No. 8703

INCREASING THE EFFICIENCY OF POOLED ESTIMATION WITH A BLOCK COVARIANCE STRUCTURE

by

Jeffery W. Gunther* and Ronald H. Schmidt*
Research Department
Federal Reserve Bank of Dallas

June 1987

Research Paper

Federal Reserve Bank of Dallas

No. 8703

INCREASING THE EFFICIENCY OF POOLED ESTIMATION WITH A BLOCK COVARIANCE STRUCTURE

by

Jeffery W. Gunther* and Ronald H. Schmidt*
Research Department
Federal Reserve Bank of Dallas

June 1987

* The authors would like to thank Tom Fomby for his helpful comments and suggestions. Views expressed are those of the authors and do not necessarily reflect views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

I. Introduction

Use of econometric models which pool microsamples has become more frequent with the increasing availability of economic data in both times series and cross-sectional form. One common approach to pooling is to assume that all coefficients are constant, with variations in individual and time series observations modeled in the specification of the disturbance term. The variance-covariance structure of the disturbance in this model typically includes different variances for each cross-sectional unit and first-order autoregressive disturbances within each cross-sectional unit. The contemporaneous covariances of the disturbance between pairs of cross-sectional units are typically assumed either to be nonzero and different for each pair or uniformly equal to zero. The model including all the covariance terms will be conveniently referred to as the "full" model and the model excluding them as the "diagonal" model.

In this paper, we compare the efficiency of pooled estimation using alternative restrictions on the disturbance variance-covariance matrix. In addition to the full and diagonal models, we examine the properties of estimators employing block covariance structures, in which only some contemporaneous disturbance correlations are assumed to equal zero.

Block covariance structures offer several potential advantages over the other models. First, in data sets where the number of cross-sectional units exceeds the number of time series observations, estimation cannot proceed using the full model because the disturbance variance-covariance matrix is singular. In such cases, the diagonal model would normally be applied. However, block covariance structures can be devised which preserve the nonsingularity of the disturbance variance-covariance matrix while incorporating more information into the estimation than the diagonal model, thereby potentially increasing efficiency.

Second, Monte Carlo results presented here indicate that in cases where the disturbances are highly correlated for only some of the cross-sectional units and the number of time series observations is relatively small, estimators employing block covariance structures are substantially more efficient than those using either the full or diagonal model.

These results, while generally applicable, may have special significance in analyzing regional data. Because cross-sectional units, particularly state-level observations, may have geographical similarities not easily captured in right hand side variables, the disturbance is likely to exhibit strong correlations between certain states and relatively minor correlations between others. Furthermore, where observations are recorded annually, it is unlikely that the number of time series observations would equal or exceed the number of available cross-sectional observations. Consequently, development of a block covariance structure may offer more efficient estimation in many cases.

This paper is organized as follows. In Section II, we describe the general model and present the computational difficulty which arises when the number of cross-sectional units is greater than the number of time series observations. Although this result has been asserted in passing elsewhere, a brief derivation is instructive.[1/] In Section III, we present a feasible GLS estimator using a block covariance structure and describe its properties. Monte Carlo results are then described in Section IV, with conclusions presented in Section V.

II. Computational Limitations of the Full Model

Consider the model

$$1)\quad Y_{m,t} = X_{m,t}\beta + \varepsilon_{m,t};\quad m = 1,\ldots M,\ t = 1,\ldots, T,$$

where Y is the dependent variable, X is a 1xK vector of independent variables, β is a Kx1 vector of coefficients, ε is the disturbance term, m refers to the cross-sectional unit, and t denotes the time period.

The full disturbance variance-covariance structure for equation 1 often incorporates cross-sectional heteroskedasticity, contemporaneous correlation of disturbances over cross-sectional units, and first-order autoregressive disturbances within cross-sectional units. For our purposes, it is possible to simplify the discussion without altering the results by dropping the consideration of autocorrelation. The full disturbance variance-covariance structure can then be written as

$$\begin{aligned} 2)\quad E(\varepsilon_{i,s}\varepsilon_{j,t}) &= \sigma_i^2 \text{ for } i=j \text{ and } s=t; \\ &= \sigma_{i,j} \text{ for } i \neq j,\ s = t; \\ &= 0 \text{ for } s \neq t, \end{aligned}$$

where i and j refer to individuals, and s and t denote time periods.

With known variance matrix, the GLS estimator of β can be written

$$3)\quad \tilde{\beta} = (X'(\Omega \otimes I_T)^{-1}X)^{-1}X'(\Omega \otimes I_T)^{-1}\, Y,$$

where Y is the time series for the dependent variable for all M individuals stacked into a MTx1 vector, X is of dimensions MTxK and contains the corresponding stacked time series observations for the K independent variables, Ω is the covariance matrix of size M, and I_T is the TxT identity matrix.

When Ω is unknown, the estimator in equation 3 must be replaced by the feasible GLS estimator of β,

$$4)\quad \hat{\beta} = (X'(\hat{\Omega} \otimes I_T)^{-1}X)^{-1}X'(\hat{\Omega} \otimes I_T)^{-1}\, Y.$$

Following Zellner (1962), elements of the feasible GLS variance-covariance matrix are estimated using the residual vector from OLS estimation of equation 1. Asymptotically efficient estimates of each element of Ω are obtained by taking the cross product of the disturbance subvectors from cross-sectional units i and j and dividing by the number of time series observations. In matrix notation, $\hat{\Omega}$ can be written as

$$5)\quad \hat{\Omega} = E'E/T, \text{ where } E = [\underline{e}_1\ \underline{e}_2\ \cdots\ \underline{e}_M]$$

and $\underline{e}_i$ is the Tx1 subvector of the OLS residual vector corresponding to cross section i.

A necessary condition for the feasibility of computing equation 4 is apparent in equation 5. The matrix E is of dimension TxM. Consequently, the maximum rank of E -- and therefore $\hat{\Omega}$ -- is the minimum of T or M. Clearly, the number of time series observations must be greater than or equal to the number of cross-sectional units for $\hat{\Omega}$ to have full rank. If this condition is violated, $\hat{\Omega}$ is singular and equation 4 cannot be calculated.

Although this problem would not typically arise in the case of seemingly unrelated regression models, around which much of the theoretical and empirical work on the full disturbance variance-covariance structure has been centered previously, data sets used in pooled estimation may often violate this condition. Panel data sets often have fewer time series observations than cross-sectional units, making the full model inappropriate.[2/]

III. Block Covariance Structures

In cases where estimation using the full model is infeasible, the usual practice has been to use the diagonal disturbance variance-covariance structure, thereby insuring nonsingularity of $\hat{\Omega}$. However, the diagonal model sacrifices potential efficiency gains in cases where contemporaneous disturbance covariances are nonzero.

A less restrictive approach is to include nonzero off diagonal elements in $\hat{\Omega}$ insofar as the singularity condition is not violated. In some cases it might be difficult to determine which disturbance covariance terms should be estimated and which should be assumed to equal zero. In most applications, however, priors are likely to exist regarding the correlations of disturbances over cross-sectional units that lead to natural groupings. Presumably, a hierarchy in correlations of disturbances among cross-sectional units is likely to exist. By grouping cross-sectional units together whose disturbances are likely to be highly correlated, $\hat{\Omega}$ can be formed as a block diagonal matrix defined by

$$6)\quad E(\varepsilon_{i,s,x}\varepsilon_{j,t,z}) = \sigma_i^2 \text{ for } i=j \text{ and } s=t;$$
$$= \sigma_{ij} \text{ for } i \neq j,\ s = t,\ x = z;$$
$$= 0 \text{ for } s \neq t \text{ or } x \neq z,$$

where i and j refer to individuals, s and t denote time periods, and x and z denote blocks. Assuming the cross-sections have been prearranged by blocks, the resulting estimator of Ω can then be written as

$$7)\quad \hat{\Omega} = \begin{bmatrix} \hat{\Omega}_1 & & & 0 \\ & \hat{\Omega}_2 & & \\ & & \ddots & \\ 0 & & & \hat{\Omega}_k \end{bmatrix}$$

where k blocks are estimated. Note that when k=M, equation 7 collapses to the diagonal structure, while setting k=1 yields the full model.

The choice of k requires judgement. Setting k to the smallest number of blocks consistent with the nonsingularity of $\hat{\Omega}$ would maximize the number of nonzero parameters in $\hat{\Omega}$. Using that strategy would imply estimating the full model whenever T exceeds M. As demonstrated below, however, there are conditions under which grouping the cross sections by blocks is preferred to the full specification. Although estimation using the maximum block size is more efficient in large samples, it is not necessarily efficient in small samples. For small samples, estimation of the maximum feasible number of contemporaneous disturbance covariances can reduce the efficiency of the estimate. In particular, when the correlation of the disturbance among cross-sectional units is low, limited degrees of freedom may yield imprecise estimates of those parameters. In such cases, more efficient estimation often can be achieved by assuming *a priori* that the covariances are equal to zero.[3/]

IV. Monte Carlo Results

To examine the effect of choosing various block sizes on the efficiency of pooled estimation, a series of Monte Carlo experiments were performed. Monte Carlo experiments were used because of the difficulty involved in deriving relative efficiency measures analytically under the general conditions being evaluated. The model used is of the form[4/]

8) $Y_{m,t} = 1 + 3X_{1,m,t} + 5X_{2m,t} + \varepsilon_{m,t}.$

The independent variables $X_{1,m,t}$ and $X_{2,m,t}$ were randomly drawn from the uniform distribution on the interval zero to one. The columns of the X matrix are therefore assumed to be uncorrelated.[5/] The same X matrix was used for all of the experiments.

A design matrix for Ω was specified as

9a) $\sigma_i^2 = \alpha\sigma_{i-1}^2$, where $\alpha \geq 1$;

9b) $\sigma_{ij} = [\sigma_i\sigma_j]\delta^{|j-i|}$, where $0 \leq \delta \leq 1$.

In equation 9a, the value of α determines the degree of cross-sectional heteroskedasticity.[6/] Equation 9b specifies a decay in disturbance correlations among cross-sectional units as the distance from the diagonal increases, with the degree of decay determined by δ. If $\delta=1$, all the disturbance correlations are equal to one, whereas, if $\delta=0$, all the disturbance correlations are equal to zero.[7/] The disturbance variances were scaled by a constant to maintain a reasonable signal to noise ratio.[8/]

By varying δ, the expected efficiency gains to estimation using a block diagonal variance-covariance structure can be illustrated. As δ falls, we would expect estimation using block specifications to become more efficient relative to estimation using the full model because of the expense involved in estimating the relatively low disturbance covariance terms far off the diagonal with a small number of time series observations.

The effects of δ on efficiency gains were analyzed using 24 cross-sectional units with 12 time series observations and 12 cross-sectional units with 24 time series observations.[9/] For each case considered, a series drawn from the normal distribution with mean zero and variance 1 was filtered through a transformation matrix to result in a disturbance vector randomly drawn from a normal population with the characteristics specified in equations 9 for that case. The resulting disturbance vector was then added to $X\beta$ to obtain a Y vector. Estimation of the coefficient vector was then carried out using OLS and GLS with various block sizes. The process was repeated 1000 times.[10/]

The relative efficiencies of GLS estimation using different block sizes are presented in the table. Relative efficiency is defined here as the ratio of the trace of the estimated mean square error matrices from a GLS procedure to that of the OLS procedure.[11/] The smaller this ratio, the greater the efficiency gain.

Several interesting results appear in the table. First, GLS methods uniformly register greater efficiency gains in the cases using 12 cross-sectional units and 24 time series observations than in those using 24 cross-sectional units and 12 time series observations. This finding is consistent with previous analytical results derived in the framework of 2 cross-sectional units which demonstrate that the relative efficiency in small samples of the seemingly unrelated regressions approach over single-equation estimation increases as degrees of freedom increase [Mehta and Swamy (1976)].

More importantly, the results show that block diagonal variance-covariance structures can offer greater efficiency gains than the diagonal model when estimation of the full model is infeasible. Estimation of the full model was infeasible for the 3 cases using 24 cross-sectional units and 12 time series observations. As shown in the table, with δ set to .55 or .9, a block size of 4 offers substantially greater efficiency gains than those achieved using the diagonal model in these cases. Only when the contemporaneous correlation of the disturbance drops off rapidly, when δ equals .2, does the diagonal model shown in the first column of the table offer the most efficient estimation.

Furthermore, the results indicate that use of less than the full specification -- even when the full model can be estimated -- leads to greater efficiency gains in a variety of cases. Estimation using the full model was feasible in the 3 cases using 12 cross-sectional units and 24 time series observations. The relative efficiency of the full model in these cases is shown in the last column of the table. In none of the cases does the full model show the highest efficiency gains. Even with δ set at .9, restricting some of the covariances to equal zero improves the efficiency of the estimator. A block diagonal variance-covariance structure offers the most efficient estimation with δ set to .55 or .9.

V. Conclusions

When estimating a pooled regression model with a large number of cross-sectional units and a relatively small number of time series observations, restrictions placed on the disturbance variance-covariance structure may yield substantial efficiency gains. As shown in this paper, estimation of the full model is not possible when the number of cross-sectional units exceeds the number of time series observations. In such cases, however, it is not necessary to totally neglect information on the disturbance covariances. Rather, strong efficiency gains can be realized over OLS or the estimator incorporating cross-sectional heteroskedasticity alone by using a block diagonal variance-covariance structure which includes only those disturbance covariances which are most likely to be significantly different from zero.

Furthermore, bias toward estimating the full GLS specification where such an estimator exists results in substantial efficiency losses in a variety of settings. If the disturbances of only some of the cross-sectional units are highly correlated and the number of time series observations is not large, a block variance-covariance structure offers greater efficiency gains as long as prior information can be used to group cross-sectional units according to the magnitude of the correlations between their disturbances.

Footnotes

1. See Theil (1971), 310.

2. Computation of the full model in these cases should be infeasible. However, we found that the procedure TSCSREG, available in SAS, still produces coefficient estimates and t-statistics. By contrast, SHAZAM does not provide estimates in these cases.

3. See Fomby, Hill, and Johnson (1984), 164-66, for a general discussion of this issue.

4. Choice of values for the β vector was arbitrary. Previous studies have used the specification presented in Kmenta and Gilbert (1968,1970), but as shown by Breusch (1980), properties of the estimators are invariant with respect to the β vector.

5. As demonstrated by Zellner (1962), the efficiency gains of estimators incorporating contemporaneous disturbance covariances are greatest when the disturbances are highly correlated and the explanatory variables are not. The correlation coefficient of X_1 and X_2 is -.009. This low degree of multicollinearity highlights efficiency gains.

6. The degree of cross-sectional heteroskedasticity was kept constant across different disturbance variance-covariance matrix dimensions by

choosing α such that the ratio of σ_1^2 to σ_M^2 was held equal to 9 as M increased.

7. This particular specification of the design covariance matrix yields a positive definite matrix.

8. The overall variance of the disturbance was set to yield an average R^2 of 0.8.

9. The general pattern of results proved invariant to a variety of settings on the number of cross-sectional and time series observations.

10. In addition to holding the X matrix constant, the same raw error stream (12x24x1000 in size) was used in every experiment.

11. The estimated mean square error matrix for a given covariance specification is defined as

$$1/n \begin{bmatrix} \underline{b}_1'\underline{b}_1 & \underline{b}_1'\underline{b}_2 & \underline{b}_1'\underline{b}_3 \\ \underline{b}_2'\underline{b}_1 & \underline{b}_2'\underline{b}_2 & \underline{b}_2'\underline{b}_3 \\ \underline{b}_3'\underline{b}_1 & \underline{b}_3'\underline{b}_2 & \underline{b}_3'\underline{b}_3 \end{bmatrix}$$

where $\underline{b}_i$ is an nx1 vector with each element defined as the difference between $\hat{\beta}_i$ and β_i in each experiment (n=1000). The relative efficiency measure is the ratio of traces of these matrices. Other measures could also be used, such as the ratio of the determinants of these matrices.

TABLE: Monte Carlo Estimates of the Relative Efficiency of Various Covariance Structures

Cross Sections	Time Series	δ	Block Size 1x1	4x4	6x6	12x12
Number of Blocks:			24	6	4	2
		.2	.727	.821	.865	.988
24	12	.55	.771	.644	.695	.977
		.9	.848	.407	.503	.958
Number of blocks:			12	3	2	1
		.2	.635	.669	.715	.843
12	24	.55	.656	.490	.516	.644
		.9	.708	.234	.256	.392

Relative efficiency = trace($EMSE_i$)/trace($EMSE_{OLS}$), where EMSE is the estimated mean square error matrix using the particular covariance structure. Number of trials: 1000.

References

Breusch, Trevor S., 1980, "Useful Invariance Results for Generalized Regression Models," Journal of Econometrics, 13, 327-340.

Fomby, T. B., Hill, R. C., and Johnson, S. R., 1984, Advanced Econometric Methods, Springer-Verlag New York Inc.

Kmenta, J. and R.F. Gilbert, 1968, "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions," Journal of the American Statistical Association 63, 1180-1200.

Kmenta, J. and R.F. Gilbert, 1970, "Estimation of Seemingly Unrelated Regressions with Autoregressive Disturbances," Journal of the American Statistical Association 65, 186-97.

Mehta, J.S. and Swamy, P.A.V.B., 1976, "Further Evidence on the Relative Efficiencies of Zellner's Seemingly Unrelated Regressions Estimator," Journal of the American Statistical Association, 71, 634-639.

Theil, H., 1971, Principles of Econometrics, John Wiley & Sons Inc.

Zellner, A., 1962, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, 57, 348-368.