Economics Bulletin

Volume 36, Issue 3

The optimality of non-optimal GMM estimation of parameters of interest and the partial asymptotic efficiency of 2SLS estimation

Heather L. Bednarek
Saint Louis University

Hailong Qian
Saint Louis University

Abstract

In this paper, we first derive a necessary and sufficient condition for generalized method of moments (GMM) estimation of a subset of parameters using a non-optimal weighting matrix to be asymptotically as efficient as the optimal GMM estimation. We then apply our result to simultaneous equations models and derive a necessary and sufficient condition for 2SLS estimation of a subset of regression coefficients to be asymptotically as efficient as the 3SLS estimation applied to the whole system. Our condition for the partial asymptotic efficiency of 2SLS estimation encompasses many existing results for the numerical equality of 2SLS and 3SLS estimation of all regression coefficients.

The authors are very grateful to the Editor, an Associate Editor (Professor Helle Bunzel) and two anonymous referees for their very constructive comments and suggestions that significantly improved the quality of the paper. However we are solely responsible for the remaining deficiency of the paper.

Citation: Heather L. Bednarek and Hailong Qian, (2016) "The optimality of non-optimal GMM estimation of parameters of interest and the partial asymptotic efficiency of 2SLS estimation", *Economics Bulletin*, Volume 36, Issue 3, pages 1636-1649
Contact: Heather L. Bednarek - bednarhl@slu.edu, Hailong Qian - qianh@slu.edu.
Submitted: November 12, 2015. **Published:** August 24, 2016.

1. Introduction

There is a large literature on the numerical equality of ordinary least squares (OLS) and generalized least squares (GLS) estimation of regression coefficients in single or system of linear regression models. For single linear regression models, examples include Zyskind (1967), Kruskal (1968), Rao (1968), Anderson (1971), Rao and Mitra (1971), Gourieroux and Monfort (1980) and Amemiya (1985). Zyskind (1967), Haberman (1975), Gourieroux and Monfort (1980), Rao and Mitra (1971) and Qian and Schmidt (2003) also provide conditions for the partial numerical equality of OLS and GLS estimation of a subset of regression coefficients.

For systems of seemingly unrelated regressions (SUR), Zellner (1962) shows that when the covariance matrix of the system is diagonal or each equation in the system includes the same explanatory variables, GLS and OLS estimation of regression coefficients are numerically identical. Motivated by Zellner's sufficient conditions, many subsequent papers derive different (but equivalent) forms of necessary and sufficient conditions for the numerical equality of OLS and GLS estimators of SUR models; see for examples, Dwivedi and Srivastava (1978), Gourieroux and Monfort (1980), Srivastava and Giles (1987), Baltagi (1988), Baksalary and Trenkler (1989) and Bartels and Fiebig (1991). Revankar (1974), Schmidt (1978) and Gourieroux and Monfort (1980) also provide conditions for the numerical equality of OLS and GLS estimation of a subset of regression coefficients under various conditions, while Qian and Bednarek (2015) using a generalized method of moments (GMM) estimation setup derive a necessary and sufficient condition for the partial asymptotic efficiency of OLS estimation of regression coefficients in a subset of equations.

Parallel to the SUR literature, there also exists a large number of published papers examining the numerical equality of two-stage least squares (2SLS) and three-stage least squares (3SLS) estimators of regression coefficients in simultaneous equations models (SEMs); for examples, Zellner and Theil (1962), Srivastava and Tiwari (1978), Gourieroux and Monfort (1980), Kapteyn and Fiebig (1981) and Baltagi (1988-89). However, surprisingly, it appears that no paper has so far considered the asymptotic equivalence of the 2SLS and 3SLS estimators of a subset of regression coefficients, though Zellner and Theil (1962), Gourieroux and Monfort (1980, Section 3E) and Schmidt (1976, Theorem 5.2.13, p.213) also provide various conditions for the numerical equality of 2SLS and 3SLS estimators of the regression coefficients in over-identified structural equations. This paper seeks to fill this gap. More precisely, this paper aims to extend the partial optimality result for OLS estimation of SUR models in Qian and Bednarek (2015) to 2SLS estimation of SEMs. That is, we will derive a necessary and sufficient condition for the equation by equation 2SLS estimation of a subset of parameters to be asymptotically as efficient as the 3SLS estimation applied to the whole system. We accomplish this goal by first casting 2SLS and 3SLS estimation as GMM estimation and deriving a necessary and sufficient condition for the asymptotic equivalence between a GMM estimation using a non-optimal weighting matrix and the optimal GMM estimation of a subset of parameters.

It is worth pointing out the main difference between the current paper and the existing literature on the comparison of 2SLS and 3SLS estimation of SEMs: our current paper is concerned with the partial asymptotic efficiency of 2SLS estimation of a subset of regression coefficients, while almost all of the published papers cited in the previous paragraph are mainly focused on the question of when 2SLS and 3SLS estimation of all regression coefficients are numerical identical, though Zellner and Theil (1962), Schmidt (1976, Theorem 5.2.13, p. 213) and Gourieroux and Monfort (1980, Section 3E) also consider the numerical equality of 2SLS and 3SLS estimation of the regression coefficients in over-identified structural equations. As a result, our necessary and sufficient condition for the partial asymptotic efficiency of the 2SLS estimator of a subset of coefficients (Theorem 2 in the next section) generalizes existing sufficient conditions for the numeral equality of 2SLS and 3SLS estimators of all regression coefficients.

The rest of the paper is organized as follows. In Section 2, we derive a necessary and sufficient condition for a non-optimal GMM estimator of a subset of parameters to be asymptotically efficient. In Section 3, we apply the necessary and sufficient condition to examine the optimality of 2SLS estimation of coefficients of interest in simultaneous equations models, while Section 4 briefly concludes the paper. The appendix provides the proofs of the main results of the paper.

2. The Optimality of Non-optimal GMM Estimation for Parameters of Interest

In this section we will derive a necessary and sufficient condition for GMM estimation of parameters of interest using a non-optimal weighting matrix (hereafter referred to as a non-optimal GMM estimation) to be asymptotically as efficient as the GMM estimation using the optimal weighting matrix.

Suppose that we have a set of moment conditions:

$$E[g(v_t,\theta_0)] = 0, \tag{1}$$

where $g(v_t, \theta)$ is a $q \times 1$ vector of moment functions, v_t is an $L \times 1$ vector of observable variables, and θ_0 is a $p \times 1$ vector of (true) parameters to be estimated. We assume that $\Omega \equiv var[g(v_t,\theta_0)]$ is positive definite (p.d.), and that $D \equiv E[\partial g(v_t,\theta_0)/\partial\theta']$ has full column rank such that the moment conditions in (1) locally identify the unknown parameter vector θ_0. For simplicity of derivation, we assume that the sample { v_1, ..., v_T } is i.i.d.

Under standard assumptions (see, for examples, Hansen 1982; Wooldridge 2010, Chapter 14), the GMM estimator $\hat{\theta}(W_T)$ of θ_0, based on moment conditions (1), using a symmetric and positive semidefinite (p.s.d.) weighting matrix W_T is consistent and asymptotically normal, $\sqrt{T}[\hat{\theta}(W_T)-\theta_0] \rightarrow N[0, V(W)]$, with $V(W) \equiv (D'WD)^{-1}D'W\Omega WD(D'WD)^{-1}$, where W (a symmetric and p.d. matrix) is the probability limit of W_T. Hansen (1982) shows that when the weighting matrix used equals the inverse of $\hat{\Omega}$ (a consistent estimate of Ω), the resulting GMM estimator $\tilde{\theta}$ achieves its (first-order) asymptotic efficiency, with its asymptotic variance matrix equal to $V^* \equiv V(\Omega^{-1}) = (D'\Omega^{-1}D)^{-1}$. ($W_T = \hat{\Omega}^{-1}$ is usually referred to as the optimal weighting matrix and the corresponding GMM estimator $\tilde{\theta}$ is called the optimal GMM estimator of θ_0, based on moment conditions (1)). Thus, $V(W) - V^*$ is always positive semidefinite. Now an interesting question is when $V(W) = V^*$; that is, under what circumstances is $\hat{\theta}(W_T)$ asymptotically as efficient as $\tilde{\theta}$? To answer this question, we first define $G \equiv W^{1/2}D$, $\Phi \equiv W^{1/2}\Omega W^{1/2}$, where $W^{1/2}$ is a symmetric and p.d matrix satisfying $W = W^{1/2}W^{1/2}$. We also define $M_{[Q]} \equiv I - Q(Q'Q)^{-1}Q'$, for any full column rank matrix Q. Then, we have Lemma 1, as follows.

Lemma 1. The GMM estimator $\hat{\theta}(W_T)$ of θ_0, based on moment conditions (1), using weighting matrix W_T is asymptotically as efficient as the optimal GMM estimator $\tilde{\theta}$, if and only if: (A) $\Phi G = GC$, with C ($p \times p$) non-singular; or equivalently, (B) $M_{[G]}\Phi G = 0$.
Proof: See Wooldridge (2010, p. 234, Problem 8.5). ■

Now suppose that we are only interested in estimating a subset of parameters. For example, in systems of linear regression models, we sometimes are only interested in estimating regression coefficients of a subset of equations; see for example, Qian and Bednarek (2015). As another example, in GMM estimation of linear panel data models with time-varying individual effects (see for example Ahn, et al. 2001), we are usually interested in the efficient estimation of regression coefficients but not the parameters in the second moments of the composite errors. As such, we wonder under what circumstances the non-optimal GMM estimation of a subset of parameters is as efficient as the optimal GMM estimation. To answer this question and without loss of generality, we now assume that we are only interested in estimating the first subset of parameters. We thus partition the parameter vector θ_0 into $(\theta_{01}', \theta_{02}')'$, with θ_{01} $p_1 \times 1$, θ_{02} $p_2 \times 1$ and $p_1 + p_2 = p$. We also partition the derivative matrix D accordingly, $D = E[\partial g(v_t, \theta_0)/\partial\theta_1', \partial g(v_t, \theta_0)/\partial\theta_2'] \equiv (D_1, D_2)$.

Using $V(W) \equiv (D'WD)^{-1}D'W\Omega WD(D'WD)^{-1} = (G'G)^{-1}G'\Phi G(G'G)^{-1}$ and the partitioned-matrix inverse formula, we can show that the asymptotic variance of the GMM estimator $\hat{\theta}_1(W_T)$ of θ_{01}, based on (1), using W_T as the weighting matrix equals:

$$V_1(W) \equiv AV[\sqrt{T}(\hat{\theta}_1(W_T) - \theta_{01})] = (G_1'M_{[G_2]}G_1)^{-1}G_1'M_{[G_2]}\Phi M_{[G_2]}G_1(G_1'M_{[G_2]}G_1)^{-1}, \quad (2)$$

where $G_1 \equiv W^{1/2}D_1$ and $G_2 \equiv W^{1/2}D_2$. To improve the flow of the text, we move the derivation of (2) to the appendix. When $W_T = \hat{\Omega}^{-1}$ and $W = \Omega^{-1}$, $V_1(\Omega^{-1})$ becomes the asymptotic variance of the optimal GMM estimator $\tilde{\theta}_1$ of θ_{01},

$$V_1^* \equiv V(\Omega^{-1}) = AV[\sqrt{T}(\tilde{\theta}_1 - \theta_{01})] = (G_1^{*}'M_{[G_2^*]}G_1^*)^{-1}, \quad (3)$$

with $G_1^* \equiv \Omega^{-1/2}D_1$ and $G_2^* \equiv \Omega^{-1/2}D_2$.

Because $\tilde{\theta}_1$ is asymptotically no less efficient than $\hat{\theta}_1(W_T)$, $V_1(W) - V_1^*$ is always p.s.d. We now wonder when $V_1(W) - V_1^* = 0$. When $V_1(W) = V_1^*$, we say that the non-optimal GMM estimator $\hat{\theta}_1(W)$ of θ_{01} is partially optimal. The following theorem provides a necessary and sufficient condition for the partial optimality of the non-optimal GMM estimator $\hat{\theta}_1(W_T)$ of θ_{01}.

Theorem 1. The GMM estimator $\hat{\theta}_1(W_T)$ of θ_{01}, based on moment conditions (1), using weighting matrix W_T is asymptotically as efficient as the optimal GMM estimator $\tilde{\theta}_1$ of θ_{01}, if and only if any one of the following equivalent conditions holds:

(A) $\Phi M_{[G_2]}G_1 = GC_1$, with C_1 $(p\times p_1)$ of full column rank;

(B) $M_{[G_2]}\Phi M_{[G_2]}G_1 = M_{[G_2]}G_1C_2$, with C_2 $(p_1\times p_1)$ nonsingular;

(C) $M_{[G]}\Phi M_{[G_2]}G_1 = 0$.

Proof: See Appendix. ■

Conditions (A) and (B) can be thought of as extension of Lemma 1 (A) for the efficient estimation of θ_{01}, while condition (C) is an extension of Lemma 1 (B). Also, note that GMM estimation of θ_0, based on $E[g(v_t,\theta_0)]=0$, using weighting matrix W_T is asymptotically equivalent to GMM estimation of θ_0, based on $E[W^{1/2}g(v_t,\theta_0)]=0$, using an identity matrix as the weighting matrix. Then, we can easily see that Conditions (A) and (B) have, respectively, the same generic forms as Conditions (G) and (C*) of Qian and Schmidt (2003, Theorem 3, pp. 388-389), though their result is for the partial numerical equality of OLS and GLS estimation of linear regression models.

3. Application to Simultaneous Equations Models

There exists a large number of papers in the literature on efficient estimation of simultaneous equations models (SEMs) that investigate when the equation by equation 2SLS estimation of regression coefficients are numerically identical to the 3SLS estimation. See for examples, Zellner and Theil (1962), Schmidt (1976, pp. 211-216), Srivastava and Tiwari (1978), Gourieroux and Monfort (1980), Kapteyn and Fiebig (1981), Baltagi (1988-89) and Wooldridge (2010, Chapter 9). However, surprisingly, it appears that no paper has so far systematically examined the asymptotic equivalence of 2SLS and 3SLS estimation of a subset of regression coefficients. In this section, we seek to fill this gap of the existing literature. More precisely, we will apply Theorem 1 of the previous section to the efficient estimation of parameters of interest in SEMs. For this purpose, we consider a system of G linear simultaneous equations:

$$y_{gt} = Y_{gt}'\gamma_g + z_{gt}'\delta_g + \varepsilon_{gt},\ \ g = 1, 2, \ldots, G, \tag{4}$$

where the subscripts g and t index equations and observations, respectively, y_{gt} is the dependent variable of the g-th equation, Y_{gt} is a $G_g\times 1$ vector of endogenous explanatory variables, z_{gt} is an $M_g\times 1$ vector of exogenous (or predetermined) explanatory variables, γ_g and δ_g are the corresponding coefficient vectors, and ε_{gt} is the disturbance term. Note, z_{gt} usually contains a unity element so that the structural equations in (4) include nonzero intercepts.

Let $x_{gt}\equiv(Y_{gt}',z_{gt}')'$. Then, (4) can be rewritten as:

$$y_{gt} = x_{gt}'\beta_g + \varepsilon_{gt}, \tag{5}$$

where $\beta_g\equiv(\gamma_g',\delta_g')'$ is $K_g\times 1$, with $K_g\equiv G_g+M_g$. Stacking over equations for a given observation t, we can rewrite (5) as:

$$y_t = X_t\beta + \varepsilon_t, \tag{6}$$

where $y_t = (y_{1t},...,y_{Gt})'$, $X_t = diag(x_{1t}',\cdots,x_{Gt}')$, $\beta = (\beta_1',...,\beta_G')'$ and $\varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{Gt})'$.

Let z_t be an $M \times 1$ vector of all distinct exogenous variables appearing in (4). Now, for the purpose of estimating the unknown regression coefficients in (4) and to be consistent with the existing literature (see for example, Chapters 8 and 9 of Wooldridge 2010), we make three standard assumptions.

Assumptions: (SEM.1) $E[(I_G \otimes z_t)\varepsilon_t] = 0$, with $\otimes$ denoting the Kronecker product.
(SEM.2) $E(z_t z_t')$ is nonsingular and $E[(I_G \otimes z_t)X_t]$ has full column rank.
(SEM.3) $E[(I_G \otimes z_t)\varepsilon_t\varepsilon_t'(I_G \otimes z_t')] = E[(I_G \otimes z_t)\Sigma(I_G \otimes z_t')]$, with $\Sigma \equiv E(\varepsilon_t\varepsilon_t')$ p.d.

Assumptions (SEM.1) and (SEM.2) imply that the regression coefficients of the SEM (4) are identified by using the instrument set z_t. Assumption (SEM.3) is a system homoscedasticity assumption; Wooldridge (2010, Sections 8.3-8.4) makes similar assumptions. Also, for simplicity of derivation, we assume that $(y_t', z_t')'$ is i.i.d. over observations from 1 to T.

Given the SEM (4), Assumption (SEM.1) implies the following set of orthogonality conditions:

$$E[g_t(\beta)] \equiv E[(I_G \otimes z_t)(y_t - X_t\beta)] = E(\begin{bmatrix} z_t(y_{1t} - x_{1t}'\beta_1) \\ \vdots \\ z_t(y_{Gt} - x_{Gt}'\beta_G) \end{bmatrix}) = 0. \qquad (7)$$

Then, it is easy to verify that the equation by equation 2SLS estimator of β in (4) is algebraically the same as the GMM estimator based on (7), using $W_T = I_G \otimes (Z'Z/T)^{-1}$ as the weighting matrix, where $Z \equiv (z_1,\cdots,z_T)'$, while the 3SLS estimator of β is also algebraically identical to the GMM estimator based on (7), using $W_T = \hat{\Sigma}^{-1} \otimes (Z'Z/T)^{-1}$ as the optimal weighting matrix, where we assumed that both the 3SLS and the GMM estimators use the same initial consistent estimate, $\hat{\Sigma}$, of the covariance matrix Σ in (4). See for example, Wooldridge (2010, Chapter 8).

Now, without loss of generality, suppose that we are only interested in estimating the regression coefficients of the first m equations in (4), with $1 \le m \le G$. Then, adopting the notation of the previous section, we have $\theta_{01} \equiv (\beta_1',\cdots,\beta_m')'$, $\theta_{02} \equiv (\beta_{m+1}',\cdots,\beta_G')'$, and $\theta_0 = (\theta_{01}',\theta_{02}')' \equiv \beta$. Now, applying Theorem 1 (C) to moment conditions (7), we obtain Theorem 2.

Theorem 2. Under Assumptions (SEM.1)-(SEM.3), the equation by equation 2SLS estimation of $\theta_{01} \equiv (\beta_1',\cdots,\beta_m')'$ in (4) is asymptotically as efficient as the 3SLS estimation applied to the whole system of (4), if and only if,

$$\sigma_{ij}M_{[A_i]}A_j = 0, \text{ for } i \neq j;\ i = 1, 2, \ldots, G;\ j = 1, 2, \ldots, m; \qquad (8)$$

where, $1 \le m \le G$, $A_i \equiv E(z_t x_{it}')$, $M_{[A_i]} = I_M - A_i(A_i'A_i)^{-1}A_i'$, and $\sigma_{ij} = cov(\varepsilon_{it}, \varepsilon_{jt})$ is the (i, j) element of the covariance matrix Σ.
Proof: See Appendix. ■

This theorem is a natural extension of the necessary and sufficient condition for the partial efficiency of OLS estimation of SUR models (Qian and Bednarek 2015, Theorem 2) to the partial asymptotic efficiency of 2SLS estimation of simultaneous equations models. Condition (8) is of course satisfied when the disturbances of the system are uncorrelated with each other or each equation in the system is just-identified (in this case, $A_i \equiv E(z_t x_{it}')$ becomes a nonsingular square matrix). In fact, given condition (8), we obtain Corollary 1.

Corollary 1. Under Assumptions (SEM.1)-(SEM.3), the equation by equation 2SLS estimation of $\theta_{01} \equiv (\beta_1', \cdots, \beta_m')'$ in (4) is asymptotically as efficient as the 3SLS estimation applied to the whole system of (4), if any one of the following sufficient conditions is satisfied:
(A) The disturbances of system are uncorrelated with each other;
(B) Each equation in the system is just-identified;
(C) The disturbances of the first m equations are uncorrelated with each other and furthermore are uncorrelated with the disturbances of the last (G-m) equations in the system;
(D) The disturbances of the first m equations are uncorrelated with each other, and the last (G-m) equations in the system are just-identified;
(E) Each of the first m equations in the system is just-identified, and the disturbances of the first m equations are uncorrelated with the disturbances of the remaining equations in the system.
Proof: It is easy to verify that any one of the five conditions listed here is sufficient for condition (8) to hold. ■

Conditions (A) and (B) are just the two well-known sufficient conditions for the numerical equality of 2SLS and 3SLS estimation of the whole system; see Zellner and Theil (1962). Conditions (C) appears new, though it is very intuitive. Note that Condition (C) is weaker than Condition (A), since it does not require the disturbances of the last (G-m) equations uncorrelated with each other. Condition (D) is an extension of Schmidt (1976, Theorem 5.2.13, p. 213) and Gorieroux and Monfort (1980, Section 3E), where they show that 3SLS estimation applied to the over-identified structural equations alone is numerically identical to 3SLS estimation applied to the whole system. Condition (E) also seems a new sufficient condition, though it is intuitive too.

The sufficient conditions in Corollary 1 are stated in the population. Now, to gain further insight of condition (8), we proceed to find an additional sufficient condition that is expressed in the sample. Thus, for a given random sample of size T, we define the following data matrices:

$$Z = (z_1, \cdots, z_T)', \; Y_{(i)} = (Y_{i1}, \cdots, Y_{iT})', \; Z_{(i)} = (z_{i1}, \cdots, z_{iT})', \; X_{(i)} = (Y_{(i)}, Z_{(i)}),$$

for i = 1, 2, …, G. We also define $\hat{X}_{(i)} = (\hat{Y}_{(i)}, Z_{(i)})$, with $\hat{Y}_{(i)} \equiv P_{[Z]} Y_{(i)}$ denoting the fitted values from the first-stage regression for the i-th structural equation in (4). Then, using the sample analogues of $M_{[A_i]}$ and A_j in condition (8), we obtain Corollary 2.

Corollary 2. Under Assumptions (SEM.1)-(SEM.3), the equation by equation 2SLS estimation of $\theta_{01} \equiv (\beta_1', \cdots, \beta_m')'$ in (4) is asymptotically as efficient as the 3SLS estimation applied to the whole system of (4), if the following sufficient condition holds in the sample:

$$\sigma_{ij} M_{[\hat{X}_{(i)}]} \hat{X}_{(j)} = 0, \text{ for } i \neq j; \; i = 1, 2, \ldots, G; \; j = 1, 2, \ldots, m, \tag{9}$$

with probability equal to 1.
Proof: See Appendix. ■

If we compare Condition (9) with the necessary and sufficient condition of Baltagi (1988-89, Eq. (11), p.167) for the numerical equality of the 2SLS and 3SLS estimators of all regression coefficients in (4) (that is, in the notation of the current paper, $\sigma^{ij} M_{[\hat{X}_{(i)}]} \hat{X}_{(j)} = 0$, for i, j = 1, 2, …, G, where σ^{ij} is the (i, j) element of Σ^{-1}), we can easily see, as expected, that his condition is sufficient but not necessary for Condition (9). Here we also note that the necessary and sufficient condition for the numerical equality of 2SLS and 3SLS estimators given in Kapteyn and Fiebig (1981, Proposition, p. 57) is actually equivalent to Baltagi's (1988-89) condition. Thus, Corollary 2 somewhat extends the necessary and sufficient conditions of Baltagi (1988-89) and Kapteyn and Fiebig (1981) for full numerical equality of 2SLS and 3SLS estimators of SEMs to the partial asymptotic equivalence of 2SLS and 3SLS estimators of a subset of coefficients.

4. Conclusions

We made three new contributions in this paper. First, we derived a necessary and sufficient condition for a non-optimal GMM estimator of a subset of parameters to be asymptotically efficient. This result appears new in the GMM estimation literature and is of course applicable to both linear and nonlinear models. Secondly, we extended the current literature on the numerical equality of 2SLS and 3SLS estimation of all regression coefficients in linear simultaneous equations models to the asymptotic equivalence of 2SLS and 3SLS estimation of a subset of regression coefficients. Thirdly, we provided several new and easily checkable sufficient conditions for the partial asymptotic efficiency of 2SLS estimation.

Appendix

Derivation of Equation (2).

First, using the definition of $G \equiv W^{1/2}D$ and $\Phi \equiv W^{1/2}\Omega W^{1/2}$, we can rewrite $V(W) \equiv (D'WD)^{-1}D'W\Omega WD(D'WD)^{-1}$ as $V(W) \equiv (G'G)^{-1}G'\Phi G(G'G)^{-1}$. Then, using $G = (G_1, G_2)$ (with $G_1 \equiv W^{1/2}D_1$ and $G_2 \equiv W^{1/2}D_2$) and the partitioned matrix inverse formula, we have:

$$
\begin{aligned}
&V(W)\\
&= \begin{bmatrix} G_1'G_1 & G_1'G_2 \\ G_2'G_1 & G_2'G_2 \end{bmatrix}^{-1} \begin{bmatrix} G_1'\Phi G_1 & G_1'\Phi G_2 \\ G_2'\Phi G_1 & G_2'\Phi G_2 \end{bmatrix} \begin{bmatrix} G_1'G_1 & G_1'G_2 \\ G_2'G_1 & G_2'G_2 \end{bmatrix}^{-1}\\
&= \begin{bmatrix} E^{-1} & -E^{-1}BF^{-1} \\ -F^{-1}CE^{-1} & F^{-1}+F^{-1}CE^{-1}BF^{-1} \end{bmatrix} \begin{bmatrix} G_1'\Phi G_1 & G_1'\Phi G_2 \\ G_2'\Phi G_1 & G_2'\Phi G_2 \end{bmatrix} \begin{bmatrix} E^{-1} & -E^{-1}BF^{-1} \\ -F^{-1}CE^{-1} & F^{-1}+F^{-1}CE^{-1}BF^{-1} \end{bmatrix},
\end{aligned}
$$

where we defined $A \equiv G_1'G_1$, $B \equiv G_1'G_2$, $C \equiv G_2'G_1$, $F \equiv G_2'G_2$ and $E \equiv A - BF^{-1}C$. Then, $V_1(W)$, the asymptotic variance of $\sqrt{T}[\hat{\theta}_1(W_T) - \theta_{01}]$, is just the upper-left block of the matrix product above. Thus,

$$V_1(W) = (E^{-1}A_{11} - E^{-1}BF^{-1}A_{21})E^{-1} - (E^{-1}A_{12} - E^{-1}BF^{-1}A_{22})F^{-1}CE^{-1}, \qquad \text{(A.1)}$$

where $A_{11} \equiv G_1'\Phi G_1$, $A_{12} \equiv G_1'\Phi G_2$, $A_{21} \equiv G_2'\Phi G_1$ and $A_{22} \equiv G_2'\Phi G_2$. We now proceed to simplify it. In fact, the right hand side of (A.1) can be factorized as,

$$V_1(W) = E^{-1}[(A_{11} - A_{12}F^{-1}C) - BF^{-1}(A_{21} - A_{22}F^{-1}C)]E^{-1}, \qquad \text{(A.2)}$$

where

$$E = A - BF^{-1}C = G_1'G_1 - G_1'G_2(G_2'G_2)^{-1}G_2'G_1 = G_1'M_{[G_2]}G_1, \qquad \text{(A.3)}$$

$$A_{11} - A_{12}F^{-1}C = G_1'\Phi G_1 - G_1'\Phi G_2(G_2'G_2)^{-1}G_2'G_1 = G_1'\Phi M_{[G_2]}G_1, \qquad \text{(A.4)}$$

$$
\begin{aligned}
BF^{-1}(A_{21} - A_{22}F^{-1}C) &= G_1'G_2(G_2'G_2)^{-1}[G_2'\Phi G_1 - G_2'\Phi G_2(G_2'G_2)^{-1}G_2'G_1]\\
&= G_1'G_2(G_2'G_2)^{-1}G_2'\Phi[I - G_2(G_2'G_2)^{-1}G_2']G_1\\
&= G_1'P_{[G_2]}\Phi M_{[G_2]}G_1. \qquad \text{(A.5)}
\end{aligned}
$$

Now, substituting (A.3)-(A.5) into (A.2), we have:

$$
\begin{aligned}
V_1(W) &= (G_1'M_{[G_2]}G_1)^{-1}[G_1'\Phi M_{[G_2]}G_1 - G_1'P_{[G_2]}\Phi M_{[G_2]}G_1](G_1'M_{[G_2]}G_1)^{-1}\\
&= (G_1'M_{[G_2]}G_1)^{-1}G_1'M_{[G_2]}\Phi M_{[G_2]}G_1(G_1'M_{[G_2]}G_1)^{-1}.
\end{aligned}
$$

This is just the expression of equation (2) in the text.

Proof of Theorem 1.

$V_1(W) - V_1^*$ being p.s.d. is equivalent to $V_1^{*-1} - V_1(W)^{-1}$ being p.s.d. Then, using (2) and (3) of Section 2, $G_1^* \equiv \Omega^{-1/2}D_1$, $G_2^* \equiv \Omega^{-1/2}D_2$, $G_1 \equiv W^{1/2}D_1$, $G_2 \equiv W^{1/2}D_2$ and $\Phi \equiv W^{1/2}\Omega W^{1/2}$, we have:

$$
\begin{aligned}
&V_1^{*-1} - V_1(W)^{-1}\\
&= G_1^{*}{}'M_{[G_2^*]}G_1^* - G_1{}'M_{[G_2]}G_1(G_1{}'M_{[G_2]}\Phi M_{[G_2]}G_1)^{-1}G_1{}'M_{[G_2]}G_1\\
&= D_1{}'\Omega^{-1/2}M_{[\Omega^{-1/2}D_2]}\Omega^{-1/2}D_1\\
&\quad - D_1{}'W^{1/2}M_{[G_2]}G_1(G_1{}'M_{[G_2]}W^{1/2}\Omega W^{1/2}M_{[G_2]}G_1)^{-1}G_1{}'M_{[G_2]}W^{1/2}D_1\\
&= D_1{}'\Omega^{-1/2}\{M_{[\Omega^{-1/2}D_2]}\\
&\quad - \Omega^{1/2}W^{1/2}M_{[G_2]}G_1(G_1{}'M_{[G_2]}W^{1/2}\Omega W^{1/2}M_{[G_2]}G_1)^{-1}G_1{}'M_{[G_2]}W^{1/2}\Omega^{1/2}\}\Omega^{-1/2}D_1\\
&= D_1{}'\Omega^{-1/2}\{M_{[\Omega^{-1/2}D_2]} - P_{[\Omega^{1/2}W^{1/2}M_{[G_2]}G_1]}\}\Omega^{-1/2}D_1 \quad (\text{using } P_{[X]} = X(X'X)^{-1}X')\\
&= D_1{}'\Omega^{-1/2}\{I - P_{[\Omega^{-1/2}D_2]} - P_{[\Omega^{1/2}W^{1/2}M_{[G_2]}G_1]}\}\Omega^{-1/2}D_1\\
&= D_1{}'\Omega^{-1/2}\{I - P_{[\Omega^{-1/2}D_2,\, \Omega^{1/2}W^{1/2}M_{[G_2]}G_1]}\}\Omega^{-1/2}D_1\\
&\qquad (\text{using } G_2 \equiv W^{1/2}D_2 \text{ and } (\Omega^{-1/2}D_2)'(\Omega^{1/2}W^{1/2}M_{[G_2]}G_1) = 0)\\
&= D_1{}'\Omega^{-1/2}M_{[\Omega^{-1/2}D_2,\, \Omega^{1/2}W^{1/2}M_{[G_2]}G_1]}\Omega^{-1/2}D_1.
\end{aligned}
$$

Thus, $V_1(W) - V_1^*$ (or equivalently, $V_1^{*-1} - V_1(W)^{-1}$) is p.s.d. and equals 0, if and only if $M_{[\Omega^{-1/2}D_2,\, \Omega^{1/2}W^{1/2}M_{[G_2]}G_1]}\Omega^{-1/2}D_1 = 0$. This is equivalent to:

$$\Omega^{-1/2}D_1 = P_{[\Omega^{-1/2}D_2,\, \Omega^{1/2}W^{1/2}M_{[G_2]}G_1]}\Omega^{-1/2}D_1, \text{ or}$$
$$\Omega^{-1/2}D_1 = (P_{[\Omega^{-1/2}D_2]} + P_{[\Omega^{1/2}W^{1/2}M_{[G_2]}G_1]})\Omega^{-1/2}D_1,$$

using $(\Omega^{-1/2}D_2)'(\Omega^{1/2}W^{1/2}M_{[G_2]}G_1) = 0$. This implies:

$$
\begin{aligned}
\Omega^{-1/2}D_1 &= \Omega^{-1/2}D_2(D_2{}'\Omega^{-1}D_2)^{-1}D_2{}'\Omega^{-1}D_1\\
&\quad + \Omega^{1/2}W^{1/2}M_{[G_2]}G_1(G_1{}'M_{[G_2]}W^{1/2}\Omega W^{1/2}M_{[G_2]}G_1)^{-1}G_1{}'M_{[G_2]}W^{1/2}D_1,
\end{aligned}
$$

or
$$
\begin{aligned}
D_1 &= D_2(D_2{}'\Omega^{-1}D_2)^{-1}D_2{}'\Omega^{-1}D_1\\
&\quad + \Omega W^{1/2}M_{[G_2]}G_1(G_1{}'M_{[G_2]}W^{1/2}\Omega W^{1/2}M_{[G_2]}G_1)^{-1}G_1{}'M_{[G_2]}W^{1/2}D_1.
\end{aligned}
$$

This can be rewritten as

$$D_1 = D_2(D_2{}'\Omega^{-1}D_2)^{-1}D_2{}'\Omega^{-1}D_1 + \Omega W^{1/2}M_{[G_2]}G_1C,$$

where $C \equiv (G_1'M_{[G_2]}W^{1/2}\Omega W^{1/2}M_{[G_2]}G_1)^{-1}G_1'M_{[G_2]}G_1$ is nonsingular. Now, pre-multiplying it by $W^{1/2}$ and post-multiplying it by C^{-1}, we have:

$$W^{1/2}D_1C^{-1} = W^{1/2}D_2(D_2'\Omega^{-1}D_2)^{-1}D_2'\Omega^{-1}D_1C^{-1} + W^{1/2}\Omega W^{1/2}M_{[G_2]}G_1,$$

which is equivalent to:

$$(W^{1/2}D_1, W^{1/2}D_2)C_1 = W^{1/2}\Omega W^{1/2}M_{[G_2]}G_1, \text{ or } GC_1 = \Phi M_{[G_2]}G_1,$$

using the definitions of $G \equiv (G_1, G_2)$, $G_1 \equiv W^{1/2}D_1$, $G_2 \equiv W^{1/2}D_2$ and $\Phi \equiv W^{1/2}\Omega W^{1/2}$, where $C_1 \equiv [C^{-1}{}', -((D_2'\Omega^{-1}D_2)^{-1}D_2'\Omega^{-1}D_1C^{-1})']'$ is $p \times p_1$ and has full column rank. Thus, we prove condition (A) of Theorem 1.

We now turn to showing that conditions (A) and (B) are equivalent. We first show that (A) implies (B). Using (A), we have:

$$M_{[G_2]}\Phi M_{[G_2]}G_1 = M_{[G_2]}GC_1 = M_{[G_2]}(G_1, G_2)C_1 = M_{[G_2]}(G_1C_2 + G_2C_3) = M_{[G_2]}G_1C_2,$$

where we defined $C_1 \equiv (C_2', C_3')'$. Since the left hand side of the expression above has full column rank (from (2) of Section 2), the $p_1 \times p_1$ square matrix C_2 must be non-singular. This proves that condition (A) implies condition (B). We now show that the reverse is also true. In fact, condition (B) can be rewritten as $M_{[G_2]}(\Phi M_{[G_2]}G_1 - G_1C_2) = 0$. This implies $(\Phi M_{[G_2]}G_1 - G_1C_2)$ is in the column space of G_2; that is, there exists a $p_2 \times p_1$ matrix C_4 such that $\Phi M_{[G_2]}G_1 - G_1C_2 = G_2C_4$. This is equivalent to $\Phi M_{[G_2]}G_1 = GC_1$, with $C_1 \equiv (C_2', C_4')'$. Because $\Phi M_{[G_2]}G_1$ has full column rank, the $p \times p_1$ matrix C_1 must have full column rank too. This completes the proof of (B) implying (A). Thus, we have proved the equivalence of conditions (A) and (B).

Lastly, we show that conditions (B) and (C) are equivalent. In fact, pre-multiplying both sides of condition (B) by G_1 and solving for C_2, we obtain $C_2 = (G_1'M_{[G_2]}G_1)^{-1}G_1'M_{[G_2]}\Phi M_{[G_2]}G_1$. Substituting it into condition (B), we can see that condition (B) is equivalent to:

$$M_{[G_2]}\Phi M_{[G_2]}G_1 = M_{[G_2]}G_1(G_1'M_{[G_2]}G_1)^{-1}G_1'M_{[G_2]}\Phi M_{[G_2]}G_1, \text{ or}$$
$$[M_{[G_2]} - M_{[G_2]}G_1(G_1'M_{[G_2]}G_1)^{-1}G_1'M_{[G_2]}]\Phi M_{[G_2]}G_1 = 0,$$

which can be rewritten as $[I - P_{[G_2]} - P_{[M_2G_1]}]\Phi M_{[G_2]}G_1 = 0$ (with $M_2 \equiv M_{[G_2]}$). This is equivalent to:

$$M_{[G_1, G_2]}\Phi M_{[G_2]}G_1 = 0 \text{ or } M_{[G]}\Phi M_{[G_2]}G_1 = 0.$$

This is just condition (C). Thus we complete the proof of this theorem. ■

Proof of Theorem 2.
As we explained in Section 3, the equation by equation 2SLS estimator of β in (4) is algebraically the same as the GMM estimator based on moment conditions (7), using $W_T = I_G \otimes (Z'Z/T)^{-1}$ as the weighting matrix, while the 3SLS estimator of β is also algebraically identical to the GMM estimator based on (7), using $W_T = \hat{\Sigma}^{-1} \otimes (Z'Z/T)^{-1}$ as the optimal weighting matrix. Thus, to prove Theorem 2, we only need to apply Theorem 1 to the case of GMM estimation of $\theta_{01} \equiv (\beta_1', \cdots, \beta_m')'$ based on (7). For this purpose, we first note that $\Phi \equiv W^{1/2}\Omega W^{1/2}$, $G \equiv W^{1/2}D$, $G_1 \equiv W^{1/2}D_1$ and $G_2 \equiv W^{1/2}D_2$. Then, it is easy to verify that Theorem 1 (C), $M_{[G]}\Phi M_{[G_2]}G_1 = 0$, is equivalent to:

$$
\begin{aligned}
& [I - W^{1/2}D(D'WD)^{-1}D'W^{1/2}]W^{1/2}\Omega W^{1/2}[I - W^{1/2}D_2(D_2'WD_2)^{-1}D_2'W^{1/2}]W^{1/2}D_1 = 0 \\
\Leftrightarrow \quad & [W^{1/2} - W^{1/2}D(D'WD)^{-1}D'W]\Omega[W - WD_2(D_2'WD_2)^{-1}D_2'W]D_1 = 0 \\
\Leftrightarrow \quad & [W - WD(D'WD)^{-1}D'W]\Omega[W - WD_2(D_2'WD_2)^{-1}D_2'W]D_1 = 0 \\
& \text{(pre-multiplying both sides by } W^{1/2}\text{, a nonsingular matrix)} \\
\Leftrightarrow \quad & [I - WD(D'WD)^{-1}D']W\Omega W[I - D_2(D_2'WD_2)^{-1}D_2'W]D_1 = 0. \qquad \text{(A.6)}
\end{aligned}
$$

Now, using the moment functions in (7), we can calculate:

$$
D \equiv E[\partial g_t(\beta)/\partial\beta'] = -E[(I_G \otimes z_t)X_t] = -\begin{bmatrix} E(z_t x_{1t}') & & \\ & \ddots & \\ & & E(z_t x_{Gt}') \end{bmatrix}.
$$

Define $A_i \equiv E(z_t x_{it}')$ for i = 1, .., G. Then, D can be rewritten as $D \equiv -\text{diag}(A_1, \cdots, A_G)$. Similarly, using the definition of $\theta_{01} \equiv (\beta_1', \cdots, \beta_m')'$ and $\theta_{02} \equiv (\beta_{m+1}', \cdots, \beta_G')'$, we have:

$$
D_1 \equiv E[\partial g_t(\beta)/\partial\theta_1'] = \begin{bmatrix} D_{11} \\ 0 \end{bmatrix} \text{ and } D_2 \equiv E[\partial g_t(\beta)/\partial\theta_2'] = \begin{bmatrix} 0 \\ D_{22} \end{bmatrix},
$$

with $D_{11} \equiv -\text{diag}(A_1, \cdots, A_m)$ and $D_{22} \equiv -\text{diag}(A_{m+1}, \cdots, A_G)$.
Note that $\Omega \equiv \text{var}[(I_G \otimes z_t)\varepsilon_t] = \Sigma \otimes B$ and $W = I_G \otimes B^{-1}$ (for 2SLS estimation), with $B \equiv E(z_t z_t')$. Then, we have:

$$
W\Omega W = (I \otimes B^{-1})(\Sigma \otimes B)(I \otimes B^{-1}) = \Sigma \otimes B^{-1}, \qquad \text{(A.7)}
$$

$$WD(D'WD)^{-1}D' = \begin{bmatrix} B^{-1}A_1(A_1'B^{-1}A_1)^{-1}A_1' & & \\ & \ddots & \\ & & B^{-1}A_G(A_G'B^{-1}A_G)^{-1}A_G' \end{bmatrix}. \quad \text{(A.8)}$$

Also, $D_2'WD_1 = (0, D_{22}')(I_G \otimes B^{-1})(D_{11}', 0)' = 0$. This implies:

$$[I - D_2(D_2'WD_2)^{-1}D_2'W]D_1 = D_1. \quad \text{(A.9)}$$

Then, by plugging (A.7)-(A.9) and $D_1 = (D_{11}', 0)'$ into (A.6), (A.6) becomes:

$$\begin{bmatrix} I_M - B^{-1}A_1(A_1'B^{-1}A_1)^{-1}A_1' & & \\ & \ddots & \\ & & I_M - B^{-1}A_G(A_G'B^{-1}A_G)^{-1}A_G' \end{bmatrix}(\Sigma \otimes B^{-1})\begin{bmatrix} D_{11} \\ 0 \end{bmatrix} = 0.$$

Now, using $\Sigma = (\sigma_{ij})_{G\times G}$ and $D_{11} = -\text{diag}(A_1, \cdots, A_m)$, we can easily verify that this equation is equivalent to:

$$[I_M - B^{-1}A_i(A_i'B^{-1}A_i)^{-1}A_i'] \cdot \sigma_{ij}B^{-1}A_j = 0, \text{ for i = 1, 2, …, G; j = 1, 2, …, m.} \quad \text{(A.10)}$$

Thus, we have shown that Theorem 1(C) is equivalent to (A.10). We now proceed to show that (A.10) is also equivalent to equation (8) of Theorem 2. In fact, (A.10) is equivalent to:

$$\sigma_{ij} = 0 \text{ or } [I_M - B^{-1}A_i(A_i'B^{-1}A_i)^{-1}A_i']B^{-1}A_j = 0,$$

for $i \neq j$; i = 1, 2, …, G; j = 1, 2, …, m. But $[I_M - B^{-1}A_i(A_i'B^{-1}A_i)^{-1}A_i']B^{-1}A_j = 0$ is the same as $A_j = A_i(A_i'B^{-1}A_i)^{-1}A_i'B^{-1}A_j$. This implies that A_j is in the column space of A_i, or $M_{[A_i]}A_j = 0$. Therefore, (A.10) is equivalent to:

$$\sigma_{ij}M_{[A_i]}A_j = 0, \text{ for } i \neq j; \text{ i = 1, 2, …, G; j = 1, 2, …, m.}$$

This is just equation (8) of Theorem 2. This completes our proof of Theorem 2. ■

Proof of Corollary 2.

Note that equation (8) of Theorem 2 is equivalent to:

$$\sigma_{ij} = 0 \text{ or } M_{[A_i]}A_j = 0, \text{ for } i \neq j; \text{ i = 1, 2, …, G; j = 1, 2, …, m.}$$

But $M_{[A_i]}A_j = 0$ is equivalent to $A_j = A_iC_{ij}$, with C_{ij} being a $K_i \times K_j$ matrix of full column rank. Recall the definition of $A_i \equiv E(z_t x_{it}')$, which can be consistently estimated by

$\hat{A}_i = T^{-1}Z'X_{(i)}$. Then, for a given sample of size T, a sufficient condition for the population condition $A_j = A_i C_{ij}$ is that the following condition holds, with probability equal to 1, in the sample:

$$T^{-1}Z'X_{(j)} = T^{-1}Z'X_{(i)}\hat{C}_{ij}, \text{ or } Z(Z'Z)^{-1}Z'X_{(j)} = Z(Z'Z)^{-1}Z'X_{(i)}\hat{C}_{ij},$$

since $Z(Z'Z)^{-1}$ has full column rank. This is just $\hat{X}_{(j)} = \hat{X}_{(i)}\hat{C}_{ij}$ or $M_{[\hat{X}_{(i)}]}\hat{X}_{(j)} = 0$, using $\hat{X}_{(i)} = P_{[Z]}X_{(i)}$ and $\hat{C}_{ij} = (\hat{X}_{(i)}'\hat{X}_{(i)})^{-1}\hat{X}_{(i)}'\hat{X}_{(j)}$. Therefore, a sufficient condition for (8) is:

$$\sigma_{ij}M_{[\hat{X}_{(i)}]}\hat{X}_{(j)} = 0, \text{ for } i \neq j; \ i = 1, 2, \ldots, G; \ j = 1, 2, \ldots, m,$$

holds with probability equal to 1. This completes the required proof. ■

References

Ahn, S., Y. Lee and P. Schmidt (2001) "GMM estimation of linear panel data models with time-varying individual effects" *Journal of Econometrics* 101, 219-255.

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

Anderson, T. (1971) *The Statistical Analysis of Time Series*, John Wiley and Sons.

Baksalary, J. and G. Trenkler (1989) "The efficiency of OLS in a seemingly unrelated regressions model" *Econometric Theory* 5, 463-465.

Baltagi, B. (1988) "The efficiency of OLS in a seemingly unrelated regressions model" *Econometric Theory* 4, 536-537.

Baltagi, B. (1988-89) "On the efficiency of two-stage and three-stage least squares estimators" *Econometric Reviews* 7, 165-169.

Bartels, R. and D. Fiebig (1991) "A simple characterization of seemingly unrelated regressions models in which OLS is BLUE" *The American Statistician* 45, 137-140.

Dwivedi, T. and V. Srivastava (1978) "Optimality of least squares in the seemingly unrelated regression equation model" *Journal of Econometrics* 7, 391-395.

Gourieroux, C. and A. Monfort (1980) "Sufficient linear structures: econometric applications" *Econometrica* 48, 1083-1097.

Haberman, S. (1975) "How much do Gauss-Markov and least squares estimates differ? A coordinate-free approach" *Annals of Statistics* 3, 982-990.

Hansen, L. (1982) "Large sample properties of generalized method of moments estimators" *Econometrica* 50, 1029-1054.

Kapteyn, A. and D. Fiebig (1981) "When are two-stage and three-stage least squares estimators identical?" *Economics Letters* 8, 53-57.

Kruskal, W. (1968) "When are Gauss-Markov and least squares estimators identical? A coordinate-free Approach" *Annals of Mathematical Statistics* 39, 70-75.

Qian, H. and P. Schmidt (2003) "Partial GLS regression" *Economics Letters* 79, 385-392.

Qian, H. and H. Bednarek (2015) "Partial efficient estimation of SUR models" *Economics Bulletin* 35 (1), 338-348.

Rao, C.R. (1968) "A note on a previous lemma in the theory of least squares and some further results" *Sankhya*, Series A, 30, 245-252.

Rao, C.R. and S. Mitra (1971) *Generalized Inverse of Matrices and Its Applications*, Wiley: New York.

Revankar, N. (1974) "Some finite sample results in the context of two seemingly unrelated regression equations" *Journal of the American Statistical Association* 69, 187-190.

Schmidt, P. (1976) *Econometrics*, Marcel Dekker: New York.

Schmidt, P. (1978) "A note on the estimation of seemingly unrelated regression systems" *Journal of Econometrics* 7, 259-261.

Srivastava, V. and R. Tiwari (1978) "Efficiency of two-stage and three-stage least squares estimators" *Econometrica* 46, 1495-1498.

Srivastava, V. and D. Giles (1987) *Seemingly Unrelated Regression Equations Models*, Marcel Dekker: New York.

Wooldridge, J. (2010) *Econometric Analysis of Cross Section and Panel Data,* The MIT Press.

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.

Zellner, A. and H. Theil (1962) "Three-stage least squares: simultaneous estimation of simultaneous equations" *Econometrica* 30, 54-78.

Zyskind, G. (1967) "On canonical forms, nonnegative covariance matrices and best and simple least squares linear estimators in linear models" *Annals of Mathematical Statistics* 38, 1092-1109.