estimate from the whole sample or to estimate the equations separately for each cross-section. The implicit assumption in both the fixed-effects and random-effects models for pooling the data is that the slope coefficients are all the same for all the cross-section units. This may not be a tenable assumption. In practice, the null hypothesis of constancy of slope parameters across the different cross-section units is often rejected. This implies that the equations should be estimated separately for each cross-section rather than obtaining an overall pooled estimate.

The problem with the two usual estimation methods of either pooling the data or obtaining separate estimates for each cross-section is that both are based on extreme assumptions. If the data are pooled, it is assumed that the parameters are all the same. If separate estimates are obtained for each cross-section, it is assumed that the parameters are all different in each cross-section. The truth probably lies somewhere in between. The parameters are not exactly the same, but there is some similarity between them. One way of allowing for the similarity is to assume that the parameters all come from a joint distribution with a common mean and a nonzero covariance matrix. It will be shown later that the resulting parameter estimates will be a weighted average of the overall pooled estimate and the separate time series estimates based on each cross-section. Thus, each cross-section estimate is "shrunk" toward the overall pooled estimate.

The idea of shrinkage occurs frequently in the literature on prediction (e.g., see Copas 1983; Rao 1987). Rubin (1980) also provided evidence on better predictions with shrinkage estimators [although Scott (1980) in her discussion of the article, disputed their usefulness]. He applied an empirical Bayes technique for law-school validation studies. The studies are primarily concerned with the prediction of first-year grades in law school from the LSAT score and undergraduate grade-point average. Traditionally, a separate admitting equation is estimated for each law school by the method of least squares by using data for students who attended that school in recent years. These least squares estimates can fluctuate wildly from year to year. The study by Rubin argued that the estimation of the admitting equations by the empirical Bayes methods provides more stable estimates and better predictions.

A shrinkage estimator sometimes suggested is the so-called Stein-rule estimator defined by

$$\tilde{\beta}_i = \left(1 - \frac{c}{F}\right)\hat{\beta}_i + \frac{c}{F}\hat{\beta}_p, \tag{2}$$

where $\hat{\beta}_i$ is the ordinary least squares (OLS) estimator and $\hat{\beta}_p$ is the estimator from the pooled regression. F is the test statistic to test the null hypothesis

$$\beta_1 = \beta_2 = \cdots = \beta_N = \beta. \tag{3}$$

Under the null hypothesis (3), F has an F distribution with degrees of freedom $(N-1)k$ and $N(T-k)$, where k is the dimension of β. The optimal value of the constant c suggested by Judge and Bock (1978, pp. 190–198) is

$$c = \frac{(N-1)k-2}{NT-Nk+2}. \tag{4}$$

This rule has, however, been derived for the case of strictly exogenous X's. Zeimer and Wetzstein (1983) applied the Stein-rule estimator to a wilderness-demand model and argued that the Stein rule gives better forecasts than the pooled or the individual cross-section estimators. The Stein rule shrinks the estimators $\hat{\beta}_i$ toward the pooled estimator $\hat{\beta}_p$. The factor c in (4) is roughly $k/(T-k)$ for large n. Thus, the higher the number of explanatory variables k relative to T, the smaller will be the shrinkage factor $(1-c/F)$ for given F.

1.1 The Random-Coefficient Model

For ease of exposition, in the following presentation we will assume that X_i does not include lagged dependent variables. In the presence of lagged dependent variables, the results hold only asymptotically.

In the random-coefficient model

$$y_i = X_i\beta_i + u_i, \tag{5}$$

we assume that

$$\beta_i \sim IN(\mu, \Sigma). \tag{6}$$

We can write this as

$$\beta_i = \mu + v_i, \tag{7}$$

where $v_i \sim N(0, \Sigma)$.

In the classical random-coefficient model, Equations (5) and (7) imply that

$$y_i = X_i\mu + w_i, \tag{8}$$

where $w_i \sim IN(0, \Omega_i)$ with $\Omega_i = (X_i \Sigma X_i' + \sigma_i^2 I)$. Assuming that the u_i are independent across the N equations, we get the GLS estimator μ^* of μ as

$$\mu^* = \left(\sum_{i=1}^{N} X_i'\Omega_i^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\Omega_i^{-1}y_i\right).$$

Using the matrix identity

$$(A + BDB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B + D^{-1})^{-1}B'A^{-1}$$

(see Rao 1973, p. 33) to decompose Ω_i, Swamy (1970) showed that

$$\mu^* = \sum_{i=1}^{N} W_i\hat{\beta}_i,$$

where $\hat{\beta}_i = (X_i'X_i)^{-1}X_i'y_i$ is the OLS estimator of β_i and

$$W_i = \left(\sum_{j=1}^{N} P_j^{-1}\right)^{-1} P_i^{-1} \quad (9)$$

with

$$P_i = (\sigma_i^2(X_i'X_i)^{-1} + \Sigma). \quad (10)$$

Two obstacles in applying the GLS procedure are the unknown parameters Σ and σ_i^2 in Ω. Swamy (1970) proposed a two-step procedure that uses the least squares estimators $\hat{\beta}_i$ of β_i and the least squares residuals to obtain unbiased estimators of Σ and σ_i^2. Of course, in dynamic models with lagged dependent variables, the unbiasedness property does not hold. We have to appeal to the consistency property. These estimators were also used by Rao (1975) in the empirical Bayes procedure. The Swamy (1970) method for estimating Σ and σ_i^2 is

$$\hat{\Sigma} = \frac{1}{N-1}\sum_i \left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right) \times \left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right)' - \frac{1}{N}\sum_i (X_i'X_i)^{-1}\hat{\sigma}_i^2, \quad (11)$$

where $\hat{\sigma}_i^2 = [1/(T-k)](y_i - X_i\hat{\beta}_i)'(y_i - X_i\hat{\beta}_i)$ and $\hat{\beta}_i$ is the OLS estimator of β_i.

One potential problem with the estimator for Σ is that it may not be positive definite. In this case Swamy suggests using only the first part of Equation (11) to estimate Σ. Hsiao (1986) adopted the same strategy. The suggested estimator, which is consistent in this case, is

$$\hat{\Sigma} = \frac{1}{N-1}\sum_i \left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right)\left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right)'. \quad (12)$$

In the random-coefficient model, interest centers on the mean parameters μ and σ_i^2 and the measure of heterogeneity Σ. Pesaran and Smith (1995) were concerned with the estimation of μ. Mairesse and Griliches (1990) were concerned with the estimation of Σ as well. The traditional estimator of μ from the random-coefficient model is a GLS, but because Ω_i involves the unknown parameters σ_i^2 and Σ, we substitute the estimates from Equation (11) and use the "feasible" GLS. Thus, it is a two-step GLS estimator based on an estimated covariance matrix. This procedure, as discussed later (Sec. 2) gives asymptotically efficient estimates only if there are no lagged dependent variables.

In the classical model it makes no sense to talk about obtaining estimators for the individual parameters β_i because they are treated as random variables. Hence, inference is based entirely on the parameters μ, Σ, and σ_i^2. One can, however, talk of predictors for the random parameters β_i. Lee and Griffiths (1978) derived the best linear unbiased predictors for β_i based on the prior likelihood approach advocated by Edwards (1969). This amounts to estimating β_i, σ_i^2, μ, and Σ by maximizing the likelihood

$$\begin{aligned} L(\beta_i, \sigma_i^2, \mu, \Sigma | y, X) &= \text{const.} - \frac{T}{2}\sum_{i=1}^{N} \ln \sigma_i^2 \\ &- \frac{1}{2}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}(y_i - X_i\beta_i)'(y_i - X_i\beta_i) \\ &- \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(\beta_i - \mu)'\Sigma^{-1}(\beta_i - \mu). \end{aligned} \quad (13)$$

The resulting estimates for β_i, σ_i^2, μ, and Σ are given as

$$\beta_i^* = \left(\frac{1}{\sigma_i^2} X_i'X_i + \Sigma^{*-1}\right)^{-1}\left(\frac{1}{\sigma_i^2} X_i'X_i\hat{\beta}_i + \Sigma^{*-1}\mu^*\right), \quad (14)$$

$$\mu^* = \frac{1}{N}\sum_{i=1}^{N}\beta_i^*, \quad (15)$$

$$\sigma_i^2 = \frac{1}{T}(y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*), \quad (16)$$

and

$$\Sigma^* = \frac{1}{N}\sum_{i=1}^{N}(\beta_i^* - \mu^*)(\beta_i^* - \mu^*)', \quad (17)$$

where $\hat{\beta}_i$ is the OLS estimate of β_i. As will be shown next, these prior likelihood estimators are closely related to the Bayesian estimators.

1.2 An Iterative Bayesian Approach

In the Bayesian framework, Equation (6) specifies the prior distribution of β_i. Because this prior distribution involves the parameters μ and Σ, if they are not known priors must be specified for these hyperparameters. One can then derive the posterior distribution for the parameters β_i.

If μ, σ_i^2, and Σ are known, the posterior distribution of β_i is normal with mean β_i^* given by

$$\beta_i^* = \left(\frac{1}{\sigma_i^2} X_i'X_i + \Sigma^{-1}\right)^{-1}\left(\frac{1}{\sigma_i^2} X'X_i\hat{\beta}_i + \Sigma^{-1}\mu\right) \quad (18)$$

and variance

$$V(\beta_i^*) = \left(\frac{1}{\sigma_i^2} X_i'X_i + \Sigma^{-1}\right)^{-1},$$

where $\hat{\beta}_i$ is the OLS estimate of β_i. Note that, with lagged dependent variables, the normality of the posterior distribution of β_i^* holds only asymptotically and under the usual regularity conditions assumed in dynamic regression models.

Assuming a noninformative prior for μ, the mean of the posterior distribution of μ is $\mu^* = (1/N)\sum_{i=1}^{N}\beta_i^*$. Because in general σ_i^2 and Σ will not be known, one needs to specify priors for these hyperparameters. Smith (1973) took the conjugate Wishart distribution for Σ^{-1} and the independent inverse χ^2 distributions for the σ_i^2. He suggested using the mode of the posterior distribution. His modal estimators are

$$\hat{\sigma}_i^2 = \frac{1}{T + v_i + 2} [v_i\lambda_i + (y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*)] \quad (19)$$

and

$$\Sigma^* = \frac{1}{N - k - 2 + \delta} \left[R + \sum_{i=1}^{N} (\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'\right]. \quad (20)$$

As discussed by Smith (1973), v_i, λ_i, R, and δ are parameters arising from the specification of the prior distributions, and k is the dimension of β. Approximations to vague priors are obtained by setting $v_i = 0, \delta = 1$, and R to be a diagonal matrix with small positive entries (e.g., .001).

The estimators are then

$$\hat{\sigma}_i^2 = \frac{1}{T+2} (Y_i - X_i\beta_i^*)'(Y_i - X_i\beta_i^*) \quad (21)$$

and

$$\Sigma^* = \frac{1}{N - k - 1} \sum_{i=1}^{N} (\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'. \quad (22)$$

Equations (21) and (22) have to be solved iteratively along with the equations for β^* and μ^* given by

$$\beta_i^* = \left(\frac{1}{\hat{\sigma}_i^2} X_i'X_i + \Sigma^{*-1}\right)^{-1} \left(\frac{1}{\hat{\sigma}_i^2} X_i'X_i\hat{\beta}_i + \Sigma^{*-1}\mu^*\right) \quad (23)$$

and

$$\mu^* = \frac{1}{N} \sum_{i=1}^{N} \beta_i^*. \quad (24)$$

Note that in Equations (21)–(24) the prior mean μ^* is an average of β_i^*, the estimate of the prior variance Σ is obtained from deviations of β_i^* from their average μ^*, and the estimate of σ_i^2 is obtained from the residual sum of squares using β_i^*. Equations (21)–(24) have to be solved iteratively, with the initial iteration using the OLS estimator $\hat{\beta}_i$ to compute μ^*, σ_i^2, and Σ^*. We take μ^* as the simple average of $\hat{\beta}_i, \hat{\sigma}_i^2$ as the usual estimate of σ_i^2 from the ith cross-section, and Σ^* as given by Equation (12). Moreover, to improve the convergence with the iterative procedure, Σ^* is computed as

$$\Sigma^* = \frac{1}{N - k - 1} \left[R + \sum_{i=1}^{N} (\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'\right], \quad (22a)$$

where R is a diagonal matrix with small positive entries (e.g., .001).

Thus, as noted by Maddala (1991), many shrinkage estimators only differ on the basis of the overall estimators toward which the individual estimators are shrunk and the estimates of variances and covariance matrices. Smith (1973) showed with some matrix manipulations that the GLS estimator of μ is related to the Bayesian estimator β_i^* by the equation

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \beta_i^*, \quad (25)$$

which is the same as the Bayesian estimator and the maximum likelihood estimator (MLE) for μ. Moreover, the MLE's given by Lee and Griffiths (1978) only differ from the GLS and Bayesian estimators in the divisors for σ_i^2 and Σ^*.

1.3 The Empirical Bayes Approach

In the empirical Bayes approach, we take Equation (18) and plug in sample-based estimates for μ, σ_i^2, and Σ. Rao (1975) suggested estimating them by

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i,$$

$$\hat{\sigma}_i^2 = \frac{1}{T-k} (y_i'y_i - y_i'X_i\hat{\beta}_i),$$

and

$$\hat{\Sigma} = \frac{1}{N-1} \sum_i (\hat{\beta}_i - \hat{\mu})(\hat{\beta}_i - \hat{\mu})' - \frac{1}{N} \sum_i (X_i'X_i)^{-1}\hat{\sigma}_i^2.$$

All of these expressions are based on the least squares estimators, $\hat{\beta}_i$. The estimators for σ_i^2 and Σ are unbiased if X_i includes exogenous variables only. As pointed out in the preceding sections, the Bayesian approach gives a different set of estimators, based on the Bayes estimators, β_i^*. Kadiyala and Oberhelman (1986) used a modification of the estimators suggested by Rao (1975). They used μ^* as an average of β_i^* rather than $\hat{\beta}_i$ as in the Bayesian approach, but they still used the same estimator for Σ as that suggested by Rao.

Both of these empirical estimators are two-step estimators and do not involve any iteration. One can think of the following empirical Bayes estimator, which, however, has to be computed iteratively:

$$\begin{aligned} \mu^* &= \frac{1}{N} \sum_{i=1}^{N} \beta_i^* \\ \hat{\sigma}_i^2 &= \frac{1}{T-k} (y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*) \\ \hat{\Sigma} &= \frac{1}{N-1} \sum_i (\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'. \end{aligned} \quad (26)$$

These expressions differ from the ones derived by Smith (1973) and by Lee and Griffiths (1978) only in the divisors in $\hat{\sigma}_i^2$ and Σ^*. In the preceding equations, the prior mean μ^* is an average of β_i^*, the estimate of the prior variance Σ^* is obtained from deviations of β_i^* from their average μ^*, and the estimate of σ_i^2 is obtained from the residual sum of squares using β_i^*, not the OLS estimator $\hat{\beta}_i$. The preceding equations are solved iteratively, with the initial iteration using the OLS estimator $\hat{\beta}_i$ to compute μ^*, σ_i^2, and Σ^* as described earlier. Moreover, to improve the convergence with the iterative procedure Σ^* is computed as

$$\Sigma^* = \frac{1}{N-1}\left[R + \sum_{i=1}^{N}(\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'\right],$$

where R is a diagonal matrix as defined previously. Hu and Maddala (1994) presented some Monte Carlo evidence to suggest that the iterative procedures for the computation of Σ^* give better estimates (in the mean squared sense) for both the overall mean μ and the heterogeneity matrix Σ than the two-step procedures.

1.4 In Summary

We have discussed the classical, Bayesian, and empirical Bayesian approaches to the estimation of individual cross-section parameters. We have also discussed two-step versus iterative methods. These are summarized in a tabular form in Table 1. In Section 2, we will discuss the theoretical justification of the iterative procedure over the two-step procedure in dynamic models. To conserve space, we will present empirical results for only the Stein-rule and (the iterative) empirical Bayes estimators.

The most important points to note are the estimation of Σ, which is computed using Formula (26) and not (11) or (12), the computation of μ^* as $(1/N)\sum_{i=1}^{N}\beta_i^*$ and the computation of σ_i^{2*} using not least squares residuals but residuals based on β_i^*. Note that, incidentally, we get an estimator of μ, which is computed in an iterative fashion, not using the two-step procedure used by Swamy (1970) and Hsiao (1986). Even if our interest centers on μ, we suggest using the iterative procedure.

2. TWO-STEP VERSUS ITERATIVE ESTIMATORS

In Section 1, we discussed several estimators for the estimation of the overall mean μ and covariance matrix Σ as well as the individual parameters β_i. The estimators for μ and β_i can be classified into two categories, (1) two-step estimators based on initial consistent estimators of Σ and σ_i^2 and (2) iterative estimators in which all the parameters μ, β_i, Σ, and σ_i^2 are estimated jointly.

In the first category we have the two-step empirical Bayes estimator for β_i and the Swamy–Hsiao estimator for μ. In

Table 1. Different Estimators for σ_i^2, Σ, β_i, and μ

Method	*Estimator*	*Note*
	For σ_i^2	
Swamy–Hsiao and two-step empirical Bayesian	$\frac{1}{T-k}(y_i - X_i\hat{\beta}_i)'(y_i - X_i\hat{\beta}_i)$	1
Prior likelihood	$\frac{1}{T}(y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*)$	2
Iterative Bayesian (Smith)	$\frac{1}{T+2}(y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*)$	
Empirical Bayes (iterative)	$\frac{1}{T-k}(y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*)$	
	For Σ	
Swamy–Hsiao	$\hat{\Sigma}_1 = \frac{1}{N-1}\sum_i\left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^N\hat{\beta}_i\right)\left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^N\hat{\beta}_i\right)'$	
Swamy–Rao (unbiased)	$\hat{\Sigma} = \hat{\Sigma}_1 - \frac{1}{N}\sum_i(X_i'X_i)^{-1}\hat{\sigma}_i^2$	3
Prior likelihood	$\Sigma^* = \frac{1}{N}\sum_{i=1}^N(\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'$	
Iterative Bayes (Smith)	$\Sigma^* = \frac{1}{N-k-1}\sum_{i=1}^N(\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'$	
Iterative empirical Bayes	$\Sigma^* = \frac{1}{N-1}\sum_{i=1}^N(\beta_i^* - \mu^*)(\beta_i^* - \mu^*)'$	
	For β_i	
Swamy	No estimators	
Stein-rule	$\tilde{\beta}_i = (1 - c/F)\hat{\beta}_i + (c/F)\hat{\beta}_p$	4
Prior likelihood and iterative Bayes	$\beta_i^* = \left(\frac{1}{\sigma_i^{*2}}X_i'X_i + \Sigma^{*-1}\right)^{-1}\left(\frac{1}{\sigma_i^{*2}}X_i'X_i\hat{\beta}_i + \Sigma^{*-1}\mu^*\right)$	
Two-step empirical Bayes	$\beta_i^* = \left(\frac{1}{\hat{\sigma}_i^2}X_i'X_i + \hat{\Sigma}^{-1}\right)^{-1}\left(\frac{1}{\hat{\sigma}_i^2}X_i'X_i\hat{\beta}_i + \hat{\Sigma}^{-1}\hat{\mu}\right)$	
	For μ	
Swamy	$\hat{\mu}_{GLS} = \sum_{i=1}^N W_i\hat{\beta}_i$	5
Prior likelihood and iterative Bayes	$\mu^* = \frac{1}{N}\sum_{i=1}^N\beta_i^*$	6

NOTE: (1) $\hat{\beta}_i$ is the OLS estimator of β. (2) β_i^* is defined later. (3) The two-step empirical Bayes method is based on these estimators of Σ. $\hat{\Sigma}_1$ is used if $\hat{\Sigma}$ is not positive definite. (4) $c = [(N-1)k - 2]/[NT - Nk + 2]$, $\hat{\beta}_p$ is the pooled estimator, and F is the F statistic to test equality of the coefficients. (5) $W_i = (\Sigma P_j^{-1})^{-1}P_i^{-1}$ with $P_i = [\sigma_i^2(X_i'X_i)^{-1} + \Sigma]$. We first get $\hat{\sigma}_i^2$ and $\hat{\Sigma}$ and then substitute these for σ_i^2 and Σ in $\hat{\mu}_{GLS}$. It is thus a two-step method. (6) Because β_i^* are computed iteratively, μ^* is an iterative estimator. If σ_i^2 and Σ are known, then $\mu^* = \hat{\mu}_{GLS}$.

the second category we have the prior likelihood, iterative empirical Bayes, and iterative Bayesian (Smith) estimators.

As is evident from Equations (15) and (18), the expressions for estimation of the overall mean μ and the individual parameters β_i involve the unknown parameters σ_i^2 and Σ. We need estimators for these parameters to implement the procedures. This is where the difference between the two-step (based on initial consistent estimators of σ_i^2 and Σ) and iterative procedures lies. Essentially, what we have here is a problem similar to that of GLS estimation with an estimated covariance matrix. As pointed out by Amemiya and Fuller (1967), Maddala (1971), and Pagan (1986), in the case that the explanatory variables X_i include only exogenous variables, the two-step estimators based on any consistent estimators of parameters σ_i^2 and Σ are as asymptotically efficient as the estimators from the joint efficient estimation of all the parameters. If the variables X_i include lagged dependent variables, however, as in the case of dynamic models, the two-step estimators based on any consistent estimates of σ_i^2 and Σ are consistent but not efficient. This suggests that even if we are interested in the estimation of the overall mean μ, in the presence of lagged dependent variables, the Swamy or Hsiao estimators of Σ do not provide efficient estimates of μ. We have to use the iterative procedure suggested in Section 1. In the case, as in this article, our interest centers on the estimation of β_i, and we again should use the iterative procedures.

Many of the applications in the statistical literature, as well as the example on Grunfeld's investment function of Swamy (1970), have no lagged dependent variables. Hence, the GLS based on the Swamy or Hsiao estimator of Σ is asymptotically efficient. This is not the case, however, with dynamic models considered by, for instance, Pesaran and Smith (1995) even though their interest was in the overall mean μ. In the example to be discussed next, we have a dynamic demand model. Hence, we used the iterative procedure described in Section 1 for the estimation of β_i as well as μ.

3. AN EMPIRICAL EXAMPLE

There have been numerous studies on the price and income elasticities of residential natural-gas and electricity demand. Short-run price elasticity estimates of both residential natural-gas and electricity demand have ranged from −.05 to −.9, and long-run price elasticities estimates have ranged from −.2 to as high as −4.6. There is a similar diversity among short-run and long-run income elasticity estimates. For a partial review of these studies, see Al-Sahlawi (1989), Bohi (1981), Bohi and Zimmerman (1984), and Taylor (1975).

In this section we present new residential gas and electricity price and income elasticity estimates using pooled cross-section and time series data from 49 U.S. states. The state of Hawaii was excluded because we could not obtain annual weather data.

The annual state residential electricity and gas price data used in this study were obtained from *The State Energy Price and Expenditure System* of the U.S. Energy Information Administration (1993). Weather data are population-weighted heating and cooling degree days by state and were taken from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Ashville, North Carolina. Annual state income data were drawn from the Bureau of Economic Affairs, and the annual Consumer Price Index for the United States was from CITIBASE. The sample period is from 1970 to 1990.

To estimate the short-run and long-run price and income demand elasticities for gas and electricity, we specified the following DLR(1, 1) model:

$$\begin{aligned} y_{it} &= \beta_{i0} + \beta_{i1} y_{i,t-1} + \beta_{i2} x_{1it} \\ &\quad + \beta_{i3} x_{1i,t-1} + \beta_{i4} x_{2it} + \beta_{i5} x_{2i,t-1} \\ &\quad + \beta_{i6} x_{3it} + \beta_{i7} x_{4it} + \beta_{i8} x_{5it} + u_{it}, \end{aligned} \tag{27}$$

where $i = 1, 2, 3, \ldots, 49$ (U.S. state subscript) and $t = 2, 3, \ldots, 21$ (year subscript).

The variables for the electricity regression are y_{it} = ln(residential electricity per capita consumption), x_{1it} = ln(real per capita personal income), x_{2it} = ln(real residential electricity price), x_{3it} = ln(real residential natural-gas price), x_{4it} = heating degree days (HDD), and x_{5it} = cooling degree days (CDD).

For the natural-gas regression, we have y_{it} = ln(residential natural gas per capita consumption), x_{1it} = ln(real per capita personal income), x_{2it} = ln(real residential natural-gas price), and x_{3it} = ln(real residential electricity price), with x_{4it} and x_{5it} unchanged.

From Equation (27) the following short-run and long-run elasticities can be derived for the ith state:

Short-run income elasticity:	$\mathrm{SR}\eta_y = \beta_{i2}$
Long-run income elasticity:	$\mathrm{LR}\eta_y = \frac{\beta_{i2}+\beta_{i3}}{1-\beta_{i1}}$
Short-run price elasticity:	$\mathrm{SR}\eta_p = \beta_{i4}$
Long-run price elasticity:	$\mathrm{LR}\eta_p = \frac{\beta_{i4}+\beta_{i5}}{1-\beta_{i1}}$
Short-run cross-price elasticity:	$\mathrm{SR}\eta_p^c = \beta_{i6}$
Long-run cross-price elasticity:	$\mathrm{LR}\eta_p^c = \frac{\beta_{i6}}{1-\beta_{i1}}$.

4. EMPIRICAL RESULTS

We here present 10 tables of results. The separate state regression coefficients will be presented in Tables 6–9. Tables 10 and 11 will present estimates of the short- and long-run elasticities.

Tables 2–5 summarize the estimates from the five approaches discussed. In Table 2 the parameter estimates from the electricity-demand regression and t statistics are presented. Column 1 contains the results from the fixed-effects model given by $y_{it} = \alpha_1 y_{i,t-1} + \alpha_2 x_{1it} + \alpha_3 x_{1i,t-1} + \alpha_4 x_{2it} + \alpha_5 x_{2i,t-1} + \alpha_6 x_{3it} + \alpha x_{4it} + \alpha_8 x_{5it} + \sum_{i=1}^{49} \delta_i + u_{it}$, and column 2 contains the pooled data model given by $y_{it} = \alpha_0 + \alpha_1 y_{i,t-1} + \alpha_2 x_{1it} + \alpha_3 x_{1i,t-1} + \alpha_4 x_{2it} + \alpha_5 x_{2i,t-1} + \alpha_6 x_{3it} + \alpha_7 x_{4it} + \alpha_8 x_{5it} + u_{it}$. The next column of Table 2 presents the average of the separate state regression estimates followed by columns with the results on the average of the Bayesian estimates and finally the average of the Stein-rule estimates.

Table 2. Parameter Estimate of the Electricity Regression (t values in parentheses)

Variables	Pooled with dummies	Pooled without dummies	OLS estimator (average)	Shrinkage estimator (average)	Stein-rule estimator (average)
$y_{i,t-1}$	.789	.903	.530	.629	.685
	(49.00)	(106.03)	(22.28)	(32.12)	(47.76)
x_{1it}	.137	.138	.429	.394	.308
	(3.25)	(3.12)	(7.80)	(11.39)	(8.32)
$x_{1i,t-1}$	−.013	−.134	−.040	−.064	−.079
	(−.31)	(−3.08)	(−.74)	(−1.80)	(−2.16)
x_{2it}	−.196	−.214	−.163	−.158	−.184
	(−7.96)	(−8.47)	(−6.19)	(−8.37)	(−9.88)
$x_{2i,t-1}$	.078	.113	.057	.061	.081
	(3.15)	(4.49)	(2.16)	(2.71)	(4.30)
x_{3it}	.035	.028	.017	.023	.022
	(4.48)	(4.86)	(1.43)	(1.80)	(2.93)
x_{4it}	.138	.084	.129	.142	.110
	(3.99)	(6.97)	(3.53)	(6.77)	(5.03)
x_{5it}	1.319	.398	1.561	1.069	1.077
	(12.58)	(9.87)	(2.75)	(10.02)	(3.24)

Table 3 contains the short- and long-run income elasticities ($SR\eta_y$ and $LR\eta_y$), price elasticities ($SR\eta_p$ and $LR\eta_p$), and cross-elasticities ($SR\eta_p^c$ and $LR\eta_p^c$) for electricity demand in the columns. The average, maximum, and minimum estimates by different methods are provided in the rows. There is less variability in the short-run elasticities compared with the long-run elasticities. The second law of demand appears to hold; the long-run elasticities are higher than the short-run elasticities except in one case. This exception appears to be due to an outlier in the $LR\eta_p$ using the state-specific regressions. The F statistic testing hypothesis (3) is 1.698, and the 5% critical value for the F distribution is 1. Thus, the Stein-rule estimator for the electricity regression is $\tilde{\beta}_i = .584\hat{\beta}_i + .416\hat{\beta}_p$. The constant in the regression is .706.

The average $SR\eta_y$ elasticity estimates for electricity presented in Table 3 range from .137 to .429 with the two pooled model estimates being approximately the same, .137 and .138. The separate OLS estimates are much higher and have the greatest variability ranging from −.83 to 1.328. The average of the Bayesian shrinkage estimates is .394 and the average of the Stein estimates is .308. The Bayesian shrinkage estimates have a much smaller spread than the Stein estimates, reflecting the greater weight for heterogeneity across the states in the Stein estimator.

The average $LR\eta_y$, $LR\eta_p$, and $LR\eta_p^c$ elasticities for electricity are presented in two ways. First we calculate the simple average of the elasticities for each state. Then the elasticities are calculated using the average based on the individual parameter estimates. The $LR\eta_y$ elasticity estimates range from .044 with the pooled data model to 1.58 for the OLS estimator. The separate OLS regression results have by far the greatest variability (−1.229 to 33.033) followed by the Stein estimator (−7.771 to 2.82) and the Bayesian shrinkage estimator (.486 to 1.966). The Bayesian shrinkage estimator produces a tighter and more sensible range for $LR\eta_y$ than the Stein estimator because the latter is influenced by the separate regression results that contain several negative long-run elasticities.

The adverse impact of ignoring cross-section specific effects is clear by comparing the pooled model with state intercept dummies to the pooled model without state inter-

Table 3. Short-Run and Long-Run Income and Price Elasticities of the Electricity Regression

Approaches	$SR\eta_y$	$LR\eta_y$	$SR\eta_p$	$LR\eta_p$	$SR\eta_p^c$	$LR\eta_p^c$
Pooled w/ dummies	.137	.584	−.196	−.561	.035	.166
Pooled w/o dummies	.138	.044	−.214	−1.031	.028	.290
OLS[a] average	.429	1.580	−.163	−.375	.017	−.135
maximum	1.328	33.033	.778	2.179	.648	.977
minimum	−.830	−1.229	−.916	−10.380	−.392	−6.454
OLS[b]	.429	.827	−.163	−.225	.017	.037
Shrk[a] average	.394	.921	−.158	−.239	.023	.054
maximum	.585	1.966	−.064	.242	.161	.447
minimum	.211	.486	−.281	−.873	−.108	−.263
Shrk[b]	.394	.890	−.158	−.263	.023	.062
Stein[a] average	.308	.609	−.184	−.276	.022	.083
maximum	.833	2.820	.365	1.526	.390	1.888
minimum	−.427	−7.771	−.624	−1.745	−.217	−.994
Stein[b]	.308	.727	−.184	−.329	.022	.069

[a] Calculated based on the individual elasticities for each state. The same is the case for maximum and minimum.
[b] Calculated by first taking the mean of the individual parameter estimates. Then the elasticities are calculated from this mean. The elasticities marked by [a] and [b] are the same for the short-run but not for the long-run.

Table 4. Parameter Estimate of the Natural-Gas Regression (t values in parentheses)

Variables	Pooled with dummies	Pooled without dummies	OLS estimator (average)	Shrinkage estimator (average)	Stein-rule estimator (average)
$y_{i,t-1}$	.628	.983	.173	.499	.440
	(25.49)	(195.83)	(5.77)	(19.67)	(21.83)
x_{1it}	.114	.048	.307	.280	.222
	(1.15)	(.45)	(2.63)	(3.42)	(2.58)
$x_{1i,t-1}$	−.075	−.040	−.583	−.314	−.404
	(−.79)	(−.38)	(−5.31)	(−4.12)	(−4.98)
x_{2it}	−.177	−.164	−.092	−.099	−.116
	(−4.68)	(−4.13)	(−2.46)	(−3.77)	(−4.08)
$x_{2i,t-1}$	.035	.141	−.073	−.041	−.003
	(.92)	(3.60)	(−1.92)	(−1.43)	(−.09)
x_{3it}	.016	.026	−.034	−.001	−.014
	(.49)	(1.74)	(−.73)	(−.03)	(−.45)
x_{4it}	.529	.027	.421	.424	.291
	(6.30)	(.96)	(4.63)	(7.91)	(4.72)
x_{5it}	−1.104	−.079	1.494	−.913	.976
	(−4.40)	(−.96)	(.42)	(−5.83)	(.41)

cept dummies. The electricity $LR\eta_y$ elasticity of .584 from the former is closer to what economic theory would suggest than the simple pooled estimate without state intercept dummies of .044. This can be explained by the near canceling of the income estimates in the pooled model without state intercept dummies. Moreover, by ignoring heterogeneity in the intercepts, the size and explanatory power of the lagged dependent variable is greatly increased in the pooled model without state intercept dummies.

The $SR\eta_p$ are very similar across the five methods (−.158 to −.214). The separate state Bayesian shrinkage estimates are closest to the average and numerically the most inelastic (−.158) and do not contain any positive elasticities as does the Stein estimator. The range for the $LR\eta_p$ is (−.239 to −1.031), and the higher estimates are from the two pooled models. This appears due to the "imposition" of homogeneity across the states.

Table 4 summarizes the parameter estimates from the five methods using the natural-gas models. The explanatory power is provided primarily by the lagged consumption term and contemporary natural-gas price. Evidence of heterogeneity across the states is again evident by the relatively large estimate for the lagged dependent variable in the pooled model with state intercept dummies (.628) and the pooled model without state intercept dummies (.983) relative to the average from the individual state OLS estimates (.173).

Short- and long-run elasticities for the natural gas models are presented in Table 5 and organized as in Table 3. The short-run price elasticity estimates are quite stable over the different estimation approaches, ranging from −.092 to −.177. The short-run income-elasticity estimates range from .048 to .307. As with electricity demand, the long-run income and price elasticity estimates for natural-gas demand vary quite a bit over the five estimation approaches. The $LR\eta_p$ estimates range from −.239 to −1.358, and the $LR\eta_y$ estimates range from −.425 to .491.

The null hypothesis in (7) is again rejected from the F statistic for the Stein-rule estimator; the statistic is 2.143 and the 5% critical value is 1. This implies that the Stein-rule estimator is given by $\tilde{\beta}_i = .671\hat{\beta}_i + .329\hat{\beta}_p$ and suggests that there is relatively greater heterogeneity in the natural-gas regression model compared with the electricity model.

One final note on the natural-gas elasticities is worth

Table 5. Short-Run and Long-Run Income and Price Elasticities of the Natural Gas Regression

Approaches	$SR\eta_y$	$LR\eta_y$	$SR\eta_p$	$LR\eta_p$	$SR\eta_p^c$	$LR\eta_p^c$
Pooled w/ dummies	.114	.104	−.177	−.381	.016	.044
Pooled w/o dummies	.048	.491	−.164	−1.358	.026	1.558
OLS[a] average	.307	−.425	−.092	−.239	−.034	.046
maximum	3.339	1.847	1.069	2.198	.783	5.506
minimum	−2.745	−5.132	−1.341	−2.454	−.963	−2.104
OLS[b]	.307	−.334	−.092	−.200	−.034	−.041
Shrk[a] average	.280	−.057	−.099	−.273	−.001	−.007
maximum	.356	.473	−.060	.085	.083	.151
minimum	.210	−.486	−.150	−.660	−.091	−.233
Shrk[b]	.280	−.068	−.099	−.280	−.001	−.002
Stein[a] average	.222	−.412	−.116	−.250	−.014	.057
maximum	2.255	1.816	.663	2.127	.534	5.185
minimum	−1.825	−5.020	−.953	−2.365	−.637	−2.031
Stein[b]	.222	−.326	−.116	−.212	−.014	−.025

[a] Calculated based on the individual elasticities for each state. The same is the case for maximum and minimum.
[b] Calculated by first taking the mean of the individual parameter estimates. Then the elasticities are calculated from this mean. The elasticities marked by [a] and [b] are the same for the short-run but not for the long-run.

Table 6. OLS Estimates and t Values of Individual States (electricity regression)

State	$y_{i,t-1}$	x_{1it}	$x_{1i,t-1}$	x_{2it}	$x_{2i,t-1}$	x_{3it}
CA	.581	.510	−.215	−.211	.087	−.036
	3.647	1.814	−.918	−2.076	1.367	−1.016
FL	.422	1.265	−.659	−.163	.009	−.073
	2.617	6.491	−2.688	−2.601	.125	−1.303
IL	.241	.558	.144	−.011	−.051	.019
	1.832	2.353	.605	−.090	−.423	.570
MI	.754	.285	−.126	−.040	.012	−.008
	6.475	2.404	−1.124	−.468	.128	−.452
NY	.810	.476	−.243	−.152	.191	−.053
	7.615	2.491	−1.194	−2.581	3.400	−1.715
OH	.637	.437	−.007	−.110	−.065	.055
	7.022	2.428	−.042	−1.271	−.602	1.613
PA	.521	.522	.034	−.207	.166	−.010
	5.632	2.788	.174	−2.686	2.137	−.653
TX	.688	.091	.028	−.444	.206	.103
	3.243	.188	.073	−2.955	1.470	.651

NOTE: t values are presented in the second line for each state. Coefficients for cooling and heating degree days are not reported.

Table 8. OLS Estimates and t Values of Individual States (natural-gas regression)

State	$y_{i,t-1}$	x_{1it}	$x_{1i,t-1}$	x_{2it}	$x_{2i,t-1}$	x_{3it}
CA	−.179	−.223	−1.686	.193	−.310	−.160
	−1.009	−.326	−2.036	1.554	−2.520	−.908
FL	.172	2.733	−3.914	−.579	−.472	.423
	1.183	3.376	−4.470	−2.703	−1.399	2.578
IL	−.512	.947	−.877	.662	−.447	−.898
	−2.905	2.588	−2.513	4.108	−2.728	−4.037
MI	−.023	−.563	.463	.359	−.263	−.506
	−.088	−.990	1.106	1.282	−.937	−1.829
NY	.023	1.088	−.738	.006	−.128	.012
	.089	1.982	−1.481	.041	−1.008	.089
OH	−.441	−.136	−.920	−.067	−.191	−.025
	−2.041	−.286	−2.345	−.419	−1.271	−.111
PA	−.133	−.010	−.410	.103	−.090	−.557
	−.450	−.017	−.793	.645	−.567	−2.677
TX	.599	−.751	−1.304	1.069	−.189	−.843
	4.700	−.772	−1.555	3.134	−.812	−2.807

NOTE: See note to Table 6.

making. The long-run income elasticity for natural gas is persistently estimated as negative with the individual OLS regressions and is nearly 0 (−.057) with the shrunken estimates. Although it seems counterintuitive that the long-run natural-gas income elasticity is smaller than the short-run natural-gas elasticity, there are several explanations for this result. First, as incomes rise, households may buy microwave ovens and will substitute away from gas cooking into microwave cooking. Second, as incomes rise, households may convert their homes to central air conditioning and households that previously used gas for heating now have the option of converting to electric heating and cooling with a heat pump. Hence, a certain subset of these households will reduce their gas consumption dramatically as incomes rise. Third, as incomes rise, households will remodel their homes. In many cases the configuration of appliances such as ranges, clothing dryers, and water heaters after remodeling are not convenient to gas lines. Again a subset of households that previously used gas for these end uses will now convert to electricity as incomes rise. Finally, natural-gas price controls had an impact on the availability of supplies. The availability of natural gas for new homes fell from 50% in 1971 to 37% in 1979 (when prices were decontrolled) and has risen to 63% in 1993. Over the same period, electricity's share of heating in new homes rose from 43% in 1971 to 57% in 1985 and then fell to 33% in 1993. In addition, the U.S. Energy Information Administration in 1987 reported that among the 2.2 million new homes that used electricity as a primary heating source, 89.3 percent of the households stated that they did not have access to natural gas in their neighborhoods. The combination of these factors can explain the income elasticity results.

Tables 6–9 present the estimates of the individual state regression coefficients with the associated t statistics and Tables 10 and 11 present the corresponding price, income, and cross-price elasticities. To conserve space, only the results for the OLS and Bayesian shrinkage estimation methods are presented for the largest eight states—California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania,

Table 7. Shrinkage Estimates and t Values of Individual States (electricity regression)

State	$y_{i,t-1}$	x_{1it}	$x_{1i,t-1}$	x_{2it}	$x_{2i,t-1}$	x_{3it}
CA	.635	.441	−.205	−.208	.090	−.018
	12.035	6.022	−3.241	−4.450	1.944	−.726
FL	.647	.585	−.296	−.233	.139	−.016
	12.760	8.387	−6.103	−5.347	3.090	−.477
IL	.561	.359	.035	−.086	.006	.014
	12.572	4.993	.528	−1.553	.117	.476
MI	.621	.338	−.050	−.120	.024	.003
	14.291	5.482	−.957	−2.499	.469	.162
NY	.715	.453	−.121	−.136	.167	−.042
	20.162	6.369	−2.166	−3.699	4.033	−1.536
OH	.679	.395	−.043	−.152	.044	.034
	19.903	6.128	−.931	−3.298	.814	1.259
PA	.679	.476	−.129	−.172	.115	−.004
	19.702	7.240	−2.452	−3.827	2.830	−.266
TX	.625	.353	.009	−.197	.045	.023
	11.576	3.864	.116	−3.223	.591	.429

NOTE: See note to Table 6.

Table 9. Shrinkage Estimates and t Values of Individual States (natural-gas regression)

State	$y_{i,t-1}$	x_{1it}	$x_{1i,t-1}$	x_{2it}	$x_{2i,t-1}$	x_{3it}
CA	.451	.240	−.417	−.128	−.154	.052
	11.531	6.998	−7.651	−4.485	−4.080	1.349
FL	.502	.225	−.409	−.135	−.115	.018
	13.056	5.436	−5.650	−3.822	−1.793	.429
IL	.481	.299	−.295	−.078	−.024	.003
	13.546	9.473	−6.176	−2.757	−.739	.094
MI	.512	.322	−.235	−.067	.029	−.027
	14.520	10.533	−5.207	−2.352	.920	−.787
NY	.523	.324	−.229	−.074	.040	−.045
	16.409	11.847	−6.159	−2.792	1.526	−1.518
OH	.461	.270	−.352	−.099	−.073	.030
	12.174	8.384	−7.246	−3.614	−2.351	.843
PA	.489	.290	−.303	−.090	−.020	−.008
	14.073	9.693	−7.138	−3.246	−.726	−.226
TX	.480	.268	−.359	−.106	−.093	.033
	12.880	7.682	−6.459	−3.522	−2.215	.837

NOTE: See note to Table 6.

Table 10. Short-Run and Long-Run Income Elasticities (η_y), Price Elasticities (η_p), and Cross-Price Elasticities (η_p^c) of Individual States (electricity regression)

Estimation methods	State	$SR\eta_y$	$LR\eta_y$	$SR\eta_p$	$LR\eta_p$	$SR\eta_p^c$	$LR\eta_p^c$
OLS							
	CA	.510	.703	−.211	−.297	−.036	−.085
	FL	1.265	1.048	−.163	−.265	−.073	−.125
	IL	.558	.925	−.011	−.082	.019	.025
	MI	.285	.648	−.040	−.117	−.008	−.032
	NY	.476	1.225	−.152	.202	−.053	−.281
	OH	.437	1.184	−.110	−.480	.055	.152
	PA	.522	1.160	−.207	−.085	−.010	−.022
	TX	.091	.382	−.444	−.765	.103	.329
Shrinkage							
	CA	.441	.647	−.208	−.324	−.018	−.049
	FL	.585	.820	−.233	−.265	−.016	−.046
	IL	.359	.898	−.086	−.181	.014	.031
	MI	.338	.760	−.120	−.253	.003	.007
	NY	.453	1.166	−.136	.109	−.042	−.147
	OH	.395	1.096	−.152	−.336	.034	.105
	PA	.476	1.083	−.172	−.176	−.004	−.013
	TX	.353	.965	−.197	−.406	.023	.063

and Texas. Detailed tables for all the 49 states as well as for the Stein-rule estimation method are available on request from us. The differences between the results from the OLS and Bayesian shrinkage methods are evident and have also been discussed earlier and hence will not be elaborated on.

5. CONCLUSIONS

The most common procedure used in the analysis of panel data is to pool with individual-specific dummies that are assumed to be fixed (fixed-effects model) or random (random-effects model). This procedure, however, assumes homogeneity of the slope coefficients. At the other extreme is the case of complete heterogeneity and separate estimation of cross-section coefficients β_i. An intermediate case between complete homogeneity and complete heterogeneity is the random-coefficient model. In this framework we would be interested in estimating (a) the mean μ and covariance matrix Σ of the cross-section coefficients β_i and (b) the coefficients β_i themselves. In the classical framework only μ and Σ are estimable; β_i are not. One can talk of predictors for β_i, however.

This article presents in a unified framework different procedures for the estimation of β_i. The Bayesian approach to this problem results in shrinkage estimators that shrink the individual OLS estimator $\hat{\beta}_i$ toward the estimator of the overall mean μ. Other shrinkage estimators have been suggested in the literature. For instance, the Stein-rule estimator shrinks the OLS estimator $\hat{\beta}_i$ toward the pooled estimator obtained under complete homogeneity. The estimator of μ is different from the pooled estimator.

Table 11. Short-Run and Long-Run Income Elasticities (η_y), Price Elasticities (η_p), and Cross-Price Elasticities (η_p^c) of Individual States (electricity regression)

Estimation methods	State	$SR\eta_y$	$LR\eta_y$	$SR\eta_p$	$LR\eta_p$	$SR\eta_p^c$	$LR\eta_p^c$
OLS							
	CA	−.223	−1.620	.193	−.099	−.160	−.136
	FL	2.733	−1.425	−.579	−1.270	.423	.511
	IL	.947	.046	.662	.142	−.898	−.593
	MI	−.563	−.097	.359	.095	−.506	−.494
	NY	1.088	.358	.006	−.125	.012	.012
	OH	−.136	−.733	−.067	−.179	−.025	−.018
	PA	−.010	−.370	.103	.011	−.557	−.491
	TX	−.751	−5.132	1.069	2.198	−.843	−2.104
Shrinkage							
	CA	.240	−.322	−.128	−.513	.052	.094
	FL	.225	−.368	−.135	−.500	.018	.037
	IL	.299	.008	−.078	−.197	.003	.007
	MI	.322	.178	−.067	−.077	−.027	−.056
	NY	.324	.199	−.074	−.073	−.045	−.095
	OH	.270	−.153	−.099	−.319	.030	.055
	PA	.290	−.025	−.090	−.216	−.008	−.015
	TX	.268	−.176	−.106	−.383	.033	.063

The article discusses the classical prediction approach as well as the empirical Bayes and Bayes approaches to this problem. It discusses two-step versus iterative estimation methods. It is argued that the latter methods should be preferred in the presence of lagged dependent variables.

The article applies these procedures to the problem of estimating short-run and long-run elasticities of residential demand for electricity and natural gas in the United States for each of 49 states. The separate estimates were hard to interpret and had several wrong signs. These are not presented here to conserve space but are available from us. The pooled estimator gave a very high coefficient of the lagged dependent variable implying long lags in adjustment. This estimator was rejected because it rests on the hypothesis of homogeneity of the coefficients, which was easily rejected by an F test. The shrinkage estimator gave much more reasonable parameter values. Because the model was dynamic, iterative estimation methods were used.

If our interest lies in obtaining the elasticity estimates for each state, as in this study, then there are three choices open:

1. Use the individual state data only—but this gave bad results.
2. Pool the data and use the estimates from the pooled estimates—but this is not valid because the hypothesis of homogeneity is convincingly rejected.
3. Use the procedure presented here of allowing some (but not complete) heterogeneity (or homogeneity), which is what the random-coefficient model implies.

In many applications, the experience of researchers will be similar to ours in this case. The comparative performance of the different procedures in the context of dynamic panel-data models is being studied by Monte Carlo experiments.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Energy Information Administration (EIA). All statements are our own and do not reflect the views of the EIA. Helpful comments of an anonymous referee, an associate editor, and the editor, George Tauchen, are gratefully acknowledged. Responsibility for any remaining errors is ours.

[Received May 1994. Revised January 1996.]

REFERENCES

Alogoskoufis, G., and Smith, R. (1991), "On Error Correction Models: Specification, Interpretation, Estimation," *Journal of Economic Surveys*, 5, 97–128.

Al-Sahlawi, M. A. (1989), "The Demand For Natural Gas: A Survey Of Price And Income Elasticities," *The Energy Journal*, 10, 77–87.

Amemiya, T., and Fuller, W. (1967), "A Comparative Study of Alternative Estimators in a Distributed Lag Model," *Econometrica*, 35, 509–529.

Berndt, E. R. (1991), *The Practice of Econometrics: Classic and Contemporary*, Reading, MA: Addison-Wesley.

Bohi, D. R. (1981), *Analyzing Demand Behavior: A Study of Energy Elasticities*, Baltimore: Johns Hopkins University Press for Resources for the Future.

Bohi, D. R., and Zimmerman, M. B. (1984), "An Update On Econometric Studies Of Energy Demand Behavior," *Annual Review of Energy*, 9, 105–154.

Copas, J. B. (1983), "Regression, Prediction, and Shrinkage" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 45, 311–354.

Edwards, A. W. F. (1969), "Statistical Methods in Scientific Inference," *Nature, Land*, 222, 1233–1237.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge, U.K.: Cambridge University Press.

Hu, W., and Maddala, G. S. (1994), "Estimation and Prediction Problems in Dynamic Heterogeneous Panel Data Models," working paper, Ohio State University, Dept. of Economics.

Judge, G. G., and Bock, M. E. (1978), *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, Amsterdam: North-Holland.

Kadiyala, K. R., and Oberhelman, D. (1986), "Estimation of Actual Realizations in Stochastic Parameter Models," in *Bayesian Inference and Decision Techniques*, eds. P. Goel and A. Zellner, Amsterdam: Elsevier, pp. 363–373.

Lee, L.-F., and Griffiths, W. E. (1978), "The Prior Likelihood and Best Linear Unbiased Prediction in Stochastic Coefficient Linear Models," unpublished manuscript, University of Minnesota, Dept. of Economics.

Maddala, G. S. (1971), "Generalized Least Squares With an Estimated Variance Covariance Matrix," *Econometrica*, 39, 23–33.

——— (1991), "To Pool or Not To Pool: That Is the Question," *Journal of Quantitative Economics*, 7, 255–264.

Mairesse, J., and Griliches, Z. (1990), "Heterogeneity in Panel Data: Are There Stable Production Functions?" in *Essays in Honor of Edmond Malinvaud* (Vol. 3), Cambridge, MA: MIT Press.

Nickell, S. (1985), "Error Correction, Partial Adjustment And All That: An Expository Note," *Oxford Bulletin of Economics and Statistics*, 47, 119–129.

Pagan, A. (1986), "Two Stage and Related Estimators and Their Application," *Review of Economics Studies*, 53, 517–538.

Pesaran, M. H., and Smith, R. (1995), "Estimating Long-Run Relationships From Dynamic Heterogeneous Panels," *Journal of Applied Econometrics*, 68, 79–113.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: John Wiley.

——— (1975), "Simultaneous Estimation of Parameters in Different Linear Models and Applications to Biometric Problems," *Biometrics*, 31, 545–554.

——— (1987), "Prediction of Future Observations in Growth Curve Models" (with discussion), *Statistical Science*, 2, 434–471.

Robertson, D., and Symons, J. (1992), "Some Strange Properties of Panel Data Estimators," *Journal of Applied Econometrics*, 7, 175–189.

Rubin, D. B. (1980), "Using Empirical Bayes Techniques in the Law School Validity Studies," *Journal of the American Statistical Association*, 75, 801–827.

Scott, E. L. (1980), Comment on "Using Empirical Bayes Techniques in the Law School Validity Studies," by D. B. Rubin, *Journal of the American Statistical Association*, 75, 821–823.

Smith, A. F. M. (1973), "A General Bayesian Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 35, 67–75.

Swamy, P. A. V. B. (1970), "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38, 311–323.

Taylor, L. D. (1975), "The Demand For Electricity: A Survey," *Bell Journal of Economics*, 6, 74–110.

U.S. Energy Information Administration (1993a), *The State Energy Price and Expenditure System*, Washington, DC: Author.

——— (1993b), *State Energy Data System*, Washington, DC: Author.

Zeimer, R. F., and Wetzstein, M. E. (1983), "A Stein-rule Method for Pooling Data," *Economics Letters*, 11, 137–143.