Division of Research
Graduate School of Business Administration
The University of Michigan

January, 1977

POOLING CROSS-SECTIONAL AND TIME SERIES DATA - A REVIEW OF STATISTICAL ESTIMATION TECHNIQUES

Working Paper No. 141

by

Terry Dielman
and
Roger L. Wright

The University of Michigan

Introduction

Several recent articles have been published examining the methods of pooling cross-sectional and time series data. The occurrence of observations of a large number of individuals over certain periods of time is becoming more common and, thus, these pooling methods are attracting more attention. When examining the current literature we find that there are three cases involving slight alterations in the models used that must be considered: (1) a model with exogenous variables but no lagged values of the dependent variable, (2) a model with both exogenous and lagged dependent variables, and (3) a model with lagged values of the dependent variable but no exogenous variables.

Simulation studies have been performed to determine small sample properties of estimation procedures for these three cases.

In this paper we present a review of these studies as well as a summary of findings from other available literature. Within our review we will note when and how the estimation procedures may best be utilized in pooling cross-section and time series data.

Linear Model And Generalized Least Squares

In general, the model we are concerned with can be written as

$$Y = X\beta + u$$

where $y_{it} = \sum_{j=1}^{K} x_{itj}\beta_j + u_{it}$ with

Y an NT X 1 vector of the dependent variable y_{it};

X an NT X K matrix of K variables which may be exogenous or lagged dependent;

β a K X 1 vector of unknown parameters;

u an NT X 1 vector of the unknown stochastic components u_{it}.

The generalized least squares estimate of the parameter β is:

$$\hat{\beta}GLS = (X'\Omega^{-1}X)^{-1}\ (X'\Omega^{-1}Y)$$

$$\text{where } \Omega = var(u) = E(uu')$$

Special Case: Error Component Model

Suppose now that $u_{it} = \mu_i + V_{it}$

where $E(\mu_i V_{i't}) = 0$ for all i, i', and t

$E(\mu_i) = 0$ and $E(V_{it}) = 0$ for all i and t

$$E(\mu_i \mu_{i'}) = \begin{cases} \sigma_\mu^2 & \text{if } i = i' \\ 0 & \text{otherwise} \end{cases}$$

and

$$E(V_{it} V_{i't'}) = \begin{cases} \sigma_\nu^2 & \text{if } i = i' \text{ and } t = t' \\ 0 & \text{otherwise} \end{cases}$$

The μ_i are time invariant individual effects and the V_{it} represent remaining effects which are assumed to vary over both individuals and time.

(Note that in our variance or error component representation of u_{it} as $\mu_i + V_{it}$, we have simplified from the more general case of

$u_{it}=\mu_i + \lambda_t + v_{it}$ where the λ_t represent the period specific and individual invariant effects. Further comment on this simplification will be made later.)

We now have

$$E(uu')=\Omega = \sigma^2 \begin{bmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A \end{bmatrix}$$

where A is the T X T matrix

$$\begin{bmatrix} 1 & p & \cdots & p \\ p & 1 & \cdots & p \\ \vdots & \vdots & \ddots & \vdots \\ p & p & \cdots & 1 \end{bmatrix}$$

with $\sigma^2 = \sigma\mu^2 + \sigma v^2$ and $\rho = \sigma\mu^2/\sigma^2$.

We can then write the GLS estimator as follows (see Appendix A for derivation):

$$\hat{\beta}GLS = [W_{xx} + \Theta B_{xx}]^{-1}[W_{xy} + \Theta B_{xy}]$$

where $T_{xx}=\Sigma X_i'X_i$ $\qquad$ $T_{xy}= \Sigma X_i'Y_i$

$B_{xx}= \frac{1}{T}\Sigma(X_i'ee'X_i)$ $\qquad$ $B_{xy}=\frac{1}{T}\Sigma(x_i'ee'Y_i)$

$W_{xx}=T_{xx} - B_{xx}$ $\qquad$ $W_{xy}=T_{xy} - B_{xy}$

T_{xx}, W_{xx} and B_{xx} are K X K; T_{xy}, B_{xy}, and W_{xy} are K X 1, e is a T X 1 vector with all elements unity and

$$\theta = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2} \quad .[3,4]$$

The problem in using the GLS estimator is that the value of θ is unknown. Therefore we must produce an estimate of θ by estimating σ_v^2 and $\sigma\mu^2$. (Note that we could also proceed by estimating $\rho = \frac{\sigma\mu^2}{\sigma\mu^2 + \sigma v^2}$

and using this in an equivalent GLS estimator expressed in terms of ρ. The Nerlove articles speak in terms of finding ρ rather than θ; see Appendix C for details.)

We now present several methods used to produce estimators of β in the articles examined (see Appendix B for a further listing):

1. The true GLS estimator with θ assumed known.

2. The ordinary least squares estimator (OLS):

$$\hat{\beta}_{OLS} = (T_{xx})^{-1} T_{xy}$$

(This is $\hat{\beta}_{GLS}$ with θ = 1.)

3. The least squares with dummy variables estimator (LSDV):

$$\hat{\beta}_{LSDV} = (W_{xx})^{-1} W_{xy}$$

(This is $\hat{\beta}_{GLS}$ with θ = 0.)

The model for the LSDV method can be written as

$$Y = X\beta + Z\mu + v$$

where

X is the matrix of exogenous and/or lagged dependent variables;

Z is an NT X N design matrix introducing the "individual dummy variables;"

μ is an N X 1 vector of coefficients for the dummy variables;

v is an NT X 1 vector of disturbance terms.

4. Maximum likelihood estimator (ML):

The value of $\hat{\theta}_{ML}$ is determined by searching for the value between zero and one that maximizes the likelihood function.

We then determine

$$\hat{\sigma}^2_{v,ML} = \frac{1}{NT}\,[W_{yy} + \hat{\Theta}_{ML}B_{yy} - (W_{xy}+\hat{\Theta}_{ML}B_{XY})'(W_{xx}+\hat{\Theta}_{ML}B_{xx})^{-1}(W_{xy}+\hat{\Theta}_{ML}B_{xy})]$$

$\hat{\sigma}\mu^2_{ML}$ can then be derived from $\hat{\theta}_{ML}$ and $\hat{\sigma}^2_{v,ML}$ since

$$\hat{\Theta}_{ML} = \frac{\hat{\sigma}^2_{v,ML}}{\hat{\sigma}^2_{v,ML}+T\hat{\sigma}^2_{\mu,ML}}$$

Therefore we have

$$\hat{\sigma}^2_{\mu,ML} = \frac{1}{T}\left(\frac{\hat{\sigma}^2_{v,ML}}{\hat{\Theta}_{ML}} - \hat{\sigma}^2_{v,ML}\right).$$

5. Nerlove's estimator (2RC):

Compute the estimate of σ_v^2 from the estimated residuals of LSDV and estimate σ_μ^2 from the estimates of coefficients of dummy variables [5].

$$\hat{\sigma}^2_{v,2RC} = \frac{1}{NT}\left[W_{YY}-W'_{XY}(W_{XX})^{-1}W_{XY}\right]$$

$$\hat{\sigma}^2_{\mu,2RC} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}_{i,LSDV} - \bar{\hat{\mu}}_{i,LSDV})^2 \text{ where } \bar{\hat{\mu}}_{i,LSDV} = \frac{\sum_{i=1}^{N}\hat{\mu}_i}{N}.$$

We now present a summary of three simulation studies examining the results of estimation procedures for the cases mentioned in the introduction.

Results of Comparison

A study by Maddala and Mount involving Monte Carlo simulation examines the first case [4].

1. Presence of exogenous variables but no lagged values of the dependent variable

The Model: $Y = X\beta + u$ where

$$y_{it} = \beta x_{it} + u_{it} \quad \text{with} \quad u_{it} = \mu_i + v_{it}$$

Comparison of Methods:

* Bias: The bias in all cases was reported to be negligible. The maximum bias was .007, less than one percent, with no systematic variation of the relative bias.

* Mean square error: No appreciable differences were found to exist between any of the two step estimators examined (ML, and 2RC). All performed equally well. Also, for overall performance, the two-step estimators outperformed OLS and LSDV.

  Examining the MSE's for the estimates, the MSE is seen to increase as θ decreases (ρ increases) for OLS and decreases as θ increases for LSDV.

* Notes: Possibilities for negative variances exist with the ML method. In this case $\hat{\sigma}_\mu^2$ was put equal to zero.

The above findings also hold for the additional methods mentioned in Appendix B.

Monte Carlo studies by Nerlove have examined properties of estimators in the two cases involving lagged dependent variables:

2. Presence of both exogenous and lagged dependent variables

The Model: $Y=X\beta+u$ where

$$Y_{it}=\beta_o+\beta_i y_{i,t-1}+\beta_2 x_{it}+u_{it} \quad \text{with}$$

$$u_{it}=\mu_i+v_{it}$$

Comparison of Methods:

* Bias: Small sample bias was known to exist in all cases in which a lagged value of the dependent variable is one of the explanatory variables. Using GLS with the true value of ρ resulted in only a slight bias in all cases.

OLS: (1) $\rho=0$ The OLS and GLS estimates coincide.

(2) $\rho\neq 0$ Severe bias encountered.

β_1 severely biased upwards;

β_2 severely biased downwards for large ρ values except when $\beta_2=0$;

σ^2 severely biased downward.

LSDV: β_1 biased downwards (less for higher values of ρ and β_2);

$\beta_o, \beta_2, \sigma^2$ and ρ are biased upwards.

2RC: Despite the upward bias in ρ using LSDV, 2RC was found to be superior over a wide range of true parameter values to all other estimates considered. Estimates of β_o, β_1, β_2 and σ^2 are not seriously biased in comparison with either OLS or LSDV. The principal source of difficulty noted is the upward bias in the estimate of σ^2.

ML: Biases in cases where a large number of boundary solutions ($\rho=0$) did not occur were slight. Where boundary solutions occurred the ML estimates were greatly affected.

Relative mean square error: Nerlove considered the mean square errors (MSEs) for the GLS method as an ideal of comparison and formed ratios of MSEs of β_o, β_1 and β_2 for each method to the corresponding MSE of GLS.

OLS: $\rho=0$: MSEs coincide with GLS or are somewhat smaller.

$\rho\neq0$: The above result no longer applies and estimates deteriorate. Note that even the LSDV estimates are superior for β_1 and β_2.

LSDV: Superior to OLS for β_1 and β_2; worse than OLS for β_o.

2RC: Lower MSEs than LSDV except when ρ is very small. Lower than ML even when this method might be expected to perform well.

2RC compares favorably with all other estimates over a wide range of parameter values.

ML: Large numbers of boundary solutions were encountered for the MLE estimates even when they were not expected to occur.

The MLE method does not compare favorably to the 2RC.

3. <u>Presence of lagged values of the dependent variable but no exogeneous variable</u>

The Model: $Y=X\beta + u$ where

$$y_{it} = \beta y_{i,t-1} + u_{it} \quad \text{with}$$

$$u_{it} = \mu_i + v_{it}$$

Comparison of Methods:

* Bias:

GLS: Small amount of bias.

OLS: Estimates of β are strongly biased upwards when $\rho \neq 0$; estimates of σ^2 are strongly biased towards zero.

LSDV: Estimates of β are biased towards zero; estimates of ρ and σ^2 are biased upwards.

ML: Estimates of σ^2 are highly erratic.

2RC: Estimates of β show some bias downwards for large true β and some bias upwards for small true β; estimates of σ^2 are less erratic for this method than for ML.

Conclusions

These simulation studies yield the following conclusions:

1. With a lagged value of the dependent variable present, Nerlove's estimator, 2RC, appears to be superior to other suggested methods of estimation.

2. With no lagged value present there is no clearly superior method among the two-step procedures in terms of the properties examined, although these are superior to methods such as OLS and

LSDV. Maddala and Mount suggest, in this case, that two of the methods be applied to the data at hand and the results compared. Widely differing estimates should be taken as an indication of a possible misspecification of the model. [1]

When looking for a single method of estimation to be used with all data sets, however, the choice would go to Nerlove's method.

Prospects for Futher Study

The literature thus far has examined cases where

$$u_{it} = \mu_i + v_{it}, \text{ i.e.,}$$

where the random time effect has been omitted from the disturbance term. This is simply a matter of choice in specifying the model.

With the time effect included, we would have the following specification:

$$Y = X\beta + u, \; u_{it} = \mu_i + \lambda_t + v_{it}.$$

The the GLS estimate of β will be

$$\hat{\beta}_{GLS} = \left[W_{xx} + \theta_1 B_{xx} + \theta_2 C_{xx}\right]^{-1} \left[W_{xy} + \theta_1 B_{xy} + \theta_2 C_{xy}\right]$$

where

$$\theta_1 = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2}$$

$$\theta_2 = \frac{\sigma_v^2}{\sigma_v^2 + N\sigma_\lambda^2}$$

and C_{xx}, C_{xy} and C_{yy} represent the between time period decomposition of the variances [3].

---

1 See Arora [2] for another study supporting use of a two-step method.

Nerlove suggests that further study in terms of small sample properties of estimators would be desirable in this case, especially in terms of the results when the λ_t are erroneously assumed absent [5].

The assumptions made concerning the u vector, i.e., that

$$E(uu') = \sigma^2 \begin{bmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A \end{bmatrix}$$

where $\sigma^2 = \sigma_\mu^{\ 2} + \sigma_\nu^{\ 2}$, $\rho = \dfrac{\sigma_\mu^{\ 2}}{\sigma^2}$ and

$$A = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

imply a specific form of serial correlation. The effect of a misspecification here needs to be investigated [5].

Another area of study concerning the pooling of cross-section and time series data is the use of the random coefficient regression (RCR) model. Study in this area has begun. For example, Hsiao has proposed the following model [2]:

$$Y_{it} = (\beta + \delta_i + \gamma_t)x_{it} + \varepsilon_{it}$$

with δ_i, γ_t and ε_{it} random disturbances,

$$E(\varepsilon_{it}\varepsilon_{jt}') = \begin{cases} \sigma_\varepsilon^2, & i = j,\ t=t' \\ 0, & \text{otherwise} \end{cases}$$

$$E(\delta_i\delta_j) = \begin{cases} \sigma_\delta^2, & i = j \\ 0, & \text{otherwise} \end{cases}$$

$$E(\gamma_t\gamma_t') = \begin{cases} \sigma_\gamma^2, & t = t' \\ 0, & \text{otherwise} \end{cases}$$

$E(\delta_i\gamma_t) = E(\delta_i\varepsilon_{jt}) = E(\gamma_t\varepsilon_{it}) = 0$; $E(\varepsilon_{it}) = E(\delta_i) = E(\gamma_t) = 0$ for all i, j and t.

Note that this case can include the error components model. Hsiao writes the model as

$$Y_{it} = \sum_{k=1}^{K} (\beta_k+\delta_{ik} + \gamma_{tk})X_{itk} + \varepsilon_{it}.$$

Now, if each element of the first column of x is equal to one, we have

$$Y_{it} = \beta_1 + \sum_{k=2}^{K} (\beta_k + \delta_{ik} + \gamma_{tk})X_{itk} + \delta_{i1} + \gamma_{1t} + \varepsilon_{it}$$

where δ_{i1}, γ_{1t}, and ε_{it} correspond to μ_i, λ_t, and V_{it} in our previous error components model. In this case, however, we also allow for the possibility of random slope coefficients.

RCR models are justifiable in economic use. The coefficient of a variable may not remain constant because of unobservable influences of the variable. Thus, we may be better off predicting the mean of some process that determines the coefficient rather than assuming the coefficient to be constant.

Since these models represent additional general instances of the error components model used in pooling cross-section and time series data, they deserve the attention of further study.

Appendix A

Derivation of GLS Estimator

We have written $\hat{\beta}_{GLS} = [W_{xx} + \theta B_{xx}]^{-1} [W_{xy} + \theta B_{xy}]$ following Maddala and Mount. The following is a verification that our form is equivalent to the GLS estimator.

$Y = X\beta + u$ where $u_{it} = \mu_i + v_{it}$

The GLS estimator for the above model is

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$$

where $\Omega = E(uu') =$

$$\sigma^2 \begin{bmatrix} A & 0 & \cdots & 0 \\ & & & \vdots \\ 0 & A & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A \end{bmatrix} \text{(NT X NT)}$$

where $A = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$ (T X T)

$\sigma^2 = \sigma^2_\mu + \sigma^2_\nu$ and $\rho = \dfrac{\sigma^2_\mu}{\sigma^2}$

To obtain Ω^{-1} we can make use of the orthogonal transformation

$C = \begin{bmatrix} e'/\sqrt{T} \\ C_1 \end{bmatrix}$ (T X T) where $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ (T X 1) and C_1 is defined so that $C_1 \cdot e=0$, $C_1 \cdot C_1' = I_{T-1}$, and $C_1' \cdot C_1 = I_T - \frac{e'e}{T}$.

$$\text{Then } CAC' = \begin{bmatrix} (1-\rho)+T\rho & 0 & \cdots & 0 \\ 0 & 1-\rho & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\rho \end{bmatrix}$$

$$\text{Therefore } \Omega^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} A^{-1} & 0 & \cdots & 0 \\ 0 & A^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A^{-1} \end{bmatrix}$$

$$\text{where } A^{-1} = C' \begin{bmatrix} \frac{1}{(1-\rho)+T\rho} & 0 & \cdots & 0 \\ 0 & \frac{1}{1-\rho} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \frac{1}{1-\rho} \end{bmatrix} C$$

Note: This formulation follows from a result in Theil [8] (pp. 27-28) which may be stated as follows: For any symmetric matrix M there exists an orthogonal matrix B such that BM = QB, BMB' = Q and M=B'QB where Q is diagonal and contains the latent roots of M along the diagonal.

$$(**) \text{ Then } A^{-1} = \frac{1}{1-\rho} I_T + \frac{1}{T}\left(\frac{1}{(1-\rho)+T\rho} - \frac{1}{1-\rho}\right) ee'$$

$$= \frac{1}{1-\rho} I_T + \frac{1}{T}\left(\frac{(1-\rho)-(1-\rho+T\rho)}{(1-\rho+T\rho)(1-\rho)}\right) ee'$$

$$= \frac{1}{1-\rho} I_T + \frac{-\rho}{(1-\rho)(1-\rho+T\rho)} ee'$$

$$= \lambda_2 I_T + \lambda_1 ee'$$

where $\lambda_1 = \dfrac{-\rho}{(1-\rho)(1-\rho+T\rho)}$ and $\lambda_2 = \dfrac{1}{1-\rho}$

We can now write the GLS estimator as follows:

$$\hat{\beta}_{GLS} = \left[X' \begin{bmatrix} A^{-1} & 0 & \cdots & 0 \\ 0 & A^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A^{-1} \end{bmatrix} X \right]^{-1} \left[X' \begin{bmatrix} A^{-1} & 0 & \cdots & 0 \\ 0 & A^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A^{-1} \end{bmatrix} Y \right]$$

$$= \left[\sum_{i=1}^{N} X_i' A^{-1} X_i \right]^{-1} \left[\sum_{i=1}^{N} X_i' A^{-1} Y_i \right]$$

where $X' = [X_1' \ X_2' \ \ldots \ X_N']$ and X_i' is K x T,

$Y' = [Y_1' \ Y_2' \ \ldots \ Y_N']$ and Y_i' is 1 x T.

$$= \left[\sum_{i=1}^{N} X_i' [\lambda_1 ee' + \lambda_2 I] X_i \right]^{-1} \left[\sum_{i=1}^{N} X_i' [\lambda_1 ee' + \lambda_2 I] Y_i \right]$$

$$= \left[\lambda_1 \sum_{i=1}^{N} X_i'ee'X_i + \lambda_2 \sum_{i=1}^{N} X_i'X_i\right]^{-1} \left[\lambda_1 \sum_{i=1}^{N} X_i'ee'Y_i + \lambda_2 \sum_{i=1}^{N} X_i'Y_i\right]$$

$$= \left[\lambda_1 T B_{xx} + \lambda_2 T_{xx}\right]^{-1} \left[\lambda_1 T B_{xy} + \lambda_2 T_{xy}\right]$$

$$= \left[\frac{\lambda_1 T}{\lambda_2} B_{xx} + T_{xx}\right]^{-1} \left[\frac{\lambda_1 T}{\lambda_2} B_{xy} + T_{xy}\right]$$

and, on setting $\theta = 1 - \frac{\lambda_1 T}{\lambda_2}$, we have

$$= \left[(\theta - 1)B_{xx} + T_{xx}\right]^{-1} \left[(\theta - 1)B_{xy} + T_{xy}\right]$$

$$= \left[\theta B_{xx} + (T_{xx} - B_{xx})\right]^{-1} \left[\theta B_{xy} + (T_{xy} - B_{xy})\right]$$

$$= \left[\theta B_{xx} + W_{xx}\right]^{-1} \left[\theta B_{xy} + W_{xy}\right]$$

where W_{xx}, B_{xx}, T_{xx}, W_{xy}, B_{xy} and T_{xy} are defined on page 4.

Note:

$$\theta = 1 + \frac{\lambda_1 T}{\lambda_2} =$$

Appendix A (Continued)

$$= 1 + \frac{\left(\frac{-\rho T}{(1-\rho)(1-\rho+T)}\right)}{\left(\frac{1}{1-\rho}\right)}$$

$$= 1 - \frac{\rho T}{1-\rho+\rho T}$$

$$= \frac{1-\rho}{1-\rho+\rho T}$$

$$= \frac{1 - \frac{\sigma_\mu^2}{\sigma^2}}{1 - \frac{\sigma_\mu^2}{\sigma^2} + \frac{T\sigma_\mu^2}{\sigma^2}}$$

$$= \frac{\sigma^2 - \sigma_\mu^2}{\sigma^2 - \sigma_\mu^2 + T\sigma_\mu^2}$$

$$= \frac{\sigma_\nu^2}{\sigma_\nu^2 + T\sigma_\mu^2}$$

Appendix B

Alternative Estimation Methods

Description of other estimators examined in certain studies are as follows:

6. The least squares between groups estimator (LSBG):

$$\hat{\beta}_{LSBG} = (B_{xx})^{-1} (B_{xy})$$

7. Wallace and Hussain's estimator (WH) [9]:

Wallace and Hussain propose using estimated OLS residuals as the true residuals to estimate σ^2_μ and σ^2_ν :

$$\hat{\sigma}^2_{\nu,WH} = \frac{1}{N(T-1)} \left[W_{xx} - 2T_{xy}'(T_{xx})^{-1} W_{xy} + T_{xy}'(T_{xx})^{-1} W_{xx}(T_{xx})^{-1} T_{xy} \right]$$

$$\hat{\sigma}^2_{\mu,WH} = \frac{1}{NT} \left[B_{yy} - 2 T_{xy}(T_{xx})^{-1} B_{xy} + T_{xy}(T_{xx})^{-1} B_{xx}(T_{xx})^{-1} T_{xy} \right] - \frac{1}{T} \hat{\sigma}^2_{\nu,WH}$$

8. Amemiya's estimator (AM) [1]:

Identical to WH except using LSDV rather than OLS in the first step:

$$\hat{\sigma}^2_{\nu,AM} = \frac{1}{N(T-1)} [W_{yy} - W_{xy}'(W_{xx})^{-1} W_{xy}]$$

$$\hat{\sigma}^2_{\mu,AM} = \frac{1}{NT} [B_{yy} - 2 W_{xy}'(W_{xx})^{-1} B_{xy} + W_{xy}(W_{xx})^{-1} B_{xx}(W_{xx})^{-1} W_{xy}] - \frac{1}{T} \hat{\sigma}^2_{\nu,AM}$$

9. Instrumental-variable estimates using $x_{i,\ t-1}$ as an instrument for $y_{i,t-1}$ (IV)[5]:

With this method ρ is estimated from the calculated residuals, $\hat{u}_{it}$, by

$$\hat{\rho} = \frac{\frac{1}{T}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{U}_{it}\right)^2 - \frac{1}{NT}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{U}_{it}^2\right)}{NT\hat{\sigma}^2}$$

where $$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{U}_{it}^2}{NT}$$

10. Two round instrumental (2RI) estimates similar to those in 2RC but using the calculated residuals from the IV method rather than LSDV[5].

11. Analysis of covariance method (ANCOVA):

$$\hat{\sigma}^2_{\nu,ANCOVA} = \frac{1}{N(T-1)-K}\left[W_{xx} - W_{xy}'(W_{xx})^{-1}W_{xy}\right]$$

$$\hat{\sigma}^2_{\mu,ANCOVA} = \frac{1}{T(N-K)}\left[B_{yy} - B_{xy}'(B_{xx})^{-1}B_{xy}\right] - \frac{1}{T}\hat{\sigma}^2_{\nu,ANCOVA}$$

where k = number of slope parameters.

12. Henderson's Method III(H3):

$$\hat{\sigma}^2_{\nu,H3} = \hat{\sigma}^2_{\nu,ANCOVA}$$

$$\hat{\sigma}^2_{\mu,H3} = a^{-1}[W_{xy}'(W_{xx})^{-1}W_{xy} - T_{xy}(T_{xx})^{-1}T_{xy} - N\hat{\sigma}^2_{\nu,H3}]$$

where $a = T[N - \text{trace}[(T_{xx})^{-1}B_{xy}]]$

13. Minimum norm quadratic unbiased estimator (MINQUE) [7]: This method (due to Rao) proposed that we minimize the Euclidean norm of the difference between the actual estimator and the "natural" estimator given μ and v in

$$Y = X\beta + Z\mu + v$$

Estimators of σ^2_ν and σ^2_μ are computed by solving the following simultaneous equation system:

$$\begin{bmatrix} \text{trace}\{[RZZ']^2\} & \text{trace}\{[RZZ'R]\} \\ \text{trace}\{[RZZ'R]^2\} & \text{trace}\{R^2\} \end{bmatrix} \begin{bmatrix} \hat{\sigma}^2_{\mu,MINQUE} \\ \hat{\sigma}^2_{\nu,MINQUE} \end{bmatrix} = \begin{bmatrix} Y'RZZ'RY \\ Y'RRY \end{bmatrix}$$

where $R = (H^{-1} - H^{-1}X(X'H^{-1}X)^{-1}X'H^{-1})$

$H^{-1} = (ZZ' + I_{NT})^{-1}$.

Studies involving the above methods and results

1. Presence of exogeneous variables but no lagged dependent variables.

   Also examined methods
   6. LSBG
   7. WH
   8. AM
   11. ANCOVA
   12. H3
   13. MINQUE

   Notes: Possibilities for negative variances exist with any of the following: MINQUE, ANCOVA, H3, WH, and AM. In such cases $\hat{\sigma}_\mu^2$ was put equal to zero.

The ANCOVA and H3 methods both lack the property of uniqueness in the estimators produced.

Results stated in main body of paper hold for these six methods also. The two step methods listed above are 7, 8, 11, 12, and 13.

2. Presence of exogenous and lagged dependent variables.

Also examined methods 9. IV
10. 2RI

Results: Bias

IV: When $\beta = 0$, these estimates are inconsistent and highly erratic behavior occurs. When $\beta = 1$, bias in all parameters is slight.

2RI: The 2RI estimates are greatly affected by the extreme variability of the estimates of ρ obtained from the IV method. The estimates of σ^2 show the greatest bias and erratic behavior.

Mean square error

IV: Ratio is high for $\beta = 0$ but falls markedly as β increases.

2RI: Erratic behavior when $\beta = 0$ because underlying estimates of ρ are so poor. Higher MSE's than 2RC even for large values of β.

3. Lagged dependent variables only.

No further methods were examined.

Appendix C

Derivation of Nerlove's GLS Estimator

To derive Nerlove's form of the GLS estimator (which involves ρ rather than θ), we note that in line (**) of Appendix A we have:

$$A^{-1} = \frac{1}{1-\rho} I_T + \frac{1}{T}\left(\frac{1}{1-\rho+T\rho} - \frac{1}{1-\rho}\right) ee'$$

$$= \frac{1}{1-\rho}\left(I_T - \frac{1}{T} ee' \right) + \frac{1}{T(1-\rho+\rho T)} ee'$$

$$\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} Y)$$

$$= \left(\sum_{i=1}^{N} X_i' A^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' A^{-1} Y_i\right)$$

$$= \left[\sum_{i=1}^{N} \frac{X_i' \left(I - \frac{ee'}{T}\right) X_i}{1-\rho} + \frac{1}{T}\sum_{i=1}^{N} \frac{X_i' ee' X_i}{1-\rho+\rho T}\right]^{-1}$$

$$\times \left[\sum_{i=1}^{N} \frac{X_i \left(I_T - \frac{ee'}{T}\right) Y_i}{1-\rho} + \frac{1}{T}\sum_{i=1}^{N} \frac{X_i' ee' Y_i}{1-\rho+\rho T}\right]$$

$$= \left[\frac{1}{1-\rho} W_{xx} + \frac{1}{1-\rho+\rho T} B_{xx}\right]^{-1} \left[\frac{1}{1-\rho} W_{xy} + \frac{1}{1-\rho+\rho T} B_{xy}\right]$$

This is the form used by Nerlove [5].

References

1. Amemiya, T. "The Estimation of the Variances in a Variance - Components Model." International Economics Review, February 1971, pp. 1-13.

2. Arora, Swarnjit S. "Error Components Models and Their Applications." Annals of Economic and Social Measurement 2/4 (1973): 451-67.

3. Hsiao, Cheng. "Some Estimation Methods for a Random Coefficient Model." Econometrica 43, 2 (March 1975): 305-25.

4. Maddala, G. S. "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data." Econometrica 39, 2 (March 1971): 341-58.

5. Maddala, G. S. and Mount, T.D. "A Comparative Study of Alternative Estimators for Variance Components Models Used in Econometric Applications." JASA 68, 342 (June 1973): 324-28.

6. Nerlove, Marc. "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections." Econometrica 39, 2 (March 1971): 359-82

7. Nerlove, Marc. "A Note on Error Components Models." Econometrica 39, 2 (March 1971): 383-96.

8. Rao, C. R. "Estimation of Variance and Covariance Components in Linear Models." JASA 67 (March 1972): 112-15.

9. Theil, Henry. Principles of Econometrics (New York, NY: John Wiley & Sons, Inc., (1971), pp. 27-28.

10. Wallace, T. D. and Hussain, A. "The Use of Error Components Models in Combining Cross-Section with Time Series Data." Econometrica 37, 1 (January 1969): 55-72.

The above findings also hold for the additional methods mentioned in Appendix B.

Monte Carlo studies by Nerlove have examined properties of estimators in the two cases involving lagged dependent variables:

2. Presence of both exogenous and lagged dependent variables

The Model: $Y = X\beta + u$ where

$$Y_{it} = \beta_o + \beta_i y_{i,t-1} + \beta_2 x_{it} + u_{it} \quad \text{with}$$

$$u_{it} = \mu_i + v_{it}$$

Comparison of Methods:

* Bias: Small sample bias was known to exist in all cases in which a lagged value of the dependent variable is one of the explanatory variables. Using GLS with the true value of ρ resulted in only a slight bias in all cases.

OLS: (1) $\rho = 0$ The OLS and GLS estimates coincide.

(2) $\rho \neq 0$ Severe bias encountered.

β_1 severely biased upwards;

β_2 severely biased downwards for large ρ values except when $\beta_2=0$;

σ^2 severely biased downward.

LSDV: β_1 biased downwards (less for higher values of ρ and β_2);

$\beta_o, \beta_2, \sigma^2$ and ρ are biased upwards.

2RC: Despite the upward bias in ρ using LSDV, 2RC was found to be superior over a wide range of true