EFFICIENT ESTIMATION OF PANEL DATA MODELS WITH STRICTLY EXOGENOUS EXPLANATORY VARIABLES

Kyung So Im
Department of Economics
Wichita State University, Wichita, KS 67260, USA

Seung C. Ahn*
Department of Economics
Arizona State University, Tempe, AZ 85287, USA

Peter Schmidt
Department of Economics
Michigan State University, E. Lansing, MI 48824, USA

Jeffrey M. Wooldridge
Department of Economics
Michigan State University, E. Lansing, MI 48824, USA

Abstract

With panel data, exogeneity assumptions imply many more moment conditions than standard estimators use. However, many of the moment conditions may be redundant, in the sense that they do not increase efficiency; if so, we may establish the standard estimators' efficiency. We prove efficiency results for GLS in a model with unrestricted error covariance matrix, and for 3SLS in models where regressors and errors are correlated, such as the Hausman-Taylor model. For models with correlation between regressors and errors, and with unrestricted error covariance structure, we provide a simple estimator based on a GLS generalization of deviations from means.

Key Words: Panel data; Strict exogeneity; Efficiency; Redundancy

JEL Classification: C13, C23

---

* Corresponding author, Department of Economics, Arizona State University, Tempe, AZ 85287-3806, USA. Phone: (602) 965-6574; FAX: (602) 965-0748; email: miniahn@asu.edu.

1. INTRODUCTION*

In this paper we consider the efficient estimation of panel data models containing unobserved individual effects. (Here, and throughout the paper, "efficient" actually means "asymptotically efficient.") The two most widely applied estimation procedures are random effects (RE) and fixed effects (FE). It is well-known that the consistency of the RE and FE estimators (as the cross section dimension tends to infinity with the time dimension fixed) requires the strict exogeneity of the regressors, but that the strict exogeneity assumption generates many more moment conditions than these estimators use. For example, in a panel data model with ten strictly exogenous time-varying regressors and six time periods, the total number of moment conditions available is 360 under RE assumptions and 300 under FE assumptions. The point of this paper is to establish the efficiency of simple estimators, like the RE and FE estimators, by providing assumptions under which the efficient GMM estimator based on the entire set of available moment conditions reduces to the simpler estimator. That is, we establish the efficiency of simple estimators by showing the redundancy of the moment conditions that they do not use. We do not provide any new estimators, but we do provide simplified versions of some existing estimators.

---

* The second author gratefully acknowledges the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business. The third author gratefully acknowledges the financial support of the National Science Foundation. We would also like to thank the editor and two anonymous referees for many helpful comments on the previous versions of this paper. All errors are our own.

In the models we consider, there will always be an estimator that is no less efficient than any of the estimators we study, namely the GMM estimator that uses all of the moment conditions and an unrestricted weighting matrix. Thus, in terms of achieving asymptotic efficiency, there really is no need to eliminate redundant moment conditions. As a practical matter, though, it is very useful to consider simpler estimators. GMM with a very large number of moment conditions is computationally very demanding, and may have poor small-sample properties. For example, see Tauchen (1986), Altonji and Segal (1996) and Andersen and Sørensen (1996) for a discussion of the small-sample bias of GMM in very overidentified problems.

The redundancy of moment conditions in GMM depends on relationships between the matrix of expected derivatives of the moment conditions and the optimal weighting matrix. We establish our results under an assumption of no conditional heteroskedasticity, which implies a simple and tractable form for the optimal weighting matrix. In this case the GMM estimator is a 3SLS estimator, as considered by Amemiya (1977), Hausman, Newey, and Taylor (1987), Schmidt (1990) and Wooldridge (1996). Our results are then given by showing the numerical equivalence of various 3SLS estimators. From these equivalences it would then be straightforward to make statements about relative asymptotic efficiencies of feasible GMM estimators.

Section 2 of the paper gives some preliminary results. We give conditions for the algebraic equivalence of 3SLS and generalized instrumental variables (GIV) in a system of linear equations, and for the redundancy of instruments in 3SLS. These results are used to prove the main theorems of the paper, but may also be useful in other settings.

Section 3 gives results applicable to models where the (composite) error is uncorrelated with all explanatory variables in all time periods, such as the random effects model under the strict exogeneity assumption. We establish the efficiency of the generalized least squares (GLS) estimator by showing its equivalence to the efficient 3SLS estimator.

Section 4 accommodates a time-invariant unobserved effect that can be correlated with the explanatory variables, which is the essential feature of the familiar fixed effects model. When the error has the usual random-effects covariance structure, and when the time-invariant effect is correlated with all of the explanatory variables, we show that the 3SLS estimator is equivalent to the within estimator. For the Hausman-Taylor model in which the time-invariant effect is correlated with some of the explanatory variables, the 3SLS estimator is equivalent to estimators proposed by Hausman and Taylor (1981), Amemiya and MaCurdy (1986), and Breusch, Mizon, and Schmidt (1989). If the time-varying error has a general covariance structure, things become more complicated. The 3SLS estimator is equivalent to an estimator of Kiefer (1980), and to several other new estimators we derive, when all explanatory variables are correlated with the individual effect. In the Hausman-Taylor model with general covariance structure, the 3SLS estimator is equivalent to a simplified estimator whose form depends on whether or not the correlation between the individual effect and the explanatory variables is time-invariant. If so, we obtain a simple GIV estimator. If not, we require a new method of handling deviations from means in order to obtain a simple estimator that is equivalent to 3SLS. This method amounts to taking the residuals from a GLS (as opposed to OLS) regression of the data on an individual-specific intercept.

Section 5 provides simulation evidence on the finite sample properties of the general GMM estimator and of some of the simpler estimators considered in the paper. All of the estimators that are theoretically efficient have small finite sample bias and reasonable finite sample efficiency. However, as would be expected, the unrestricted GMM estimator that uses all of the moment conditions yields standard errors that are overly optimistic; that is, they seriously understate the finite sample variability of the estimates, and thus lead to seriously incorrect inference.

2. PRELIMINARIES

2.1. Setup

We are interested in a linear model of the form

$$y_i = X_i\beta + u_i , \tag{2.1}$$

where i=1,...,N indexes the cross-sectional observations, $y_i = (y_{i1},...,y_{iT})'$ is a T×1 vector, X_i is a T×k matrix, β is the k×1 parameter vector of interest, and u_i is the T×1 vector of errors. Because our results are algebraic, there is no real need to introduce sampling assumptions. Nevertheless, it is useful to think of $\{(y_i,X_i): i=1,..., N\}$ as constituting an independent, identically distributed sequence of size N drawn from the population. When we refer to asymptotic results, this is the sampling assumption we have in mind. Importantly, this puts no restrictions on the dependence within the elements of y_i and X_i.

Throughout the paper, for any T×p matrix M_i, we denote $M \equiv (M_1',..., M_N')'$, which has dimension NT×p. Thus the matrix M is the stacked matrix corresponding to M_i. Then we can write (2.1) for the entire sample as $y = X\beta + u$.

A general approach to estimating β is to find a set of (say) T×h instruments W_i that are orthogonal to u_i, and then to apply a method of moments procedure. A natural orthogonality condition is

$$E(W_i' u_i) = 0 \ . \tag{2.2}$$

Let $\Sigma = E(u_i u_i')$ denote the T×T variance matrix of the errors, and define $\Omega \equiv I_N \otimes \Sigma$. Further, let X, y, and W denote the data matrices. Then the three stage least squares (3SLS) estimator of β is defined as

$$\hat{\beta}_{3SLS} = [X'W(W'\Omega W)^{-1}W'X]^{-1}X'W(W'\Omega W)^{-1}W'y \ . \tag{2.3}$$

Under the orthogonality condition (2.2), an identification assumption, and standard regularity conditions, $\text{plim}_{N\to\infty}\ \hat{\beta}_{3SLS} = \beta$. For purposes of algebraic equivalence results, we simply assume that $\hat{\beta}_{3SLS}$ exists.

Another familiar estimator is the generalized instrumental variables (GIV) estimator. This estimator is obtained by premultiplying (2.1) by $\Sigma^{-1/2}$ to prewhiten the errors, and then applying instrumental variables with instruments $\Sigma^{-1/2}W_i$. This leads to the estimator

$$\hat{\beta}_{GIV} = [X'\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}X]^{-1}X'\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}y \ . \tag{2.4}$$

Generally, the GIV estimator is not consistent under (2.2). The weakest orthogonality condition that implies consistency (when coupled with identification assumptions) of GIV is $E(W_i'\Sigma^{-1}u_i) = 0$. This condition is not generally implied by (2.2), since Σ^{-1} mixes instruments and errors over time. In some models (e.g., the dynamic panel data model), some of the instruments are only weakly exogenous, in the sense that $W_{i,tj}$ is uncorrelated with u_{is} only for $s \geq t$. Then $E(W_i'u_i) = 0$ holds but $E(W_i'\Sigma^{-1}u_i) = 0$ does not.

The optimal GMM estimator based on the orthogonality condition (2.2) uses the weighting matrix $E(W_i'u_i u_i'W_i)$, and the 3SLS estimator is asymptotically equivalent to the optimal GMM estimator under the assumption

$$E(W_i' u_i u_i' W_i) = E(W_i' \Sigma W_i) \ . \tag{NCH}$$

We refer to this assumption as "NCH", for no conditional heteroskedasticity. It has been used recently by Wooldridge (1996) to study 3SLS in the context of a general system of equations. This assumption will play a very significant role in our analyses, because it allows us to limit our focus to 3SLS estimators. Perhaps a more proper use of words would be to use NCH to refer to the stronger assumptions

$$E(u_i \mid W_i) = 0 \quad , \quad E(u_i u_i' \mid W_i) = \Sigma \tag{2.5}$$

which imply our assumption NCH. Chamberlain (1987) characterized the optimal instrumental variables under (2.5), where only the orthogonality condition $E(u_i \mid W_i) = 0$ is used in estimation. He showed that the optimal instruments are $\Sigma^{-1}E(X_i \mid W_i)$, and the IV estimator using these instruments is no less efficient than the 3SLS or GIV estimator. However, our orthogonality condition (2.2) is stated in terms of zero correlation only, not the stronger condition (2.5), and so it seems natural to rely on the weaker condition NCH.

2.2. Some Preliminary Results

We now give some preliminary results that will be useful in proving our later results. The first is a general result on redundancy of instruments in 3SLS. It is the algebraic equivalence analog of the asymptotic redundancy result of White (1984, Proposition 4.50).

THEOREM 2.1: Let $W = [W_1, W_2]$. Suppose $W_2'X = W_2'\Omega W_1(W_1'\Omega W_1)^{-1}W_1'X$. Then W_2 is redundant: $\hat{\beta}_{3SLS}$ using $W_i = [W_{i1}, W_{i2}]$ is the same as $\hat{\beta}_{3SLS}$ using W_{i1} only.

The proofs of all theorems are given in the Appendix.

We now turn to the relationship between 3SLS and GIV. Even under assumption

(2.5), neither the 3SLS estimator nor the GIV estimator can be shown to generally dominate the other (see, for example, Bowden and Turkington (1984, p. 72) and White (1984, pp. 83-105)). We seek conditions under which the 3SLS and GIV estimators are identical. We first give an algebraic result, which is the generalization of the familiar result that OLS = GLS if there is a nonsingular matrix R such that $\Omega^{-1}X = XR$.

THEOREM 2.2: Suppose that there exists a nonsingular matrix B such that $\Omega^{-1}W = WB$ (that is, $\Sigma^{-1}W_i = W_iB$ for all i). Then $\hat{\beta}_{3SLS} = \hat{\beta}_{GIV}$.

We first consider the case of common instruments; that is, where the same set of instruments is used for all t. This is given by:

ASSUMPTION 2.1 (Common Instruments): $W_i = I_T \otimes w_i^o$, where w_i^o is some $1 \times q$ vector.

(Note that then W_i is $T \times h$ where $h = Tq$.) As we will see, Assumption 2.1 is applicable to many panel data models with strictly exogenous regressors since the regressors in each time period are orthogonal to the errors in all time periods. Other models such as the standard simultaneous equation model or SUR model also have common instruments.

THEOREM 2.3: For estimating (2.1) under Assumption 2.1 (common instruments), the 3SLS and GIV estimators are the same.

The GIV estimator is just the IV estimator of (2.1) with instruments $\Sigma^{-1}W_i$. Using essentially the same proof, it is easy to see that estimators using instruments of the form $\Sigma^{\delta}W_i$ for any δ also reduce to the 3SLS estimator. The result of Schmidt, Ahn and Wyhowski (1992) corresponds to $\delta = -1/2$.

In a SUR system, OLS = GLS either if there are common regressors in each equation, or if Σ is diagonal. Theorem 2.3 is a generalization of the first of these results. The generalization of the second result is as follows. We can use Theorem 2.2 to show that, when Σ is diagonal and the instruments are of unrestricted form, GIV and 3SLS are identical. Details are given in an earlier version of this paper (Im, Ahn, Schmidt and Wooldridge (1996), Theorem 2.3).

3. REGRESSORS UNCORRELATED WITH ERRORS

We now study estimators of β in model (2.1) under the assumption that each element of X_i is orthogonal to each element of u_i; thus, we have in mind that

$$E(X_i \otimes u_i) = 0 . \tag{3.1}$$

Letting x_{it} denote the t^{th} row of X_i, we can choose the instruments for each t to be the nonredundant elements of $x_i^o = (x_{i1},x_{i2},...,x_{iT})$. Thus, we make the following assumption.

ASSUMPTION 3.1: $W_i = I_T \otimes w_i^o$, where w_i^o contains all nonredundant elements of x_i^o.

If there are no time-invariant elements in x_{it} then $w_i^o = x_i^o$. Any time-invariant variables (e.g., the constant term) only appear once in w_i^o. At this point we make no assumptions about the form of the error covariance matrix Σ.

In this case we would expect the efficient estimator to be GLS, even though our assumptions are weaker than those required for GLS to be best linear unbiased. This is shown by Arellano and Bover (1995, p. 34) for the special case that Σ has the random effects covariance structure. For completeness we give the general result.

THEOREM 3.1: The 3SLS estimator using the full set of instruments $W_i = I_T \otimes w_i^o$ equals the GLS estimator.

Even though it is unsurprising, this is a strong result. The strict exogeneity assumption implies T^2 moment conditions for each time-varying regressor, and T moment conditions for each time-invariant regressor. The GLS estimator exploits only one moment condition for each regressor: $\Sigma^{-½}X_i$ is uncorrelated with $\Sigma^{-½}u_i$. The GLS moment conditions are a linear combination of the strict-exogeneity conditions, and the remaining linear combinations are therefore redundant.

Under the NCH assumption, the 3SLS estimator is the optimal GMM estimator and our result implies the efficiency of GLS. Without the NCH assumption, GLS makes little sense and it is not surprising that it is dominated by the efficient GMM estimator.

An important special case is the random effects covariance structure.

ASSUMPTION 3.2 (Random Effects): The T×T matrix Σ can be written as

$$\Sigma = \sigma_\phi^2 e_T e_T' + \sigma_\epsilon^2 I_T , \tag{3.2}$$

where σ_ε^2 and σ_ϕ^2 are positive scalars and e_T is the T×1 vector with each element unity.

This form of Σ arises from the variance components structure in which $u_{it} = \phi_i + \varepsilon_{it}$, where the ε_{it} are assumed to be serially uncorrelated with variance σ_ε^2, and uncorrelated with ϕ_i. Then $\Sigma = (\sigma_\varepsilon^2 + T\sigma_\phi^2)P_T + \sigma_\varepsilon^2 Q_T$, where $P_T \equiv e_T(e_T'e_T)^{-1}e_T$ and $Q_T \equiv I_T - P_T$.

Theorem 3.1 implies that the efficient estimator is GLS. However, one might not wish to use GLS, perhaps because of doubt as to whether the NCH assumption and random effects covariance structure are correct. The following result may then be useful.

THEOREM 3.2: Under Assumption 3.2 (random effects), the 3SLS estimator using instruments $(P_T X_i, Q_T X_i)$ equals the 3SLS estimator using the entire instrument set $W_i = I_T \otimes w_i^o$, and therefore also equals the random effects GLS estimator.

This result shows that we can reduce the number of moment conditions significantly, while ensuring that the GMM estimator is no less efficient than the random effects GLS estimator. In particular, the GMM estimator using instruments $(P_T X_i, Q_T X_i)$ and an unrestricted weighting matrix will be just as efficient as the random effects GLS estimator if the NCH assumption and random effects covariance structure hold. If these assumptions do not hold, this new GMM estimator will generally be more efficient than GLS. It would generally be inefficient relative to the GMM estimator using the full set of moment conditions, but much more attractive computationally, and perhaps in terms of the finite sample properties of the estimators.

4. REGRESSORS CORRELATED WITH A TIME-INVARIANT ERROR COMPONENT

In this section we consider two models where the time-invariant unobserved effect may be correlated with some or all of the regressors. In the next subsection we consider the model in which all regressors are time-varying and possibly correlated with the unobserved effect. Depending on the error covariance structure, this corresponds to the traditional fixed effects model, or to a model of Kiefer (1980). In the following subsection we then consider extensions of the Hausman-Taylor (1981) model, in which some regressors are assumed to be uncorrelated with the unobserved effect. We establish the efficiency of some existing estimators, but also suggest some new estimators.

4.1. *Fixed Effects-Type Models*

In this subsection we consider the unobserved effect model in which $u_{it} = \phi_i + \varepsilon_{it}$. However, we now allow ϕ_i and x_{it} to be correlated, so that the model is of the fixed effects variety. We assume that the regressors x_{it} are strictly exogenous with respect to the time-varying errors ε_{it}, so that $E(X_i \otimes \varepsilon_i) = 0$. Under these assumptions, only coefficients on time-varying regressors are identified; thus, for this subsection, x_{it} contains only time-varying regressors.

As before, we consider estimation of (2.1) by instrumental variables. Now, only certain linear combinations of X_i are guaranteed to be uncorrelated with u_i. Define the T×(T-1) differencing matrix as

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}. \tag{4.1}$$

Then, define $W_i = L \otimes x_i^o$, where as before $x_i^o = (x_{i1},\dots,x_{iT})$ is a $1 \times Tk$ row vector; thus, W_i is $T \times T(T-1)k$. It is well known (e.g., Schmidt, Ahn and Wyhowski (1992)) that the set of orthogonality conditions available for estimating β is $E(W_i'u_i) = 0$.

Under the assumption that Σ has the standard random effects covariance structure, the within estimator would commonly be used. The instruments used by the within estimator are just the deviations from individual means, $Q_T X_i$ (where as before Q_T is the T×T demeaning matrix). The following theorem establishes the efficiency of this estimator.

THEOREM 4.1: Under Assumption 3.2 (random effects structure), the 3SLS estimator using the instruments $W_i = L \otimes x_i^o$ is the same as the within estimator.

This result was also stated without proof by Arellano and Bover (1995, p. 34). It is the analog of the result of Chamberlain (1992) on the efficiency of the within estimator under conditional mean assumptions.

We next wish to relax the assumption that Σ has the random effects covariance structure. Thus we allow $\Sigma_\varepsilon = E(\varepsilon_i \varepsilon_i')$ to have an unrestricted form, which implies that $\Sigma = E(u_i u_i')$ is also unrestricted. This effectively gives the setup of Kiefer (1980).

We can apply Theorem 2.1 to show that some of the instruments used in 3SLS are redundant. Define $\bar{x}_i = T^{-1}\Sigma_{t=1}^{T} x_{it}$ and $\ddot{x}_i = (x_{i1}-\bar{x}_i, \ldots, x_{i,T-1}-\bar{x}_i)$. Then it is clear that $(\bar{x}_i, \ddot{x}_i)$ is a nonsingular linear transformation of x_i^o, and so $(\bar{x}_i, \ddot{x}_i)$ and x_i^o are equivalent as instruments.

THEOREM 4.2: The 3SLS estimator using instruments $L \otimes x_i^o$ equals the 3SLS estimator using instruments $L \otimes \ddot{x}_i$.

This theorem shows that, among the T(T-1)k instruments in $L \otimes x_i^o$, the (T-1)k instruments $L \otimes \bar{x}_i$ are redundant. This result may be more understandable if one writes out the terms involved in the Kronecker products above. A typical element of the expression $(L \otimes x_i^o)' u_i$ is $x_i^{o\prime}(u_{it}-u_{i,t+1})$, and similarly for $(L \otimes \ddot{x})' u_i$ and $(L \otimes \bar{x})' u_i$. The statement that $E[\bar{x}_i'(u_{it}-u_{i,t+1})] = 0$ is correct, but it is redundant because means are irrelevant once the equation has implicitly been first-differenced.

As with Theorem 3.2, this result can be used to construct a GMM estimator (using instruments $L \otimes \ddot{x}_i$ and general weighting matrix) that is at least as efficient as 3SLS using the same set of instruments, but that is robust to violations of the NCH assumption.

Under these assumptions, other simple estimators are available. One can apply GLS to the first differenced or demeaned equations. Kiefer (1980) proposed GLS applied to the demeaned data using $(Q_T \Sigma Q_T)^-$, a generalized inverse of the error covariance of the demeaned data. Given the singularity of the variance matrix of the demeaned errors, it seems reasonable that no information is lost by deleting any one equation in the demeaned data. Also, demeaning and differencing should preserve the same information. To state a precise result, let $\hat{\beta}_{3SLS}$ denote the 3SLS estimator from Theorem 4.2, let $\hat{\beta}_{KF}$ denote Kiefer's estimator, let $\hat{\beta}_{DM}$ be the GLS estimator in the demeaned equation after deleting any one time period, and let $\hat{\beta}_{DF}$ be the GLS estimator in the differenced set of equations.

THEOREM 4.3: $\hat{\beta}_{3SLS} = \hat{\beta}_{KF} = \hat{\beta}_{DM} = \hat{\beta}_{DF}$.

4.2. Hausman and Taylor-Type Models

Hausman and Taylor (1981) considered a model where certain explanatory variables are uncorrelated with the unobserved effect. This model offers a middle ground between the pure random effects and pure fixed effects approaches. Write the model for all T time periods as

$$y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + (e_T \otimes z_{i1})\gamma_1 + (e_T \otimes z_{i2})\gamma_2 + \phi_i e_T + \epsilon_i \, , \tag{4.2}$$

where X_{i1} and X_{i2} are $T \times k_1$ and $T \times k_2$ matrices, respectively, of time-varying explanatory variables, and z_{i1} and z_{i2} are $1 \times g_1$ and $1 \times g_2$ vectors, respectively, of time-invariant explanatory variables. Other definitions are as before.

The explanatory variables are all strictly exogenous with respect to the time-varying error ε_i, but X_{i2} and z_{i2} can be correlated with the unobserved effect ϕ_i. Writing $u_i = \phi_i e_T + \varepsilon_i$ as before, consider the following orthogonality conditions:

$$E(X_{i1} \otimes u_i) = 0 \ ; \quad E(X_{i2} \otimes \epsilon_i) = 0 \ ; \quad E(z_{i1} \otimes u_i) = 0 \ ; \quad E(z_{i2} \otimes \epsilon_i) = 0 \tag{4.3}$$

Given the orthogonality conditions (4.3), we have the moment conditions $E(W_i^{(1)\prime} u_i) = 0$, with the instruments $W_i^{(1)}$ defined as follows:

$$W_i^{(1)} \equiv (I_T \otimes x_{i1}^o, L \otimes x_{i2}^o, I_T \otimes z_{i1}, L \otimes z_{i2}) \ . \tag{4.4}$$

This set of instruments can be expressed in a number of different ways. An equivalent instrument set (in the sense of leading to the same projection and estimators) is

$$W_i^{(2)} = [L \otimes (x_{i1}^o, x_{i2}^o, z_{i1}, z_{i2}), e_T \otimes (x_{i1}^o, z_{i1})] \ . \tag{4.5}$$

This distinguishes the instruments in deviations from means space from those in means space. All of our redundancy results involve the first set of instruments.

Breusch, Mizon and Schmidt (1989) suggested an additional assumption:

$$E(x_{it2}^{\prime} \phi_i) \textit{ is the same for } t = 1, 2, \ldots, T. \tag{BMS}$$

This means that, even though the unobserved effect might be correlated with x_{it2}, the covariance does not change with time. We will refer to this as the BMS assumption. Under the BMS assumption the additional instruments $e_T \otimes \ddot{x}_{i2}$ become valid.

If we assume the random effects covariance structure, it is well known that all of the instruments in deviations from means space are redundant, except for the deviations themselves, $Q_T X_i$. This was shown by Breusch, Mizon and Schmidt (1989); see also Arellano and Bover (1995, Appendix) and Ahn and Schmidt (1995, p. 19). For completeness we state this result (without proof).

THEOREM 4.4: Under Assumption 3.2 (random effects), the 3SLS estimator using the full set of instruments $W_i^{(2)}$ in (4.5) equals the 3SLS estimator using the instruments $W_i^{(3)} = [Q_T X_i, e_T \otimes (x^o_{i1}, z_{i1})]$. Similarly, the 3SLS estimator using instruments $W_i^{(4)} = [L \otimes (x^o_{i1}, x^o_{i2}, z_{i1}, z_{i2}), e_T \otimes (x^o_{i1}, \ddot{x}_{i2}, z_{i1})]$ equals the 3SLS estimator using instruments $W_i^{(5)} = [Q_T X_i, e_T \otimes (x^o_{i1}, \ddot{x}_{i2}, z_{i1})]$.

We now relax the assumption that Σ has the random effects structure. Instead, we allow $\Sigma = E(u_i u_i')$ to be unrestricted. Interestingly, the BMS assumption now becomes fundamentally important. Not only does the BMS assumption add to the instrument set, but it also affects the ways in which the instruments can be used to yield a consistent estimator.

We first consider the case in which the BMS assumption is assumed to hold. The set of available instruments is $W_i^{(4)}$ as given in the statement of Theorem 4.4. This set of instruments can be rewritten in a number of ways that are equivalent, in the sense of leading to the same projection. One such equivalent instrument set is the following.

$$W_i^{(6)} = [I_T \otimes (x^o_{i1}, \ddot{x}_{i2}, z_{i1}), L \otimes (\bar{x}_{i2}, z_{i2})] \ . \tag{4.6}$$

We can also consider the smaller instrument set

$$W_i^{(7)} = I_T \otimes (x^o_{i1}, \ddot{x}_{i2}, z_{i1}) \ . \tag{4.7}$$

This leads to the following redundancy result, which is a weaker version of a result given by Arellano and Bover (1995, p. 38).

THEOREM 4.5: The 3SLS estimator using the instrument set $W_i^{(4)}$ — or $W_i^{(6)}$ — equals the 3SLS estimator using the instrument set $W_i^{(7)}$. That is, the instruments $L \otimes (\bar{x}_{i2}, z_{i2})$ are redundant.

We still have a large number of instruments in deviations from means space, and the question arises whether we can reduce these simply to Q_TX_i, as is true under the random effects covariance structure. Explicitly, consider the instrument set $W_i^{(5)}$ = $[Q_TX_i,e_T\otimes(x^o_{i1},\ddot{x}_{i2},z_{i1})]$, which yields the efficient 3SLS estimator when Σ has the random effects covariance structure, as claimed in Theorem 4.4. Unfortunately, when Σ does not have the random effects structure, it is not the case that 3SLS based on $W_i^{(5)}$ is the same as 3SLS based on the larger instrument sets $W_i^{(4)}$ or $W_i^{(6)}$. We can still derive an efficient estimator using the instrument set $W_i^{(5)}$, by considering a GIV estimator instead of 3SLS.

THEOREM 4.6: The 3SLS estimator using the instruments $W_i^{(2)}$ — or $W_i^{(4)}$ or $W_i^{(6)}$ — equals the GIV estimator using the instruments $W_i^{(5)} = [Q_TX_i,e_T\otimes(x^o_{i1},\ddot{x}_{i2},z_{i1})]$.

Theorem 4.6 is a novel and practically useful result. It shows that, in the case of unrestricted covariance matrix, we can still obtain an efficient estimator using the same number of instruments as were necessary when Σ has the random effects structure. Since the GIV estimator using instruments $W_i^{(5)}$ is a 3SLS estimator using instruments $\Sigma^{-1}W_i^{(5)}$, we do obtain a 3SLS estimator with a very reduced instrument set; the key to the successful treatment is filtering of the instruments by Σ^{-1}.

If the model does not contain x_{i1} or z_i, then it reduces to the fixed effects-type model of section 4.1 with general covariance matrix but with the additional BMS assumption. Then, the instruments $e_T\otimes\ddot{x}_{i2}$ are relevant, and the GIV estimator using instruments $(Q_Tx_i,e_T\otimes\ddot{x}_{i2})$ is more efficient than the estimators discussed in Theorem 4.3.

We now relax the BMS assumption, while allowing Σ to be unrestricted. Without the BMS assumption, the instruments $e_T\otimes\ddot{x}_{i2}$ are no longer legitimate. The set of available instruments can be written in a number of ways; e.g., see $W_i^{(1)}$ and $W_i^{(2)}$ above.

A more subtle implication of the relaxation of the BMS assumption is that a GIV estimator with Q_TX_i in the instrument set will (in general) be inconsistent. To see why, observe that the demeaned regressors Q_TX_i are legitimate instruments for GIV only if $E(X_i'Q_T\Sigma^{-1}u_i) = 0$. Clearly

$$E(X_i'Q_T\Sigma^{-1}u_i) = E(X_i'Q_T\Sigma^{-1}e_T\phi_i) + E(X_i'Q_T\Sigma^{-1}\epsilon_i) \; . \tag{4.8}$$

The last term in (4.8) equals zero given strict exogeneity of X_i with respect to ε_i. The random effects structure implies $Q_T\Sigma^{-1}e_T = 0$ so that both terms on the right-hand side of (4.8) equal zero. However, for general Σ, $Q_T\Sigma^{-1}e_T \neq 0$. The BMS assumption implies that every element of Q_TX_i is uncorrelated with ϕ_i, so that $E(X_i'Q_T\Sigma^{-1}e_T\phi_i) = 0$ and the expression in (4.8) equals zero. However, with general Σ and without the BMS assumption, $E(X_i'Q_T\Sigma^{-1}u_i) \neq 0$ in general, and Q_TX_i is not legitimate for GIV estimation.

A simple solution to this problem is to replace Q_T by a different matrix that removes the effects. Define the T×T idempotent (though not symmetric) matrix R_Σ:

$$R_\Sigma = I_T - e_T(e_T'\Sigma^{-1}e_T)^{-1}e_T'\Sigma^{-1} \; . \tag{4.9}$$

We can note that R_Σ corresponds to taking residuals in a GLS (as opposed to OLS) regression on an individual-specific intercept. Clearly $R_\Sigma e_T = 0$, and also $R_\Sigma'\Sigma^{-1}e_T = 0$. Thus

$X_i'R_\Sigma'\Sigma^{-1}e_T\phi_i = 0$, and $E(X_i'R_\Sigma'\Sigma^{-1}u_i) = E(X_i'R_\Sigma'\Sigma^{-1}\varepsilon_i) = 0$ given strict exogeneity of X_i with respect to ε_i. Thus, $R_\Sigma X_i$ are legitimate instruments for GIV.

This discussion motivates the GIV estimator based on the instrument set

$$W_i^{(8)} = [R_\Sigma X_i, e_T \otimes (x_{i1}^o, z_{i1})] \ . \qquad \textbf{(4.10)}$$

We will call this a modified GIV (or MGIV) estimator, where the name reflects the modified form of taking deviations from means. The following theorem establishes its equivalence to the efficient 3SLS estimator.

THEOREM 4.7: The GIV estimator using the instrument set $W_i^{(8)}$ of equation (4.10) equals the 3SLS estimator using the instrument set $W_i^{(1)}$ of equation (4.4).

The GIV estimator based on $W_i^{(8)}$ is not the same as the 3SLS estimator based on $W_i^{(8)}$. However, it is obviously the same as the 3SLS estimator based on $\Sigma^{-1}W_i^{(8)}$.

Amemiya and MaCurdy (1986, pp. 877-878) provide an alternative estimator for the case we are considering — namely, Σ is unrestricted and the orthogonality assumptions in (4.3) hold, but the BMS assumption is not maintained. Define $\ddot{u}_i = (u_{i1}-\bar{u}_i,\ldots,u_{i,T-1}-\bar{u}_i) = (Q_T^* u_i)'$, where Q_T^* is Q_T with the last row deleted. Let f be the (T-1)×1 vector such that $\mathrm{Proj}(\bar{u}_i|\ddot{u}_i) = f'\ddot{u}_i' = f'Q_T^* u_i$. Then the AM estimator is the GIV estimator using instruments

$$W_i^{(9)} = [X_{i1}, (Q_T + e_T f' Q_T^*)X_{i2}, e_T \otimes (x_{i1}^o, z_{i1})] \qquad \textbf{(4.11)}$$

Amemiya and MaCurdy provide conditions under which this estimator is efficient, which are stronger than our assumptions above.

There is no apparent comparison between the instrument sets $W_i^{(8)}$ in (4.10) and $W_i^{(9)}$ in (4.11). However, perhaps surprisingly, they lead to the same GIV estimator.

THEOREM 4.8: The GIV estimator based on the instrument set $W_i^{(9)}$ — that is, the Amemiya-MaCurdy estimator — equals the GIV estimator based on $W_i^{(8)}$.

5. SIMULATIONS

In this section we report the results of limited Monte Carlo simulations that compare the finite sample properties of some of the estimators discussed above. The basis of the experiment is a regression model of the form of equation (4.2) above. For simplicity we take $k_1 = k_2 = g_1 = g_2 = 1$. We will consider estimators that do not impose the BMS assumption. The various estimators rely on three different sets of instruments, as follows: the *big instrument set* $W_i^{(2)}$, defined in equation (4.5), which contains $2T^2+T-1$ instruments; the *small instrument set* $W_i^{(3)}$, defined in Theorem 4.4, which contains T+3 instruments; and the *MGIV instrument set* $W_i^{(8)}$, defined in equation (4.10), which also contains T+3 instruments. We consider the following estimators. (i) *GMM-BS*, which is the GMM estimator using the big instrument set and an unrestricted weighting matrix; (ii) *GMM-SS*, which is the same as GMM-BS except that it uses the small instrument set; (iii) *3SLS*, which uses the small instrument set and an unrestricted Σ matrix (but imposes the NCH assumption on the weighting matrix); (iv) *3SLS-RE*, which is the same as 3SLS except that it imposes the random effects structure on Σ; and (v) *MGIV*, which uses the MGIV instrument set and an unrestricted Σ matrix.[1] Several related estimators are not considered separately because they are the same as those above. For example, 3SLS using the big instrument set and unrestricted Σ is the same as MGIV, by Theorem 4.7; 3SLS using the big instrument set and Σ of random effects structure is the same as 3SLS-RE, by Theorem 4.4; and MGIV with Σ of

---

[1] This footnote describes estimation of Σ. Let e_i be the T×1 vector of 2SLS residuals. The unrestricted estimate of Σ is $S = N^{-1}\Sigma_{i=1}^{N} e_i e_i'$. For the random effects estimates, the estimate of σ_ϕ^2 is $s_\phi^2 = [T(T-1)]^{-1}$ times the sum of the off-diagonal elements of S. The estimate of σ_ε^2 is T^{-1} times the sum of the diagonal elements of S, minus s_ϕ^2.

random effects structure is the same as 3SLS-RE, by Theorems 4.4 and 4.7.

We pick $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 1$; N = 200 and 500; and T = 5, 8 and 10.[2] z_1 is an intercept ($z_{1i} \equiv 1$). Let ϕ_i, f_i, ξ_i, η_{it1} and η_{it2} be independent series, with independence also over all i and t, where ϕ is N(0,1), while f, ξ, η_1 and η_2 are uniform on [-2,2]. Then $x_{it1} = 0.7x_{i,t-1,1} + f_i + \eta_{it1}$, $x_{it2} = 0.7x_{i,t-1,2} + \phi_i + \eta_{it2}$, and $z_{i2} = f_i + \phi_i + \xi_i$.[3] Thus x_2 and z_2 are correlated with each other and with the error ($u_{it} = \phi_i + \varepsilon_{it}$) through joint dependence on α, while x_1 is correlated with z_2 through joint dependence on f.

To construct the time-varying errors, let v_{it} be independent draws from N(0,2), and define $\varepsilon_{i1} = v_{i1}/(1-\rho^2)^{½}$ and $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + v_{it}[(1-b)+bx_{1it}/se(x_{1it})]$ for t = 2,...,T, where $se(x_{1it})$ is the sample standard error of the x_1. We consider three cases: (i) b=ρ=0, so NCH and the random effects structure hold; (ii) b=0, ρ=0.5, so that NCH holds but there is serial correlation in the idiosyncratic errors; and (iii) b=1, ρ=0.5, so that the NCH condition does not hold.

Calculations were done in GAUSS. The number of replications was 2000.

For each of the four coefficients, we report the bias and root mean square error

---

[2] We consider T=10 only for N = 500, because with T= 10 there are 209 instruments in the big instrument set, and with N=200 this would imply singularity of the unrestricted weighting matrix for GMM-BS. The results for N=500 and T=10 are not displayed in this paper, but are in supplemental tables available from the authors on request.

[3] Initial values x_{i11} and x_{i12} were created in such a way as to imply covariance stationarity of x_1 and x_2. Thus, for example, if η_{i11} is drawn as above, we created $x_{i11} = \eta_{i11}/(1-0.7^2)^{½} + \alpha_i/(1-0.7)$. We note that this implies that the BMS assumption holds in the data generating process, even though we do not impose it in estimation.

(RMSE). These numbers are of obvious interest to see if any of the estimators are substantially biased, and to compare their precision. We also report the "average s.e." which is the average over the replications of the conventionally-calculated standard error, in an attempt to see whether the standard errors indicated by asymptotic theory are reliable in finite samples. The "robust s.e." is similarly the average over the replications of the heteroskedasticity-robust standard error (relevant for the 3SLS, 3SLS-RE and MGIV estimators only). Finally, "5% size" is the fraction of rejections of the null hypothesis that the coefficient equals its true value, based on the asymptotic normal statistic given by the ratio of the coefficient estimate to its (non-robust) asymptotic standard error. If asymptotic theory is reasonably reliable, we should find the average standard error to be nearly equal to RMSE and the 5% size close to 0.05.

Table 1 gives the results for the case that the random effects structure holds ($b=\rho=0$). For this case our earlier results imply that all of the estimators are equally efficient. None of the estimators shows any substantial bias. In terms of RMSE, the estimators are quite similar except that GMM-BS tends to have somewhat larger RMSE, especially for N=200; presumably this is due to its use of a large number of instruments that are redundant for this case. 3SLS-RE, which correctly imposes the random effects structure, does not seem to be significantly better than the other estimators that use the small instrument set. The average standard error is usually close to the RMSE, indicating the reliability of the asymptotic standard errors, except for GMM-BS, for which the average standard error is considerably smaller than RMSE. Thus the asymptotically-valid standard errors are a reasonably accurate guide to the finite sample variability of the estimates, except for GMM-BS, for which they are substantially too small. For example, for N=200 and T=8, the asymptotic standard error of GMM-BS is on average only about half of its RMSE, though the results are more favorable for larger N or smaller T. The unreliability of the asymptotic standard errors for GMM with an unrestricted weighting matrix in heavily overidentified problems is not a surprise, given the results of earlier studies such as Tauchen (1986), Altonji and Segal (1996) and Andersen and Sørensen (1996). This problem is also evident in the results for 5% size, which indicate too many rejections of the true null hypothesis for t-tests based on GMM-BS.

Table 2 gives the results for the case that the NCH assumption holds, but there is serial correlation in the ε_{it} and so the random effects structure does not hold (b=0, ρ=0.5). In this case, our earlier results indicate that GMM-BS and MGIV are equally efficient, while the other three estimators are inefficient. The bias of GMM-BS is often larger than that of the other estimators, but none of the estimators is seriously biased. In terms of RMSE, MGIV is usually best, which is not surprising given its theoretical efficiency. The three inefficient estimators (GMM-SS, 3SLS and 3SLS-RE) have larger RMSE than MGIV, by about equal amounts. 3SLS-RE, which incorrectly imposes the random effects structure, does not seem any worse than the other inefficient estimators, even though it is theoretically less efficient than GMM-SS or 3SLS-SS, which use the optimal weighting matrix given their instrument set. GMM-BS, which is theoretically efficient, has considerably higher RMSE than MGIV, and is often worse than the inefficient estimators. Once again the average standard errors are quite close to the RMSE, except for GMM-BS, for which the standard errors substantially underestimate RMSE. Correspondingly, the t-tests for GMM-BS have size much larger than 5%.

Table 3 gives the results for the case that the NCH assumption fails (b=1, ρ=0.5); there are both serial correlation and conditional heteroskedasticity. In this case only

GMM-BS is efficient. Once again, none of the estimators seems seriously biased. In terms of RMSE, GMM-BS is indeed best, but it is generally not much better than the other estimators. Its superiority in terms of RMSE must be balanced against the fact that its standard errors still seriously understate the RMSE, and correspondingly its t-tests reject too often. For the other estimators, the standard errors are on average quite close to RMSE. For the 3SLS and MGIV estimators, only the robust standard errors are asymptotically correct, and they are more accurate but not much more accurate than the non-robust standard errors.

The main results of the Monte Carlo experiments can be summarized as follows. The GMM estimator using the big instrument set and unrestricted weighting matrix (GMM-BS) has an acceptably small bias and good efficiency properties, but its standard errors are unreliable. All of the other estimators do quite well when the NCH assumption and random effects structure hold. The MGIV estimator is best when the NCH assumption holds but the random effects structure does not. When the NCH assumption fails, the MGIV estimator with robust standard errors is still a good alternative to GMM-BS, because it is computationally simpler, it is nearly as efficient in finite samples, and it generates much more reliable inferences.

6. CONCLUDING REMARKS

In this paper we have considered 3SLS and GIV estimation of popular panel data models, under the assumption of strict exogeneity of the regressors with respect to the time-varying error. The 3SLS and GIV estimators are often equivalent, and we give a general condition for this equivalence. Under an assumption of no conditional heteroskedasticity of the errors, we provide a systematic treatment of the estimation problem, starting with 3SLS based on the entire set of moment conditions implied by strict exogeneity, and then reducing the number of moment conditions without loss of efficiency. Thus the simple 3SLS or GIV estimators that we provide are efficient, even though they rely on only a small subset of the available moment conditions.

These results are important because the strict exogeneity assumption generates a very large number of moment conditions. Empirically relevant models can have hundreds of available instruments, and this can cause computational problems and may call into question the finite sample properties of the estimates. Our results show the redundancy of most of these moment conditions. Our focus on 3SLS and GIV is motivated by the fact that the redundancy results depend on the relationship between the regressors, instruments and weighting matrix; with a general weighting matrix, such results are generally not possible.

An obvious extension of our paper is to relax the strict exogeneity assumption. Some results under the assumption of weak exogeneity can be found in Keane and Runkle (1992), Schmidt, Ahn and Wyhowski (1992), and Im (1994).

APPENDIX

We will use the following notation. First, for any row vector d_i, we define $D \equiv (d_1',...,d_N')'$. Second, for any q×p matrix A of full column rank, we define the projection matrix onto the column space of A by P(A) and the projection matrix onto the null space of A by Q(A); that is, $P(A) = A(A'A)^{-1}A'$ and $Q(A) = I_q - P(A)$.

PROOF OF THEOREM 2.1: Since the assumption implies $W_2'\Omega^{½}Q(\Omega^{½}W_1)\Omega^{-½}X = 0$, we obtain

$$P(\Omega^{1/2}W)\Omega^{-1/2}X = [P(\Omega^{1/2}W_1)+P(Q(\Omega^{1/2}W_1)\Omega^{1/2}W_2)]\Omega^{-1/2}X = P(\Omega^{1/2}W_1)\Omega^{-1/2}X\ ,$$

where the first equality results from Amemiya (1985, p. 461). Thus, we can show

$$\begin{aligned} W(W'\Omega W)^{-1}W'X &= \Omega^{-½}P(\Omega^{½}W)\Omega^{-½}X \\ &= \Omega^{-½}P(\Omega^{½}W_1)\Omega^{-½}X = W_1(W_1'\Omega W_1)^{-1}W_1'X\ . \end{aligned}$$ ■

PROOF OF THEOREM 2.2: The desired result follows since

$$\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}X = WB(B'W'\Omega WB)^{-1}B'W'X = W(W'\Omega W)^{-1}W'X\ .$$ ■

PROOF OF THEOREM 2.3: $\Sigma^{-1}(I_T\otimes w_i^o) = (I_T\otimes w_i^o)(\Sigma^{-1}\otimes I_q)$. ■

PROOF OF THEOREM 3.1: By Theorem 2.3, 3SLS = GIV when W_i is defined as in Assumption 3.1. Thus, all we need to show is that GIV = GLS; i.e., $P(\Omega^{-½}W)\Omega^{-½}X = \Omega^{-½}X$. This is clearly the case since X is in the column space of W. ■

PROOF OF THEOREM 3.2: Without loss of generality, we set $\sigma_\varepsilon^2 + T\sigma_\alpha^2 = 1$ and denote $a = \sigma_\varepsilon^2$. Observe that

$$\Sigma^{-1}(P_TX_i,Q_TX_i) = (P_TX_i,a^{-1}Q_TX_i) = (P_TX_i,Q_TX_i)\mathrm{diag}(I_k,a^{-1}I_k)\ .$$

Thus, Theorem 2.2 applies and the 3SLS estimator using (P_TX_i,Q_TX_i) equals the GIV estimator using the same instruments. Thus, our proof can be completed if we can show GIV = GLS; that is, $P[\Omega^{-½}(PX,QX)]\Omega^{-½}X = \Omega^{-½}X$, where $P = I_N\otimes P_T$ and $Q = I_N\otimes Q_T$. This is true since X is in the column space of (PX,QX). ■

PROOF OF THEOREM 4.1: It is well known that L and Q_T span the same space. Thus we have $P(L) = Q_T$, $Q_T L = L$, $P_T L = 0$, $Q_T W_i = W_i$ and $P_T W_i = 0$. With $\Sigma = aQ_T + P_T$, this implies $W_i'\Sigma W_i = aW_i'W_i$ or $W'\Omega W = aW'W$. Thus,

$$\hat{\beta}_{3SLS} = [X'P(W)X]^{-1}X'P(W)y = [X'QP(W)QX]^{-1}X'QP(W)Qy \ .$$

However, it is easy to see that $Q_T X_i$ is in the space spanned by $W_i = L\otimes x_i^o$; that is, $P(W)QX = QX$. ■

The following lemma is useful for the proof of Theorem 4.2:

LEMMA 1: $[P(\ddot{X})\otimes Q_T]X = [P(\ddot{X})\otimes Q_T]QX = QX$.

PROOF: Note that when k = 1, $X = \mathrm{vec}(X^{o\prime})$ and $\ddot{X} = X^o Q_T$. Hence,

$$[P(\ddot{X})\otimes Q_T]X = [P(\ddot{X})\otimes Q_T]\mathrm{vec}(X^{o\prime}) = \mathrm{vec}[Q_T X^{o\prime} P(\ddot{X})] = \mathrm{vec}(Q_T X^{o\prime}) = QX \ ,$$

which is also valid when k > 1 by applying this argument to each of the regressors separately. The third equality results because $\ddot{X}$ consists of the first (T-1)k columns of $X^o Q_T$ which in turn span the last k columns of $X^o Q_T$. ■

PROOF OF THEOREM 4.2: From Theorem 2.1, it is sufficient to show that

$$(\bar{X}\otimes L)'X = (\bar{X}\otimes L)'(I_N\otimes\Sigma)(\ddot{X}\otimes L)[(\ddot{X}\otimes L)'(I_N\otimes\Sigma)(\ddot{X}\otimes L)]^{-1}(\ddot{X}\otimes L)'X \ .$$

Using Lemma 1 and the fact that $L = Q_T L$, we can show that the right-hand side equals

$$(\bar{X}'P(\ddot{X})\otimes L')X = (\bar{X}'P(\ddot{X})\otimes L'Q_T)X = (\bar{X}\otimes L)'(P(\ddot{X})\otimes Q_T)X = (\bar{X}\otimes L)'X \ . \quad ■$$

PROOF OF THEOREM 4.3: To prove $\hat{\beta}_{3SLS} = \hat{\beta}_{DF}$, it is sufficient to show

$$(\ddot{X}\otimes L)[(\ddot{X}\otimes L)'(I_N\otimes\Sigma)(\ddot{X}\otimes L)]^{-1}(\ddot{X}\otimes L)'X = (I_N\otimes L)(I_N\otimes L'\Sigma L)^{-1}(I_N\otimes L)'X \ .$$

But, the left-hand side equals

$$[P(\ddot{X})\otimes L(L'\Sigma L)^{-1}L']X = [I_N\otimes L(L'\Sigma L)^{-1}L'][P(\ddot{X})\otimes Q_T]X = [I_N\otimes L(L'\Sigma L)^{-1}L']X \ ,$$

where the second equality follows from Lemma 1. To prove $\hat{\beta}_{DF} = \hat{\beta}_{KR} = \hat{\beta}_{DM}$, we need to show that $L(L'\Sigma L)^{-1}L' = Q_T(Q_T\Sigma Q_T)^- Q_T = Q_T^{d\prime}(Q_T^d\Sigma Q_T^{d\prime})^{-1}Q_T^d$, where Q_T^d denotes any $T\times(T-1)$ matrix that equals Q_T with one row deleted. Note that $Q_T(Q_T\Sigma Q_T)^- Q_T$ is invariant for any choice of g-inverse since $\text{Rank}(Q_T\Sigma Q_T) = \text{Rank}(Q_T)$ (see Rao and Mitra (1971, Lemma 2.2.6(g))). It is also easy to verify that both $L(L'\Sigma L)^{-1}L'$ and $Q_T^{d\prime}(Q_T^d\Sigma Q_T^{d\prime})^{-1}Q_T^d$ are g-inverses of $Q_T\Sigma Q_T$. Thus, we have

$$Q_T(Q_T\Sigma Q_T)^- Q_T = Q_T L(L'\Sigma L)^{-1}L'Q_T = L(L'\Sigma L)^{-1}L' .$$

Similarly, $Q_T(Q_T\Sigma Q_T)^- Q_T = Q_T^{d\prime}(Q_T^d\Sigma Q_T^{d\prime})^{-1}Q_T^d$. ■

PROOF OF THEOREM 4.5: Let $R = (X,Z\otimes e_T)$, $H = (\bar{X}_2,Z_2)$ and $G = (X_1^o,\ddot{X}_2,Z_1)$. From Theorem 2.1, it is sufficient to show that

$$(H\otimes L)'(I_N\otimes\Sigma)(G\otimes I_T)[(G\otimes I_T)'(I_N\otimes\Sigma)(G\otimes I_T)]^{-1}(G\otimes I_T)'R = (H\otimes L)'R .$$

But the left-hand side equals

$$(H\otimes L)'[P(G)\otimes I_T]R = (H\otimes L)'[P(G)\otimes Q_T]'R = (H\otimes L)'R ,$$

where the second equality results from Lemma 1 and the fact that $[P(G)\otimes Q_T](Z\otimes e_T) = 0 = (H\otimes L)'(Z\otimes e_T)$. ■

PROOF OF THEOREM 4.6: Let $R = (X,Z\otimes e_T)$ and $G = (X_1^o,\ddot{X}_2,Z_1)$. For the 3SLS estimator using the instruments $(G\otimes I_T)$,

$$(G\otimes I_T)[(G\otimes I_T)'(I_N\otimes\Sigma)(G\otimes I_T)]^{-1}(G\otimes I_T)R = [P(G)\otimes\Sigma^{-1}]R .$$

Define $P_\Sigma = \Sigma^{-1}e_T(e_T'\Sigma^{-1}e_T)^{-1}e_T'\Sigma^{-1}$. Then, for the GIV estimator using the instruments $(QX,G\otimes e_T)$, a tedious but straightforward calculation shows that

$$(I_N\otimes\Sigma^{-1})(QX,G\otimes e_T)[(QX,G\otimes e_T)'(I_N\otimes\Sigma^{-1})(QX,G\otimes e_T)]^{-1}$$
$$\times (QX,G\otimes e_T)'(I_N\otimes\Sigma^{-1})R = [MQXD^{-1}XQM + P(G)\otimes P_\Sigma]R ,$$

where $M = (I_N\otimes\Sigma^{-1}) - (P(G)\otimes P_\Sigma)$ and $D = X'QMQX$ (for more details, see Im (1994)). Thus, we can complete the proof by showing $[MQXD^{-1}XQM + P(G)\otimes P_\Sigma]R = [P(G)\otimes\Sigma^{-1}]R$.

Lemma 1 implies

$$MQX = [(I_N \otimes \Sigma^{-1})-(P(G)\otimes P_\Sigma)]QX = [P(G)\otimes(\Sigma^{-1}-P_\Sigma)]QX \ . \tag{A.1}$$

Using the fact that $P_\Sigma P_T = \Sigma^{-1}P_T$, we can also show

$$MQX = [P(G)\otimes\{(Q_T+P_T)(\Sigma^{-1}-P_\Sigma)\}]QX = QMQX \ . \tag{A.2}$$

Finally, using (A.1), (A.2) and the equality $P_\Sigma P_T = \Sigma^{-1}P_T$, we can show

$$\begin{aligned}
[MQXD^{-1}XQM + P(G)\otimes P_\Sigma]R &= [MQXD^{-1}X'QMQ + (P(G)\otimes P_\Sigma)](X,Z\otimes e_T) \\
&= [MQX+(P(G)\otimes P_\Sigma)X,\ (P(G)\otimes P_\Sigma)(Z\otimes e_T)] \\
&= [MQX+(P(G)\otimes P_\Sigma)QX+(P(G)\otimes P_\Sigma)PX,\ (P(G)\otimes\Sigma^{-1})(Z\otimes e_T)] \\
&= [(P(G)\otimes\Sigma^{-1})QX+(P(G)\otimes\Sigma^{-1})PX,\ \ (P(G)\otimes\Sigma^{-1})(Z\otimes e_T)] = [P(G)\otimes\Sigma^{-1}]R \ . \qquad \blacksquare
\end{aligned}$$

The following lemma can be used to prove the next two theorems.

LEMMA 2: Let $R_\Sigma = I_T - e_T(e_T'\Sigma^{-1}e_T)^{-1}e_T'\Sigma^{-1}$. Then, $L(L'\Sigma L)^{-1}L' = \Sigma^{-1}R_\Sigma$.

PROOF: Note that $I_T = P(\Sigma^{½}L,\Sigma^{-½}e_T) = P(\Sigma^{½}L) + P(\Sigma^{-½}e_T)$, since $(\Sigma^{½}L,\Sigma^{-½}e_T)$ is nonsingular and $L'e_T = 0$. Thus,

$$L(L'\Sigma L)^{-1}L' = \Sigma^{-½}P(\Sigma^{½}L)\Sigma^{-½} = \Sigma^{-½}[I_T-P(\Sigma^{-½}e_T)]\Sigma^{-½} = \Sigma^{-1}R_\Sigma \ . \qquad \blacksquare$$

PROOF OF THEOREM 4.7: Let $b_{i1} = (x^o_{i1},z_{i1})$, $b_{i2} = (x^o_{i2},z_{i2})$ and $b_i = (b_{i1},b_{i2})$. We wish to show that 3SLS using $[\Sigma^{-1}R_\Sigma X_i,\Sigma^{-1}(e_T\otimes b_{i1})]$ is equivalent to 3SLS using $W_i = [L\otimes b_i,e_T\otimes b_{i1}]$. Arellano and Bover (1995, p. 38) show that 3SLS using $[L\otimes b_i,\Sigma^{-1}(e_T\otimes b_{i1})]$ is equivalent to 3SLS using W_i. Thus, we can complete our proof by showing that the two sets of instruments $[L\otimes b_i,\Sigma^{-1}(e_T\otimes b_{i1})]$ and $[\Sigma^{-1}R_\Sigma X_i,\Sigma^{-1}(e_T\otimes b_{i1})]$ lead to the same 3SLS estimator. Define $W_{i1} = L\otimes b_i$, $W_{i2} = \Sigma^{-1}(e_T\otimes b_{i1})$ and $W_{i3} = \Sigma^{-1}R_\Sigma X_i$; and let $W^A_i = (W_{i1},W_{i2})$ and $W^B_i = (W_{i3},W_{i2})$. Then, what we wish to show is

$$W^A(W^{A\prime}\Omega W^A)^{-1}W^{A\prime}(X,Z\otimes e_T) = W^B(W^{B\prime}\Omega W^B)^{-1}W^{B\prime}(X,Z\otimes e_T) \ . \tag{A.3}$$

This condition can be further simplied. Since $W_{i1}'\Sigma W_{i2} = 0$ and $W_{i3}'\Sigma W_{i2} = 0$, for any i,

$$W^A(W^{A\prime}\Omega W^A)^{-1}W^{A\prime} - W^B(W^{B\prime}\Omega W^B)^{-1}W^{B\prime} = W_1(W_1'\Omega W_1)^{-1}W_1' - W_3(W_3'\Omega W_3)^{-1}W_3' .$$

Further, $W_1'(Z\otimes e_T) = 0$ and $W_3'(Z\otimes e_T) = 0$. Thus, the condition (A.3) is satisfied if

$$W_1(W_1'\Omega W_1)^{-1}W_1'X = W_3(W_3'\Omega W_3)^{-1}W_3'X .$$

But $W_3(W_3'\Omega W_3)^{-1}W_3'X = (I_N\otimes\Sigma^{-1}R_\Sigma)X$. Further, since $W_1 = B\otimes L$, we have

$$W_1(W_1'\Omega W_1)^{-1}W_1'X = [P(B)\otimes L(L'\Sigma L)^{-1}L']X = [I_N\otimes L(L'\Sigma L)^{-1}L'Q_T]X$$

$$= [I_N\otimes L(L'\Sigma L)^{-1}L']X = (I_N \otimes\Sigma^{-1}R_\Sigma)X ,$$

where the second and fourth equalities come from Lemmas 1 and 2, respectively. ■

PROOF OF THEOREM 4.8: We will complete the proof by showing (i) that the GIV estimator using $W_i^{(8)}$ is not affected if we replace X_{i1} by $R_\Sigma X_{i1}$, and (ii) that $Q_T + e_Tf'Q_T^* = R_\Sigma$. Since x_{i1}^o includes all the variables in X_{i1}, we can easily see that there exists a conformable matrix C such that $(e_T'\Sigma^{-1}e_T)^{-1}e_T'\Sigma^{-1}X_{i1} = x_{i1}^oC$. Therefore, we have

$$X_{i1} = X_{i1} - e_T(e_T'\Sigma^{-1}e_T)^{-1}e_T'\Sigma^{-1}X_{i1} + e_T(e_T'\Sigma^{-1}e_T)^{-1}e_T'\Sigma^{-1}X_{i1} = R_\Sigma X_{i1} - (e_T\otimes x_{i1}^o)C .$$

This implies that, since $W_i^{(8)}$ includes $e_T\otimes x_{i1}^o$, X_{i1} and $R_\Sigma X_{i1}$ can replace each other in GIV using $W_i^{(8)}$. Thus, we have completed the proof of (i). We now proceed to prove (ii). By definition, $f = (1/T)[E(Q_T^*u_iu_i'Q_T^{*\prime})]^{-1}E(Q_T^*u_iu_i'e_T) = (1/T)(Q_T^*\Sigma Q_T^{*\prime})^{-1}Q_T^*\Sigma e_T$. Thus,

$$Q_T + e_Tf'Q_T^* = Q_T + (1/T)e_Te_T'\Sigma Q_T^{*\prime}(Q_T^*\Sigma Q_T^{*\prime})^{-1}Q_T^* = Q_T + (1/T)e_Te_T'\Sigma\Sigma^{-1}R_\Sigma = R_\Sigma ,$$

where the second equality results from the proof of Theorem 4.3 and Lemma 2. ■

REFERENCES

Ahn, S.C. and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, Journal of Econometrics 68, 5-27.

Altonji, J.G. and L.M. Segal, 1996, Small-sample bias in GMM estimation of covariance structures, Journal of Business and Economic Statistics 14, 353-366.

Amemiya, T., 1977, The maximum likelihood and the nonlinear three stage least squares estimator in the general nonlinear simultaneous equation model, Econometrica 45, 955-968.

Amemiya, T., 1985, Advanced Econometrics (Harvard University Press, Cambridge,MA).

Amemiya, T. and T.E. MaCurdy, 1986, Instrumental-variables estimation of an error-components model, Econometrica 54, 869-880.

Andersen, T.G. and R.E. Sørensen, 1996, GMM estimation of a stochastic volatility model: A Monte Carlo Study, Journal of Business & Economic Statistics 14, 328-352.

Arellano, M. and O. Bover, 1995, Another look at the instrumental variables estimation of error-component models, Journal of Econometrics 68, 29-51.

Bowden, R.J. and D.A. Turkington, 1984, Instrumental variables (Cambridge University Press, NY).

Breusch, T.S., G.E. Mizon and P. Schmidt, 1989, Efficient estimation using panel data, Econometrica 57, 695-700.

Chamberlain, G.C., 1987, Asymptotic efficiency in estimation with conditional moment restrictions, Journal of Econometrics 34, 305-334.

Chamberlain, G., 1992, Efficiency bounds for semiparametric regression, Econometrica 60, 567-596.

Hausman, J.A. and W.E. Taylor, 1981, Panel data and unobservable individual effects, Econometrica 49, 1377-1398.

Hausman, J.A., W.K. Newey and W.E. Taylor, 1987, Efficient estimation and identification of simultaneous equation models with covariance restrictions, Econometrica 55, 849-874.

Im, K.S., 1994, Estimation with panel data, unpublished Ph.D. dissertation, Michigan State University.

Im, K.S., S.C. Ahn, P. Schmidt and J.M. Wooldridge, 1996, Efficient estimation of panel data models with strictly exogenous explanatory variables, Econometrics and economic theory working paper 9600, Michigan State University.

Keane, M.P. and D.E. Runkle, 1992, On the estimation of panel data models with serial correlation when instruments are not strictly exogenous, Journal of Business and Economic Statistics 10, 1-10.

Kiefer, N.M., 1980, Estimation of fixed effect models for time series of cross-sections with arbitrary intertemporal covariance, Journal of Econometrics 14, 195-202.

Rao, C.R. and S.K. Mitra, 1971, Generalized inverse of matrices and its applications (Wiley, New York, NY).

Schmidt, P., 1990, Three stage least squares with different instruments in different equations, Journal of Econometrics 43, 389-394.

Schmidt, P., S.C. Ahn and D. Wyhowski, 1992, Comment, Journal of Business and Economic Statistics 10, 10-14.

Tauchen, G., 1986, Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data, Journal of Business and Economic Statistics 4, 397-416.

White, H., 1984, Asymptotic theory for econometricians (Academic Press, San Diego, CA).

Wooldridge, J.M., 1996, Estimating systems of equations with different instruments for different equations, Journal of Econometrics 74, 387-405.