ANNALS OF ECONOMICS AND FINANCE **2**, 1–30 (2001)

A Comparative Study of Different Shrinkage Estimators for Panel Data Models

G. S. Maddala*

Department of Economics, Ohio State University, USA

Hongyi Li†

Department of Decision Sciences and Managerial Economics
Chinese University of Hong Kong, Hong Kong

and

V. K. Srivastava

Lucknow University, India

The present paper uses small-sigma asymptotics to show that in general the shrinkage estimators have superior properties among the individual least squares estimators, the simple average estimators, the weighted average estimators, estimators obtained by shrinking towards the simple average, and estimators obtained by shrinking towards the weighted average.

The shrinkage estimators are used to derive short-run and long-run price and income elasticities for residential natural gas demand and electricity demand in the US based on panel data covering 49 states over 21 years (1970-90). They are also used for out of sample forecasting.

Key Words: Bayesian shrinkage estimator; Parameter heterogeneity; Stein-rule estimator.

JEL Classification Numbers: C12, C13, C23.

* The first author wishes to acknowledge research support from the U.S. Energy Information Administration (EIA). The EIA is, however, not responsible for the opinions expressed.

† Corresponding author: Dr. Li Hongyi, 302 KK Leung Building, Department of Decision Sciences and Managerial Economics, Shatin, Chinese University of Hong Kong, NT, Hong Kong

1529-7373/2001

1. INTRODUCTION

In the analysis of panel data, a question often asked is whether to estimate the model separately for the different individual cross-section units or to estimate the model by pooling the entire data set. Often in empirical work, we are in a situation where we can get individual forecasts from either the pooled equation or from the individual equations; we can get aggregate forecasts either by adding individual forecasts from the pooled equation or by adding individual forecasts from the individual equations.

The problem that we are concerned with in this paper is that of estimation of cross-sectional parameters when there is parameter heterogeneity among the different cross-section units. Robertsan and Symons (1992) and Pesaran and Smith (1993) discuss the biases that are likely to occur if the parameter heterogeneity is ignored and data are pooled. The main problem they are concerned with is the estimation of aggregate parameters. The focus of this paper, by contrast, is the estimation of the individual heterogeneous parameters.

The parameters of any individual cross-sectional unit can be estimated by OLS from just the data for that particular unit. Alternatively, one could pool the data and use the parameters from the pooled regression. These two alternatives represent two extremes: the first says that all cross-section units are different. The other says that they are all identical in their behavior. The truth is perhaps somewhere in between, that the parameters have some common elements. In an earlier study of the prediction of the performance of students admitted to law schools, Rubin (1980) found that one could get better predictions by using the data on the students admitted to the other law schools and assuming that there are some common elements in the different cross-section units. In this paper we use similar empirical bayesian procedures in the estimation of short-run and long-run elasticities for residential demand for electricity and natural gas for each of the 49 states considered. In general, these estimators turn out to be some form of shrinkage estimators where the individual cross-sectional estimators are "shrunk" towards an average estimator. The question is what this average should be and what the shrinkage factor should be.

As reviewed in Maddala (1991), the shrinkage estimators appear to be better than either the pooled estimator or the individual least squares estimators. For further discussion, see Maddala and Hu (1996). There are cases where the pooled regression gives inconsistent results, while the shrinkage estimator gives consistent estimates. In some other cases, the shrinkage estimators are capable of stabilizing the resulting estimates from each individual equations. See the results in Maddala et al. (1997) and Choi and Li (2000).

In this paper, we investigate the properties of shrinkage estimators in panel data models. In particular, the small disturbance asymptotic approximations are developed for the shrinkage estimators. It is shown that in general the shrinkage estimators have superior properties among the individual least squares estimators, the simple average estimators, the weighted average estimators, estimators obtained by shrinking towards the simple average, and estimators obtained by shrinking towards the weighted average.

This paper is organized as follows. In Section 2, we discuss the properties concerning the small disturbance asymptotic approximations of the shrinkage estimators. The relevant derivations are summarized in the Mathematical Appendix. The estimators considered in this section are based on the assumption of homoskedastic errors and exogenous explanatory variables. Since the empirical application involves a lagged dependent variable we discuss this case in Section 3. In Section 4, we provide an empirical illustration of the implementation of the shrinkage estimators. Section 5 presents the conclusions.

2. THE SHRINKAGE ESTIMATORS

The N-sector cross-section equations are

$$y_i = X_i\beta_i + \sigma u_i, \qquad i = 1, 2, \cdots, N$$

where y_i is $T \times 1$, X_i is $T \times k$, β_i is $k \times 1$ and u_i is $T \times 1$, T being the number of observations. X_i is a set of exogenous variables. Assume that u_i are normally distributed errors with mean zero, $E(u_i) = 0$. Consider the simplest model where

$$E(u_i u_j') = \begin{cases} I_T & if\ i = j \\ 0 & if\ i \neq j \end{cases} .$$

We are, thus, ruling out any heteroskedasticity. Writing all the equations compactly, we have

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix} + \sigma \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}$$

or

$$y = X\beta + \sigma u$$

where y is $NT \times 1$, X is $NT \times Nk$, u is $NT \times 1$. Let us first define $J' = (I_k I_k \cdots I_k)$, an $NT \times k$ matrix. Thus define the following parameter vectors

$$\overline{\beta} = N^{-1} \sum \beta_i = N^{-1} J' \beta \qquad \text{(simple average)}$$

$\overline{\beta}_W = \left(\sum X_i' X_i\right)^{-1} \left(\sum X_i' X_i \beta_j\right) = (J'X'XJ)^{-1} J'X'X\beta$ (weighted average, or equivalently the parameter in the pooled regression without assuming either the fixed effect or random effect) together with the following two matrices

$$R = I_{NK} - J(J'X'XJ)^{-1} J'X'X$$

$$D = I_{NK} - N^{-1} JJ'$$

where R and D are idempotent matrices, which will be repeatedly used later. All the sumations $\sum$ are over $i = 1, 2, \cdots, N$ unless otherwise specified.

The null hypothesis being tested is $H_0 : \beta_1 = \beta_2 = \cdots = \beta_N (= \overline{\beta} = \overline{\beta}_W)$. Consider the following equivalent formulation of the null

$$\begin{aligned} &H_0 : (\beta_1 - \overline{\beta}_W, \beta_2 - \overline{\beta}_W, \cdots, \beta_N - \overline{\beta}_W)' \\ &= \beta - J\overline{\beta}_W = (I_{NK} - J(J'X'XJ)^{-1} J'X'X)\beta = R\beta = 0 \end{aligned}$$

The conventional F-ratio test for testing $H_0 : R\beta = 0$ (without adjustment for the degrees of freedom) is

$$f = \frac{\hat{\beta}' R' (R(X'X)^{-1} R')^{-1} R\hat{\beta}}{(y - X\hat{\beta})'(y - X\hat{\beta})}$$

where

$$\hat{\beta} = (X'X)^{-1} X'y = (\hat{\beta}_1 \hat{\beta}_2 \cdots \hat{\beta}_N)'$$

which is the simple OLS estimate of the individual equations.

To the question of whether or not to pool, the general answer from the prediction point of view is that shrinking the individual least squares estimators towards a common mean has been found to be better than the use of either individual least squares estimators or the pooled estimator. As for what the common mean ought to be, there have been many suggestions in the literature. See the survey in Maddala (1991). Here, we discuss the following five predictors for $E(y) = X\beta$. They are P_0, P_A, P_W, P_{AS} and P_{WS} which are closely related to each other and are defined below. The main results are presented here.

I. Predictor ignoring H_0:

$$P_0 = X\hat{\beta}.$$

II. Predictors incorporating H_0:

$$P_A = XJ(N^{-1}\sum\hat{\beta}_i) = N^{-1}XJJ'\hat{\beta}$$

which uses the simple average estimator and

$$P_W = XJ(\sum X_i'X_i)^{-1}(\sum X_i'X_i\hat{\beta}_i) = XJ(J'X'XJ)^{-1}J'X'X\hat{\beta}$$

which uses the weighted average estimator (or the pooled estimator).

III. Predictors shrinking $\hat{\beta}$ towards simple average $(N^{-1}\sum\hat{\beta}_i)$ and weighted average $(\sum X_i'X_i)^{-1}(\sum X_i'X_i\hat{\beta}_i)$, respectively:

$$\begin{aligned} P_{AS} &= X(N^{-1}JJ'\hat{\beta} + (1 - f^{-1}C_a)(\hat{\beta} - N^{-1}JJ'\hat{\beta})) \\ &= P_A + (1 - f^{-1}C_a)(P_0 - P_A) \\ &= X\hat{\beta} - f^{-1}C_aXD\hat{\beta} = P_0 - f^{-1}C_a(P_0 - P_A) \end{aligned}$$

and

$$\begin{aligned} P_{WS} &= X(J(J'X'XJ)^{-1}J'X'X\hat{\beta} \\ &+ (1 - f^{-1}C_w)(\hat{\beta} - J(J'X'XJ)^{-1}J'X'X\hat{\beta})) \\ &= P_0 - f^{-1}C_w(P_0 - P_W) \end{aligned}$$

where C_a and C_w are positive characterizing scalars.

Now let us study the performance properties of the five predictors P_0, P_A, P_W, P_{AS} and P_{WS} for $X\beta$. We give the analytical results of the predictive bias vector PB and the predictive mean squared error PM for any predictor $\hat{P}$ for $X\beta$ defined as

$$PB(\hat{P}) = E(\hat{P} - X\beta)$$

$$PM(\hat{P}) = E(\hat{P} - X\beta)'(\hat{P} - X\beta) = PV(\hat{P})$$

where $PV(\hat{P})$ is the predictive variance if $\hat{P}$ is unbiased for $X\beta$. We obtain the following results which are derived in the Appendix.

$$PB(P_0) = 0 \tag{1}$$
$$PV(P_0) = \sigma^2 NK \tag{2}$$
$$PB(P_A) = -XD\beta \tag{3}$$
$$PM(P_A) = \theta_A + \sigma^2 N^{-2} tr(J'(X'X)^{-1}JJ'X'XJ) \tag{4}$$
$$PB(P_W) = -XR\beta \tag{5}$$
$$PM(P_W) = \theta_W + \sigma^2 K \tag{6}$$

where

$$\theta_A = \beta' D' X' X D \beta \qquad \text{and} \qquad \theta_W = \beta' R' X' X R \beta.$$

If the null hypothesis H_0 is true, it is easy to show that $D\beta = R\beta = 0$. All the three predictors are unbiased. Further, we have

$$PV(P_W) \leq PV(P_0).$$

To compare P_A and P_W, employing result (2) of Theorem 2 on page 201 in Magnus and Neudecker (1988), we observe that

$$\begin{aligned} & tr(J'(X'X)^{-1}JJ'X'XJ) \\ = \ & tr(J'(X'X)^{-1/2}(X'X)^{-1/2}JJ'(X'X)^{1/2}(X'X)^{1/2}J) \\ \geq \ & tr(J'(X'X)^{-1/2}(X'X)^{1/2}J)^2 = tr(J'J)^2 = tr(\sum I_K)^2 = N^2K \end{aligned}$$

or

$$N^{-2}tr(J'(X'X)^{-1}JJ'X'XJ) \geq K \tag{7}$$

where it follows that $PV(P_W) \leq PV(P_A)$.

Thus when H_0 is tenable, P_W is the best choice among P_0, P_A and P_W. On the other hand, when H_0 is not tenable, P_0 is unbiased since H_0 is not incorporated while P_A and P_W are generally biased. So let us compare the biased predictors P_A and P_W. Examining the norms of predictive bias vectors, we see that

$$\begin{aligned} & \|PB(P_A)\| - \|PB(P_W)\| \\ = \ & \beta' D X' X D \beta - \beta' R' X' X R \beta (= \theta_A - \theta_W) \\ = \ & \delta'(J'X'XJ)^{-1}\delta \geq 0 \end{aligned}$$

where $\delta = J'X'X(I_{NK} - N^{-1}JJ')\beta$. Thus P_W is better than P_A with respect to the criterion of norm of predictive bias vector. Also we find that

$$\theta_A - \theta_W \geq 0. \tag{8}$$

Using (7) and (8), it follows from (4) and (6) that P_W has smaller predictive mean squared error in comparison to P_A.

The following results are small disturbance asymptotic approximations. Let us assume that disturbances are small and normally distributed, $u \sim N(0, I_{NK})$. For the predictors P_{AS} and P_{WS}, the predictive bias vectors to the order $O(\sigma^2)$ are

$$PB(P_{AS}) = -\sigma^2 \frac{C_a N(T-K)}{\theta_W} X D \beta \quad (9)$$

$$PB(P_{WS}) = -\sigma^2 \frac{C_w N(T-K)}{\theta_W} X R \beta \quad (10)$$

while the predictive mean squared errors to the order $O(\sigma^4)$ are given by

$$PM(P_{AS} = \sigma^2 NK - \sigma^4 \frac{N(T-K)(N(T-K)+2)}{\theta_W} C_a \left(2\lambda - \frac{\theta_A}{\theta_W} C_a\right) \quad (11)$$

$$PM(P_{WS}) = \sigma^2 NK - \sigma^4 \frac{N(T-K)(N(T-K)+2)}{\theta_W} C_w (2\lambda - C_w) \quad (12)$$

where

$$\lambda = [(N-1)K - 2]/[N(T-K)+2]. \quad (13)$$

Comparing with respect to the criterion of norm of predictive bias vector to order $O(\sigma^2)$, we find P_{WS} to be better than P_{AS} when

$$\left(\frac{C_w}{C_a}\right)^2 < \frac{\theta_A}{\theta_W}$$

which will be satisfied at least so long as $C_w < C_a$ by virtue of (8) (i.e., $\theta_A/\theta_W \geq 1$).

It is interesting to note that when H_0 is not true, according to small disturbance asymptotic theory, P_0 is unbiased and consistent, P_A and P_W are neither unbiased nor consistent, P_{AS} and P_{WS} are not unbiased but consistent.

From (2), (11) and (12), it is seen that P_{AS} and P_{WS} have smaller predictive mean squared errors, to the order $O(\sigma^4)$, in comparison to the predictive variance of P_0 when

$$0 < C_a < \frac{2\lambda\theta_W}{\theta_A}; \lambda > 0 \quad (14)$$

$$0 < C_w < 2\lambda; \lambda > 0 \quad (15)$$

Notice that the range of C_w is wider than the range of C_a, using (8).

If we minimize (11) with respect to C_a, the optimal value of C_a is $\lambda\theta_W/\theta_A$ and then the associated value of $PM(P_{AS})$ is

$$PM(P_{AS})_{opt} = \sigma^2 NK - \sigma^4 \frac{N(T-K)(N(T-K)+2)\lambda^2}{\theta_A}. \quad (16)$$

Similarly, the optimal value of $PM(P_{WS})$ is

$$PM(P_{WS})_{opt} = \sigma^2 NK - \sigma^4 \frac{N(T-K)(N(T-K)+2)\lambda^2}{\theta_W}. \tag{17}$$

From (8) we observe that $PM(P_{WS})_{opt} \leq PM(P_{AS})_{opt}$, implying that P_{WS} with optimal C_w is the best choice among P_0, P_A, P_W, P_{AS} and P_{WS}.

The general conclusion is to use the shrinkage estimator which shrinks towards the weighted average $\overline{\beta}_W$.

3. THE CASE OF LAGGED DEPENDENT VARIABLES

In the previous section, we considered the case of homoskedastic errors and exogenous explanatory variables. In empirical applications, however, we often have to include lagged dependent variables. There is considerable evidence that estimation of static panel data models from dynamic panel data introduces substantial biases. See, for instance, Doel and Kiviet (1993) and Ridder and Wansbeek (1990).

However, we have found that extending the results in the previous section on small-sigma asymptotics to the case of lagged dependent variable is extremely cumbersome, and no analytical results can be obtained. The results from the Monte Carlo study in Hu and Maddala (1994), however, indicate that the results in the previous section probably carry over to the case where lagged dependent variables are present. They consider the following model:

$$y_{it} = \alpha + \lambda_i y_{i,t-1} + \beta_{1i} x_{it} + \beta_{2i} x_{i,t-1} + u_{it}$$

for $i = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$. In their simulation they consider the case $N = 49$ and $T = 21$ which corresponds to the empirical application reported in section 4 here. All variables are assumed to be in logs so that the short-run elasticity is given by β_{1i} and the long-run elasticity is given by $(\beta_{1i} + \beta_{2i})/(1 - \lambda_i)$. They consider the following estimators for the parameter vector $\theta_i = (\alpha, \lambda_i, \beta_{1i}, \beta_{2i})$

(i) θ_i estimated from time series data on the i-th cross-section unit.

(ii) θ_i estimated from the pooled cross-section and time-series data. This assumes $\theta_1 = \theta_2 = \cdots = \theta_N$ and thus neglects parameter heterogeneity.

(iii) The Stein-rule estimator

$$\tilde{\theta}_i = \left(1 - \frac{c}{F}\right)\hat{\theta}_i + \frac{c}{F}\hat{\theta}_p$$

where $\hat{\theta}_i$ is the OLS estimator of θ_i from the data on the i-th cross-section unit and $\hat{\theta}_p$ is the estimator of θ from the pooled data. F is the F-statistic for testing the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_N$ and

$$c = [(N-1)k - 2]/[N(T-k) + 2]$$

where k is the number of explanatory variables.

(iv) The shrinkage estimator considered in Smith (1973) and Maddala (1991) which is discussed in the next section.

(v) The within group estimator: This is the pooled regression with dummy variables.

(vi) The between group estimator: This is based on time averages. The parameter estimates from this regression are usually interpreted as measuring the long-run effect. See for instance Baltagi and Griffin (1984).

(vii) Estimator obtained from aggregated data.

Their conclusions from a Monte Carlo study based on 2,000 replications and using the root mean squared error (RMSE) as the criterion of choice among the different estimators, is that the shrinkage estimator (iv) dominates all the rest for both the estimation of short-run and long-run elasticities, as well as for out of sample prediction.

Thus, for the estimation of the short-run and long-run elasticities for each cross-section unit, the shrinkage estimator does the best. It also does the best so far as prediction is concerned. This leads us to conjecture that the conclusions from the small-sigma results presented in the previous section carry over to the case of lagged dependent variables.

4. AN EMPIRICAL APPLICATION

In this section we apply the shrinkage estimators to regressions of the per capita electricity residential demand and the per capita natural gas residential demand in the US. We are interested in estimating the short run and long run income elasticities and price elasticities of the U. S. (for each of the individual states). As discussed earlier, possible ways to do this can be based on individual regressions, pooled regressions, and shrinkage regressions. Consider the following cross-section regressions

$$y_i = X_i\beta_i + u_i \tag{18}$$

for $i = 1, 2, \cdots, N$. β_i is the regression parameter with dimension K. y_i is the per capita electricity residential demand or the per capita natural gas residential demand. For each $y_i, t = 1, 2, \cdots, T$. X_i is a set of exogenous variables, such as price, real per capita income, heating degree days and cooling degree days. Usually the lagged dependent variables are also included in this type of models.

The traditional approach to estimating regression coefficients β_i with either pooled cross-section and time series data or with panel data is a dichotomy of either estimating β_i from the data on the i^{th} cross-section unit or from the pooled sample. As we have discussed, shrinking each individual β_i from the i^{th} cross-section towards the weighted average $\overline{\beta}_W$ proves to be the best among the predictors given above. Assume that

$$y_i \sim N(X_i\beta_i, V_i) \tag{19}$$

in which we are interested in the special case that $V_i = \sigma_i^2 I$.

The shrinkage estimator discussed in Smith (1973) and Maddala (1991) is given as

$$\beta_i^* = \left(\frac{1}{s_i^2}X_i'X_i + {\Sigma^*}^{-1}\right)^{-1}\left(\frac{1}{s_i^2}X_i'X_i\hat{\beta}_i + {\Sigma^*}^{-1}\hat{\beta}_W\right) \tag{20}$$

where

$$s_i^2 = \frac{1}{T+2}(y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*) \tag{21}$$

and

$$\Sigma^* = \frac{1}{N-K-1}\sum(\beta_i^* - \hat{\beta}_W)(\beta_i^* - \hat{\beta}_W)' \tag{22}$$

Further, $\hat{\beta}_i$ is the OLS estimator based on each separate cross-section unit, i.e., $\hat{\beta}_i = (X_i'X_i)^{-1}X_i'y_i$. Note that $\hat{\beta}_W = N^{-1}\sum\beta_i^*$ is the simple average of the shrinkage estimates. Equation (20) also shows that β_i^* is a weighted average of the OLS estimators $\hat{\beta}_i$ and an estimator for the prior mean $\hat{\beta}_W$ with the weights inversely proportional to the variances.

Equations (20), (21), and (22) have to be solved iteratively. The initial iteration uses the OLS estimate $\hat{\beta}_i$ to compute $\hat{\beta}_W$, s_i^2 and Σ^*. As a matter of fact, the initial values of $\hat{\beta}_W$ can be set as either $\hat{\beta}_W = N^{-1}\sum\hat{\beta}_i$, or $\overline{\beta}_W = (\sum X_i'X_i)^{-1}(\sum X_i'X_i\hat{\beta}_i)$, or even the pooled regression estimates. We have found that all give the same results if the iteration converges. In practice, (21) and (22) are estimated as

$$s_i^2 = \frac{1}{T+2+\nu_i}\left(\nu_i\lambda_i + (y_i - X_i\beta_i^*)'(y_i - X_i\beta_i^*)\right) \tag{21$'$}$$

and

$$\Sigma^* = \frac{1}{N-K-1+\delta}\left(R + \sum(\beta_i^* - \hat{\beta}_W)(\beta_i^* - \hat{\beta}_W)'\right) \tag{22$'$}$$

where, as discussed in Smith (1973), ν_i, λ_i, R and δ are parameters arising from prior specifications. Approximations to vague priors are obtained by setting $\nu_i = 0, \delta = 1$, and R to be a diagonal matrix with small positive entries (e.g., 0.001).

Now, let us turn to the problem of estimating the short-run and long-run elasticities of demand for electricity and natural gas for each of the states. The data are annual from 1970 to 1990 for the 49 states in the U. S. First a brief description of the data. Annual data on state residential electricity and gas price, residential electricity and gas consumption, and population used in this study were obtained from the State Energy Data System of the Energy Information Administration (1993). Weather data were obtained from the Local Climatological Data Series, prepared by the National Oceanographic and Atmospheric Administration (1993). Annual state income data were drawn from the Bureau of Labor Statistics and the annual Consumer Price Index for the U. S. was from the CITIBASE.

With only 21 time series observations, the sample may be rather small and would result in very few degrees of freedom when we include more right hand side variables. Thus, the variables are lagged only one period if lagged variables are included. The regressions takes the following form:

$$y_{it} = \alpha_{i0}+\alpha_{i1}y_{i,t-1}+\alpha_{i2}x_{it}^1+\alpha_{i3}x_{i,t-1}^1+\alpha_{i4}x_{it}^2+\alpha_{i5}x_{i,t-1}^2+\alpha_{i6}x_{it}^3+\alpha_{i7}x_{it}^4+u_{it}$$

or

$$y_{it} = X_{it}\alpha_i + u_{it}$$

for individual states $i = 1, 2, \cdots, 49$ and years $t = 1, 2, \cdots, 21$ with the first year being 1970. Note that $\alpha_i = (\alpha_{i0}, \alpha_{i1}, \cdots, \alpha_{i7})'$, the coefficient vector, and X_{it} is a matrix of all the right-hand side variables.

The variables for the electricity regression are
y_{it} = ln(Residential electricity per capita consumption);
x_{it}^1 = ln(Per capita personal income);
x_{it}^2 = ln(Residential electricity price);
x_{it}^3 = Heating degree days (HDD);
x_{it}^4 = Cooling degree days (CDD).

For the natural gas regression, we have
y_{it} = ln(Residential natural gas per capita consumption);
x_{it}^1 = ln(Per capita personal income);
x_{it}^2 = ln(Residential natural gas price);

with x_{it}^3 and x_{it}^4 unchanged.

For the i^{th} cross section unit the elasticities are calculated as
Short run income elasticity: $i_{SR} = \alpha_{i2}$
Long run income elasticity: $i_{LR} = \frac{\alpha_{i2}+\alpha_{i3}}{1-\alpha_{i1}}$
Short run price elasticity: $p_{SR} = \alpha_{i4}$
Long run price elasticity: $p_{LR} = \frac{\alpha_{i4}+\alpha_{i5}}{1-\alpha_{i1}}$

respectively. First, the null $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{49}$ is tested using the F test, which is given by

$$F = [(RRSS - URSS)/J]/[URSS/(NT - NK)]$$

where J is number of linear restrictions imposed. The regressions are estimated separately for each individual states. Then the pooled regression (without adding dummy variables) is estimated imposing the null that the coefficient estimates are the same for the different states. For the electricity regression, based on the above unrestricted and restricted estimates, the F test-statistic for testing H_0 is $F((49-1)\times 8, 49\times(20-8)) = 1.655$. The pooled regression is also estimated with 49 intercept dummies assuming fixed effects, i.e.,

$$y_{it} = X_{it}\alpha_i + \sum_{s=1}^{49} \delta_s d_s + u_{it}$$

where

$$d_{st} = \begin{cases} 1 & \text{if } t \in (T\times(s-1)+1, T\times s) \\ 0 & \text{otherwise.} \end{cases}$$

and there is no overall constant term in the regression matrix X_{it}. In this case the F test is $F((49-1)\times 7, 49\times(20-8)) = 1.134$. The 5% critical values of the F statistic is 1.00. In both cases, the null hypothesis is rejected for the electricity regression. The two test F statistics for the natural gas regression are 2.312 and 1.591, respectively. Again, the null hypothesis is rejected. As discussed earlier, when H_0 is not true, P_{WS} is the best choice among P_0, P_A, P_W, P_{AS} and P_{WS}. For comparison purposes, we report the regression results in three cases: (1) individual OLS estimates, (2) pooled regression (without dummies), and (3) shrinkage estimates. The parameter estimates and the t- values for the OLS and shrinkage estimator are reported in Table A1 to A4 in the Appendix. In general, the OLS estimates are quite different from each other, while the shrinkage estimates are shrunk to the converged common mean with very small variations.

Table 1 gives the simple average estimates, the mean of the shrinkage estimates, and the pooled regression estimates. The elasticities based on the average parameter estimates from Table 1 are calculated. Note that in order to smooth out abnormal values we do not take the average of the individual elasticities. These elasticities are summarized in Table 2.

From Table 2, we observe that for both regressions, the elasticities based on simple average and the shrinkage are approximately the same, while the pooled regression gives quite different results. There is, however, one exception. For the long run income elasticity in the natural gas regression, the simple average gives a negative value, and the pooled regression gives a number greater than one. The long run income elasticity calculated from the shrinkage estimate is 0.0583. Of course, one can explain the results from different angles. Consider natural gas being mainly used for heating purposes, it seems that the long run income elasticity based on the shrinkage estimates is preferred.

TABLE 1.

Parameter estimates by different approaches

	ELECTRICITY REGRESSION			NATURAL GAS REGRESSION		
VARIABLE	(1)	(2)	(3)	(1)	(2)	(3)
constant	−5.2607	−4.2474	−0.2490	−0.6341	−2.0339	−0.2649
$y_{i,t-1}$	0.6089	0.6888	0.9282	0.2277	0.5504	0.9850
$x^1_{i,t}$	0.3949	0.3337	0.1461	0.2812	0.3110	0.0841
$x^1_{i,t-1}$	−0.0198	−0.0318	−0.1372	−0.5386	−0.2848	−0.0632
$x^2_{i,t}$	−0.1334	−0.1312	−0.1960	−0.1102	−0.1033	−0.0983
$x^2_{i,t-1}$	0.0727	0.0795	0.1285	−0.0527	−0.0361	0.0861
$x^3_{i,t}$	0.1812	0.1670	0.0632	0.5180	0.4066	0.0367
$x^4_{i,t}$	1.0099	1.0198	0.3074	0.4640	−0.9341	−0.0474

Note
(1) Simple average estimates of individual OLS estimates;
(2) Mean of shrinkage estimates;
(3) Weighted average estimates, or pooled regression estimates.

TABLE 2.

Income and price elasticities

	ELECTRICITY REGRESSION			NATURAL GAS REGRESSION		
	(1)	(2)	(3)	(1)	(2)	(3)
i_{SR}	0.3949	0.3337	0.1461	0.2812	0.3110	0.0841
i_{LR}	0.9591	0.9701	0.1240	−0.3333	0.0583	1.3933
p_{SR}	−0.1334	−0.1312	−0.1960	−0.1102	−0.1033	−0.0983
p_{LR}	−0.1552	−0.1661	−0.9401	−0.2109	−0.3101	−0.8133

Note
(1) Based on simple average estimates of individual OLS estimates;
(2) Based on mean of shrinkage estimates;
(3) Based on weighted average estimates, or the pooled regression estimates.

TABLE 3.

Forecast (last two observations) RMSE's of the electricity regression

	Period	Pooled regression	OLS Estimator	Shrinkage Estimator
Electricity	$T-1$	0.000451	0.000718	0.000505
regression	T	0.000706	0.001153	0.000911
Natural gas	$T-1$	0.002486	0.001309	0.001585
regression	T	0.001403	0.001703	0.002174

We also compare the forecasting results of the energy demand at the state level using different estimators. For forecast purposes, the last two observations of each state are excluded when estimating the regressions. The forecast root mean squared errors are reported in Table 3. For the electricity regression the pooled regression gives the smallest root mean squared errors, while for the natural gas regression the OLS gives the smallest root mean squared errors, although all the three approaches have root mean squared errors with little difference. In both cases, the shrinkage estimator improves upon only either the OLS or the pooled regression. But, as shown in the estimation of elasticities, both the OLS and the pooled regression give bad elasticity estimates. On the other hand the regression parameters based on the shrinkage estimator are better. It is not uncommon to find (particularly in the presence of multicollinearity) that a method that gives bad estimates of individual parameters, nevertheless can do well when it comes to forecasting. Given that the shrinkage estimator considered here has been found (in Monte Carlo studies) to dominate the other estimators and that in the empirical example considered, its performance in forecasting is not uniformly dominated by the others, the shrinkage method can be recommended for the estimation of individual heterogeneous parameters.

When applying the shrinkage estimators, we found that the shrinkage estimates converge at different rates. For some variables, such as the lagged dependent variable in the elasticity regression, the convergence is rather slow, while other shrinkage estimates have already reached their "common mean". We programmed the iteration to be stopped when condition

$$\sum_{n=1}^{8} \mathrm{var}(\alpha^*_{i,n}) \leq 0.001, \quad \text{for } i = 1, 2, \cdots, 49,$$

is satisfied, where $\alpha^*_{i,n}$ is the shrinkage estimate of the n^{th} parameter. For the electricity regression, it required 13 iterations, whereas, it required 117 iterations for the natural gas regression. If the iteration is done leaving out R completely, soon $\sum^*$ will become singular because the distance between the shrinkage estimates and the common mean towards which the shrinkage estimates are approaching disappears. The calculation stops without proper convergence. By adding $R, \sum^*$ is always positive definite. The selection of R will affect the number of iterations. However, once the convergence criterion is set, the final results are basically the same.

5. CONCLUSIONS

The paper discusses the role of the shrinkage estimators in panel data models. In particular, the small disturbance asymptotic approximations are developed for the shrinkage estimators. It is shown that in general the shrinkage estimators have superior properties than other choices in the estimation of panel data models.

As an illustration example, we applied the shrinkage estimator to estimate the income elasticities and price elasticities in the short run and long run, for the residential electricity and residential natural gas demand equations. The estimates were quite different for the different states. The shrinkage estimators were close to the mean of the estimators for the individual states. The estimates from the pooled model were however, much different. The theoretical analysis presented in Section 2 and the tests for equality of the coefficients, however, suggest that the pooled estimates should not be used. The estimates of the long-run price elasticities were about 4-5 times higher from the pooled estimator than the average of the estimates for the individual states. These elasticities are very misleading for policy purposes. Further analysis based on prediction errors showed that the shrinkage estimators should be preferred to the individual estimators or the pooled estimators.

APPENDIX

In this Appendix, the derivation of our main results are presented. For the sake of completeness and ready reference, we give the exact expressions related to P_0, P_A, and P_W although they are extremely straightforward.

First we observe that

$$E(u) = 0 \quad \text{and} \quad E(u'u) = I_{NK}.$$

Now we have

$$P_0 - X\beta = X(\hat{\beta} - \beta) = \sigma X(X'X)^{-1}X'u$$

so that

$$E(P_0 - X\beta) = 0$$

$$
\begin{aligned}
E(P_0 - X\beta)'(P_0 - X\beta) &= \sigma^2 tr X(X'X)^{-1}X' \\
&= \sigma^2 tr I_{NK} \\
&= \sigma^2 NK
\end{aligned}
$$

which are the results (1) and (2).

$$
\begin{aligned}
P_A - X\beta &= N^{-1}XJJ'\hat{\beta} - \beta \\
&= N^{-1}XJJ'(\hat{\beta} - \beta) - X(I_{NK} - N^{-1}JJ')\beta \\
&= \sigma N^{-1}XJJ'(X'X)^{-1}X'u - XD\beta
\end{aligned}
$$

so that

$$E(P_A - X\beta) = -XD\beta$$

$$
\begin{aligned}
& E(P_A - X\beta)'(P_A - X\beta) \\
= \;& \sigma^2 N^{-2} tr XJJ'(X'X)^{-1}X'E(uu')X(X'X)^{-1}JJ'X' + \beta'DX'XD\beta
\end{aligned}
$$

which lead to results (3) and (4).

Similarly, we have

$$
\begin{aligned}
P_W - X\beta &= XJ(J'X'XJ)^{-1}J'X'X\hat{\beta} - \beta \\
&= XJ(J'X'XJ)^{-1}J'X'X(\hat{\beta} - \beta) \\
&- X(I_{NK} - J(J'X'XJ)^{-1}J'X'X)\beta \\
&= \sigma XJ(J'X'XJ)^{-1}J'X'u - XR\beta
\end{aligned}
$$

so that

$$E(P_W - X\beta) = -XR\beta$$

$$
\begin{aligned}
& E(P_W - X\beta)'(P_W - X\beta) \\
= \;& \sigma^2 tr XJ(J'X'XJ)^{-1}J'X'E(uu')XJ(J'X'XJ)^{-1}J'X' + \beta'R'X'XR\beta \\
= \;& \sigma^2 tr I_K + \beta'R'X'XR\beta
\end{aligned}
$$

which provide the results (5) and (6).

Now consider shrinkage predictors. From the definition of F-ratio on page 4 we notice that

$$\begin{aligned}
-\frac{1}{f} &= -\frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{\hat{\beta}'R'(R(X'X)^{-1}R')^{-1}R\hat{\beta}} \\
&= -\frac{\sigma^2 u'Mu}{(\beta+\sigma(X'X)^{-1}X'u)'\Omega(\beta+\sigma(X'X)^{-1}X'u)} \\
&= -\frac{\sigma^2 u'Mu}{\theta_w + 2\sigma\beta'\Omega(X'X)^{-1}X'u + \sigma^2 u'X(X'X)^{-1}\Omega(X'X)^{-1}X'u} \\
&= -\sigma^2\frac{u'Mu}{\theta_w}\left(1+2\sigma\frac{\beta'\Omega(X'X)^{-1}X'u}{\theta_w}+\sigma^2\frac{u'X(X'X)^{-1}\Omega(X'X)^{-1}X'u}{\theta_w}\right)^{-1}
\end{aligned}$$

or

$$-\frac{1}{f} = -\sigma^2\frac{u'Mu}{\theta_w} + 2\sigma^3\frac{u'Mu\cdot\beta'\Omega(X'X)^{-1}X'u}{\theta_w^2}+\cdots$$

where

$$\begin{aligned}
M &= I_{NT} - X(X'X)^{-1}X' \\
\Omega &= R'(R(X'X)^{-1}R')^{-1}R
\end{aligned}$$

and

$$\begin{aligned}
\theta_w &= \beta'\Omega\beta \\
&= \beta'R'X'XR\beta \\
&= \beta'(X'X - X'XJ(J'X'XJ)^{-1}J'X'X)\beta
\end{aligned}$$

Next, we observe that

$$\begin{aligned}
P_{AS} - X\beta &= X(\hat{\beta}-\beta) - f^{-1}C_a XD\hat{\beta} \\
&= \sigma X(X'X)^{-1}X'u - f^{-1}C_a(XD\beta + \sigma XD(X'X)^{-1}X'u)
\end{aligned}$$

Substituting the expression for $(-f^{-1})$ and retaining terms up to order $O_p(\sigma^3)$, we find

$$\begin{aligned}
P_{AS} - X\beta &= \sigma X(X'X)^{-1}X'u - \sigma^2\frac{C_a u'Mu}{\theta_w}XD\beta \\
&\quad -\sigma^3\frac{C_a u'Mu}{\theta_w}XD\left(I_{NK} - \frac{2}{\theta_w}\beta\beta'\Omega\right)(X'X)^{-1}X'u
\end{aligned}$$

so that the predictive bias vector to the order $O(\sigma^2)$ is

$$\begin{aligned} PB(P_{AS}) &= \sigma X(X'X)^{-1}X'E(u) - \sigma^2 \frac{C_a}{\theta_w} E(u'Mu)XD\beta \\ &= -\sigma^2 \frac{C_a N(T-K)}{\theta_w} XD\beta \end{aligned}$$

which is the result (9).

Similarly, the predictive mean squared error to order $O(\sigma^4)$ is given by

$$\begin{aligned} & PM(P_{AS}) = E(P_{AS} - X\beta)'(P_{AS} - X\beta) \\ = \ & \sigma^2 E(u'X(X'X)^{-1}X'u) - 2\sigma^3 \frac{C_a}{\theta_w} E(u'Mu \cdot \beta' DX'u) \\ - \ & \sigma^4 \frac{C_a}{\theta_w} E\left(2u'Mu \cdot u'XD\left(I_{NK} - \frac{2}{\theta_w}\beta\beta'\Omega\right)(X'X)^{-1}X'u \right. \\ - \ & \left. \frac{C_a \beta' DX'XD\beta}{\theta_w}(u'Mu)^2\right) \end{aligned}$$

Using the following result for any two fixed matrices A_1 and A_2 (assume A_1 to be symmetric)

$$E(u'A_1u \cdot u'A_2u) = (trA_1)(trA_2) + 2(trA_1A_2)$$

and employing the normality of $u(u \sim N(0, I_{NT}))$, we see that

$$\begin{gathered} E(u'X(X'X)^{-1}X'u) = NK \\ E(u'Mu \cdot \beta' DX'u) = 0 \end{gathered}$$

$$\begin{aligned} & E\left(u'Mu \cdot u'XD\left(I_{NK} - \frac{2}{\theta_w}\beta\beta'\Omega\right)(X'X)^{-1}X'u\right) \\ = \ & (trM)\left(trXD\left(I_{NK} - \frac{2}{\theta_w}\beta\beta'\Omega\right)(X'X)^{-1}X'\right) \\ = \ & (trM)\left(trD - \frac{2}{\theta_w}\beta'\Omega D\beta\right) \\ = \ & N(T-K)((N-1)K-2) \end{aligned}$$

because

$$\begin{aligned}
\Omega D &= R'(R(X'X)^{-1}R')RD \\
&= R'(R(X'X)^{-1}R')(I_{NK} - J(J'X'XJ)^{-1}J'X'X)(I_{NK} - N^{-1}JJ') \\
&= R'(R(X'X)^{-1}R')(I_{NK} - J(J'X'XJ)^{-1}J'X'X) \\
&= \Omega \\
E(u'Mu)^2 &= (trM)^2 + 2(trM) \\
&= N(T-K)(N(T-K)+2)
\end{aligned}$$

Substituting these terms we find the desired result. Similar results for P_{WS} can be obtained by just replacing D by R.

APPENDIX A

Table A1.

OLS estimates and t-values of individual states (Electricity regression)

STATE	y_{t-1}	x_t^1	x_{t-1}^1	x_t^2	x_{t-1}^2
AK	0.766	0.221	0.178	0.031	−0.069
	7.698	1.610	0.911	0.237	−0.523
AL	0.028	0.657	0.320	−0.126	−0.119
	0.124	1.919	0.717	−0.897	−0.784
AR	0.454	0.560	0.242	0.250	−0.516
	2.447	0.983	0.482	0.734	−1.508
AZ	0.633	0.808	−0.342	−0.086	−0.050
	4.998	3.066	−1.252	−0.585	−0.312
CA	0.668	0.382	−0.169	−0.274	0.099
	4.841	1.483	−0.717	−3.327	1.550
CO	0.493	−0.003	0.611	0.159	−0.029
	2.532	−0.004	0.971	0.584	−0.093
CT	0.638	0.331	−0.065	−0.069	−0.015
	6.706	1.750	−0.336	−1.251	−0.278
DE	0.697	−0.624	0.784	−0.638	0.558
	3.719	−0.591	0.670	−2.490	2.359
FL	0.556	1.225	−0.758	−0.164	0.008
	4.305	6.111	−3.125	−2.520	0.096
GA	0.326	0.779	−0.060	0.068	−0.087
	1.936	2.902	−0.195	0.565	−0.625
IA	0.628	0.153	0.321	0.141	−0.092
	5.064	0.840	1.924	0.575	−0.350

Table A1—*Continued*

STATE	y_{t-1}	x_t^1	x_{t-1}^1	x_t^2	x_{t-1}^2
ID	0.681	−0.327	0.645	0.281	−0.459
	7.823	−1.486	2.328	1.942	−2.983
IL	0.245	0.555	0.151	0.026	−0.053
	1.849	2.323	0.633	0.244	−0.431
IN	0.672	0.249	0.084	−0.005	−0.005
	6.088	1.908	0.554	−0.060	−0.046
KS	0.890	0.267	−0.193	−0.443	0.461
	4.377	0.603	−0.472	−2.143	2.009
KY	0.414	0.749	0.480	−0.112	0.258
	2.120	2.030	1.245	−0.589	1.081
LA	0.994	1.216	−1.379	0.031	0.179
	6.208	2.768	−2.552	0.166	1.086
MA	0.514	0.607	−0.265	−0.091	−0.069
	3.723	2.743	−1.018	−1.168	−0.948
MD	0.714	0.533	−0.157	−0.107	0.109
	6.677	2.605	−0.649	−1.530	1.616
ME	0.832	0.516	−0.463	−0.180	0.154
	20.545	3.179	−2.814	−2.768	2.475
MI	0.753	0.273	−0.124	−0.045	−0.003
	6.434	2.350	−1.103	−0.521	−0.037
MN	0.708	−0.106	0.366	0.005	0.159
	4.103	−0.547	2.154	0.039	0.914
MO	0.698	0.286	0.149	0.245	−0.295
	4.590	0.736	0.369	0.867	−0.896
MS	0.261	0.775	0.085	−0.277	−0.114
	1.268	2.664	0.222	−2.068	−0.952
MT	0.961	0.026	0.098	−0.602	0.528
	9.110	0.085	0.350	−3.160	2.420
NE	0.620	0.773	−0.365	−0.188	0.098
	7.849	6.788	−2.613	−2.606	1.247
NC	0.918	−0.076	0.079	−0.227	0.033
	17.254	−0.666	0.809	−0.788	0.103

Table A1—*Continued*

STATE	y_{t-1}	x_t^1	x_{t-1}^1	x_t^2	x_{t-1}^2
ND	0.787	0.361	−0.009	−0.054	0.042
	5.797	1.877	−0.053	−0.360	0.310
NH	0.700	0.661	−0.529	−0.094	−0.061
	10.432	3.609	−2.828	−1.506	−0.906
NJ	0.762	−0.036	0.127	−0.315	0.263
	10.422	−0.175	0.585	−5.889	5.064
NM	0.503	1.134	−0.391	−0.857	0.733
	2.934	2.212	−0.705	−4.129	4.293
NV	0.301	0.726	−0.296	−0.102	−0.089
	1.237	1.815	−0.825	−0.661	−0.490
NY	0.829	0.361	−0.204	−0.191	0.139
	7.323	1.884	−0.941	−3.271	2.744
OH	0.670	0.437	0.007	−0.081	0.043
	7.141	2.282	0.037	−0.899	0.477
OK	0.634	−0.093	0.524	−0.593	0.361
	1.882	−0.152	0.780	−2.044	1.287
OR	0.056	0.123	0.155	−0.272	0.141
	0.277	0.560	0.725	−2.680	1.191
PA	0.523	0.517	0.021	−0.221	0.162
	5.603	2.737	0.105	−2.968	2.069
RI	0.783	0.349	−0.222	−0.187	0.079
	8.099	1.164	−0.710	−3.106	1.287
SC	0.602	0.752	−0.255	−0.092	0.065
	3.419	2.833	−0.937	−0.670	0.418
SD	0.877	0.039	0.048	−0.065	−0.066
	10.000	0.211	0.313	−0.236	−0.273
TN	0.395	0.251	0.025	−0.003	−0.028
	1.986	0.610	0.065	−0.020	−0.174
TX	0.620	0.307	0.089	−0.403	0.232
	3.327	0.855	0.237	−2.931	1.712
UT	0.576	−0.481	0.731	−0.069	0.155
	3.710	−1.194	1.595	−0.487	1.038
VA	0.496	1.056	−0.487	0.094	−0.131
	3.603	2.721	−1.162	0.934	−1.205
VT	0.473	−0.093	−0.116	−0.486	0.531
	2.571	−0.104	−0.127	−1.302	1.433
WA	0.359	0.208	0.166	0.348	−0.430
	1.506	0.271	0.237	1.332	−1.738
WI	0.473	0.810	−0.275	−0.163	0.350
	2.360	2.302	−0.814	−0.878	1.548
WV	0.840	0.475	−0.296	−0.063	−0.062
	13.302	3.138	−1.643	−1.138	−0.858
WY	0.814	0.683	−0.037	−0.271	0.465
	19.307	3.181	−0.208	−2.290	3.387

Table A2.

Shrinkage estimates and t-values of individual states (Electricity regression)

STATE	y_{t-1}	x_t^1	x_{t-1}^1	x_t^2	x_{t-1}^2
AK	0.734	0.316	−0.009	−0.107	0.093
	28.315	12.382	−0.296	−3.470	3.089
AL	0.655	0.350	−0.047	−0.158	0.061
	18.870	13.598	−1.304	−5.351	2.117
AR	0.676	0.337	−0.032	−0.133	0.072
	16.629	12.383	−0.828	−3.864	2.182
AZ	0.666	0.339	−0.057	−0.142	0.077
	27.401	12.984	−1.900	−4.764	2.595
CA	0.582	0.354	−0.120	−0.188	0.050
	21.633	13.966	−3.694	−6.834	1.839
CO	0.718	0.325	−0.014	−0.117	0.084
	19.829	12.046	−0.365	−3.331	2.507
CT	0.611	0.338	−0.095	−0.138	0.063
	33.120	13.349	−3.272	−5.453	2.566
DE	0.700	0.333	−0.024	−0.138	0.079
	14.428	12.028	−0.543	−3.845	2.417
FL	0.621	0.357	−0.094	−0.163	0.076
	20.762	13.723	−2.956	−5.701	2.715
GA	0.690	0.332	−0.028	−0.122	0.079
	29.479	12.805	−0.902	−4.034	2.632
IA	0.728	0.316	0.006	−0.111	0.088
	23.444	12.019	0.181	−3.318	2.698
ID	0.718	0.335	0.008	−0.123	0.068
	19.274	12.674	0.201	−3.582	2.054
IL	0.673	0.322	−0.042	−0.122	0.081
	21.750	12.451	−1.205	−4.091	2.694
IN	0.699	0.326	−0.019	−0.112	0.088
	34.472	12.905	−0.639	−3.911	2.932
KS	0.738	0.321	−0.002	−0.114	0.096
	19.079	12.041	−0.064	−3.450	2.998
KY	0.784	0.316	0.043	−0.087	0.098
	33.390	12.096	1.339	−2.737	3.092
LA	0.728	0.326	−0.006	−0.114	0.090
	25.194	12.370	−0.186	−3.615	2.870
MA	0.616	0.350	−0.092	−0.172	0.043
	28.482	13.641	−3.062	−6.253	1.619
MD	0.747	0.322	0.006	−0.119	0.104
	53.696	12.596	0.219	−4.347	4.078
ME	0.685	0.330	−0.045	−0.116	0.105
	39.161	12.632	−1.511	−3.776	3.490
MI	0.645	0.335	−0.060	−0.147	0.055
	32.520	13.431	−2.081	−5.277	2.005

Table A2—*Continued*

MN	0.692	0.316	−0.027	−0.106	0.092
	21.943	11.888	−0.810	−3.092	2.825
MO	0.737	0.322	0.008	−0.112	0.085
	22.247	11.753	0.223	−3.134	2.485
MS	0.658	0.353	−0.045	−0.164	0.056
	16.642	13.380	−1.188	−5.084	1.805
MT	0.768	0.334	0.031	−0.130	0.074
	21.838	12.557	0.854	−3.665	2.225
NE	0.679	0.340	−0.043	−0.126	0.103
	40.549	13.355	−1.483	−4.405	3.567
NC	0.783	0.319	0.045	−0.106	0.088
	20.318	12.042	1.222	−2.910	2.581
ND	0.771	0.318	0.030	−0.099	0.088
	26.341	12.104	0.909	−2.994	2.720
NH	0.551	0.352	−0.139	−0.158	0.049
	29.519	13.808	−4.866	−6.033	1.875
NJ	0.632	0.330	−0.084	−0.150	0.099
	41.613	13.041	−2.963	−6.137	4.283
NM	0.698	0.332	−0.033	−0.139	0.084
	19.135	12.531	−0.864	−4.240	2.625
NV	0.581	0.358	−0.109	−0.168	0.051
	11.463	13.612	−2.534	−5.471	1.774
NY	0.717	0.326	−0.028	−0.136	0.099
	51.651	13.040	−0.983	−5.932	4.530
OH	0.728	0.331	−0.006	−0.124	0.081
	58.152	13.447	−0.217	−4.998	3.164
OK	0.703	0.333	−0.019	−0.132	0.081
	15.552	12.139	−0.463	−3.742	2.427
OR	0.567	0.361	−0.105	−0.174	0.062
	12.470	14.354	−2.570	−5.896	2.158
PA	0.704	0.343	−0.029	−0.148	0.082
	56.200	13.529	−1.039	−5.399	3.087
RI	0.651	0.336	−0.066	−0.147	0.069
	33.929	13.111	−2.224	−5.785	2.819
SC	0.715	0.334	−0.012	−0.120	0.095
	40.960	12.962	−0.395	−4.045	3.157
SD	0.756	0.319	0.023	−0.103	0.088
	22.488	12.131	0.681	−3.006	2.667
TN	0.617	0.352	−0.071	−0.146	0.066
	14.787	13.646	−1.814	−4.771	2.221
TX	0.691	0.333	−0.027	−0.147	0.074
	21.404	12.686	−0.794	−4.785	2.419
UT	0.686	0.322	−0.038	−0.114	0.095
	23.162	12.462	−1.092	−3.915	3.208

Table A2—*Continued*

VA	0.679	0.331	−0.039	−0.115	0.088
	37.650	12.878	−1.339	−4.074	3.179
VT	0.657	0.340	−0.053	−0.143	0.073
	10.871	12.235	−1.024	−3.755	2.111
WA	0.642	0.352	−0.052	−0.148	0.058
	15.049	13.793	−1.296	−4.941	1.947
WI	0.694	0.328	−0.036	−0.111	0.097
	22.918	12.389	−1.057	−3.359	3.053
WV	0.759	0.342	0.026	−0.129	0.058
	63.605	13.785	0.960	−4.944	2.187
WY	0.820	0.315	0.061	−0.092	0.111
	33.886	12.244	1.949	−2.900	3.680

Table A3.

OLS estimates and t-values of individual states (Natural gas regression)

STATE	y_{t-1}	x_t^1	x_{t-1}^1	x_t^2	x_{t-1}^2
AK	0.055	0.766	0.511	0.287	0.326
	0.204	0.806	0.467	0.475	0.510
AL	0.343	0.413	−0.690	−0.059	−0.085
	1.763	0.835	−1.434	−0.406	−0.508
AR	0.160	0.116	−0.494	−0.150	−0.027
	0.998	0.302	−1.490	−1.652	−0.292
AZ	0.274	−0.323	−0.760	−0.005	−0.325
	1.857	−0.628	−1.291	−0.042	−2.177
CA	−0.240	0.008	−2.087	0.147	−0.308
	−1.442	0.012	−2.921	1.271	−2.457
CO	0.059	0.273	−0.955	−0.313	0.200
	0.276	0.346	−1.506	−2.251	1.067
CT	−0.069	0.390	0.255	−0.073	−0.164
	−0.354	0.808	0.563	−0.730	−1.238
DE	0.166	0.593	−0.915	−0.320	0.036
	0.765	0.976	−1.416	−2.475	0.253
FL	0.327	1.846	−2.607	−0.581	−0.053
	2.145	2.182	−3.163	−2.348	−0.156
GA	0.469	0.757	−0.939	−0.446	0.387
	1.782	1.142	−1.697	−1.604	1.152
IA	−0.137	−0.466	−0.111	0.236	−0.414
	−0.739	−1.960	−0.479	1.529	−2.569

Table A3—*Continued*

ID	0.528	−2.722	2.039	−0.609	0.088
	2.967	−2.695	1.914	−1.751	0.243
IL	−0.088	1.165	−1.105	0.509	−0.597
	−0.461	2.390	−2.382	2.413	−2.776
IN	−0.044	−0.251	0.031	0.214	−0.314
	−0.183	−0.730	0.103	1.303	−1.680
KS	0.611	0.030	−0.451	−0.004	−0.050
	3.703	0.055	−0.999	−0.035	−0.394
KY	0.181	1.001	−1.694	0.081	−0.265
	1.913	3.557	−5.614	0.872	−2.499
LA	0.662	0.509	−0.132	−0.030	−0.250
	4.629	0.623	−0.130	−0.119	−0.972
MA	0.105	0.927	−0.344	0.388	−0.424
	0.609	2.014	−0.760	2.622	−2.609
MD	0.068	1.637	−1.803	−0.238	0.119
	0.315	2.968	−3.820	−2.685	1.149
ME	0.843	−1.291	1.167	−0.567	0.458
	6.451	−1.818	1.633	−4.080	2.597
MI	0.214	−0.423	0.315	0.027	−0.012
	0.868	−0.695	0.711	0.119	−0.045
MN	−0.242	−0.191	0.063	0.048	−0.229
	−1.034	−0.502	0.185	0.281	−1.142
MO	0.489	−0.253	−0.245	−0.054	−0.001
	2.839	−0.596	−0.613	−0.449	−0.007
MS	0.430	0.158	−0.918	0.088	−0.047
	2.117	0.178	−1.114	0.247	−0.129
MT	0.158	−0.057	−0.868	0.019	−0.344
	0.799	−0.103	−1.833	0.102	−1.693
NE	0.488	0.991	−0.947	−0.517	0.387
	3.150	1.849	−2.026	−2.386	1.597
NC	0.150	−0.242	0.224	0.113	−0.115
	0.613	−0.972	1.092	0.431	−0.465
ND	0.100	−0.681	−0.308	0.178	−0.267
	0.536	−2.094	−0.943	0.930	−1.297
NH	0.792	0.228	−0.090	−0.367	0.290
	4.236	0.347	−0.141	−2.710	1.550
NJ	0.368	1.596	−1.160	−0.092	−0.090
	1.639	1.893	−1.433	−0.481	−0.408
NM	0.298	−0.696	−0.124	−0.410	0.271
	1.018	−0.439	−0.078	−1.244	0.965
NV	−0.006	−0.510	0.799	0.122	−0.466
	−0.027	−0.683	1.175	0.602	−1.901
NY	0.025	1.061	−0.717	0.013	−0.129
	0.097	2.328	−1.627	0.116	−1.023

Table A3—*Continued*

OH	−0.442	−0.111	−0.934	−0.079	−0.188
	−2.047	−0.265	−2.524	−0.667	−1.273
OK	0.410	0.537	−1.072	0.325	−0.330
	2.292	0.950	−1.659	1.450	−1.772
OR	0.191	−2.307	2.449	−1.459	0.803
	1.007	−1.594	1.568	−3.637	2.075
PA	0.404	0.699	−0.903	−0.145	0.084
	1.588	1.191	−1.604	−0.960	0.497
RI	−0.120	0.397	0.760	0.085	−0.144
	−0.576	0.595	1.179	1.017	−1.546
SC	0.374	3.053	−3.064	−1.280	0.643
	2.904	2.905	−3.047	−2.553	1.168
SD	0.246	−0.318	0.086	−0.088	−0.087
	1.246	−1.085	0.292	−0.387	−0.353
TN	0.173	0.754	−0.725	−0.103	−0.139
	0.595	1.092	−1.150	−0.482	−0.509
TX	0.521	0.431	−1.583	0.320	−0.109
	3.545	0.417	−1.611	1.277	−0.400
UT	0.295	−0.098	−0.202	−0.507	0.380
	1.107	−0.036	−0.092	−1.118	0.907
VA	−0.036	2.433	−2.256	−0.197	−0.196
	−0.169	3.600	−3.674	−1.592	−1.323
VT	0.665	0.793	−0.322	−0.082	−0.104
	3.129	1.242	−0.448	−0.563	−0.832
WA	−0.146	0.393	−0.995	−0.032	−0.505
	−0.678	0.401	−0.943	−0.161	−2.018
WI	−0.028	0.734	−0.381	0.189	−0.357
	−0.088	1.103	−0.684	0.876	−1.346
WV	0.686	0.493	−0.841	−0.170	0.093
	5.655	1.237	−2.096	−2.102	1.051
WY	0.457	−0.346	−1.336	0.365	−0.191
	2.707	−0.442	−1.582	1.027	−0.689

Table A4.

Shrinkage estimates and t-values of individual states (Natural gas regression)

STATE	y_{t-1}	x_t^1	x_{t-1}^1	x_t^2	x_{t-1}^2
AK	0.552	0.303	−0.298	−0.115	−0.027
	14.606	12.310	−10.345	−3.110	−0.508
AL	0.535	0.317	−0.290	−0.118	−0.046
	14.969	12.811	−10.651	−4.458	−1.549
AR	0.528	0.319	−0.285	−0.105	−0.033
	14.594	13.028	−10.433	−4.031	−1.093
AZ	0.536	0.322	−0.298	−0.168	−0.141
	14.551	13.069	−10.550	−5.659	−3.540
CA	0.524	0.322	−0.295	−0.152	−0.112
	13.875	13.247	−10.710	−5.287	−3.035
CO	0.537	0.317	−0.286	−0.116	−0.024
	14.441	13.086	−10.607	−4.243	−0.769
CT	0.614	0.319	−0.264	−0.088	0.056
	19.792	13.185	−10.131	−3.096	1.715
DE	0.537	0.308	−0.299	−0.129	−0.050
	14.628	12.501	−10.772	−4.455	−1.385
FL	0.571	0.294	−0.312	−0.134	−0.048
	16.116	11.366	−10.424	−3.662	−0.890
GA	0.542	0.322	−0.278	−0.095	0.001
	15.297	13.139	−10.294	−3.391	0.044
IA	0.538	0.318	−0.279	−0.103	−0.019
	14.362	13.080	−10.187	−3.456	−0.486
ID	0.554	0.300	−0.308	−0.147	−0.080
	14.973	12.005	−10.637	−4.239	−1.590
IL	0.531	0.328	−0.274	−0.095	−0.015
	14.728	13.720	−10.379	−3.431	−0.472
IN	0.543	0.322	−0.274	−0.093	−0.000
	14.922	13.270	−10.219	−3.345	−0.008
KS	0.539	0.330	−0.279	−0.121	−0.061
	14.599	13.695	−10.344	−4.464	−1.886
KY	0.530	0.316	−0.298	−0.135	−0.079
	14.588	12.909	−10.855	−5.007	−2.490
LA	0.539	0.330	−0.283	−0.143	−0.091
	15.071	13.524	−10.359	−5.152	−2.735
MA	0.603	0.324	−0.263	−0.082	0.040
	19.350	13.514	−10.053	−2.671	1.040
MD	0.543	0.316	−0.287	−0.094	0.009
	15.300	12.975	−10.696	−3.358	0.269
ME	0.607	0.256	−0.342	−0.150	−0.044
	19.497	9.704	−11.722	−5.213	−1.205
MI	0.557	0.328	−0.263	−0.077	0.031
	15.498	13.633	−9.950	−2.746	0.968

Table A4—*Continued*

MN	0.546	0.313	−0.283	−0.097	−0.002
	14.953	12.869	−10.422	−3.269	−0.064
MO	0.539	0.325	−0.284	−0.119	−0.062
	14.660	13.434	−10.504	−4.362	−1.952
MS	0.547	0.317	−0.287	−0.114	−0.034
	15.140	12.731	−10.206	−3.763	−0.865
MT	0.535	0.311	−0.300	−0.137	−0.075
	14.170	12.699	−10.704	−4.560	−1.867
NE	0.570	0.302	−0.293	−0.094	0.020
	17.246	12.068	−10.676	−3.192	0.548
NC	0.563	0.299	−0.284	−0.092	0.012
	15.710	12.121	−10.266	−2.890	0.285
ND	0.533	0.319	−0.282	−0.113	−0.043
	14.188	13.156	−10.263	−3.822	−1.121
NH	0.622	0.304	−0.282	−0.103	0.038
	19.920	12.301	−10.455	−3.229	0.919
NJ	0.592	0.329	−0.260	−0.082	0.048
	17.896	13.715	−9.883	−2.651	1.213
NM	0.529	0.316	−0.290	−0.124	−0.044
	14.106	12.888	−10.343	−4.009	−1.059
NV	0.533	0.315	−0.288	−0.118	−0.044
	14.454	12.855	−10.421	−3.759	−1.042
NY	0.571	0.322	−0.272	−0.084	0.040
	17.072	13.329	−10.390	−3.199	1.474
OH	0.524	0.325	−0.283	−0.117	−0.048
	13.876	13.438	−10.512	−4.334	−1.524
OK	0.540	0.328	−0.277	−0.113	−0.045
	14.852	13.512	−10.237	−4.028	−1.328
OR	0.549	0.297	−0.307	−0.125	−0.038
	14.746	11.843	−10.598	−3.504	−0.728
PA	0.539	0.318	−0.285	−0.105	−0.011
	15.045	13.100	−10.672	−3.961	−0.385
RI	0.611	0.325	−0.259	−0.075	0.060
	18.820	13.374	−9.732	−2.382	1.481
SC	0.543	0.304	−0.304	−0.133	−0.060
	14.928	12.064	−10.479	−3.752	−1.168
SD	0.541	0.302	−0.293	−0.111	−0.026
	14.761	12.291	−10.465	−3.613	−0.628
TN	0.557	0.309	−0.291	−0.108	−0.007
	15.822	12.418	−10.499	−3.811	−0.202
TX	0.540	0.327	−0.283	−0.125	−0.063
	14.807	13.354	−10.225	−4.064	−1.586

Table A4—*Continued*

UT	0.530	0.323	−0.283	−0.117	−0.039
	13.918	13.252	−10.041	−3.370	−0.788
VA	0.545	0.306	−0.298	−0.113	−0.018
	15.670	12.392	−10.920	−3.758	−0.488
VT	0.635	0.298	−0.292	−0.116	0.010
	19.678	11.772	−10.303	−3.571	0.234
WA	0.538	0.294	−0.316	−0.148	−0.083
	14.654	11.769	−11.152	−4.878	−2.046
WI	0.550	0.317	−0.279	−0.091	0.011
	15.603	13.038	−10.495	−3.265	0.334
WV	0.543	0.322	−0.287	−0.134	−0.060
	14.793	13.195	−10.539	−5.248	−2.044
WY	0.551	0.316	−0.285	−0.106	−0.019
	14.591	12.966	−10.150	−3.240	−0.422

REFERENCES

Baltagi, B. H. and J. M. Griffin, 1984, Short and long run effects in pooled models. *International Economic Review* **25**, 631-645.

Choi, H. and H. Li, 2000, Economic development and growth convergence in China. *Journal of International Trade and Economic Development* **9**, 37-54.

Doel, I. T. van den and J. F. Kiviet, 1993, Neglected dynamics in panel data models: Consequences and detection in finite samples, Working Paper, Tinbergen Institute, University of Amsterdam.

Energy Information Administration, 1983, *The State Energy Price and Expenditure System*.

Hu, W. and G. S. Maddala, 1994, A Monte Carlo study of alternative methods for estimation and prediction from dynamic heterogeneous panel data, manuscript, Ohio State University.

Maddala, G. S. 1991, To pool or not to pool: That is the question. *Journal of Quantitative Economics* **7**, 255-264.

Maddala, G. S. and W. Hu, 1996, The pooling problem. In L. Matyas and P. Sevestre (eds.), *Econometrics of Panel Data*, 307-322, (2nd ed. Kluwer, 1996).

Maddala, G. S., R. P. Trost, H. Li and F. Joutz, 1997, Estimation of short run and long run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics* **15**, 90-100.

Magnus, J. and H. Neudecker, 1988, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Pesaran, M. H. and R. Smith, 1995, Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* **68**, 79-113.

Ridder, G. and T. J. Wansbeek, 1990, Dynamic models for panel data. In: van der Ploeg (ed.), *Advanced lectures in Quantitative Economics*, Academic Press, London.

Robertson, D. and J. Symons, 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.

Smith, A. F. M., 1973, A general Bayesian linear model. *Journal of the Royal Statistical Society* **B 35**, 67-75.