This is to certify that the

dissertation entitled

IMPROVED GENERALIZED METHOD
OF MOMENTS ESTIMATORS

presented by

HAILONG QIAN

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Economics

Peter Schmidt
Major professor

Date July 27, 1995

IMPROVED GENERALIZED METHOD OF MOMENTS ESTIMATORS

By

HAILONG QIAN

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1995

ABSTRACT

Improved Generalized Method of Moments Estimators

By

Hailong Qian

This thesis introduces a new method to improve Generalized Method of Moments estimators, given extra observable information. Monte Carlo simulation for a simple model with intercept only confirms the accuracy of the asymptotic results obtained in this thesis even when the sample size is quite small. The three-stage least squares estimator of a system of equations is shown to be asymptotically equivalent to an iterative two-stage least squares estimator applied to each equation, augmented with the residuals from the other equations.

ACKNOWLEDGEMENTS

It is impossible for me to express sufficient gratitude to my major professor, Peter Schmidt, who masterly taught me econometrics, suggested this dissertation topic, carefully guided me through each step of my research, meticulously read and corrected my drafts time after time, and kindly offered me research assistantships for Summer, 1993, and the 1994-95 academic year. Without Professor Schmidt's valuable help, it would have been impossible for me to complete this dissertation.

I am also extremely grateful to Professors Richard Baillie, Ching-Fan Chung, Robert Rasche and Jeffrey Wooldridge for their valuable help with this dissertation. I consider myself extremely fortunate to have the opportunity to associate with such an excellent group of econometricians. Their intelligence, diligence and success have inspired me so much.

I am heavily indebted to Professor Daniel Suits and Mrs. Adelaide Suits. It was Professor Suits who helped get me admitted into this program, and most of all helped me to get a scholarship from the Ford Foundation for my first year here. Without their help, I would not have been here in the first place, and everything else would not be possible either.

I also owe a debt of special gratitude to Mrs. Dottie Schmidt both for her help with my spoken English and for her extreme kindness. Her generous help given to me and numerous foreign students and scholars attending MSU makes our otherwise homesick lives more enjoyable.

It is also my great pleasure to express my appreciation to Dan Hansen, Bill Horrace, Young min Kwon, Hyung-Seung Lee, Paul Loetscher, Jess Reaser, Leslie

Schenk and Jen Tracey for all the help and entertainment they have given me in the last five years. Their freindship has made my life here enjoyable.

I would also like to express my thanks to Ms. Ann Feldman and Terie Snyder for their help in many ways in the last five years.

I owe my greatest debt to my family. Even though my parents are uneducated farmers, through example they taught me the virtues of diligence, self-sufficiency, decency and dignity. Though they often could not afford to buy what they needed, their sacrifice for my education and their wishes for me to have a better life were unconditional. I also want to thank my four older brothers from my heart for all the things they have done for me.

Finally, I gratefully acknowledge the scholarship for my first year here from the U.S. Committee on Economics Education and Research in China sponsored by the Ford Foundation.

TABLE OF CONTENTS

LIST OF TABLES viii

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: IMPROVED GMM ESTIMATORS FOR THE LINEAR REGRESSION MODEL 4

2.1. Introduction 4
2.2. GMM with Moment Conditions Not Containing Unknown Parameters 6
2.3. The Linear Model 8
2.4. Monte Carlo Results 16
2.5. Concluding Remarks 20

CHAPTER 3: IMPROVED GMM AND 3SLS ESTIMATORS FOR SYSTEM OF EQUATIONS 23

3.1. Introduction 23
3.2. Model and Notation 24
3.3. Improved GMM estimators 25
3.4. Concluding Remarks 42

CHAPTER 4: IMPROVED GMM ESTIMATORS FOR SYSTEM OF NONLINEAR EQUATIONS 44

4.1. Introduction 44
4.2. Improved GMM Estimators 44
4.3. Concluding Remarks 57

CHAPTER 5: THE ASYMPTOTIC EQUIVALENCE BETWEEN THE ITERATED IMPROVED 2SLS ESTIMATOR AND THE 3SLS ESTIMATOR 58

5.1. Introduction 58
5.2. Improved 2SLS Estimator 59
5.3. Iterated I2SLS Estimator 63
5.4. The Convergence and Asymptotic Efficiency of the Iterated I2SLS Estimators 71

CHAPTER 6: CONCLUSION 97

REFERENCES 99

LIST OF TABLES

TABLE 1 21

TABLE 2 22

CHAPTER 1

INTRODUCTION

Suppose that we have a set of moment conditions $E[\phi_1(y_t^*, \theta_0)] = 0$ which identify the unknown parameter θ_0, so that the generalized method of moments (GMM) estimation of θ_0 is feasible. However, suppose that we also have available a set of additional moment conditions $E[\phi_2(y_t^*)] = 0$, where ϕ_2 is observable because it depends only on the observed data y_t^*. Then the question is how to utilize these additional moment conditions in a simple way to improve the estimation of θ_0. This is possible when ϕ_2 is correlated with ϕ_1.

Problems of this type have been considered previously by Imbens (1992, 1993) and Imbens and Lancaster (1994). Imbens (1992, footnote 3) considered estimation of $\mu_0 = E(y_t)$. The sample mean, based on the moment condition $E(y_t - \mu_0) = 0$, is less efficient than the GMM estimate based on the moment conditions $E[(y_t - \mu_0), u_t]' = 0$, if u_t is observable, with $E(u_t) = 0$ and $\text{cov}[(y_t - \mu_0), u_t] \neq 0$. Imbens (1992) and Imbens and Lancaster (1994) analyze some other specific problems that lead to GMM estimation with additional moment conditions that do not depend on the parameters of interest.

In this dissertation, we prove that the usual GMM estimator of θ_0, say $\hat{\theta}$, using the moment conditions $E[\phi_1(y_t^*, \theta_0)] = 0$ and weighting matrix $C_{11}^{-1} = \{\lim_{T\to\infty} E[T^{-1/2} \sum_{t=1}^{T} \phi_1(y_t^*, \theta_0)][T^{-1/2} \sum_{t=1}^{T} \phi_1(y_t^*, \theta_0)]'\}^{-1}$, can be improved by using the observed extra moment conditions $E[\phi_2(y_t^*)] = 0$. Specifically, we prove that the usual GMM estimator $\hat{\theta}$ is no more efficient than the augmented GMM (AGMM) estimator, say $\tilde{\theta}$, defined as the GMM estimator of θ_0 using the moment conditions $E[\phi(y_t^*, \theta_0)] = E[\phi_1(y_t^*, \theta_0)', \phi_2(y_t^*)']' = 0$ and weighting matrix

$$C^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^{-1} = \{\lim_{T\to\infty} E[T^{-1/2}\sum_{t=1}^{T}\phi(y_t^*,\theta_0)][T^{-1/2}\sum_{t=1}^{T}\phi(y_t^*,\theta_0)]'\}^{-1}.$$

We further show that the AGMM estimator $\tilde{\theta}$ is numerically the same as the improved GMM (IGMM) estimator, say $\hat{\theta}$, defined as the GMM estimator using the moment conditions $E[\phi_1(y_t^*,\theta_0) - C_{12}C_{22}^{-1}\phi_2(y_t^*)] = 0$ and weighting matrix $C^{11} = (C_{11} - C_{12}C_{22}^{-1}C_{21})^{-1}$.

The structure of the dissertation is as follows. In chapter 2, we first provide a brief general treatment of GMM estimation with additional moment conditions not containing unknown parameters. We then give some more detailed results for the linear regression model. In the case of the linear regression model with conditional homoskedasticity and uncorrelatedness, we show that the IGMM estimate is an improved 2SLS (IV) estimate using as a new set of instruments the part of the original instruments that is orthogonal to the observed extra variables, whereas the usual GMM estimate is just an ordinary 2SLS (IV) estimate using the original set of instruments. We also provide some other estimators that can be written in closed form and that are asymptotically equivalent to the IGMM estimator. For the special case of a simple regression model with intercept only, we provide some Monte Carlo evidence on the finite sample performance of some specific improved estimators. For this simple model, the efficiency gains predicted by asymptotic theory are realized even for quite small sample sizes.

In chapter 3, we extend the general results on improved GMM to the case of a system of linear equations. Under the assumptions of conditional homoskedasticity and uncorrelatedness, we obtain explicit expressions for several asymptotically equally efficient improved GMM estimators. While the usual GMM estimator is just an ordinary 3SLS estimator, we prove that the IGMM is an improved 3SLS estimator. The improved 3SLS estimator differs from the usual 3SLS estimator in two ways. First, the covariance matrix of the residuals of the projection of the original model disturbances onto the observed extra variables is used as the relevant error covariance matrix. Second, it uses as its instruments the part of original instruments orthogonal to the observed extra variables.

In chapter 4, we extend the IGMM results from chapter 3 to the case of a system of nonlinear equations. Under suitable regularity conditions and some "high-level" assumptions, we show that essentially the same results as those in chapter 3 still hold for this case.

In chapter 5, we further extend the improved GMM idea of previous chapters to the case where the extra variables are not observed but consistently estimated. We investigate this problem in the context of a system of linear equations. We show that 3SLS applied to the entire equation system is asymptotically equivalent to iterated 2SLS applied to each equation, augmented by the residuals from the other equations. This result generalizes a result of Telser (1964) for the case of seemingly unrelated regressions. It also provides an interesting example of a setting in which the improved GMM estimator arises natually as an efficient estimator.

The final chapter concludes the dissertation with some brief comments on further possible work in this line of research.

CHAPTER 2

IMPROVED GMM ESTIMATORS FOR THE LINEAR REGRESSION MODEL

2.1. Introduction

In this chapter, we provide (in section 2.2) a brief general treatment of GMM estimation with additional moment conditions not containing unknown parameters. We also give (in section 2.3) some more detailed results for the linear regression model. Specifically, we consider the standard regression model

$$(2.1) \qquad y_t = x_t'\beta + \varepsilon_t, \quad t = 1, 2, \ldots\ldots, T,$$

with instruments z_t satisfying $E(z_t\varepsilon_t) = 0$. These moment conditions are the basis of GMM estimation of β; under a conditional homoscedasticity assumption for ε_t, the GMM estimator is the usual instrumental variables (IV) estimator. If we also have available a vector of observable variables u_t that are uncorrelated with z_t but correlated with ε_t, the additional moment conditions $E(u_t \otimes z_t) = 0$ will improve the efficiency of estimation of β. This principle applies in linear or nonlinear models, but in the linear case we obtain very simple explicit results for the improved estimators.

We believe that these results are empirically relevant, notably in the estimation of rational expectations models. In many empirical rational expectations models, the orthogonality conditions used in estimation assert that a forecast error, written as a function of data and parameters, is uncorrelated with variables in the information set at the time the forecast was made. Thus ε_t is the error made in forecasting some variable at time t, based on information available at time t-1, and z_t consists of information available at time t-1, so that it is uncorrelated with ε_t. In this setting, u_t can be the observable (ex post) error in the forecast of a set of variables at time t based on information available at time t-1.

As a specific example, suppose that s_t is a spot exchange rate at time t and f_t is the one period forward rate. Many papers have tested the unbiasedness hypothesis that $f_{t-1} = E(s_t \mid \Omega_{t-1})$, where Ω_{t-1} is the information set at time t-1. Thus we should have $\alpha = 0$ and $\beta = 1$ in the regression model

(2.2) $$s_t = \alpha + \beta f_{t-1} + \varepsilon_t.$$

When s_t and f_{t-1} contain unit roots but are cointegrated, the above regression is often replaced by a regression in stationary variables:

(2.3) $$(s_t - s_{t-1}) = \alpha + \beta(f_{t-1} - s_{t-1}) + \varepsilon_t$$

where again $\alpha = 0$ and $\beta = 1$ under the unbiasedness hypothesis. Because the forecast error ε_t is uncorrelated with variables in Ω_{t-1}, (2.2) or (2.3) can be estimated by GMM or IV, where the instruments z_t are variables in Ω_{t-1}. This is a standard applied econometric excercise. However, the estimate can be improved by using other observable variables u_t that are correlated with the forecast error ε_t but uncorrelated with z_t. Such variables will typically be forecast errors in other related variables. An obvious example would be the change in a security price from time t-1 to t. We might reasonably expect ε_t and u_t to be correlated if spot exchange rates and security prices respond to the same unforecastable economic shocks.

For Imbens's model of the estimation of the sample mean, we provide (in section

2.4) some Monte Carlo evidence on the finite sample performance of some specific improved estimators. For this simple model, the efficiency gains predicted by asymptotic theory are realized even for quite small sample sizes. The final section concludes the chapter with some comments.

2.2. GMM With Moment Conditions Not Containing Unknown Parameters

Let θ_0 be a $K \times 1$ vector of parameters to be estimated, and y_t^*, t = 1, 2, ..., T, be observed data. Suppose that the following moment conditions hold:

$$E\phi(y_t^*,\theta_0) = E\begin{bmatrix}\phi_1(y_t^*,\theta_0)\\ \phi_2(y_t^*)\end{bmatrix} = 0, \quad t = 1, \ldots, T, \tag{2.4}$$

where ϕ_1 is $N \times 1$, with $N \geq K$, and ϕ_2 is $H \times 1$. We want to compare GMM based on ϕ_1 only with GMM based on $\phi = (\phi_1{}',\phi_2{}')'$. Note that ϕ_2 does not depend on θ_0.

Define the following notation:

$$\phi_T(\theta) = \begin{bmatrix}\phi_{T1}(\theta)\\ \phi_{T2}\end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}\phi(y_t^*,\theta) \tag{2.5A}$$

$$C = \begin{bmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{bmatrix} = \lim_{T\to\infty}[T\cdot E\phi_T(\theta_0)\phi_T(\theta_0)'] \tag{2.5B}$$

$$D = \begin{bmatrix}D_1\\ 0\end{bmatrix} = \lim_{T\to\infty}\frac{\partial\phi_T(\theta_0)}{\partial\theta'}\ . \tag{2.5C}$$

(The block "zero" in D arises because ϕ_{T2} does not depend on θ.) For identification of θ_0

we require D_1 to be of full column rank. In the case that the y_t^* are iid, $C = V[\phi(y_t^*,\theta_0)]$ and $D = E[\partial\phi(y_t^*,\theta_0)/\partial\theta']$.

Let $\hat{\theta}$ denote the GMM estimator of θ_0 based on the moment conditions ϕ_1 only, using weighting matrix C_{11}^{-1}; and let $\tilde{\theta}$ denote the augmented GMM (AGMM) estimator of θ_0 based on the moment conditions $\phi = (\phi_1', \phi_2')'$, using weighting matrix C^{-1}. Under suitable regularity conditions, standard GMM results indicate that these estimators are consistent, with $AV[\sqrt{T}(\hat{\theta}-\theta_0)] = (D_1'C_{11}^{-1}D_1)^{-1}$ and $AV[\sqrt{T}(\tilde{\theta}-\theta_0)] = (D'C^{-1}D)^{-1}$ $= (D_1'C^{11}D_1)^{-1}$, where $C^{11} = (C_{11} - C_{12}C_{22}^{-1}C_{21})^{-1}$ is the block of C^{-1} corresponding to ϕ_1 (i.e., the upper left block). For discussions of regularity conditions, see, e.g., Hansen (1982) or Gallant and White (1988). The AGMM estimator is efficient relative to the GMM estimator, since $(D_1'C_{11}^{-1}D_1)^{-1} - (D_1'C^{11}D_1)^{-1}$ is positive semidefinite. In fact, a little algebra reveals that $D_1'C^{11}D_1 - D_1'C_{11}^{-1}D_1 = (C_{21}C_{11}^{-1}D_1)'(C_{22} - C_{21}C_{11}^{-1}C_{12})^{-1}\cdot$ $(C_{21}C_{11}^{-1}D_1)$, so that the condition for no gain in efficiency is $C_{21}C_{11}^{-1}D_1 = 0$. There is no efficiency gain when $C_{21} = 0$ (ϕ_1 and ϕ_2 are uncorrelated); when $C_{21} \neq 0$, the AGMM estimator is generally (but not necessarily) strictly better than the GMM estimator.

We can also write the augmented GMM estimator as follows. Consider the moment conditions

$$E[\phi_1(y_t^*,\theta_0) - C_{12}C_{22}^{-1}\phi_2(y_t^*)] = 0, \tag{2.6}$$

which essentially deal with the residuals from a regression of ϕ_1 on ϕ_2. If $V[\phi(y_t^*,\theta_0)] = C$, then $V[\phi_1(y_t^*,\theta_0) - C_{12}C_{22}^{-1}\phi_2(y_t^*)] = C_{11} - C_{12}C_{22}^{-1}C_{21} = (C^{11})^{-1}$. With this motivation, we define the improved GMM (IGMM) estimator, say $\breve{\theta}$, as the GMM estimator using the moment conditions (2.6) and weighting matrix $C^{11} = (C_{11} - C_{12}C_{22}^{-1}C_{21})^{-1}$. It is then not difficult to show that the IGMM estimator $\breve{\theta}$ and the AGMM estimator $\tilde{\theta}$ are the same. This can be seen by noting that $\breve{\theta}$ satisfies the first order condition

(2.7A) $$D_{T1}(\bar{\theta})'C^{11}[\phi_{T1}(\bar{\theta})-C_{12}C_{22}^{-1}\phi_{T2}]=0$$

where $D_{T1}(\theta)=\partial\phi_{T1}(\theta)/\partial\theta'$; $\tilde{\theta}$ satisfies the first order condition

(2.7B) $$D_{T1}(\tilde{\theta})'C^{11}\phi_{T1}(\tilde{\theta})+D_{T1}(\tilde{\theta})'C^{12}\phi_{T2}=0.$$

But (2.7A) and (2.7B) are seen to be the same with the substitution $C^{12}=-C^{11}C_{12}C_{22}^{-1}$ in (2.7B).

The above discussion treats the weighting matrix C as known. Assuming suitable regularity conditions, the superiority of the AGMM or IGMM estimator to the GMM estimator will still hold asymptotically if C is replaced by a consistent estimate $\hat{C}$. The numerical equivalence of the AGMM and IGMM estimators would require that the same estimate $\hat{C}$ be used for both estimators.

2.3. The Linear Regression Model

In this section we will apply the general results of the previous section to the case of the linear regression model. For this case we can give an explicit formula for the IGMM estimator. When the errors are conditionally homoskedastic, further simplications are possible and the IGMM estimator is related to some previous results.

The model considered in this section is as given in equation (2.1) above, which we rewrite slightly as

(2.8) $$y_t = x_t'\theta_0+\varepsilon_t,\ t=1, 2, \ldots\ldots, T,$$

where y_t is the dependent variable , x_t is a $K \times 1$ vector of explanatory variables, ε_t is the disturbance term, and θ_0 is the parameter vector to be estimated. Suppose that we have available an $M \times 1$ vector of instruments z_t satisfying $M \geq K$ and $E(z_t \varepsilon_t) = 0$, so that the GMM estimation of θ_0 based on the moment conditions $E(z_t \varepsilon_t) = 0$ is feasible. However, suppose that we also have available an $L \times 1$ observable vector u_t satisfying $E(u_t \otimes z_t) = 0$ and $E(u_t \varepsilon_t) \neq 0$. The observable data vector is $y_t^* = (y_t, x_t', z_t', u_t')'$, and in our previous notation we have moment conditions $E\phi(y_t^*, \theta_0) = 0$, with

(2.9A) $$\phi_1(y_t^*, \theta_0) = z_t(y_t - x_t'\theta_0)$$

(2.9B) $$\phi_2(y_t^*) = u_t \otimes z_t = (I_L \otimes z_t)u_t.$$

As a matter of notation, let $Z = (z_1, \cdots, z_T)'$; $X = (x_1, \cdots, x_T)'$; $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_T)'$; $U = (u_1, \cdots, u_T)'$; $y = (y_1, \cdots, y_T)'$; $u_{(j)} = (u_{j1}, \cdots, u_{jT})'$ for j = 1, ..., L; and $u^* = (u_{(1)}', \cdots, u_{(L)}')' = \text{vec}(U)$. Then straightforward calculation yields

(2.10A) $$\phi_{T1} = T^{-1}Z'(y - X\theta)$$

(2.10B) $$\phi_{T2} = T^{-1}(I_L \otimes Z')u^* = T^{-1}\text{vec}(Z'U)$$

(2.10C) $$D_{T1}(\theta) = -T^{-1}Z'X$$

Using the first order condition (2.7A) above, with these substitutions we arrive at the IGMM estimator

(2.11) $$\tilde{\theta} = (X'ZC^{11}Z'X)^{-1}X'ZC^{11}[Z'y - C_{12}C_{22}^{-1}\text{vec}(Z'U)].$$

To proceed further, we need to put more structure on C. This is possible under the assumption of no conditional heteroskedasticity or autocorrelation: suppose that, conditional on $\Omega_t = \{z_t; \varepsilon_{t-1}, u_{t-1}, z_{t-1}; \ldots\}$, the $(\varepsilon_t, u_t')'$ are mutually uncorrelated, z_t is stationary, and that

(2.12) $$V(\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} | z_t) \equiv \Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix}.$$

Then

(2.13) $$C = E\phi(y_t^*, \theta_0)\phi(y_t^*, \theta_0)' = \Sigma \otimes E(z_t z_t')$$

for which a consistent estimate is

(2.14) $$\hat{C} = \hat{\Sigma} \otimes T^{-1}Z'Z,$$

where $\hat{\Sigma}$ is any consistent estimate of Σ. Then $\hat{C}^{11} = \hat{\sigma}^{\varepsilon\varepsilon} \cdot (T^{-1}Z'Z)^{-1}$ with $\hat{\sigma}^{\varepsilon\varepsilon} = (\hat{\sigma}_\varepsilon^2 - \hat{\Sigma}_{\varepsilon u}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{u\varepsilon})^{-1}$, $\hat{C}_{12} = \hat{\Sigma}_{\varepsilon u} \otimes T^{-1}Z'Z$, $\hat{C}_{22} = \hat{\Sigma}_{uu} \otimes T^{-1}Z'Z$, $\hat{C}_{12}C_{22}^{-1} = (\hat{\Sigma}_{\varepsilon u}\hat{\Sigma}_{uu}^{-1}) \otimes I_M$. With these substitutions in (2.11), and noting that $\hat{\sigma}^{\varepsilon\varepsilon}$ cancels, we obtain

(2.15) $$\bar{\theta} = (X'P_Z X)^{-1}X'Z(Z'Z)^{-1}\{Z'y - [(\hat{\Sigma}_{\varepsilon u}\hat{\Sigma}_{uu}^{-1}) \otimes I_M]\text{vec}(Z'U)\},$$

where $P_Z = Z(Z'Z)^{-1}Z'$. More generally, if A is any matrix, we will define P_A as the projection onto A, so that $P_A = A(A'A)^{-1}A'$ if A has full column rank. Similarly, we define $M_A = I - P_A$. Obviously the first term in this expression is just $(X'P_Z X)^{-1}X'P_Z y$, the IV (2SLS) estimator, which is GMM based on ϕ_1, given the assumption of no conditional heteroskedasticity.

Using the matrix fact $\text{vec}(BC) = (C' \otimes I)\text{vec}(B)$, (2.15) can be rewritten as

(2.16) $$\bar{\theta} = (X'P_Z X)^{-1}X'Z(Z'Z)^{-1}[Z'y - Z'U\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{u\varepsilon}].$$

It is reasonable to consider $\hat{\Sigma}_{uu} = T^{-1}U'U$, $\hat{\Sigma}_{u\varepsilon} = T^{-1}U'(y - X\hat{\theta})$, where $\hat{\theta}$ is any consistent estimate of θ. Then (2.16) becomes

$$(2.17) \qquad \bar{\theta} = (X'P_Z X)^{-1} X'P_Z[y - P_U(y - X\hat{\theta})].$$

Finally, while (2.17) is defined for any consistent estimate $\hat{\theta}$, we may as well consider $\hat{\theta} = \bar{\theta}$. Then (2.17) implies $(X'P_Z X)\bar{\theta} = X'P_Z y - X'P_Z P_U y + X'P_Z P_U X\bar{\theta}$; solving for $\bar{\theta}$, we obtain

$$(2.18) \qquad \bar{\theta} = (X'P_Z M_U X)^{-1} X'P_Z M_U y.$$

The IGMM estimator $\bar{\theta}$ is very similar to an estimator considered by Schmidt (1986, 1988):

$$(2.19) \qquad \dddot{\theta} = (X'P_{[M_U Z]}X)^{-1} X'P_{[M_U Z]}y.$$

This is IV of the regression equation (2.8), using as instruments $M_U Z$, the part of Z orthogonal to U. Schmidt also notes that $\dddot{\theta}$ can be derived as IV of the augmented equation

$$(2.20) \qquad y_t = x_t'\theta_0 + u_t'\xi + v_t$$

using (Z, U) as instruments. Equation (2.20) is instructive because, speaking loosely, the effect of adding the variable u_t is to reduce the relevant variance from σ_ε^2 to $\sigma_v^2 = \sigma_{\varepsilon|u}^2 = \sigma_\varepsilon^2 - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}$. This result is closely related to the result of Wooldridge (1993), who essentially considers the case $x_t = z_t$ (in our notation).

To be more precise about the sense in which $\bar{\theta}$ and $\ddot{\theta}$ dominate the simple IV estimator, and to exhibit some other asymptotically efficient estimators, we make some more explicit assumptions. To make the asymptotic theory as simple as possible, we will make the following "high level" assumptions.

(A2.1) $$\text{plim}\,\frac{1}{T}[X,Z,\varepsilon,U]'[X,Z,\varepsilon,U] = \begin{bmatrix} A_{xx} & A_{xz} & A_{x\varepsilon} & A_{xu} \\ A_{zx} & A_{zz} & 0 & 0 \\ A_{\varepsilon x} & 0 & \sigma_\varepsilon^2 & \Sigma_{\varepsilon u} \\ A_{ux} & 0 & \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix} \text{ exists.}$$

(A2.2) A_{xx}, A_{zz} and Σ_{uu} are nonsingular; A_{zx} is of full column rank.

(A2.3a) $T^{-1/2}\text{vec}[Z'(\varepsilon,U)] \rightarrow N[0,\Psi]$.

(A2.3b) $\Psi = \Sigma \otimes A_{zz}$.

These high-level assumptions are derivable from various sets of more basic assumptions. For example, in the rational expectations context, define $e_t = (\varepsilon_t, u_t')'$ and let Ω_t be the information (sub)set $\Omega_t = \{z_t; z_{t-1}, e_{t-1}; z_{t-2}, e_{t-2}; \ldots\ldots\}$. Then (A2.1)-(A2.3) follow from the assumptions that $E(e_t|\Omega_t) = 0$, $V(e_t|\Omega_t) = \Sigma$, and x_t and z_t are covariance stationary.

Let $\hat{\theta}$ be the usual IV estimator using Z as instruments: $\hat{\theta} = (X'P_Z X)^{-1} X'P_Z y$. Under (A2.1)-(A2.3) we have the standard result:

(2.21) $$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow N[0, \sigma_\varepsilon^2 (A_{xz} A_{zz}^{-1} A_{zx})^{-1}];$$

this is consistent with the general GMM result $AV(\hat{\theta}) = (D_1' C_{11}^{-1} D_1)^{-1}$ presented earlier. We now turn to the IGMM estimator $\bar{\theta}$, Schmidt's estimator $\ddot{\theta}$, and the following additional estimators

(2.22A) $$\breve{\theta} = (X'M_U P_Z M_U X)^{-1} X'M_U P_Z M_U y$$

(2.22B) $$\ddot{\theta} = (X'P_Z X)^{-1} X'P_Z(y - U\lambda)$$

where $\lambda = \Sigma_{uu}^{-1}\Sigma_{u\varepsilon}$. In practice, (2.22B) will require a consistent estimate of λ.

We wish to show that the estimators $\widehat{\theta}$, $\dddot{\theta}$, $\breve{\theta}$ and $\ddot{\theta}$ are asymptotically equivalent, with asymptotic variance matrix $\sigma^2_{\varepsilon|u}(A_{xz}A_{zz}^{-1}A_{zx})^{-1}$, where as above $\sigma^2_{\varepsilon|u} = \sigma^2_{\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}$. This is consistent with the result of Schmidt (1988, Appendix C.4) for $\dddot{\theta}$. Comparing to the asymptotic variance matrix of $\hat{\theta}$ in (2.21) above, the inequality $\sigma^2_{\varepsilon|u} \le \sigma^2_{\varepsilon}$ establishes the asymptotic efficiency of $\widehat{\theta}$, $\dddot{\theta}$, $\breve{\theta}$ and $\ddot{\theta}$ relative to $\hat{\theta}$.

<u>LEMMA 2.1</u>: $\text{plim}\, T^{-1}X'M_U Z = \text{plim}\, T^{-1}X'Z = A_{xz}$

$$\text{plim}\, T^{-1}Z'M_U Z = \text{plim}\, T^{-1}Z'Z = A_{zz}$$

Proof: $\text{plim}\, T^{-1}X'M_U Z = \text{plim}\, [T^{-1}X'Z - T^{-1}X'U(T^{-1}U'U)^{-1}T^{-1}U'Z]$

$$= A_{xz} - A_{xu}\Sigma_{uu}^{-1}\cdot 0 = A_{xz},$$

and similarly for $\text{plim}\, T^{-1}Z'M_U Z$.

■

<u>LEMMA 2.2</u>: $\text{plim}\, T^{-1}X'P_{[M_U Z]}X = \text{plim}\, T^{-1}X'M_U P_Z M_U X$

$$= \text{plim}\, T^{-1}X'P_Z M_U X$$

$$= \text{plim}\, T^{-1}X'P_Z X$$

$$= A_{xz}A_{zz}^{-1}A_{zx}$$

Proof: $\text{plim}\, T^{-1}X'P_{[M_U Z]}X = \text{plim}\, T^{-1}X'M_U Z(\text{plim}\, T^{-1}Z'M_U Z)^{-1}\text{plim}\, T^{-1}Z'M_U X$

$$= A_{xz}A_{zz}^{-1}A_{zx} \text{ using Lemma 2.1.}$$

The proofs for $\text{plim}\, T^{-1}X'M_U P_Z M_U X$ and $\text{plim}\, T^{-1}X'P_Z M_U X$ are similar.

■

<u>LEMMA 2.3</u>: $\text{plim}\, T^{-1}X'P_{[M_U Z]}\varepsilon = \text{plim}\, T^{-1}X'M_U P_Z M_U \varepsilon$

$$= \text{plim}\, T^{-1}X'P_Z M_U \varepsilon$$

$$= p\lim T^{-1}X'P_Z(y - U\lambda)$$
$$= p\lim T^{-1}X'P_Z\varepsilon$$
$$= 0$$

Proof: $p\lim T^{-1}X'P_{[M_U Z]}\varepsilon = A_{xz}A_{zz}^{-1}\cdot p\lim T^{-1}Z'M_U\varepsilon$ using Lemma 2.1.

But $p\lim T^{-1}Z'M_U\varepsilon = p\lim T^{-1}Z'\varepsilon - p\lim T^{-1}Z'U(p\lim T^{-1}U'U)^{-1}p\lim T^{-1}U'\varepsilon$

$$= 0 - 0\cdot\Sigma_{uu}^{-1}\Sigma_{u\varepsilon} = 0$$

since $p\lim T^{-1}Z'\varepsilon = 0$ and $p\lim T^{-1}Z'U = 0$. The proofs for the other cases are similar.

■

Lemmas 2.2 and 2.3 imply that the estimators $\hat{\theta}$, $\ddot{\theta}$, $\breve{\theta}$ and $\ddot{\theta}$ are consistent. For example,

$$p\lim \ddot{\theta} = \theta_0 + [A_{xz}A_{zz}^{-1}A_{zx}]^{-1}\cdot 0 = \theta_0 \tag{2.23}$$

using Lemmas 2.2 and 2.3; similar simple arguments apply to the other estimators. It is interesting in Lemma 2.3 that the orthogonality of ε with $(M_U Z)$ occurs because ε is orthogonal to Z <u>and</u> Z is orthogonal to U.

<u>LEMMA 2.4</u>: $T^{-1/2}X'P_{[M_U Z]}\varepsilon$, $T^{-1/2}X'M_U P_Z M_U\varepsilon$, $T^{-1/2}X'P_Z M_U\varepsilon$ and $T^{-1/2}X'P_Z(y - U\lambda)$ each converge in distribution to $N[0, \sigma_{\varepsilon|u}^2 A_{xz}A_{zz}^{-1}A_{zx}]$.

Proof: We will give the proof for $T^{-1/2}X'P_{[M_U Z]}\varepsilon$. The other proofs are quite similar.

$$\begin{aligned} T^{-1/2}X'P_{[M_U Z]}\varepsilon &= T^{-1/2}X'M_U Z(Z'M_U Z)^{-1}Z'M_U\varepsilon \\ &= (T^{-1}X'M_U Z)(T^{-1}Z'M_U Z)^{-1}T^{-1/2}Z'M_U\varepsilon \\ &= A_{xz}A_{zz}^{-1}(T^{-1/2}Z'M_U\varepsilon) + o_p(1). \end{aligned} \tag{2.24}$$

So we consider

$$T^{-1/2}Z'M_U\varepsilon = T^{-1/2}Z'[I - U(U'U)^{-1}U']\varepsilon$$

$$= T^{-1/2}Z'\varepsilon - (T^{-1/2}Z'U)(T^{-1}U'U)^{-1}(T^{-1}U'\varepsilon)$$
$$= T^{-1/2}Z'(\varepsilon - U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) + o_p(1).$$

Combining expressions, we have

(2.25) $$T^{-1/2}X'P_{[M_U Z]}\varepsilon = A_{xz}A_{zz}^{-1}T^{-1/2}Z'(\varepsilon - U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) + o_p(1).$$

But

(2.26) $$T^{-1/2}Z'(\varepsilon - U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})$$
$$= T^{-1/2}\mathrm{vec}[Z'(\varepsilon - U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})]$$
$$= T^{-1/2}\mathrm{vec}\{Z'(\varepsilon, U)\begin{bmatrix} 1 \\ -\Sigma_{uu}^{-1}\Sigma_{u\varepsilon} \end{bmatrix}\}$$
$$= \{[1, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}] \otimes I_M\}T^{-1/2}\mathrm{vec}[Z'(\varepsilon, U)].$$

But according to assumption (A2.3) above, $T^{-1/2}\mathrm{vec}[Z'(\varepsilon, U)] \rightarrow N[0, \Sigma \otimes A_{zz}]$.

Therefore

(2.27) $$T^{-1/2}Z'(\varepsilon - U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) \rightarrow N(0, B)$$

where

(2.28) $$B = \{[1, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}] \otimes I_M\}(\Sigma \otimes A_{zz})\{[1, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}] \otimes I_M\}'$$
$$= \{[1, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}]\Sigma[1, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}]'\} \otimes A_{zz}$$
$$= (\sigma_\varepsilon^2 - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})A_{zz} = \sigma_{\varepsilon|u}^2 A_{zz}.$$

Using (2.27)-(2.28) in (2.25), we conclude

(2.29) $$T^{-1/2}X'P_{[M_U Z]}\varepsilon \rightarrow N[0, A_{xz}A_{zz}^{-1}\sigma_{\varepsilon|u}^2 A_{zz} \cdot A_{zz}^{-1}A_{zx}] = N[0, \sigma_{\varepsilon|u}^2 A_{xz}A_{zz}^{-1}A_{zx}].$$

■

<u>THEOREM 2.1</u>: $\sqrt{T}(\bar{\theta} - \theta_0)$, $\sqrt{T}(\ddot{\theta} - \theta_0)$, $\sqrt{T}(\breve{\theta} - \theta_0)$ and $\sqrt{T}(\ddot{\theta} - \theta_0)$ each converge in distribution to $N[0, \sigma_{\varepsilon|u}^2(A_{xz}A_{zz}^{-1}A_{zx})^{-1}]$.

Proof: $\sqrt{T}(\dddot{\theta}-\theta_0)=(T^{-1}X'P_{[M_UZ]}X)^{-1}T^{-1/2}X'P_{[M_UZ]}\varepsilon$.

Then using Lemmas 2.2 and 2.4, $\sqrt{T}(\dddot{\theta}-\theta_0)\rightarrow N[0,\Lambda]$, with

$$\Lambda=(A_{xz}A_{zz}^{-1}A_{zx})^{-1}\cdot\sigma_{\varepsilon|u}^2(A_{xz}A_{zz}^{-1}A_{zx})\cdot(A_{xz}A_{zz}^{-1}A_{zx})^{-1}$$
$$=\sigma_{\varepsilon|u}^2(A_{xz}A_{zz}^{-1}A_{zx})^{-1}.$$

The proofs for the other estimators are essentially identical. ■

Thus the asymptotic variance matrix of each of the above estimators is $\sigma_{\varepsilon|u}^2(\text{plim}\,T^{-1}X'P_ZX)^{-1}$. As noted above, this is less than the corresponding asymptotic variance matrix for the ordinary IV estimator, $\sigma_\varepsilon^2(\text{plim}\,T^{-1}X'P_ZX)^{-1}$, so long as $\Sigma_{\varepsilon u}\neq 0$. To achieve an efficiency gain, the additional variables u must be uncorrelated with the instruments z <u>and</u> correlated with the errors ε.

The estimator $\ddot{\theta}$ in (2.22B) is infeasible because it depends on $\lambda=\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}$. We can define a feasible version of it, say

$$\dot{\theta}=(X'P_ZX)^{-1}X'P_Z(y-U\hat{\lambda}), \tag{2.30}$$

where $\hat{\lambda}$ is a consistent estimate of λ. Specifically, $\hat{\lambda}=(U'U)^{-1}U'\hat{\varepsilon}$ with $\hat{\varepsilon}=y-X\hat{\theta}$, where $\hat{\theta}$ is any consistent estimate of θ_0. It is easy to show that $\dot{\theta}$ is consistent and has the same asymptotic distribution as $\ddot{\theta}$ (and, therefore, the same asymptotic distribution as $\bar{\theta}$, $\dddot{\theta}$ and $\breve{\theta}$).

2.4. Monte Carlo Results

In the previous section we have considered four improved IV (IIV) or improved GMM estimators. Each is consistent and asymptotically more efficient than the usual

IV/GMM estimator $\hat{\theta} = (X'P_Z X)^{-1} X'P_Z y$. The asymptotic efficiency gain for each of the IIV estimators over the usual IV estimator is $\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon} \cdot (p\lim T^{-1}X'P_Z X)^{-1}$, which obviously depends on the strength of the correlation between ε and u.

A natural question to ask is whether our IIV/IGMM estimators are still more efficient than the usual IV or ordinary GMM estimator in finite samples. In order to answer this question, we performed a Monte Carlo simulation on a very simple model. In the simulation we considered our IIV estimators and also some estimators of Imbens (1993) that are similar to GMM estimators. Our simulation plan is as follows. The assumed regression model is:

(2.31) $$y_t = \theta_0 + \varepsilon_t, \quad t = 1, 2, \ldots\ldots, T,$$

where θ_0 is a scalar parameter and ε_t is iid N(0,1). Thus we are estimating $\theta_0 = E(y_t)$. Further we assume that we observe a random variable u_t, which is also iid N(0,1). Let ρ denote the correlation between ε_t and u_t. This simple model has also been considered by Imbens (1993). An efficiency gain is possible here because the mean of u_t is known to be zero. Our DGP is therefore as follows:

(2.32A) $$y_t = 1 + \varepsilon_t,$$

(2.32B) $$u_t = \rho\varepsilon_t + \sqrt{1-\rho^2}\,\eta_t,$$

where ε_t is iid N(0,1), η_t is also iid N(0,1) and ε_t is independent of η_t. Thus $\theta_0 = 1$. Our results do not depend on this choice of θ_0, nor do they depend on the choice of the variance of ε_t and u_t equal to one.

The following six estimators of θ_0 are considered in our simulation:

(1) Sample mean ($\hat{\theta}_1$): $\hat{\theta}_1 = \bar{y}$. This is the GMM estimator based on $E(y_t - \theta_0) = 0$.

(2) **Infeasible GMM ($\hat{\theta}_2$):** $\hat{\theta}_2 = \arg\min_\theta\{\phi_T(\theta)'C^{-1}\phi_T(\theta)\} = \bar{y} - \rho\bar{u}$, where

$$\phi_T(\theta) = (\bar{y}-\theta, \quad \bar{u})', \text{ and } C = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = V(\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}).$$

(3) **Feasible GMM ($\hat{\theta}_3$):** $\hat{\theta}_3 = \arg\min_\theta\{\phi_T(\theta)'\hat{C}^{-1}\phi_T(\theta)\} = \bar{y} - \hat{\rho}\bar{u}$, where

$$\phi_T(\theta) = (\bar{y}-\theta, \quad \bar{u})'; \ \hat{C} = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_t\hat{e}_t' = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} \\ \hat{C}_{21} & \hat{C}_{22} \end{bmatrix} \text{ with } \hat{e}_t = \begin{bmatrix} y_t - \hat{\theta}_1 \\ u_t \end{bmatrix}; \text{ and}$$

$\hat{\rho} = \hat{C}_{21} / \hat{C}_{22}$.

(4) **IIV estimator ($\hat{\theta}_4$):** $\hat{\theta}_4 = (i'M_u i)^{-1} i'M_u y$, where $i \equiv (1, \ \cdots, \ 1)'_{T\times 1}$.

(5) **Imbens's first estimator ($\hat{\theta}_5$):** $\hat{\theta}_5$ is the pseudo maximum likelihood (PML) estimator defined by Imbens (1993) as the first part of the solution to

$$g(\theta,\delta) = \textstyle\sum_{t=1}^{T}\rho(y_t, u_t, \theta, \delta) = 0,$$

where $\rho(y,u,\theta,\delta) = (\frac{y-\theta}{1+\delta u}, \frac{u}{1+\delta u})'$ and δ is an artificial parameter.

(6) **Imbens's third estimator ($\hat{\theta}_6$):** $\hat{\theta}_6$ is defined by Imbens (1993) as the first part of the solution to

$$g(\theta,\delta,\mu) = \textstyle\sum_{t=1}^{T}\bar{\rho}(y_t, u_t, \theta, \delta, \mu) = 0,$$

where $\bar{\rho}(y,u,\theta,\delta,\mu) = ((y-\theta)\exp(\mu-\delta u), u\cdot\exp(\mu-\delta u), 1-\exp(\mu-\delta u))'$.

Notice that in our special case of a regression model with only intercept, some estimators that are different in general become identical. The infeasible GMM estimator

$(\hat{\theta}_2)$ is the same as the infeasible IIV/IGMM estimator $\ddot{\theta} = (x'P_z x)^{-1} x'P_z(y - u\lambda) = \bar{y} - \rho\bar{u}$ defined in (2.22B) above. The feasible GMM estimator $(\hat{\theta}_3)$ is the same as the feasible IIV/IGMM estimator $\dot{\theta} = (x'P_z x)^{-1} x'P_z(y - u\hat{\lambda})$ defined in (2.30) above, where $\hat{\lambda} = \hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{us} = (U'U)^{-1}[U'(y - \hat{\theta}_1)] = \hat{C}_{21} / \hat{C}_{22}$, provided that the initial consistent estimator for θ_0 is $\hat{\theta}_1$ in both cases. The three IIV/IGMM estimators $\bar{\theta} = (x'P_z M_U x)^{-1} x'P_z M_U y$ defined in (2.18), $\dddot{\theta} = (x'P_{[M_U z]} x)^{-1} x'P_{[M_U z]} y$ defined in (2.19) and $\breve{\theta} = (x'M_U P_z M_U x)^{-1} x'M_U P_z M_U y$ defined in (2.22A) are the same and equal to $\hat{\theta}_4 = (i'M_U i)^{-1} i'M_U y$, when $x = z = i = (1, \;\cdots,\; 1)'_{T\times 1}$. The second estimator of Imbens (1993), defined as the first part of the solution to $g(\theta,\delta) = \sum_{t=1}^{T} \tilde{\rho}(y_t, u_t, \theta, \delta) = 0$ with $\tilde{\rho}(y, u, \theta, \delta) = ((y - \theta)(1 - \delta u), u(1 - \delta u))'$, is also the same as the first three IIV/IGMM estimators (equal to $\hat{\theta}_4$). This leaves us with the six distinct estimators listed above.

$\hat{\theta}_1 = \bar{y}$ is unbiased and var$(\hat{\theta}_1)$ = 1/T. $\hat{\theta}_2 = \bar{y} - \rho\bar{u}$ is unbiased and var$(\hat{\theta}_2)$ = $(1 - \rho^2)/T$. For the remaining four estimators, finite sample properties are unknown, but the estimators are consistent and their asymptotic variance is $(1 - \rho^2)/T$.

Our simulation results are based on 20,000 replications. The simulations were performed in GAUSS 2.0 and used its random number generator.

Table 1 gives the means of the six distinct estimators, while Table 2 gives their mean squared errors (MSE). In each case the estimators are nearly unbiased and MSE is nearly the same as variance. For convenience we actually present MSE multiplied by sample size (T), and the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ is given as the value for $T = \infty$. For the sample mean $\hat{\theta}_1$, $T\cdot MSE_1$ should equal 1.0 apart from sampling error for all values of T and ρ, and deviations from unity in the first column labelled $T\cdot MSE_1$ give an indication of the sample variability in the experiment. Similarly, for the infeasible GMM estimator $\hat{\theta}_2$, $T\cdot MSE_2$ should equal $(1 - \rho^2)$ apart from sampling error for all T and ρ. For the other estimators $T\cdot MSE$ should converge to $(1 - \rho^2)$ for large T.

The result in Table 2 are in close agreement with the asymptotic theory, and the agreement is very close for $T \geq 50$. $T\cdot MSE$ is nearly equal to its asymptotic value

$(1-\rho^2)$ for all estimators, all values of ρ, and all sample sizes except T = 25 and occasionally T = 50. The IIV/IGMM estimators are better than the sample mean in all cases except $\rho = .1$ and T = 25 or T = 50; as expected, the size of the efficiency gain depends on ρ.

For this simple model, at least, the differences among the various IIV/IGMM estimators are quite small. As might be expected, the infeasible GMM estimator ($\hat{\theta}_2$) is usually the best. The IIV estimator $\hat{\theta}_4$ (also equal to Imbens's second estimator) is somewhat better than Imbens's first and third estimators ($\hat{\theta}_5$ and $\hat{\theta}_6$). The feasible GMM estimator ($\hat{\theta}_3$) seems to be slightly better than the IIV estimator when ρ is small, and slightly worse when ρ is larger. However, we repeat that the finite sample differences among the asymptotically equivalent estimators are quite small. The main message of the simulations is that we can indeed improve on the usual IV estimator in finite samples, and asymptotic theory is a reliable guide to the variability of these improved estimators. At least this is so in the simple model we have considered.

2.5. Concluding Remarks

In this chapter we have shown how to improve on ordinary GMM (IV or 2SLS) estimators, given observable extra variables which are uncorrelated with the instruments but correlated with the error in the equation being estimated. The difference between the improved 2SLS (IV) estimators and the ordinary 2SLS (IV) estimators is that the projection matrix P_Z in ordinary 2SLS (IV) is replaced by $P_{[M_U Z]}$, $M_U P_Z M_U$, or $P_Z M_U$, so that the 2SLS "fitted values" are constructed differently. For example, $\dddot{\theta}$ uses $M_U Z$, the part of Z orthogonal to U, as the regressors in the "first stage" regression, whereas the ordinary 2SLS estimator $\hat{\theta}$ just uses Z.

TABLE 1

Means of Alternative Estimators

ρ	T	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$
.1	25	.9998	.9998	.9992	.9992	.9992	.9992
	50	.9996	.9996	.9994	.9994	.9995	.9994
	100	.9999	1.0000	1.0001	1.0001	1.0001	1.0000
	200	.9996	.9997	.9997	.9997	.9997	.9997
	500	.9999	.9999	.9999	.9999	.9999	.9999
	∞	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
.3	25	.9998	.9998	.9993	.9993	.9942	.9993
	50	.9996	.9996	.9995	.9995	.9996	.9995
	100	.9999	1.0002	1.0002	1.0002	1.0003	1.0003
	200	.9996	.9998	.9998	.9998	.9998	.9998
	500	.9999	.9999	.9999	.9999	.9999	.9999
	∞	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
.5	25	.9998	.9998	.9995	.9996	.9975	.9996
	50	.9996	.9996	.9997	.9997	.9998	.9997
	100	.9999	1.0003	1.0003	1.0004	1.0004	1.0004
	200	.9996	1.0000	.9999	.9999	.9999	.9999
	500	.9999	1.0000	1.0000	1.0000	1.0000	1.0000
	∞	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
.7	25	.9998	.9998	.9997	.9999	.9993	.9999
	50	.9996	.9997	.9998	.9999	.9998	.9998
	100	.9999	1.0004	1.0004	1.0004	1.0004	1.0004
	200	.9996	1.0001	1.0000	1.0000	1.0000	1.0000
	500	.9999	1.0000	1.0000	1.0000	1.0000	1.0000
	∞	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
.9	25	.9998	.9999	.9997	1.0000	1.0001	1.0000
	50	.9996	.9999	.9999	1.0000	.9995	1.0000
	100	.9999	1.0003	1.0003	1.0003	1.0003	1.0003
	200	.9996	1.0001	1.0001	1.0001	1.0001	1.0001
	500	.9999	1.0000	1.0000	1.0000	1.0000	1.0000
	∞	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 2

Mean Square Errors of Alternative Estimators

ρ	T	$T\cdot MSE_1$	$T\cdot MSE_2$	$T\cdot MSE_3$	$T\cdot MSE_4$	$T\cdot MSE_5$	$T\cdot MSE_6$
.1	25	.9965	.9970	1.0345	1.0447	1.0586	1.0470
	50	.9985	.9945	1.0123	1.0147	1.0181	1.0157
	100	1.0073	.9953	1.0047	1.0052	1.0063	1.0055
	200	1.0008	.9888	.9937	.9938	.9940	.9938
	500	1.0120	1.0003	1.0027	1.0027	1.0025	1.0025
	∞	1.0000	.9900	.9900	.9900	.9900	.9900
.3	25	.9965	.9350	.9684	.9775	.9963	.9781
	50	.9985	.9249	.9406	.9432	.9469	.9439
	100	1.0073	.9111	.9196	.9198	.9207	.9200
	200	1.0008	.9051	.9094	.9094	.9096	.9098
	500	1.0120	.9163	.9180	.9180	.9180	.9180
	∞	1.0000	.9100	.9100	.9100	.9100	.9100
.5	25	.9965	.7831	.8113	.8156	.8188	.8159
	50	.9985	.7699	.7829	.7846	.7873	.7849
	100	1.0073	.7478	.7556	.7556	.7564	.7558
	200	1.0008	.7433	.7468	.7468	.7468	.7468
	500	1.0120	.7524	.7535	.7535	.7535	.7535
	∞	1.0000	.7500	.7500	.7500	.7500	.7500
.7	25	.9965	.5370	.5615	.5580	.5600	.5592
	50	.9985	.5266	.5372	.5365	.5377	.5367
	100	1.0073	.5068	.5130	.5130	.5136	.5132
	200	1.0008	.5043	.5068	.5068	.5068	.5068
	500	1.0120	.5099	.5109	.5108	.5110	.5110
	∞	1.0000	.5100	.5100	.5100	.5100	.5100
.9	25	.9965	.1983	.2199	.2066	.2192	.2069
	50	.9985	.1955	.2035	.1995	.2003	.1998
	100	1.0073	.1887	.1918	.1911	.1913	.1912
	200	1.0008	.1881	.1893	.1891	.1892	.1890
	500	1.0120	.1896	.1902	.1901	.1900	.1900
	∞	1.0000	.1900	.1900	.1900	.1900	.1900

CHAPTER 3

IMPROVED GMM AND 3SLS ESTIMATORS FOR SYSTEM OF EQUATIONS

3.1. Introduction

In section 2.2 of Chapter 2 we defined the improved GMM (IGMM) estimator as the GMM estimator using moment conditions $E[\phi_1(y_t^*,\theta_0)-C_{12}C_{22}^{-1}\phi_2(y_t^*)]=0$ and weighting matrix $C^{11}=(C_{11}-C_{12}C_{22}^{-1}C_{21})^{-1}$. In the definition, we intentionally did not specify the functional forms of ϕ_1 and ϕ_2, nor did we require the observations $\{\phi(y_t^*,\theta)=(\phi_1(y_t^*,\theta)',\phi_2(y_t^*)')',\, t=1,2,\cdots\}$ to be conditionally homoskedastic or serially uncorrelated so long as they satisfy suitable regularity conditions. In section 2.3 of Chapter 2 we applied the general results of IGMM estimator to the case of the linear regression model. Assuming conditional homoskedasticity and serial uncorrelation, and imposing the regularity conditions (A.2.1)-(A.2.3), we obtained an explicit formula for the IGMM estimator and related it to other previously known estimators, such as the estimator of Schmidt (1988).

In this chapter we will provide a similar analysis for a system of linear equations. We will first set up the model and make "high-level" assumptions of regularity conditions. Under these assumptions we derive an explicit formula for the IGMM estimator and several other asymptotically equivalent estimators, and demonstrate the efficiency of these estimators relative to the usual three-stage least sqares (3SLS) estimator.

3

(

v

c

p

o

(3

(3

(3

(3

Th

3.2. Model and Notation

The model considered in this chaper is

(3.1) $$y_{tg} = x_{tg}'\theta_{0g} + \varepsilon_{tg}, \quad g = 1, 2, \ldots\ldots, G;\ t = 1, 2, \ldots\ldots, T,$$

where y_{tg} is the dependent variable of equation g at observation t, x_{tg} is the $K_g \times 1$ vector of explanatory variables of equation g at observation t, θ_{0g} is the $K_g \times 1$ unknown parameter vector of equation g, and ε_{tg} is the model disturbance of equation g at observation t. We assume that in general $\mathrm{cov}(x_{tg}, \varepsilon_{tg}) \neq 0$ for g = 1, 2, ..., G.

We define the following notation:

(3.2A) $$y_t = \begin{bmatrix} y_{t1} \\ \vdots \\ y_{tG} \end{bmatrix}, \ X_t = \begin{bmatrix} x_{t1}' & & \\ & \ddots & \\ & & x_{tG}' \end{bmatrix}, \ \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_G \end{bmatrix}, \ \varepsilon_t = \begin{bmatrix} \varepsilon_{t1} \\ \vdots \\ \varepsilon_{tG} \end{bmatrix}, \ t = 1, 2, \ldots, T;$$

(3.2B) $$y_{(g)} = \begin{bmatrix} y_{1g} \\ \vdots \\ y_{Tg} \end{bmatrix}, \ X_{(g)} = \begin{bmatrix} x_{1g}' \\ \vdots \\ x_{Tg}' \end{bmatrix}, \ \varepsilon_{(g)} = \begin{bmatrix} \varepsilon_{1g} \\ \vdots \\ \varepsilon_{Tg} \end{bmatrix}, \ g = 1, 2, \ldots, G;$$

(3.2C) $$y_* = \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(G)} \end{bmatrix}, \ X_* = \begin{bmatrix} X_{(1)} & & \\ & \ddots & \\ & & X_{(G)} \end{bmatrix}, \ \varepsilon_* = \begin{bmatrix} \varepsilon_{(1)} \\ \vdots \\ \varepsilon_{(G)} \end{bmatrix};$$

(3.2D) $$\varepsilon = \begin{bmatrix} \varepsilon_1' \\ \vdots \\ \varepsilon_T' \end{bmatrix}.$$

Then (3.1) can be rewritten as

(3.3A) $$y_t = X_t\theta_0 + \varepsilon_t, \quad t = 1, 2, \ldots\ldots, T,$$

or as

(3.3B) $$y_* = X_*\theta_0 + \varepsilon_*.$$

3.3. Improved GMM Estimators

Suppose that we have available an $M \times 1$ vector of instruments z_t satisfying the moment conditions $E(\varepsilon_t \otimes z_t) = 0$, with $E[(I_G \otimes z_t)X_t]$ having full column rank. Suppose that we also have available an $L \times 1$ vector of observable variables u_t satisfying $E(u_t \otimes z_t) = 0$ and $E(u_t\varepsilon_t') = \Sigma_{u\varepsilon} \neq 0$. Then the additional moment conditions $E(u_t \otimes z_t) = 0$ will help us to improve the efficiency of estimation of θ_0. Using the notation of Chapter 2, we have

(3.4) $$\phi(y_t^*, \theta) = \begin{bmatrix} \phi_1(y_t^*, \theta) \\ \phi_2(y_t^*) \end{bmatrix} = \begin{bmatrix} (I_G \otimes z_t)(y_t - X_t\theta) \\ (I_L \otimes z_t)u_t \end{bmatrix}$$

where the observed data vector is $y_t^* = (y_t'; x_{t1}', \ldots, x_{tG}'; u_t'; z_t')'$. Then

(3.5) $$\phi_T(\theta) = \begin{bmatrix} \phi_{T1}(\theta) \\ \phi_{T2} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}\phi(y_t^*, \theta) = \begin{bmatrix} T^{-1}(I_G \otimes Z')(y_* - X_*\theta) \\ T^{-1}(I_L \otimes Z')u_* \end{bmatrix}$$

where $Z = (z_1, \ldots, z_T)'$ and $u_* = \text{vec}(U) \equiv (u_{(1)}', \ldots, u_{(L)}')'$ with $U = (u_1, \ldots, u_T)'$. Then

(3.6A) $$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \lim_{T\to\infty}[T \cdot E\phi_T(\theta_0)\phi_T(\theta_0)']$$

$$= \lim_{T\to\infty} E\begin{bmatrix} T^{-1}(I_G \otimes Z')\varepsilon_*\varepsilon_*'(I_G \otimes Z) & T^{-1}(I_G \otimes Z')\varepsilon_* u_*'(I_L \otimes Z) \\ T^{-1}(I_L \otimes Z')u_*\varepsilon_*'(I_G \otimes Z) & T^{-1}(I_L \otimes Z')u_* u_*'(I_L \otimes Z) \end{bmatrix}$$

(3.6B) $$D_{T1} = \frac{\partial \phi_{T1}(\theta)}{\partial \theta'} = -\frac{1}{T}(I_G \otimes Z')X_* \,.$$

Substituting (3.5) and (3.6) into the first order condition (2.7A) in Chapter 2 for the IGMM estimator $\hat{\theta}$, and replacing C^{11}, C_{12} and C_{22}^{-1} by consistent estimates $\hat{C}^{11}$, $\hat{C}_{12}$ and $\hat{C}_{22}^{-1}$ respectively, we arrive at:

(3.7) $$[X_*'(I_G \otimes Z)]\hat{C}^{11}[(I_G \otimes Z')(y_* - X_*\hat{\theta}) - \hat{C}_{12}\hat{C}_{22}^{-1}(I_L \otimes Z')u_*] = 0.$$

Solving for $\hat{\theta}$, we obtain

(3.8) $$\hat{\theta} = [X_*'(I_G \otimes Z)\hat{C}^{11}(I_G \otimes Z')X_*]^{-1} X_*'(I_G \otimes Z)\hat{C}^{11} \cdot[(I_G \otimes Z')y_* - \hat{C}_{12}\hat{C}_{22}^{-1}(I_L \otimes Z')u_*].$$

In order to simplify the above expression further, we need to put more structure on C. This is possible under the assumption of conditional homoskedasticity and uncorrelatedness. Suppose that conditional on $\Omega_t = \{z_t ; \varepsilon_{t-1}, u_{t-1}, z_{t-1} ; \ldots\}$, the $(\varepsilon_t', u_t')'$ are mutually uncorrelated and that:

(3.9) $$V(\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} | z_t) = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix} \equiv \Sigma.$$

Then $C = \Sigma \otimes E(z_t z_t')$ and $\hat{C} = \hat{\Sigma} \otimes T^{-1}Z'Z$ is a consistent estimate of C, where $\hat{\Sigma}$ is any consistent estimate of Σ. For the moment we will treat Σ as known, for simplicity. Therefore we have

(3.10A) $\hat{C}^{11} = \Sigma^{\varepsilon\varepsilon} \otimes (T^{-1}Z'Z)^{-1}$ with $\Sigma^{\varepsilon\varepsilon} = (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})^{-1}$

(3.10B) $\hat{C}_{12} = \Sigma_{\varepsilon u} \otimes T^{-1}Z'Z$

(3.10C) $\hat{C}_{22}^{-1} = \Sigma_{uu}^{-1} \otimes (T^{-1}Z'Z)^{-1}$

(3.10D) $\hat{C}_{12}\hat{C}_{22}^{-1} = \Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes I_M$.

Substituting (3.10A) and (3.10D) into (3.8), we get

(3.11)
$$\tilde{\theta} = [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)X_*]^{-1}X_*'[\Sigma^{\varepsilon\varepsilon} \otimes Z(Z'Z)^{-1}] \cdot [(I_G \otimes Z')y_* - (\Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes I_M)(I_L \otimes Z')u_*].$$

Noticing that

(3.12)
$$\begin{aligned}(\Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes I_M)(I_L \otimes Z')u_* &= (\Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes Z')\mathrm{vec}(U)\\ &= \mathrm{vec}(Z'U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})\\ &= (I_G \otimes Z'U)\mathrm{vec}(\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})\\ &= (I_G \otimes Z')(I_G \otimes U)\mathrm{vec}(\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}),\end{aligned}$$

and substituting into (3.11), we obtain an explicit formula for the IGMM estimator $\tilde{\theta}$:

(3.13)
$$\tilde{\theta} = [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)X_*]^{-1}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)[y_* - (I_G \otimes U)\lambda]$$

where $\lambda = \mathrm{vec}(\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) \equiv (\lambda_1', \lambda_2', \cdots, \lambda_G')'$. Thus, for i = 1, 2, ..., G, $\lambda_i = \Sigma_{uu}^{-1}$ times the i^{th} column of $\Sigma_{u\varepsilon}$; equivalently, $\lambda_i = (\mathrm{p}\lim T^{-1}U'U)^{-1}\mathrm{p}\lim T^{-1}U'\varepsilon_{(i)}$.

We can compare the IGMM estimator in (3.13) with the usual GMM (3SLS) estimator

$$\hat{\theta} = [X_*'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes P_Z)X_*]^{-1} X_*'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes P_Z)y_* \tag{3.14}$$

based on moment conditions $E[\phi_1(y_t^*, \theta_0)] = 0$. We see that the only difference between the 3SLS estimator and the IGMM estimator is that $\Sigma_{\varepsilon\varepsilon}^{-1}$ and y_* in (3.14) are replaced by $\Sigma^{\varepsilon\varepsilon}$ and $[y_* - (I_G \otimes U)\lambda]$ respectively. It is interesting to notice that $[y_* - (I_G \otimes U)\lambda] = [(y_{(1)} - U\lambda)', \cdots, (y_{(G)} - U\lambda)']'$ is just a vector of residuals from the linear projection of $y_{(g)}$ onto U. Thus the IGMM estimator $\tilde{\theta}$ in (3.13) can also be regarded as a purged GMM (PGMM) estimator.

We will now consider a specific form for a consistent estimate of λ. Define

$$\hat{\lambda} = (\hat{\lambda}_1', \cdots, \hat{\lambda}_G')' \tag{3.15}$$

with $\hat{\lambda}_g = (T^{-1}U'U)^{-1}T^{-1}U'(y_{(g)} - X_{(g)}\hat{\theta}_g)$, where $\hat{\theta}_g$ is any consistent estimate of θ_g, for $g = 1, \ldots, G$. Then

$$(I_G \otimes U)\hat{\lambda} = (I_G \otimes U)\begin{bmatrix} (U'U)^{-1}U'(y_{(1)} - X_{(1)}\hat{\theta}_{(1)}) \\ \vdots \\ (U'U)^{-1}U'(y_{(G)} - X_{(G)}\hat{\theta}_{(G)}) \end{bmatrix} \tag{3.16}$$

$$= (I_G \otimes U)(I_G \otimes (U'U)^{-1}U')(y_* - X_*\hat{\theta})$$

$$= (I_G \otimes P_U)(y_* - X_*\hat{\theta})$$

where $\hat{\theta} = (\hat{\theta}_1', \ldots, \hat{\theta}_G')'$. Substituting the above expression into (3.13), we get

$$\tilde{\theta} = [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)X_*]^{-1} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)[y_* - (I_G \otimes P_U)(y_* - X_*\hat{\theta})]. \tag{3.17}$$

This expression still depends on $\Sigma^{\varepsilon\varepsilon}$, and we will discuss its consistent estimation later.

While (3.17) is defined for any consistent estimate $\hat{\theta}$, we may as well consider the special case that $\hat{\theta} = \tilde{\theta}$. Then (3.17) implies

$$(3.18) \qquad [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)X_*]\tilde{\theta} = X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)[y_* - (I_G \otimes P_U)(y_* - X_*\tilde{\theta})].$$

Solving for $\tilde{\theta}$, we obtain

$$(3.19) \qquad \tilde{\theta} = [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)X_*]^{-1} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)y_*.$$

This is an obvious generalization of the single-equation IGMM estimator $\tilde{\theta}$ of Chapter 2.

In order to be more precise about the sense in which the IGMM estimator $\tilde{\theta}$ dominates the usual GMM (3SLS) estimator $\hat{\theta}$, and to introduce some other equally asymptotically efficient estimators, we make some more explicit assumptions. To make the asymptotics as simple as possible, we will make the following "high level" assumptions:

$$(A3.1) \qquad \text{plim} \frac{1}{T} \begin{bmatrix} X_{(1)}' \\ \vdots \\ X_{(G)}' \\ Z' \\ \varepsilon' \\ U' \end{bmatrix} \begin{bmatrix} X_{(1)} & \cdots & X_{(G)} & Z & \varepsilon & U \end{bmatrix}$$

$$= \begin{bmatrix} A_{11} & \cdots & A_{1G} & A_{1z} & A_{1\varepsilon} & A_{1u} \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ A_{G1} & \cdots & A_{GG} & A_{Gz} & A_{G\varepsilon} & A_{Gu} \\ A_{z1} & \cdots & A_{zG} & A_{zz} & 0 & 0 \\ A_{\varepsilon 1} & \cdots & A_{\varepsilon G} & 0 & \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ A_{u1} & \cdots & A_{uG} & 0 & \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix} \text{ exists.}$$

(A3.2) A_{zz}, Σ_{uu} and $\Sigma = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix}$ are nonsingular; A_{zg} has full column rank for g = 1, 2, ..., G.

(A3.3) $$\frac{1}{\sqrt{T}}(I_{G+L} \otimes Z')\begin{bmatrix} \varepsilon_* \\ u_* \end{bmatrix} \to N(0, \Sigma \otimes A_{zz}).$$

As was the case in Chapter 2, these high-level assumptions are derivable from various sets of more basic assumptions. For example, let $e_t = (\varepsilon_t', u_t')'$ and $\Omega_t = \{z_t; z_{t-1}, e_{t-1}; z_{t-2}, e_{t-2}; ......\}$; then (A3.1)-(A3.3) follow from the assumptions that $E(e_t|\Omega_t) = 0$, $V(e_t|\Omega_t) = \Sigma$, and X_t and z_t are covariance stationary.

It is well known that under (A3.1)-(A3.3), the usual GMM (3SLS) estimator $\hat{\theta}$ defined in (3.14) has the following asymptotic variance:

(3.20) $$AV[\sqrt{T}(\hat{\theta} - \theta_0)] = [p\lim \frac{1}{T} X_*'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes P_Z)X_*]^{-1} = [A'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes A_{zz}^{-1})A]^{-1}$$

where $A = p\lim \frac{1}{T}(I_G \otimes Z')X_* = \begin{bmatrix} A_{z1} & & \\ & \ddots & \\ & & A_{zG} \end{bmatrix}$.

We now wish to show that several estimators are asymptotically equally efficient, and that they are efficient relative to the 3SLS estimator. One such estimator is the IGMM estimator $\hat{\theta}$ defined in (3.19). The other such estimator is the PGMM estimator defined in (3.13) with λ known. In order to distinguish the PGMM estimator from the IGMM estimator, we now denote the PGMM estimator by $\ddot{\theta}$. We will also consider the following two additional estimators

(3.21A) $\dddot{\theta} = [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})X_*]^{-1} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})y_*$

(3.21B) $\breve{\theta} = [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)X_*]^{-1} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)y_*$.

We will show that $\bar{\theta}$, $\ddot{\theta}$, $\dddot{\theta}$ and $\breve{\theta}$ are asymptotically equivalent, with asymptotic variance matrix equal to

(3.22) $[\text{p}\lim \frac{1}{T} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)X_*]^{-1} = [A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A]^{-1} \equiv B^{-1}$.

Comparing to the asymptotic variance matrix of $\hat{\theta}$ in (3.20) above, the fact that the matrix $\{[A'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes A_{zz}^{-1})A]^{-1} - [A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A]^{-1}\}$ is positive semidefinite (shown later in Theorem 3.3) establishes the asymptotic efficiency of $\bar{\theta}$, $\ddot{\theta}$, $\dddot{\theta}$ and $\breve{\theta}$ relative to $\hat{\theta}$. We now turn to a rigorous proof of these results.

<u>LEMMA 3.1</u>: $\text{p}\lim T^{-1}Z'M_U X_{(g)} = \text{p}\lim T^{-1}Z'X_{(g)} = A_{zg}$, for g = 1, 2, ..., G.

$$\text{p}\lim T^{-1}Z'M_U Z = \text{p}\lim T^{-1}Z'Z = A_{zz}$$

Proof: The proof is similar to the proof of Lemma 2.1 of Chapter 2. For example,

$$\begin{aligned}\text{p}\lim T^{-1}Z'M_U X_{(g)} &= \text{p}\lim[T^{-1}Z'X_{(g)} - (T^{-1}Z'U)(T^{-1}U'U)^{-1}(T^{-1}U'X_{(g)})] \\ &= A_{zg} - 0 \cdot \Sigma_{uu}^{-1} A_{ug} = A_{zg},\end{aligned}$$

using (A3.1) and (A3.2).

■

<u>LEMMA 3.2</u>: $\text{p}\lim T^{-1}X_{(h)}'P_Z M_U X_{(g)} = \text{p}\lim T^{-1}X_{(h)}'P_{[M_U Z]}X_{(g)}$

$$= p\lim T^{-1}X_{(h)}'M_U P_Z M_U X_{(g)}$$

$$= p\lim T^{-1}X_{(h)}'P_Z X_{(g)}$$

$$= A_{hz}A_{zz}^{-1}A_{zg}$$

for h, g = 1, 2, ..., G.

Proof: The proof is essentially the same as the proof of Lemma 2.2 of Chapter 2. For example,

$$p\lim T^{-1}X_{(h)}'P_Z M_U X_{(g)} = p\lim[T^{-1}X_{(h)}'Z](T^{-1}Z'Z)^{-1}[T^{-1}Z'M_U X_{(g)}]$$
$$= A_{hz}A_{zz}^{-1}A_{zg}$$

using (A3.1), (A3.2) and Lemma 3.1.

■

<u>LEMMA 3.3</u>: $p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_Z M_U)X_*] = p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_{[M_U Z]})X_*]$

$$= p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes M_u P_Z M_U)X_*]$$

$$= p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_Z)X_*]$$

$$= A'(\Sigma^{\varepsilon\varepsilon}\otimes A_{zz}^{-1})A \equiv B,$$

and B is nonsingular.

Proof: Let $\Sigma^{\varepsilon\varepsilon} = (\sigma^{ij})_{G\times G}$. Then

$$p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_{[M_U Z]})X_*]$$

$$= p\lim T^{-1}\begin{bmatrix} X_{(1)}' & & \\ & \ddots & \\ & & X_{(G)}' \end{bmatrix}\begin{bmatrix} \sigma^{11}P_{[M_U Z]} & \cdots & \sigma^{1G}P_{[M_U Z]} \\ \vdots & \vdots & \vdots \\ \sigma^{G1}P_{[M_U Z]} & \cdots & \sigma^{GG}P_{[M_U Z]} \end{bmatrix}\begin{bmatrix} X_{(1)} & & \\ & \ddots & \\ & & X_{(G)} \end{bmatrix}$$

$$= p\lim T^{-1}\begin{bmatrix} X_{(1)}'\sigma^{11}P_{[M_U Z]}X_{(1)} & \cdots & X_{(1)}'\sigma^{1G}P_{[M_U Z]}X_{(G)} \\ \vdots & \vdots & \vdots \\ X_{(G)}'\sigma^{G1}P_{[M_U Z]}X_{(1)} & \cdots & X_{(G)}'\sigma^{GG}P_{[M_U Z]}X_{(G)} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11}A_{1z}A_{zz}^{-1}A_{z1} & \cdots & \sigma^{1G}A_{1z}A_{zz}^{-1}A_{zG} \\ \vdots & \vdots & \vdots \\ \sigma^{G1}A_{Gz}A_{zz}^{-1}A_{z1} & \cdots & \sigma^{GG}A_{Gz}A_{zz}^{-1}A_{zG} \end{bmatrix} \text{(using Lemma 3.2)}$$

$$= \begin{bmatrix} A_{1z} & & \\ & \ddots & \\ & & A_{Gz} \end{bmatrix} \begin{bmatrix} \sigma^{11}A_{zz}^{-1} & \cdots & \sigma^{1G}A_{zz}^{-1} \\ \vdots & \cdots & \vdots \\ \sigma^{G1}A_{zz}^{-1} & \cdots & \sigma^{GG}A_{zz}^{-1} \end{bmatrix} \begin{bmatrix} A_{z1} & & \\ & \ddots & \\ & & A_{zG} \end{bmatrix}$$

$$= A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A$$

using the definition of A in (3.20). The probability limits for the other cases involve essentially the same arguments. Finally, $B = A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A$ is nonsingular bacause Σ, Σ_{uu} and A_{zz} are nonsingular, which implies $\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1}$ nonsingular, and because A has full column rank (see (A3.2)).

■

LEMMA 3.4: $$\begin{aligned} \text{plim}\, T^{-1}X_{(h)}'P_Z M_U \varepsilon_{(g)} &= \text{plim}\, T^{-1}X_{(h)}'P_{[M_U Z]}\varepsilon_{(g)} \\ &= \text{plim}\, T^{-1}X_{(h)}'M_U P_Z M_U \varepsilon_{(g)} \\ &= \text{plim}\, T^{-1}X_{(h)}'P_Z(\varepsilon_{(g)} - U\lambda) \\ &= \text{plim}\, T^{-1}X_{(h)}'P_Z \varepsilon_{(g)} \\ &= 0 \end{aligned}$$

where h, g = 1, 2, ..., G.

Proof: The proof is similar to the proof of Lemma 2.3 of Chapter 2. For example,

$$\begin{aligned} \text{plim}\, T^{-1}X_{(h)}'M_U P_Z M_U \varepsilon_{(g)} &= \text{plim}[T^{-1}X_{(h)}'M_U Z](T^{-1}Z'Z)^{-1}[T^{-1}Z'M_U\varepsilon_{(g)}] \\ &= A_{hz}A_{zz}^{-1} \cdot \text{plim}\, T^{-1}Z'M_U\varepsilon_{(g)} \end{aligned}$$

using Lemma 3.1. But

$$\begin{aligned} \text{plim}\, T^{-1}Z'M_U\varepsilon_{(g)} &= \text{plim}\, T^{-1}Z'\varepsilon_{(g)} - \text{plim}(T^{-1}Z'U)(T^{-1}U'U)^{-1}[T^{-1}U'\varepsilon_{(g)}] \\ &= 0 - 0 \cdot \Sigma_{uu}^{-1}\Sigma_{u\varepsilon,g} = 0 \end{aligned}$$

using (A3.1) and (A3.2), where $\Sigma_{u\varepsilon,g}$ is the g-th column of $\Sigma_{u\varepsilon}$. Therefore

$$p\lim T^{-1}X_{(h)}'M_U P_Z M_U \varepsilon_{(g)} = A_{hz}A_{zz}^{-1}\cdot 0 = 0.$$

The proofs for the other cases are similar.

■

<u>LEMMA 3.5</u>: $p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_Z M_U)\varepsilon_*]$

$$= p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_{[M_U Z]})\varepsilon_*]$$

$$= p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes M_U P_Z M_U)\varepsilon_*]$$

$$= p\lim\{T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_Z)[\varepsilon_* - (I_G\otimes U)\lambda]\}$$

$$= p\lim[T^{-1}X_*'(\Sigma_{\varepsilon\varepsilon}^{-1}\otimes P_Z)\varepsilon_*]$$

$$= 0$$

Proof: Let $\Sigma^{\varepsilon\varepsilon} = (\sigma^{ij})_{G\times G}$ as above. Then

$$p\lim[T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon}\otimes P_{[M_U Z]})\varepsilon_*]$$

$$= p\lim T^{-1}\begin{bmatrix} X_{(1)}' & & \\ & \ddots & \\ & & X_{(G)}' \end{bmatrix}\begin{bmatrix} \sigma^{11}P_{[M_U Z]} & \cdots & \sigma^{1G}P_{[M_U Z]} \\ \vdots & \vdots & \vdots \\ \sigma^{G1}P_{[M_U Z]} & \cdots & \sigma^{GG}P_{[M_U Z]} \end{bmatrix}\begin{bmatrix} \varepsilon_{(1)} \\ \vdots \\ \varepsilon_{(G)} \end{bmatrix}$$

$$= p\lim T^{-1}\begin{bmatrix} \sum_{g=1}^{G} X_{(1)}'\sigma^{1g}P_{[M_U Z]}\varepsilon_{(g)} \\ \vdots \\ \sum_{g=1}^{G} X_{(G)}'\sigma^{Gg}P_{[M_U Z]}\varepsilon_{(g)} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{g=1}^{G} \sigma^{1g} p\lim T^{-1}X_{(1)}'P_{[M_U Z]}\varepsilon_{(g)} \\ \vdots \\ \sum_{g=1}^{G} \sigma^{Gg} p\lim T^{-1}X_{(G)}'P_{[M_U Z]}\varepsilon_{(g)} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{g=1}^{G}\sigma^{1g}\cdot 0 \\ \vdots \\ \sum_{g=1}^{G}\sigma^{Gg}\cdot 0 \end{bmatrix} \quad \text{(using Lemma 3.4)}$$

$= 0.$

The proofs for the other cases are similar.

■

THEOREM 3.1: The improved GMM estimators $\hat{\theta}$, $\ddot{\theta}$, $\dddot{\theta}$ and $\breve{\theta}$ are consistent.

Proof: We will give the proof for $\hat{\theta}$. The other proofs are quite similar.

$$\begin{aligned}
\text{p}\lim \hat{\theta} &= \text{p}\lim[X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)X_*]^{-1} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)y_* \\
&= \text{p}\lim[X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)X_*]^{-1} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)(X_*\theta_0 + \varepsilon_*) \\
&= \theta_0 + [\text{p}\lim X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)X_*]^{-1}[\text{p}\lim X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)\varepsilon_*] \\
&= \theta_0 + B^{-1} \cdot 0 \quad \text{(using Lemmas 3.3 and 3.5)} \\
&= \theta_0.
\end{aligned}$$

■

LEMMA 3.6: $T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_* \to N[0, (\Sigma^{\varepsilon\varepsilon})^{-1} \otimes A_{zz}]$

where $\Sigma^{\varepsilon\varepsilon} = (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})^{-1}$.

Proof:
$$\begin{aligned}
T^{-1/2}Z'M_U\varepsilon_{(g)} &= T^{-1/2}Z'\varepsilon_{(g)} - (T^{-1/2}Z'U)(T^{-1}U'U)^{-1}(T^{-1}U'\varepsilon_{(g)}) \\
&= T^{-1/2}Z'\varepsilon_{(g)} - (T^{-1/2}Z'U)\lambda_g + o_p(1) \\
&= T^{-1/2}Z'(\varepsilon_{(g)} - U\lambda_g) + o_p(1)
\end{aligned}$$

for g = 1, 2, ..., G, where $\lambda_g = \Sigma_{uu}^{-1} \cdot E(u_t\varepsilon_{tg})$. Then

$$T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_* = \begin{bmatrix} T^{-1/2}Z'M_U\varepsilon_{(1)} \\ \vdots \\ T^{-1/2}Z'M_U\varepsilon_{(G)} \end{bmatrix}$$

$$= \begin{bmatrix} T^{-1/2}Z'(\varepsilon_{(1)} - U\lambda_1) \\ \vdots \\ T^{-1/2}Z'(\varepsilon_{(G)} - U\lambda_G) \end{bmatrix} + o_p(1)$$

$$= T^{-1/2}(I_G \otimes Z')[\varepsilon_* - (I_G \otimes U\Sigma_{uu}^{-1})\mathrm{vec}(\Sigma_{u\varepsilon})] + o_p(1)$$

(using the definition of $\lambda_g = \Sigma_{uu}^{-1} \cdot E(u_t \varepsilon_{tg})$)

$$= T^{-1/2}(I_G \otimes Z')[\mathrm{vec}(\varepsilon) - \mathrm{vec}(U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})] + o_p(1)$$

$$= T^{-1/2}(I_G \otimes Z')\mathrm{vec}[\varepsilon - U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}] + o_p(1)$$

$$= T^{-1/2}\mathrm{vec}[Z'(\varepsilon - U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})] + o_p(1)$$

$$= T^{-1/2}\mathrm{vec}\{Z'(\varepsilon, U)\begin{bmatrix} I_G \\ -\Sigma_{uu}^{-1}\Sigma_{u\varepsilon} \end{bmatrix}\} + o_p(1)$$

$$= T^{-1/2}([I_G, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}] \otimes I_M)\mathrm{vec}[Z'(\varepsilon, U)] + o_p(1)$$

$$= T^{-1/2}([I_G, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}] \otimes I_M)(I_{G+L} \otimes Z')\mathrm{vec}(\varepsilon, U)$$

$$+ o_p(1)$$

$$= ([I_G, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}] \otimes I_M)\{T^{-1/2}(I_{G+L} \otimes Z')\begin{bmatrix} \varepsilon_* \\ u_* \end{bmatrix}\}$$

$$+ o_p(1).$$

But according to assumption (A3.3) above, $T^{-1/2}(I_{G+L} \otimes Z')\begin{bmatrix} \varepsilon_* \\ u_* \end{bmatrix} \to N(0, \Sigma \otimes A_{zz})$.

Therefore

$$T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_* \to N(0, W)$$

with

$$W = ([I_G, -\Sigma_{\varepsilon u}\Sigma_{uu}^{-1}] \otimes I_M)(\Sigma \otimes A_{zz})(\begin{bmatrix} I_G \\ -\Sigma_{uu}^{-1}\Sigma_{u\varepsilon} \end{bmatrix} \otimes I_M)$$

$$= (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) \otimes A_{zz}$$

$$= (\Sigma^{\varepsilon\varepsilon})^{-1} \otimes A_{zz}.$$

■

<u>LEMMA 3.7</u>: $T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)\varepsilon_*$, $T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})\varepsilon_*$, $T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)\varepsilon_*$ and $T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)[\varepsilon_* - (I_G \otimes U)\lambda]$ each converge in distribution to $N[0, A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A]$ with $A = \mathrm{diag}(A_{z1}, \ldots\ldots, A_{zG})$ as defined in (3.20).

Proof: We give the proof for $T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)\varepsilon_*$. The proofs for the other cases are quite similar.

$$T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)\varepsilon_*$$

$$= T^{-1/2}\begin{bmatrix} X_{(1)}' & & \\ & \ddots & \\ & & X_{(G)}' \end{bmatrix}\begin{bmatrix} \sigma^{11}M_U P_Z M_U & \cdots & \sigma^{1G}M_U P_Z M_U \\ \vdots & \vdots & \vdots \\ \sigma^{G1}M_U P_Z M_U & \cdots & \sigma^{GG}M_U P_Z M_U \end{bmatrix}\begin{bmatrix} \varepsilon_{(1)} \\ \vdots \\ \varepsilon_{(G)} \end{bmatrix}$$

$$= T^{-1/2}\begin{bmatrix} \sum_{g=1}^{G} X_{(1)}'\sigma^{1g}M_U P_Z M_U \varepsilon_{(g)} \\ \vdots \\ \sum_{g=1}^{G} X_{(G)}'\sigma^{Gg}M_U P_Z M_U \varepsilon_{(g)} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{g=1}^{G} \sigma^{1g} T^{-1/2} X_{(1)}' M_U P_Z M_U \varepsilon_{(g)} \\ \vdots \\ \sum_{g=1}^{G} \sigma^{Gg} T^{-1/2} X_{(G)}' M_U P_Z M_U \varepsilon_{(g)} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{g=1}^{G} \sigma^{1g} (T^{-1}X_{(1)}' M_U Z)(T^{-1}Z'Z)^{-1}[T^{-1/2}Z'M_U \varepsilon_{(g)}] \\ \vdots \\ \sum_{g=1}^{G} \sigma^{Gg} (T^{-1}X_{(G)}' M_U Z)(T^{-1}Z'Z)^{-1}[T^{-1/2}Z'M_U \varepsilon_{(g)}] \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{g=1}^{G} \sigma^{1g} A_{1z}A_{zz}^{-1} \cdot T^{-1/2}Z'M_U \varepsilon_{(g)} \\ \vdots \\ \sum_{g=1}^{G} \sigma^{Gg} A_{Gz}A_{zz}^{-1} \cdot T^{-1/2}Z'M_U \varepsilon_{(g)} \end{bmatrix} + o_p(1) \quad \text{(using Lemma 3.1)}$$

$$= \begin{bmatrix} A_{1z}A_{zz}^{-1} & & \\ & \ddots & \\ & & A_{Gz}A_{zz}^{-1} \end{bmatrix}(\Sigma^{\varepsilon\varepsilon} \otimes I_M)[T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_*] + o_p(1)$$

$$= A'(I_G \otimes A_{zz}^{-1})(\Sigma^{\varepsilon\varepsilon} \otimes I_M)[T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_*] + o_p(1)$$

$$= A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})[T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_*] + o_p(1).$$

But according to Lemma 3.6: $T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_* \to N[0,(\Sigma^{\varepsilon\varepsilon})^{-1} \otimes A_{zz}]$. Then

$$T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)\varepsilon_* \to N[0, V]$$

where

$$V = A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})[(\Sigma^{\varepsilon\varepsilon})^{-1} \otimes A_{zz}][A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})]' = A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A$$

which is the same as the matrix B defined in (3.22) above.

■

<u>THEOREM 3.2</u>: $\sqrt{T}(\overset{\frown}{\theta}-\theta_0)$, $\sqrt{T}(\ddot{\theta}-\theta_0)$, $\sqrt{T}(\dddot{\theta}-\theta_0)$ and $\sqrt{T}(\breve{\theta}-\theta_0)$ each converge in distribution to $N(0, B^{-1})$ with $B = A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A = \text{p}\lim T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)X_*$.

Proof: We will give the proof for $\sqrt{T}(\overset{\frown}{\theta}-\theta_0)$. The proofs for the other cases are essentially identical.

$$\begin{aligned}\sqrt{T}(\overset{\frown}{\theta}-\theta_0) &= [T^{-1}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)X_*]^{-1} T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)\varepsilon_* \\ &= B^{-1} \cdot T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)\varepsilon_* + o_p(1)\end{aligned}$$

using Lemma 3.3. But according to Lemma 3.7:

$$T^{-1/2}X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)\varepsilon_* \to N(0, B).$$

Therefore

$$\sqrt{T}(\overset{\frown}{\theta}-\theta_0) \to N(0, \Lambda)$$

with

$$\Lambda = B^{-1}B(B^{-1})' = (B^{-1})' = B^{-1}.$$

■

<u>THEOREM 3.3</u>: The improved GMM estimators: $\overset{\frown}{\theta}$, $\ddot{\theta}$, $\dddot{\theta}$ and $\breve{\theta}$ are asymptotically efficient relative to the 3SLS estimator $\hat{\theta}$. They are strictly more efficient than the 3SLS estimator if $\Sigma_{u\varepsilon} \neq 0$.

Proof: From Theorem 3.2, the asymptotic variance matrix of each of the IGMM estimators is $B^{-1} = [A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A]^{-1}$. The asymptotic variance matrix of the 3SLS estimator is $Q^{-1} = [A'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes A_{zz}^{-1})A]^{-1}$. We wish to show that $(Q^{-1} - B^{-1})$ is positive semidefinite (psd), and positive definite (pd) when $\Sigma_{u\varepsilon} \neq 0$. This is equivalent to showing that (B - Q) is psd, and pd when $\Sigma_{u\varepsilon} \neq 0$. But

$$B - Q = A'[(\Sigma^{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon}^{-1}) \otimes A_{zz}^{-1}]A.$$

When $\Sigma_{u\varepsilon} = 0$, $\Sigma^{\varepsilon\varepsilon} = \Sigma_{\varepsilon\varepsilon}^{-1}$ and B - Q = 0; there is no efficiency gain for IGMM relative to 3SLS. But when $\Sigma_{u\varepsilon} \neq 0$, $\Sigma^{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon}^{-1}$ is pd; A_{zz}^{-1} is pd; this implies that $(\Sigma^{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon}^{-1}) \otimes A_{zz}^{-1}$ is pd; and B - Q is pd because $(\Sigma^{\varepsilon\varepsilon} - \Sigma_{\varepsilon\varepsilon}^{-1}) \otimes A_{zz}^{-1}$ is pd and A has full column rank.

■

Theorem 3.3 is best understood by the intuitive explanation of the following theorem.

THEOREM 3.4: Consider the augmented system

(3.23) $$y_{tg} = x_{tg}'\theta_{0g} + u_t'\lambda_g + v_{tg}, \quad g = 1, 2, ..., G; t = 1, 2, ..., T,$$

where $\lambda_g = [E(u_t u_t')]^{-1} E(u_t \varepsilon_{tg})$ is the linear projection coefficient of ε_{tg} onto u_t, and $v_{tg} = \varepsilon_{tg} - u_t'\lambda_g$ is the error in the linear projection. Define $\hat{\theta}_{3SLS}$ as the estimator of $\theta_0 = (\theta_{01}', \ldots, \theta_{0G}')'$ when the system (3.23) is estimated by 3SLS, using $(z_t', u_t')'$ as instruments. Then $\hat{\theta}_{3SLS}$ is numerically the same as the IGMM estimator $\dddot{\theta}$ defined in (3.21A) above.

Proof: Let $v_t = (v_{t1}, \ldots, v_{tG})'$. Then

$$v_t = \begin{bmatrix} \varepsilon_{t1} - u_t'\lambda_1 \\ \vdots \\ \varepsilon_{tG} - u_t'\lambda_G \end{bmatrix} = \varepsilon_t - \begin{bmatrix} \lambda_1' \\ \vdots \\ \lambda_G' \end{bmatrix} u_t = \varepsilon_t - \begin{bmatrix} E(\varepsilon_{t1} u_t')[E(u_t' u_t)]^{-1} \\ \vdots \\ E(\varepsilon_{tG} u_t')[E(u_t' u_t)]^{-1} \end{bmatrix} u_t$$

$$= \varepsilon_t - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}u_t.$$

Therefore

$$V(v_t) = \Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon} = (\Sigma^{\varepsilon\varepsilon})^{-1}.$$

As before, equation (3.23) can be rewritten as

(3.24) $$y_* = X_*\theta_0 + (I_G \otimes U)\lambda + v_* = [X_*, (I_G \otimes U)]\begin{bmatrix}\theta_0 \\ \lambda\end{bmatrix} + v_*,$$

where $\lambda = (\lambda_1', \ldots, \lambda_G')'$ and $V(v_*) = (\Sigma^{\varepsilon\varepsilon})^{-1} \otimes I_T$.

Applying 3SLS formula to (3.24) using (Z, U) as instruments, we obtain:

(3.25) $$\begin{bmatrix}\hat{\theta}_{3SLS} \\ \hat{\lambda}_{3SLS}\end{bmatrix} = \left\{\begin{bmatrix}X_*' \\ (I_G \otimes U)'\end{bmatrix}(\Sigma^{\varepsilon\varepsilon} \otimes P_{[Z,U]})[X_*, (I_G \otimes U)]\right\}^{-1} \cdot \begin{bmatrix}X_*' \\ (I_G \otimes U)'\end{bmatrix}(\Sigma^{\varepsilon\varepsilon} \otimes P_{[Z,U]})y_*.$$

Because $P_{[Z,U]} = P_U + P_{[M_U Z]}$ and $P_{[Z,U]}U = U$, (3.25) becomes

(3.25) $$\begin{bmatrix}\hat{\theta}_{3SLS} \\ \hat{\lambda}_{3SLS}\end{bmatrix} = \begin{bmatrix}X_*'[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})]X_* & X_*'(\Sigma^{\varepsilon\varepsilon} \otimes U) \\ (\Sigma^{\varepsilon\varepsilon} \otimes U')X_* & \Sigma^{\varepsilon\varepsilon} \otimes U'U\end{bmatrix}^{-1} \cdot \begin{bmatrix}X_*'[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})]y_* \\ (\Sigma^{\varepsilon\varepsilon} \otimes U')y_*\end{bmatrix}.$$

Using the partitioned inverse rule, we get

(3.26) $$\hat{\theta}_{3SLS} = E^{-1} \cdot X_*'[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})]y_* - E^{-1}BD^{-1} \cdot (\Sigma^{\varepsilon\varepsilon} \otimes U')y_*$$

where $B = X_*'(\Sigma^{\varepsilon\varepsilon} \otimes U)$, $D = \Sigma^{\varepsilon\varepsilon} \otimes U'U$, and $E = A - BD^{-1}C$ with $A = X_*'[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})]X_*$ and $C = (\Sigma^{\varepsilon\varepsilon} \otimes U')X_*$.

But

$$(3.27) \qquad \begin{aligned} E &= A - BD^{-1}C \\ &= X_*'[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})]X_* \\ &\qquad - X_*'(\Sigma^{\varepsilon\varepsilon} \otimes U)(\Sigma^{\varepsilon\varepsilon} \otimes U'U)^{-1}(\Sigma^{\varepsilon\varepsilon} \otimes U')X_* \\ &= X_*'\{[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})] - (\Sigma^{\varepsilon\varepsilon} \otimes P_U)\}X_* \\ &= X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})X_*, \end{aligned}$$

and

$$(3.28) \qquad BD^{-1} = X_*'(\Sigma^{\varepsilon\varepsilon} \otimes U)(\Sigma^{\varepsilon\varepsilon} \otimes U'U)^{-1} = X_*'[I_G \otimes U(U'U)^{-1}].$$

Substituting (3.27) and (3.28) into (3.26), we obtain:

$$(3.30) \quad \begin{aligned} \hat{\theta}_{3SLS} &= [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})X_*]^{-1} \\ &\qquad \cdot\{X_*'[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})]y_* - X_*'[I_G \otimes U(U'U)^{-1}](\Sigma^{\varepsilon\varepsilon} \otimes U')y_*\} \\ &= [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})X_*]^{-1} X_*'\{[\Sigma^{\varepsilon\varepsilon} \otimes (P_U + P_{[M_U Z]})] - (\Sigma^{\varepsilon\varepsilon} \otimes P_U)\}y_* \\ &= [X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})X_*]^{-1} X_*'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})y_* \\ &= \dddot{\theta}. \end{aligned}$$

■

Equation (3.23) is instructive because, speaking loosely, the effect of adding the variable u_t is to reduce the relevant variance of ε_t from $V(\varepsilon_t) = \Sigma_{\varepsilon\varepsilon}$ to $V(\varepsilon_t | u_t) = \Sigma_{\varepsilon u} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}$. Obviously, this result is a direct extension of the similar result of Schmidt (1986), and is also closely related to the result of Wooldridge (1993).

Our discussion so far has assumed that the covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix}$ is known. Since we generally do not know Σ, the estimators $\bar{\theta}$, $\ddot{\theta}$, $\dddot{\theta}$ and $\breve{\theta}$ are infeasible. However we can define feasible versions of them, say

(3.31A) $$\bar{\theta}_F = [X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_Z M_U)X_*]^{-1} X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_Z M_U)y_*$$

(3.31B) $$\ddot{\theta}_F = [X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_Z)X_*]^{-1} X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_Z)[y_* - (I_G \otimes U)\hat{\lambda}]$$

(3.31C) $$\dddot{\theta}_F = [X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})X_*]^{-1} X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})y_*$$

(3.31D) $$\breve{\theta}_F = [X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)X_*]^{-1} X_*'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)y_*$$

where $\hat{\Sigma}^{\varepsilon\varepsilon}$ and $\hat{\lambda}$ are consistent estimates of $\Sigma^{\varepsilon\varepsilon}$ and λ respectively. Specifically, $\hat{\Sigma}^{\varepsilon\varepsilon} = (\hat{\Sigma}_{\varepsilon\varepsilon} - \hat{\Sigma}_{\varepsilon u}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{u\varepsilon})^{-1}$ and $\hat{\lambda} = \text{vec}(\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{u\varepsilon})$ with $\hat{\Sigma}_{\varepsilon\varepsilon} = T^{-1}\hat{\varepsilon}'\hat{\varepsilon}$, $\hat{\Sigma}_{u\varepsilon}' = \hat{\Sigma}_{\varepsilon u} = T^{-1}\hat{\varepsilon}'U$ and $\hat{\Sigma}_{uu} = T^{-1}U'U$, where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_T)' = (y_1 - X_1\hat{\theta}, \cdots, y_T - X_T\hat{\theta})'$ with $\hat{\theta}$ being any consistent estimate of θ_0; for example, the usual 2SLS estimator. Then it is not difficult to show that $\bar{\theta}_F$, $\ddot{\theta}_F$, $\dddot{\theta}_F$ and $\breve{\theta}_F$ have the same asymptotic distribution as $\bar{\theta}$, $\ddot{\theta}$, $\dddot{\theta}$ and $\breve{\theta}$.

3.4. Concluding Remarks

This chapter provided some improved versions of 3SLS, and extended the results in Chapter 2 on improved versions of IV (2SLS). The improved 3SLS estimators differ from the usual 3SLS estimator in two ways. The first difference is that the projection matrix P_Z in 3SLS is replaced by $P_Z M_U$, $M_U P_Z M_U$, or $P_{[M_U Z]}$, so that the 3SLS "fitted values" are constructed differently. For example, $\dddot{\theta}$ uses $M_U Z$, the part of Z orthogonal to U, as the regressors in the "first stage" regressions, whereas the 3SLS estimator $\hat{\theta}$ just uses Z. This is exactly the same as the difference between the improved and ordinary

2SLS estimators in Chapter 2. However, there is a second difference between the usual and improved 3SLS estimators that did not arise in the case of 2SLS. Where 3SLS estimator uses $\Sigma_{\varepsilon\varepsilon}^{-1}$, the improved 3SLS estimator uses $\Sigma^{\varepsilon\varepsilon} = (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})^{-1}$. Thus 3SLS estimator uses the inverse of $V(\varepsilon_t)$, while the improved 3SLS estimtors use the inverse of $V(\varepsilon_t|u_t)$, in the final "stage" of estimation.

CHAPTER 4

IMPROVED GMM ESTIMATORS FOR SYSTEM OF NONLINEAR EQUATIONS

4.1. Introduction

In this chapter we will extend the results of Chapter 3 to the case of a system of nonlinear equations. The structure of the chapter is as follows. In the next section, we will first define several improved GMM (IGMM) estimators, and then show that these IGMM estimators are asymptotically equally efficient and efficient relative to the usual GMM estimator. The final section gives some concluding remarks.

4.2. Improved GMM Estimators

The model considered in this chaper is

(4.1) $$f(y^*_{tg},\theta_{0g}) = \varepsilon_{tg}, \quad g = 1, 2, \ldots, G; t = 1, 2, \ldots, T,$$

with $y^*_{tg} = (y_{tg}, Y_{tg}', x_{tg}')'$, where y_{tg} is the dependent variable of equation g at observation t, Y_{tg} is the t^{th} observation on the $M_g \times 1$ vector of other endogenous variables included in equation g, x_{tg} is the $N_g \times 1$ vector of exogenous variables of equation g at observation t, θ_{0g} is the $K_g \times 1$ unknown parameter vector of equation g, and ε_{tg} is the model disturbance of equation g at observation t.

Suppose that we have available an $M \times 1$ vector of instruments z_t satisfying the moment conditions $E[(I_G \otimes z_t)(\varepsilon_{t1},\ldots,\varepsilon_{tG})'] = 0$, and $E[z_t \partial f(y^*_{tg},\theta_{0g})/\partial\theta_g']$ has full column rank, g = 1, 2, ..., G. Then the usual GMM estimation of $(\theta_{01}',\ldots,\theta_{0G}')'$ in (4.1) is feasible. However, suppose that we also have available an $L \times 1$ vector of observable variables u_t satisfying $E[(I_L \otimes z_t)u_t] = 0$ and $E(u_t\varepsilon_t') = \Sigma_{u\varepsilon} \neq 0$. Then the additional moment conditions $E[(I_L \otimes z_t)u_t] = 0$ will help us to improve the efficiency of the estimation of $(\theta_{01}',\ldots,\theta_{0G}')'$.

We define the following notation:

(4.2A) $$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_G \end{bmatrix};$$

(4.2B) $$y^*_t = \begin{bmatrix} y^*_{t1} \\ \vdots \\ y^*_{tG} \end{bmatrix},\ f(y^*_t,\theta) = \begin{bmatrix} f(y^*_{t1},\theta_1) \\ \vdots \\ f(y^*_{tG},\theta_G) \end{bmatrix},\ \varepsilon_t = \begin{bmatrix} \varepsilon_{t1} \\ \vdots \\ \varepsilon_{tG} \end{bmatrix},\ t = 1, \ldots, T;$$

(4.2C) $$y^*_{(g)} = \begin{bmatrix} y^*_{1g} \\ \vdots \\ y^*_{Tg} \end{bmatrix},\ f_{(g)}(y^*_{(g)},\theta_g) = \begin{bmatrix} f(y^*_{1g},\theta_g) \\ \vdots \\ f(y^*_{Tg},\theta_g) \end{bmatrix},\ \varepsilon_{(g)} = \begin{bmatrix} \varepsilon_{1g} \\ \vdots \\ \varepsilon_{Tg} \end{bmatrix},\ g = 1, \ldots, G;$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1' \\ \vdots \\ \varepsilon_T' \end{bmatrix},\ U = \begin{bmatrix} u_1' \\ \vdots \\ u_T' \end{bmatrix},\ Z = \begin{bmatrix} z_1' \\ \vdots \\ z_T' \end{bmatrix};$$

(4.2D) $$f_*(\theta) = \begin{bmatrix} f_{(1)}(y^*_{(1)},\theta_1) \\ \vdots \\ f_{(G)}(y^*_{(G)},\theta_G) \end{bmatrix},\ u_* = \mathrm{vec}(U),\ \varepsilon_* = \mathrm{vec}(\varepsilon) = \begin{bmatrix} \varepsilon_{(1)} \\ \vdots \\ \varepsilon_{(G)} \end{bmatrix};$$

(4.2E) $$D_g(\theta_g) = \frac{\partial f_{(g)}(y^*_{(g)}, \theta_g)}{\partial \theta_g'}, \quad g = 1, \ldots, G;$$

(4.2F) $$D(\theta) = \begin{bmatrix} D_1(\theta_1) & & \\ & \ddots & \\ & & D_G(\theta_G) \end{bmatrix};$$

(4.2G) $$P_X = X(X'X)^{-1}X', \quad M_X = I - P_X \quad \text{for any matrix X with full column rank.}$$

Then (4.1) can be rewritten as

(4.3A) $$f(y^*_t, \theta_0) = \varepsilon_t, \quad t = 1, 2, \ldots, T,$$

or

(4.3B) $$f_{(g)}(y^*_{(g)}, \theta_{0g}) = \varepsilon_{(g)}, \quad g = 1, 2, \ldots, G.$$

Using the notation of Chapter 2, we have

(4.4) $$\phi(y^*_t, \theta) = \begin{bmatrix} \phi_1(y^*_t, \theta) \\ \phi_2(y^*_t) \end{bmatrix} = \begin{bmatrix} (I_G \otimes z_t) f(y^*_t, \theta) \\ (I_L \otimes z_t) u_t \end{bmatrix}.$$

Then

(4.5) $$\phi_T(\theta) = \begin{bmatrix} \phi_{T1}(\theta) \\ \phi_{T2} \end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \phi(y^*_t, \theta) = \begin{bmatrix} T^{-1}(I_G \otimes Z') f_*(\theta) \\ T^{-1}(I_L \otimes Z') u_* \end{bmatrix}.$$

Therefore

(4.6A) $$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \lim_{T \to \infty} [T \cdot E\phi_T(\theta_0)\phi_T(\theta_0)']$$

$$= \lim_{T\to\infty} E\begin{bmatrix} T^{-1}(I_G \otimes Z')\varepsilon_*\varepsilon_*'(I_G \otimes Z) & T^{-1}(I_G \otimes Z')\varepsilon_* u_*'(I_L \otimes Z) \\ T^{-1}(I_L \otimes Z')u_*\varepsilon_*'(I_G \otimes Z) & T^{-1}(I_L \otimes Z')u_* u_*'(I_L \otimes Z) \end{bmatrix}$$

(4.6B) $$D_{T1} = \frac{\partial \phi_{T1}(\theta)}{\partial \theta'} = \frac{1}{T}(I_G \otimes Z')\frac{\partial f_*(\theta)}{\partial \theta'} = \frac{1}{T}(I_G \otimes Z')D(\theta).$$

According to the general results of Chapter 2, we know that the augmented GMM (AGMM) estimator ($\tilde{\theta}$) using moment conditions $E\{(I_{G+L} \otimes z_t)[f(y_t^*,\theta_0)', u_t']'\} = 0$ and weighting matrix C^{-1} is numerically equivalent to the IGMM estimator ($\hat{\theta}$) using the moment conditions $E[(I_G \otimes z_t)f(y_t^*,\theta_0) - C_{12}C_{22}^{-1}(I_L \otimes z_t)u_t] = 0$ and weighting matrix $C^{11} = (C_{11} - C_{12}C_{22}^{-1}C_{21})^{-1}$, since both $\tilde{\theta}$ and $\hat{\theta}$ satisfy the same first order condition (2.7A) in Chapter 2.

Under suitable regularity conditions, standard GMM results also tell us that both $\tilde{\theta}$ and $\hat{\theta}$ are consistent. For discussions of regularity conditions, see, e.g., Hansen (1982), Gallant and White (1988) or Amemiya (1985).

Substituting (4.5) and (4.6) into the first order condition (2.7A) in Chapter 2 for the IGMM estimator $\hat{\theta}$, and replacing C^{11}, C_{12} and C_{22}^{-1} by consistent estimates $\hat{C}^{11}$, $\hat{C}_{12}$ and $\hat{C}_{22}^{-1}$ respectively, we arrive at:

(4.7) $$D(\hat{\theta})'(I_G \otimes Z)\hat{C}^{11}[(I_G \otimes Z')f_*(\hat{\theta}) - \hat{C}_{12}\hat{C}_{22}^{-1}(I_L \otimes Z')u_*] = 0.$$

In order to simplify (4.7) further, we need to put more structure on C (or $\hat{C}$). To do so, and to allow us to investigate the asymptotic properties of our estimators, we make the following "high-level" assumptions:

(A4.1)
$$\text{p}\lim \frac{1}{T}\begin{bmatrix} D_1(\theta_{01})' \\ \vdots \\ D_G(\theta_{0G})' \\ Z' \\ \varepsilon' \\ U' \end{bmatrix}\left[D_1(\theta_{01}), \;\cdots\; D_G(\theta_{0G}) \;\; Z \;\; \varepsilon \;\; U\right]$$

$$= \begin{bmatrix} A_{11} & \cdots & A_{1G} & A_{1z} & A_{1\varepsilon} & A_{1u} \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ A_{G1} & \cdots & A_{GG} & A_{Gz} & A_{G\varepsilon} & A_{Gu} \\ A_{z1} & \cdots & A_{zG} & A_{zz} & 0 & 0 \\ A_{\varepsilon 1} & \cdots & A_{\varepsilon G} & 0 & \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ A_{u1} & \cdots & A_{uG} & 0 & \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix} \text{ exists.}$$

(A4.2) A_{zz}, Σ_{uu} and $\Sigma = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix}$ are nonsingular; A_{zg} has full column rank for g = 1, 2, ..., G.

(A4.3)
$$\frac{1}{\sqrt{T}}(I_{G+L} \otimes Z')\begin{bmatrix} \varepsilon_* \\ u_* \end{bmatrix} \to N(0, \Sigma \otimes A_{zz}).$$

Then

(4.8A)
$$C = \Sigma \otimes E(z_t z_t') = \Sigma \otimes A_{zz}.$$

Note that (A4.3), and hence (4.8A), implicitly reflect an assumption of no conditional heteroskedasticity. Furthermore,

(4.8B)
$$\hat{C} = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} \\ \hat{C}_{21} & \hat{C}_{22} \end{bmatrix} = \hat{\Sigma} \otimes T^{-1}Z'Z = \begin{bmatrix} \hat{\Sigma}_{\varepsilon\varepsilon} \otimes T^{-1}Z'Z & \hat{\Sigma}_{\varepsilon u} \otimes T^{-1}Z'Z \\ \hat{\Sigma}_{u\varepsilon} \otimes T^{-1}Z'Z & \hat{\Sigma}_{uu} \otimes T^{-1}Z'Z \end{bmatrix}$$

is a consistent estimate of C, where $\hat{\Sigma}$ is any consistent estimate of Σ. For the moment we will treat Σ as known, for simplicity. Therefore we have

(4.9A) $$\hat{C}^{11} = (\hat{C}_{11} - \hat{C}_{12}\hat{C}_{22}^{-1}\hat{C}_{21})^{-1} = \Sigma^{\varepsilon\varepsilon} \otimes (T^{-1}Z'Z)^{-1}$$

$$\text{with } \Sigma^{\varepsilon\varepsilon} = (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})^{-1}$$

(4.9B) $$\hat{C}_{12}\hat{C}_{22}^{-1} = \Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes I_M.$$

Substituting (4.9A) and (4.9B) into (4.7), we get

(4.10) $$D(\hat{\theta})'(I_G \otimes Z)[\Sigma^{\varepsilon\varepsilon} \otimes (T^{-1}Z'Z)^{-1}] \cdot$$
$$\cdot[(I_G \otimes Z')f_*(\hat{\theta}) - (\Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes I_M)(I_L \otimes Z')u_*] = 0.$$

Noticing that

(4.11) $$\begin{aligned}(\Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes I_M)(I_L \otimes Z')u_* &= (\Sigma_{\varepsilon u}\Sigma_{uu}^{-1} \otimes Z')\text{vec}(U) \\ &= \text{vec}(Z'U\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) \\ &= (I_G \otimes Z'U)\text{vec}(\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) \\ &= (I_G \otimes Z')(I_G \otimes U)\text{vec}(\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}),\end{aligned}$$

and substituting this expression into (4.10), we get

(4.12) $$D(\hat{\theta})'(I_G \otimes Z)[\Sigma^{\varepsilon\varepsilon} \otimes (Z'Z)^{-1}](I_G \otimes Z')[f_*(\hat{\theta}) - (I_G \otimes U)\lambda] = 0$$

or

(4.13) $$D(\hat{\theta})'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)[f_*(\hat{\theta}) - (I_G \otimes U)\lambda] = 0$$

where $\lambda = \text{vec}(\Sigma_{uu}^{-1}\Sigma_{u\varepsilon}) \equiv (\lambda_1', \lambda_2', \cdots, \lambda_G')'$. Thus, for i = 1, 2, ..., G, $\lambda_i = \Sigma_{uu}^{-1}$ times the i^{th} column of $\Sigma_{u\varepsilon}$; equivalently, $\lambda_i = (\text{p}\lim T^{-1}U'U)^{-1}\text{p}\lim T^{-1}U'\varepsilon_{(i)}$.

It is well known that the first order condition for the usual GMM estimator ($\hat{\theta}$) using moment conditions $E[\phi_1(y_t^*,\theta_0)] = E[(I_G \otimes z_t)f(y_t^*,\theta_0)] = 0$ and weighting matrix $\Sigma_{\varepsilon\varepsilon}^{-1}$ is

$$D(\hat{\theta})'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes P_Z)f_*(\hat{\theta}) = 0. \tag{4.14}$$

Comparing (4.14) with (4.13), we see that the only difference between the first order condition of the usual GMM estimator and the first order condition of the IGMM estimator is that $\Sigma_{\varepsilon\varepsilon}^{-1}$ and $f_*(\theta)$ in (4.14) are replaced by $\Sigma^{\varepsilon\varepsilon}$ and $[f_*(\theta) - (I_G \otimes U)\lambda]$ respectively. It is interesting to notice that $[f_*(\theta_0) - (I_G \otimes U)\lambda] = [(f_{(1)}(y_{(1)}^*,\theta_{01}) - U\lambda)', \cdots, (f_{(G)}(y_{(G)}^*,\theta_{0G}) - U\lambda)']'$ is just a vector of residuals from the linear projection of $f_{(g)}(y_{(g)}^*,\theta_{0g})$ onto U. Thus the IGMM estimator $\tilde{\theta}$ implicitly determined by (4.13) can also be regarded as a purged GMM (PGMM) estimator, as in Chapter 3.

Under assumptions (A4.1)-(A4.3), it is not difficult to show that the usual GMM estimator $\hat{\theta}$ implicitly determined by (4.14) has the following asymptotic variance:

$$\begin{aligned} AV[\sqrt{T}(\hat{\theta}-\theta_0)] &= [\operatorname{plim}\frac{1}{T}D(\theta_0)'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes P_Z)D(\theta_0)]^{-1} \\ &= [A'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes A_{zz}^{-1})A]^{-1} \end{aligned} \tag{4.15}$$

where

$$A = \operatorname{plim}\frac{1}{T}(I_G \otimes Z')D(\theta_0) = \begin{bmatrix} A_{z1} & & \\ & \ddots & \\ & & A_{zG} \end{bmatrix} \tag{4.16}$$

has full column rank because of assumption (A4.2).

We now wish to consider the IGMM estimator $\tilde{\theta}$ and the following three more estimators

(4.17A) $$\dddot{\theta} = \arg\min_{\theta}\{f_*(\theta)'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})f_*(\theta)\}$$

(4.17B) $$\breve{\theta} = \arg\min_{\theta}\{f_*(\theta)'(\Sigma^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)f_*(\theta)\}$$

(4.17C) $$\ddot{\theta} = \arg\min_{\theta}\{f_*(\theta)'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z M_U)f_*(\theta)\}.$$

We will show that these estimators are asymptotically equally efficient, with asymptotic variance matrix equal to

(4.18) $$[\operatorname{plim}\frac{1}{T}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)D(\theta_0)]^{-1} = [A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A]^{-1} \equiv B^{-1}.$$

We will presume that the estimators are consistent, because each can be written as a GMM estimator that exploits valid moment conditions.

Comparing (4.18) with the asymptotic variance matrix of $\hat{\theta}$ in (4.15) above, the fact that the matrix $\{[A'(\Sigma_{\varepsilon\varepsilon}^{-1} \otimes A_{zz}^{-1})A]^{-1} - [A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A]^{-1}\}$ is positive semidefinite (shown later in Theorem 4.2) establishes the asymptotic efficiency of $\tilde{\theta}$, $\dddot{\theta}$, $\breve{\theta}$ and $\ddot{\theta}$ relative to $\hat{\theta}$. We now turn to a proof of these results.

<u>LEMMA 4.1</u>: $\operatorname{plim} T^{-1}Z'M_U Z = \operatorname{plim} T^{-1}Z'Z = A_{zz}$;

$\operatorname{plim} T^{-1}Z'M_U D_i(\theta_{0i}) = \operatorname{plim} T^{-1}Z'D_i(\theta_{0i}) = A_{zi}$, for i = 1, 2, ..., G.

Proof: The proof is similar to the proof of Lemma 3.1 of Chapter 3. For example,

$$\begin{aligned}
&\operatorname{plim} T^{-1}Z'M_U D_i(\theta_{0i}) \\
&= \operatorname{plim}[T^{-1}Z'D_i(\theta_{0i}) - (T^{-1}Z'U)(T^{-1}U'U)^{-1}(T^{-1}U'D_i(\theta_{0i}))] \\
&= A_{zi} - 0 \cdot \Sigma_{uu}^{-1}A_{ui} \\
&= A_{zi}
\end{aligned}$$

where the second equality is implied by (A4.1).

■

LEMMA 4.2: $$\operatorname{plim} T^{-1}D_i(\theta_{0i})'P_{[M_UZ]}D_j(\theta_{0j}) = \operatorname{plim} T^{-1}D_i(\theta_{0i})'M_UP_ZM_UD_j(\theta_{0j})$$
$$= \operatorname{plim} T^{-1}D_i(\theta_{0i})'P_ZM_UD_j(\theta_{0j})$$
$$= \operatorname{plim} T^{-1}D_i(\theta_{0i})'P_ZD_j(\theta_{0j})$$
$$= A_{iz}A_{zz}^{-1}A_{zj}$$

for i, j = 1, 2, ..., G.

Proof: We will give the proof for $\operatorname{plim} T^{-1}D_i(\theta_{0i})'M_UP_ZM_UD_j(\theta_{0j}) = A_{iz}A_{zz}^{-1}A_{zj}$. The proofs for the other cases are quite similar.

$$\operatorname{plim} T^{-1}D_i(\theta_{0i})'M_UP_ZM_UD_j(\theta_{0j})$$
$$= \operatorname{plim}[T^{-1}D_i(\theta_{0i})'M_UZ](T^{-1}Z'Z)^{-1}[T^{-1}Z'M_UD_j(\theta_{0j})]$$
$$= A_{iz}A_{zz}^{-1}A_{zj}$$

using (A4.1), (A4.2) and Lemma 4.1.

■

LEMMA 4.3: $$\operatorname{plim}[T^{-1}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_UZ]})D(\theta_0)]$$
$$= \operatorname{plim}[T^{-1}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes M_UP_ZM_U)D(\theta_0)]$$
$$= \operatorname{plim}[T^{-1}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_ZM_U)D(\theta_0)]$$
$$= \operatorname{plim}[T^{-1}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_Z)D(\theta_0)]$$
$$= A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A \equiv B,$$

and B is nonsingular.

Proof: Let $\Sigma^{\varepsilon\varepsilon} = (\sigma^{ij})_{G\times G}$. Then

$$\operatorname{plim}[T^{-1}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_UZ]})D(\theta_0)]$$

$$= \operatorname{plim} T^{-1}\begin{bmatrix} D_1(\theta_{01})' & & \\ & \ddots & \\ & & D_G(\theta_{0G})' \end{bmatrix}\begin{bmatrix} \sigma^{11}P_{[M_UZ]} & \cdots & \sigma^{1G}P_{[M_UZ]} \\ \vdots & & \vdots \\ \sigma^{G1}P_{[M_UZ]} & \cdots & \sigma^{GG}P_{[M_UZ]} \end{bmatrix}.$$

$$\cdot\begin{bmatrix} D_1(\theta_{01}) & & \\ & \ddots & \\ & & D_G(\theta_{0G}) \end{bmatrix}$$

$$= \text{p}\lim T^{-1}\begin{bmatrix} D_1(\theta_{01})'\sigma^{11}P_{[M_U Z]}D_1(\theta_{01}) & \cdots & D_1(\theta_{01})'\sigma^{1G}P_{[M_U Z]}D_G(\theta_{0G}) \\ \vdots & \vdots & \vdots \\ D_G(\theta_{0G})'\sigma^{G1}P_{[M_U Z]}D_1(\theta_{01}) & \cdots & D_G(\theta_{0G})'\sigma^{GG}P_{[M_U Z]}D_G(\theta_{0G}) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11}A_{1z}A_{zz}^{-1}A_{z1} & \cdots & \sigma^{1G}A_{1z}A_{zz}^{-1}A_{zG} \\ \vdots & \vdots & \vdots \\ \sigma^{G1}A_{Gz}A_{zz}^{-1}A_{z1} & \cdots & \sigma^{GG}A_{Gz}A_{zz}^{-1}A_{zG} \end{bmatrix} \text{ (using Lemma 4.2)}$$

$$= \begin{bmatrix} A_{1z} & & \\ & \ddots & \\ & & A_{Gz} \end{bmatrix} \begin{bmatrix} \sigma^{11}A_{zz}^{-1} & \cdots & \sigma^{1G}A_{zz}^{-1} \\ \vdots & \cdots & \vdots \\ \sigma^{G1}A_{zz}^{-1} & \cdots & \sigma^{GG}A_{zz}^{-1} \end{bmatrix} \begin{bmatrix} A_{z1} & & \\ & \ddots & \\ & & A_{zG} \end{bmatrix}$$

$$= A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A \equiv B.$$

using the definition of A in (4.16). The probability limits for the other cases involve essentially the same arguments. Finally, $B = A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A$ is nonsingular bacause Σ, Σ_{uu} and A_{zz} are nonsingular, which implies $\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1}$ nonsingular, and because A has full column rank (see (A4.2)).

■

<u>LEMMA 4.4</u>: $\text{p}\lim T^{-1}(I_G \otimes Z'M_U)D(\theta_0) = \text{p}\lim T^{-1}(I_G \otimes Z')D(\theta_0) = A$.

Proof: $\text{p}\lim T^{-1}(I_G \otimes Z'M_U)D(\theta_0)$

$$= \operatorname{plim} T^{-1} \begin{bmatrix} Z'M_U & & \\ & \ddots & \\ & & Z'M_U \end{bmatrix} \begin{bmatrix} D_1(\theta_{01}) & & \\ & \ddots & \\ & & D_G(\theta_{0G}) \end{bmatrix}$$

$$= \begin{bmatrix} \operatorname{plim} T^{-1}Z'M_U D_1(\theta_{01}) & & \\ & \ddots & \\ & & \operatorname{plim} T^{-1}Z'M_U D_G(\theta_{0G}) \end{bmatrix}$$

$$= \begin{bmatrix} \operatorname{plim} T^{-1}Z'D_1(\theta_{01}) & & \\ & \ddots & \\ & & \operatorname{plim} T^{-1}Z'D_G(\theta_{0G}) \end{bmatrix} \text{(using Lemma 4.1)}$$

$$= \begin{bmatrix} A_{z1} & & \\ & \ddots & \\ & & A_{zG} \end{bmatrix} \text{(using (A4.1))}$$

$$= A$$

using the definition of A in (4.16).

■

LEMMA 4.5: $T^{-1/2}(I_G \otimes Z')[\varepsilon_* - (I_G \otimes U)\lambda]$ and $T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_*$ each converge in distribution to $N[0, (\Sigma^{\varepsilon\varepsilon})^{-1} \otimes A_{zz}]$ with $\Sigma^{\varepsilon\varepsilon} = (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon u}\Sigma_{uu}^{-1}\Sigma_{u\varepsilon})^{-1}$.

Proof: See the proof of Lemma 3.6 in Chapter 3.

■

THEOREM 3.1: $\sqrt{T}(\hat{\theta} - \theta_0)$, $\sqrt{T}(\dddot{\theta} - \theta_0)$, $\sqrt{T}(\breve{\theta} - \theta_0)$ and $\sqrt{T}(\ddot{\theta} - \theta_0)$ each converge in distribution to $N(0, B^{-1})$, where $B = A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A$ with $A = \operatorname{diag}(A_{z1}, \ldots, A_{zG})$ as defined in (4.16).

Proof: We will give the proof for $\sqrt{T}(\dddot{\theta}-\theta_0)$. The proofs for the other cases are essentially identical. Using the definition in (4.17A), we know that $\dddot{\theta}$ satisfies the following first order condition:

$$(4.19) \qquad D(\dddot{\theta})'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})f_*(\dddot{\theta}) = 0.$$

Using the first order Taylor expansion formula, we have

$$(4.20) \qquad f(\dddot{\theta}) = f(\theta_0) + D(\theta^*)(\dddot{\theta}-\theta_0) = \varepsilon_* + D(\theta^*)(\dddot{\theta}-\theta_0)$$

where θ^* is between θ_0 and $\dddot{\theta}$. Substituting (4.20) into (4.19), we get

$$(4.21) \qquad D(\dddot{\theta})'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})D(\theta^*)(\dddot{\theta}-\theta_0) + D(\dddot{\theta})'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})\varepsilon_* = 0.$$

(4.21) is equivalent to

$$(4.22) \qquad T^{1/2}(\dddot{\theta}-\theta_0) = -[T^{-1}D(\dddot{\theta})'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})D(\theta^*)]^{-1} \cdot T^{-1/2}D(\dddot{\theta})'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})\varepsilon_*.$$

Because $\dddot{\theta}$ is consistent, (4.22) can be rewritten as

$$(4.23) \qquad T^{1/2}(\dddot{\theta}-\theta_0) = -[\text{p}\lim T^{-1}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})D(\theta_0)]^{-1} \cdot \cdot T^{-1/2}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})\varepsilon_* + o_p(1) = -B^{-1} \cdot T^{-1/2}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})\varepsilon_* + o_p(1)$$

using Lemma 4.3. But

$$(4.24) \qquad T^{-1/2}D(\theta_0)'(\Sigma^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})\varepsilon_*$$

$$= T^{-1/2}D(\theta_0)'[\Sigma^{\varepsilon\varepsilon} \otimes M_U Z(Z'M_U Z)^{-1} Z'M_U]\varepsilon_*$$
$$= T^{-1/2}D(\theta_0)'(I_G \otimes M_U Z)[\Sigma^{\varepsilon\varepsilon} \otimes (Z'M_U Z)^{-1}](I_G \otimes Z'M_U)\varepsilon_*$$
$$= [T^{-1}D(\theta_0)'(I_G \otimes M_U Z)][\Sigma^{\varepsilon\varepsilon} \otimes (T^{-1}Z'M_U Z)^{-1}] \cdot$$
$$\cdot T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_*$$
$$= A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1}) \cdot T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_* + o_p(1)$$

using Lemmas 4.4 and 4.1.

Substituting (4.24) into (4.23), we arrive at

(4.25) $$T^{1/2}(\dddot{\theta} - \theta_0) = -B^{-1}A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1}) \cdot T^{-1/2}(I_G \otimes Z'M_U)\varepsilon_* + o_p(1)$$
$$\rightarrow N(0,W)$$

using Lemma 4.5, where

(4.26) $$W = B^{-1}A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1}) \cdot [(\Sigma^{\varepsilon\varepsilon})^{-1} \otimes A_{zz}] \cdot [B^{-1}A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})]'$$
$$= B^{-1}A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1}) \cdot [(\Sigma^{\varepsilon\varepsilon})^{-1} \otimes A_{zz}] \cdot (\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})AB^{-1}$$
$$= B^{-1} \cdot A'(\Sigma^{\varepsilon\varepsilon} \otimes A_{zz}^{-1})A \cdot B^{-1}$$
$$= B^{-1} \cdot B \cdot B^{-1} \text{ (using the definition of B in (4.18))}$$
$$= B^{-1}.$$

■

<u>THEOREM 4.2</u>: The improved GMM estimators: $\tilde{\theta}$, $\dddot{\theta}$, $\breve{\theta}$ and $\ddot{\theta}$ are asymptotically efficient relative to the usual GMM estimator $\hat{\theta}$. They are strictly more efficient than the usual GMM estimator $\hat{\theta}$ if $\Sigma_{u\varepsilon} \neq 0$.

Proof: See the proof of Theorem 3.3 in Chapter 3.

■

Our discussion so far has assumed that the covariance matrix $\Sigma = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{bmatrix}$ is known. Since we generally do not know Σ, the estimators $\hat{\theta}$, $\dddot{\theta}$, $\breve{\theta}$ and $\ddot{\theta}$ are infeasible. However we can define feasible versions of them, say

(4.27A) $\hat{\theta}_F = \arg\min_\theta \{[f_*(\theta) - (I_G \otimes U)\hat{\lambda}]'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_Z)[f_*(\theta) - (I_G \otimes U)\hat{\lambda}]\}$

(4.27B) $\dddot{\theta}_F = \arg\min_\theta \{f_*(\theta)'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_{[M_U Z]})f_*(\theta)\}$

(4.27C) $\breve{\theta}_F = \arg\min_\theta \{f_*(\theta)'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes M_U P_Z M_U)f_*(\theta)\}$

(4.27D) $\ddot{\theta}_F = \arg\min_\theta \{f_*(\theta)'(\hat{\Sigma}^{\varepsilon\varepsilon} \otimes P_Z M_U)f_*(\theta)\}$

where $\hat{\Sigma}^{\varepsilon\varepsilon}$ and $\hat{\lambda}$ are consistent estimates of $\Sigma^{\varepsilon\varepsilon}$ and λ respectively. Specifically, $\hat{\Sigma}^{\varepsilon\varepsilon} = (\hat{\Sigma}_{\varepsilon\varepsilon} - \hat{\Sigma}_{\varepsilon u}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{u\varepsilon})^{-1}$ and $\hat{\lambda} = \text{vec}(\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{u\varepsilon})$ with $\hat{\Sigma}_{\varepsilon\varepsilon} = T^{-1}\hat{\varepsilon}'\hat{\varepsilon}$, $\hat{\Sigma}_{u\varepsilon}' = \hat{\Sigma}_{\varepsilon u} = T^{-1}\hat{\varepsilon}'U$ and $\hat{\Sigma}_{uu} = T^{-1}U'U$, where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_T)' = [f(y_1^*, \hat{\theta}), \cdots, f(y_T^*, \hat{\theta})]'$ with $\hat{\theta}$ being any consistent estimate of θ_0; for example, the usual GMM estimator. Then it is not difficult to show that $\hat{\theta}_F$, $\dddot{\theta}_F$, $\breve{\theta}_F$ and $\ddot{\theta}_F$ have the same asymptotic distribution as $\hat{\theta}$, $\dddot{\theta}$, $\breve{\theta}$ and $\ddot{\theta}$.

4.3. Concluding Remarks

In this chapter, we have generalized the results of the previous chapter for 3SLS-type estimators to the case of a nonlinear model. We have simplified the analysis by making high-level assumptions, and by not giving a rigorous proof of consistency. Given these simplifications, the extension from linear to nonlinear 3SLS is fairly straightforward. As in the linear case, the improved estimators use a different projection matrix than the usual nonlinear 3SLS (e.g. the projection onto $M_U Z$ instead of onto Z), and they use the inverse of $V(\varepsilon_t | u_t)$ instead of $V(\varepsilon_t)$ in weighting equations.

CHAPTER 5

THE ASYMPTOTIC EQUIVALENCE BETWEEN THE ITERATED IMPROVED 2SLS ESTIMATOR AND THE 3SLS ESTIMATOR

5.1. Introduction

In Chapters 2, 3 and 4, we showed that we can improve on the usual IV (2SLS) and 3SLS estimators provided that we have available an extra vector of observable variables u_t which is uncorrelated with the instruments and correlated with the disturbances of the model being estimated. In this chapter, we will extend the improved IV (2SLS) idea to the case when the extra information u_t is consistently estimated instead of observed. Because the asymptotic distribution of the estimated u_t depends on the the model structure from which it is estimated, we choose to consider our extension in the context of a system of linear equations.

It is well known that the only difference between equation-by-equation 2SLS and 3SLS is that the 3SLS estimator utilizes information about the relationships among the disturbances of different equations, but 2SLS does not. Then a natural question one wants to ask is whether it is possible for us, on one hand, to still keep the simplicity of the equation-by-equation 2SLS estimator; on the other hand, to also utilize the information contained in the covariance structure of the model disturbances, such that the new equation-by-equation estimator has the same asymptotic efficiency as the 3SLS estimator applied to the entire equation system.

Telser (1964) has addressed essentially the same question in the context of a seemingly uncorrelated equation (SURE) system:

(5.1) $$y_{tg} = x_{tg}'\theta_{0g} + \varepsilon_{tg}, \qquad g = 1, 2, ..., G;\ t = 1, 2, ..., T,$$

where y_{tg} is the dependent variable of equation g at observation t, x_{tg} is the $K_g \times 1$ vector of explanatory variables of equation g at observation t, ε_{tg} is the model disturbance of equation g at observation t, θ_{og} is the $K_g \times 1$ unknown parameter vector of equation g, and x_{tg} is strictly exogenous. Telser proved that the iterated LS estimators of the augmented equations

(5.2) $$y_{tg} = x_{tg}'\theta_{0g} + \varepsilon_{t(g)}'\lambda_{0g} + v_{tg}, \quad g = 1, 2, ..., G;\ t = 1, 2, ..., T,$$

where $\varepsilon_{t(g)} = (\varepsilon_{t1}, \ldots, \varepsilon_{t,g-1}, \varepsilon_{t,g+1}, \ldots, \varepsilon_{tG})'$, $\lambda_{0g} = [E(\varepsilon_{t(g)}\varepsilon_{t(g)}')]^{-1}E(\varepsilon_{t(g)}\varepsilon_{tg})$ and $v_{tg} = \varepsilon_{tg} - \varepsilon_{t(g)}'\lambda_{og}$, are asymptotically as efficient as the SURE estimator of the system (5.1). It turns out that the analogous result is also true for a simultaneous equation system.

The rest of the chapter is organized as follows. Section 5.2 defines the improved 2SLS (I2SLS) estimator. Section 5.3 describes the iterated I2SLS estimators. Section 5.4 proves that the iterated I2SLS estimators converge to a limit, and that their limit is asymptotically as efficient as the 3SLS estimator.

5.2. Improved 2SLS Estimator

The model considered in this chapter is

(5.3) $$y_{tg} = x_{tg}'\theta_{0g} + \varepsilon_{tg}, \qquad g = 1, 2, ..., G;\ t = 1, 2, ..., T,$$

where y_{tg} is the dependent variable of equation g at observation t, x_{tg} is the $K_g \times 1$ vector of explanatory variables of equation g at observation t, ε_{tg} is the model disturbance of equation g at observation t, and θ_{og} is the $K_g \times 1$ unknown parameter vector of equation g. We assume that in general $\text{cov}(x_{tg}, \varepsilon_{tg}) \neq 0$ for all g. Suppose that we have available an $M \times 1$ vector of instruments z_t satisfying $E(z_t \varepsilon_{tg}) = 0$, with $E(z_t x_{tg}')$ having full column rank for g = 1, 2, ..., G.

We define the following notation:

(5.4A) $$y_g = \begin{bmatrix} y_{1g} \\ \vdots \\ y_{Tg} \end{bmatrix},\ X_g = \begin{bmatrix} x_{1g}' \\ \vdots \\ x_{Tg}' \end{bmatrix},\ \varepsilon_g = \begin{bmatrix} \varepsilon_{1g} \\ \vdots \\ \varepsilon_{Tg} \end{bmatrix},\ Z = \begin{bmatrix} z_1' \\ \vdots \\ z_T' \end{bmatrix};$$

(5.4B) $$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_G \end{bmatrix},\ \varepsilon = [\varepsilon_1, \ldots, \varepsilon_G],\ \varepsilon_* = \text{vec}(\varepsilon);$$

(5.4C) $$\varepsilon_{t(g)} = (\varepsilon_{t1}, \ldots, \varepsilon_{t,g-1}, \varepsilon_{t,g+1}, \ldots, \varepsilon_{tG})';$$

(5.4D) $$\varepsilon_{(g)} = \begin{bmatrix} \varepsilon_{1(g)}' \\ \vdots \\ \varepsilon_{T(g)}' \end{bmatrix} = (\varepsilon_1, \ldots, \varepsilon_{g-1}, \varepsilon_{g+1}, \ldots, \varepsilon_G);$$

(5.4E) $P_Z = Z(Z'Z)^{-1}Z'$ provided that Z'Z is nonsingular;

(5.4F) $\hat{X}_g = P_Z X_g$ for g = 1, 2, ..., G;

(5.4G) $L(h|W) = W\lambda_0$ = the linear projection of h onto W, with $\lambda_0 = [E(W'W)]^{-1}E(W'h)$ defined for any $s \times 1$ vector h and any $s \times m$ matrix W as long as E(W'W) is nonsingular.

It is important to note the distinction between ε_g (T observations for the error of equation g) and $\varepsilon_{(g)}$ (T observations for the errors of all the equations except equation g).

With this notation, (5.3) can be rewritten as

$$\text{(5.5)} \qquad y_g = X_g\theta_{0g} + \varepsilon_g, \quad g = 1, 2, \ldots, G.$$

We make the following "high level" assumptions.

$$\text{(A5.1)} \qquad \text{plim}\,\frac{1}{T}\begin{bmatrix} X_1{}' \\ \vdots \\ X_G{}' \\ Z' \\ \varepsilon_1{}' \\ \vdots \\ \varepsilon_G{}' \end{bmatrix}[X_1,\ldots,X_G,Z,\varepsilon_1,\ldots,\varepsilon_G]$$

$$= \begin{bmatrix} A_{11} & \cdots & A_{1G} & A_{1z} & A_{x\varepsilon,11} & \cdots & A_{x\varepsilon,1G} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ A_{G1} & \cdots & A_{GG} & A_{Gz} & A_{x\varepsilon,G1} & \cdots & A_{x\varepsilon,GG} \\ A_{z1} & \cdots & A_{zG} & A_{zz} & 0 & \cdots & 0 \\ A_{\varepsilon x,11} & \cdots & A_{\varepsilon x,1G} & 0 & \sigma_{11} & \cdots & \sigma_{1G} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ A_{\varepsilon x,G1} & \cdots & A_{\varepsilon x,GG} & 0 & \sigma_{G1} & \cdots & \sigma_{GG} \end{bmatrix} \text{ exists.}$$

(A5.2) A_{zg} has full column rank for every $g = 1, 2, \ldots, G$;

$$A_{zz} \text{ and } \Sigma \equiv \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1G} \\ \vdots & & \vdots \\ \sigma_{G1} & \cdots & \sigma_{GG} \end{bmatrix} \text{ are nonsingular.}$$

$$\text{(A5.3)} \qquad T^{-1/2}(I_G \otimes Z')\varepsilon_* \rightarrow N(0, \Sigma \otimes A_{zz}).$$

(A5.4) $T^{-1/2}\varepsilon_{(i)}'v_i \to N(0,\Omega_i)$

with $v_i = \varepsilon_i - L(\varepsilon_i|\varepsilon_{(i)})$ for i = 1, 2, ..., G.

$$\text{Define } \Sigma_{ii} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1,i-1} & \sigma_{1,i+1} & \cdots & \sigma_{1G} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sigma_{i-1,1} & \cdots & \sigma_{i-1,i-1} & \sigma_{i-1,i+1} & \cdots & \sigma_{i-1,G} \\ \sigma_{i+1,1} & \cdots & \sigma_{i+1,i-1} & \sigma_{i+1,i+1} & \cdots & \sigma_{i+1,G} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sigma_{G1} & \cdots & \sigma_{G,i-1} & \sigma_{G,i+1} & \cdots & \sigma_{GG} \end{bmatrix} i = 1, 2, ..., G.$$

Assumption (A5.2) implies that Σ_{ii} is nonsingular for all i.

These high-level assumptions are derivable from various sets of more basic assumptions. For example, in a time series context, define $e_t = (\varepsilon_{t1},...,\varepsilon_{tG})'$ and $\Psi_t = \{z_t; z_{t-1}, e_{t-1}; z_{t-2}, e_{t-2}; ...\}$. Then (A5.1)-(A5.4) follow from the assumptions that $E(e_t|\Psi_t) = 0$, $V(e_t|\Psi_t) = \Sigma$, $(x_{t1}',...,x_{tG}', z_t')'$ is covariance stationary, and the fourth moments of e_t exist.

We now define our I2SLS estimator as the IV estimator of the augmented equation

(5.6) $$y_g = X_g\theta_{0g} + \varepsilon_{(g)}\lambda_{og} + v_g, \quad g = 1, 2, ..., G,$$

using $(\hat{X}_g, \varepsilon_{(g)})$ as instruments, where

(5.7A) $$\lambda_{0g} = [E(\varepsilon_{(g)}'\varepsilon_{(g)})]^{-1}E(\varepsilon_{(g)}'\varepsilon_g) \equiv (\lambda_{0g1},...,\lambda_{0g,g-1},\lambda_{0g,g+1},...,\lambda_{0gG})'$$

(5.7B) $$v_g = \varepsilon_g - L(\varepsilon_g|\varepsilon_{(g)}) = \varepsilon_g - \varepsilon_{(g)}\lambda_{0g}.$$

If we knew $\varepsilon_{(g)}$, then as pointed out in Chapter 3, adding $\varepsilon_{(g)}$ to (5.6) reduces the variance of the model disturbance in (5.6) from $V(\varepsilon_g)$ to $V(\varepsilon_g | \varepsilon_{(g)})$, and IV applied to (5.6) would be more efficient than IV applied to (5.3). However, the IV estimation of (5.6) is infeasible because we do not observe $\varepsilon_{(g)}$. In the next section we will define the iterated feasible I2SLS estimator.

5.3. Iterated I2SLS Estimator

Our iterated feasible I2SLS estimator of $\theta_0 = (\theta_{01}', \ldots, \theta_{0G}')'$ in (5.3) is defined as follows.

<u>Round 0</u>: Apply the usual IV (2SLS) estimation method to (5.3) equation by equation, using Z as instruments, to get the initial consistent estimate $\hat{\theta}(0) = [\hat{\theta}_1(0)', \ldots, \hat{\theta}_G(0)']'$ of θ_0. Then we estimate the model disturbance ε_g by

$$(5.8) \qquad \varepsilon_g(0) = y_g - X_g \hat{\theta}_g(0)$$

for g = 1, 2, ..., G.

<u>Round 1:</u> <u>For equation 1</u>, apply the usual IV estimation method to

$$(5.9) \qquad y_1 = X_1 \theta_{01} + \varepsilon_{(1)}(1) \lambda_{01} + v_1(1)$$

using $[\hat{X}_1, \varepsilon_{(1)}(1)]$ as instruments, where λ_{01} is defined in (5.7A), and

$$(5.10A) \qquad \varepsilon_{(1)}(1) = [\varepsilon_2(0), \ldots, \varepsilon_G(0)]$$

$$(5.10B) \qquad v_1(1) = y_1 - X_1 \theta_{01} - \varepsilon_{(1)}(1) \lambda_{01} = \varepsilon_1 - \varepsilon_{(1)}(1) \lambda_{01}.$$

Denote the IV estimate of (5.9) by $[\hat{\theta}_1(1)', \hat{\lambda}_1(1)']'$. We then update the estimate of ε_1 by

(5.11) $$\varepsilon_1(1) = y_1 - X_1\hat{\theta}_1(1).$$

<u>For equation 2</u>: Apply the usual IV estimation method to

(5.12) $$y_2 = X_2\theta_{02} + \varepsilon_{(2)}(1)\lambda_{02} + v_2(1)$$

using $[\hat{X}_2, \varepsilon_{(2)}(1)]$ as instruments, where λ_{02} is defined in (5.7A), and

(5.13A) $$\varepsilon_{(2)}(1) = [\varepsilon_1(1), \varepsilon_3(0), \ldots, \varepsilon_G(0)]$$

(5.13B) $$v_2(1) = \varepsilon_2 - \varepsilon_{(2)}(1)\lambda_{02}.$$

Denote the resulting IV estimate of (5.12) by $[\hat{\theta}_2(1)', \hat{\lambda}_2(1)']'$. We then update the estimate of ε_2 by

(5.14) $$\varepsilon_2(1) = y_2 - X_2\hat{\theta}_2(1).$$

Generally, <u>in Round 1 for equation g</u>: We apply the usual IV estimation method to

(5.15) $$y_g = X_g\theta_{0g} + \varepsilon_{(g)}(1)\lambda_{0g} + v_g(1)$$

using $[\hat{X}_g, \varepsilon_{(g)}(1)]$ as instruments, where λ_{0g} is defined in (5.7A), and

(5.16A) $$\varepsilon_{(g)}(1) = [\varepsilon_1(1), \ldots, \varepsilon_{g-1}(1), \varepsilon_{g+1}(0), \ldots, \varepsilon_G(0)]$$

(5.16B) $$v_g(1) = \varepsilon_g - \varepsilon_{(g)}(1)\lambda_{0g}.$$

Denote the resulting IV estimate of (5.15) by $[\hat{\theta}_g(1)', \hat{\lambda}_g(1)']'$. We then update the estimate of ε_g by

(5.17) $$\varepsilon_g(1) = y_g - X_g\hat{\theta}_g(1).$$

Round 2: For equation 1, apply the usual IV estimation method to

(5.18) $$y_1 = X_1\theta_{01} + \varepsilon_{(1)}(2)\lambda_{01} + v_1(2)$$

using $[\hat{X}_1, \varepsilon_{(1)}(2)]$ as instruments, where λ_{01} is defined in (5.7A), and

(5.19) $$\varepsilon_{(1)}(2) = [\varepsilon_2(1), \ldots, \varepsilon_G(1)]$$

(5.20) $$v_1(2) = y_1 - X_1\theta_{01} - \varepsilon_{(1)}(2)\lambda_{01} = \varepsilon_1 - \varepsilon_{(1)}(2)\lambda_{01}.$$

Denote the IV estimate of (5.18) by $[\hat{\theta}_1(2)', \hat{\lambda}_1(2)']'$. We then update the estimate of ε_1 by

(5.21) $$\varepsilon_1(2) = y_1 - X_1\hat{\theta}_1(2).$$

For equation 2: Apply the usual IV estimation method to

(5.22) $$y_2 = X_2\theta_{02} + \varepsilon_{(2)}(2)\lambda_{02} + v_2(2)$$

using $[\hat{X}_2, \varepsilon_{(2)}(2)]$ as instruments, where λ_{02} is defined in (5.7A), and

(5.23) $$\varepsilon_{(2)}(2) = [\varepsilon_1(2), \varepsilon_3(1), \ldots, \varepsilon_G(1)]$$

(5.24) $$v_2(2) = \varepsilon_2 - \varepsilon_{(2)}(2)\lambda_{02}.$$

Denote the resulting IV estimate of (5.22) by $[\hat{\theta}_2(2)', \hat{\lambda}_2(2)']'$. We then update the estimate of ε_2 by

(5.25) $$\varepsilon_2(2) = y_2 - X_2\hat{\theta}_2(2).$$

Generally, <u>in Round 2 for equation g</u>: We apply the usual IV estimation method to

(5.26) $$y_g = X_g\theta_{0g} + \varepsilon_{(g)}(2)\lambda_{0g} + v_g(2)$$

using $[\hat{X}_g, \varepsilon_{(g)}(2)]$ as instruments, where λ_{0g} is defined in (5.7A), and

(5.27) $$\varepsilon_{(g)}(2) = [\varepsilon_1(2), \ldots, \varepsilon_{g-1}(2), \varepsilon_{g+1}(1), \ldots, \varepsilon_G(1)]$$

(5.28) $$v_g(2) = \varepsilon_g - \varepsilon_{(g)}(2)\lambda_{0g}.$$

Denote the resulting IV estimate of (5.26) by $[\hat{\theta}_g(2)', \hat{\lambda}_g(2)']'$. We then update the estimate of ε_g by

(5.29) $$\varepsilon_g(2) = y_g - X_g\hat{\theta}_g(2).$$

Further rounds continue in the same fashion. Generally, <u>in Round n for equation g</u>, we apply the usual IV estimation method to

(5.30) $$y_g = X_g\theta_{0g} + \varepsilon_{(g)}(n)\lambda_{0g} + v_g(n)$$

using $[\hat{X}_g, \varepsilon_{(g)}(n)]$ as instruments, where λ_{0g} is defined in (5.7A), and

(5.31A) $\varepsilon_{(g)}(n) = [\varepsilon_1(n),\ldots,\varepsilon_{g-1}(n),\varepsilon_{g+1}(n-1),\ldots,\varepsilon_G(n-1)]$

(5.31B) $v_g(n) = \varepsilon_g - \varepsilon_{(g)}(n)\lambda_{0g}.$

Denote the resulting IV estimate by $[\hat{\theta}_g(n)',\hat{\lambda}_g(n)']'$. We then update the estimate of ε_g by

(5.32) $\varepsilon_g(n) = y_g - X_g\hat{\theta}_g(n).$

Thus, we have finished describing the iterative procedure of the feasible I2SLS estimator. This process is essentially the same as defined by Telser (1964) for the SURE model, except that IV is used here where OLS was used in the SURE model. We may note that other similar iterative processes are also possible. For example, for equation g in round n, we could augmented the equation and instruments set by

$$\varepsilon^*_{(g)}(n) = [\varepsilon_1(n-1),\ldots,\varepsilon_{g-1}(n-1),\varepsilon_{g+1}(n-1),\ldots,\varepsilon_G(n-1)]$$

instead of $\varepsilon_{(g)}(n)$ as in (5.31A); this amounts to using the estimates of ε from round n-1 to estimate all equations in round n. This would change our algebra but not any of our conclusions since, if the interative processes based on $\varepsilon_{(g)}(n)$ and $\varepsilon^*_{(g)}(n)$ both converge, they obviously converge to the same limit.

We now show that the iterated I2SLS estimators defined above are consistent and asymptotically normal in every round of iteration.

<u>LEMMA 5.1</u>: $T^{1/2}[\hat{\theta}_g(n)-\theta_{0g}] \rightarrow N[0,\Omega_g(n)]$, where $\Omega_g(n)$ is a finite positive definite (pd) matrix, for n = 0, 1, 2, ..., and g = 1, 2, ..., G.

Proof: We will only prove the case of G = 2. The proof for general G is essentially the same.

When n = 0, $\hat{\theta}_g(0)$ is just the usual 2SLS estimate of θ_{0g} in (5.3). Then it is well known that

$$(5.33) \qquad T^{1/2}[\hat{\theta}_g(0)-\theta_{0g}] \rightarrow N[0, \sigma_{gg}(A_{gz}A_{zz}^{-1}A_{zg})^{-1}]$$

using (A5.1)-(A5.3), for every g.

When n = 1, using the definition that $[\hat{\theta}_g(1)', \hat{\lambda}_g(1)']'$ is the usual IV estimate of (5.9) using $[\hat{X}_1, \varepsilon_{(1)}(1)]$ as instruments, we have

$$(5.34) \quad \begin{bmatrix} T^{1/2}[\hat{\theta}_1(1)-\theta_{01}] \\ T^{1/2}[\hat{\lambda}_1(1)-\lambda_{01}] \end{bmatrix} = \begin{bmatrix} T^{-1}\hat{X}_1'X_1 & T^{-1}\hat{X}_1'\varepsilon_{(1)}(1) \\ T^{-1}\varepsilon_{(1)}(1)'X_1 & T^{-1}\varepsilon_{(1)}(1)'\varepsilon_{(1)}(1) \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\hat{X}_1'v_1(1) \\ T^{-1/2}\varepsilon_{(1)}(1)'v_1(1) \end{bmatrix}.$$

But

$$(5.35) \qquad \begin{aligned} T^{-1}\hat{X}_1'X_1 &= (T^{-1}X_1'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z'X_1) \\ &= A_{1z}A_{zz}^{-1}A_{z1} + o_p(1) \end{aligned}$$

$$(5.36) \qquad \begin{aligned} T^{-1}\hat{X}_1'\varepsilon_{(1)}(1) &= T^{-1}\hat{X}_1'\varepsilon_2(0) \quad \text{(using the definition of } \varepsilon_{(1)}(1) \text{ in (5.10A))} \\ &= T^{-1}\hat{X}_1'[y_2 - X_2\hat{\theta}_2(0)] \\ &= T^{-1}\hat{X}_1'[(y_2 - X_2\theta_{02}) + X_2(\theta_{02} - \hat{\theta}_2(0))] \\ &= T^{-1}\hat{X}_1'[\varepsilon_2 - X_2(\hat{\theta}_2(0) - \theta_{02})] \\ &= T^{-1}\hat{X}_1'\varepsilon_2 - (T^{-1}\hat{X}_1'X_2)[\hat{\theta}_2(0) - \theta_{02}] \\ &= (T^{-1}X_1'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z'\varepsilon_2) \\ &\quad -(T^{-1}X_1'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z'X_2)[\hat{\theta}_2(0) - \theta_{02}] \\ &= A_{1z}A_{zz}^{-1}\cdot o_p(1) - A_{1z}A_{zz}^{-1}A_{z2}\cdot o_p(1) + o_p(1) \end{aligned}$$

(using (A5.1) and the consistency of $\hat{\theta}_2(0)$)

$$= o_p(1)$$

(5.37)
$$T^{-1}\varepsilon_{(1)}(1)'X_1 = T^{-1}\varepsilon_2(0)'X_1$$
$$= T^{-1}[y_2 - X_2\hat{\theta}_2(0)]'X_1$$
$$= T^{-1}[\varepsilon_2 - X_2(\hat{\theta}_2(0) - \theta_{02})]'X_1$$
$$= T^{-1}\varepsilon_2'X_1 - (\hat{\theta}_2(0) - \theta_{02})'(T^{-1}X_2'X_1)$$
$$= A_{\varepsilon x,21} - o_p(1)\cdot A_{21} + o_p(1)$$

(using (A5.1) and the consistency of $\hat{\theta}_2(0)$)

$$= A_{\varepsilon x,21} + o_p(1)$$

(5.38)
$$T^{-1}\varepsilon_{(1)}(1)'\varepsilon_{(1)}(1) = T^{-1}[\varepsilon_2 - X_2(\hat{\theta}_2(0) - \theta_{02})]'[\varepsilon_2 - X_2(\hat{\theta}_2(0) - \theta_{02})]$$
$$= \sigma_{22} + o_p(1)$$

using (A5.1) and the consistency of $\hat{\theta}_2(0)$. Combining (5.35)-(5.38), we have

(5.39)
$$\begin{bmatrix} T^{-1}\hat{X}_1'X_1 & T^{-1}\hat{X}_1'\varepsilon_{(1)}(1) \\ T^{-1}\varepsilon_{(1)}(1)'X_1 & T^{-1}\varepsilon_{(1)}(1)'\varepsilon_{(1)}(1) \end{bmatrix} = \begin{bmatrix} A_{1z}A_{zz}^{-1}A_{z1} & 0 \\ A_{\varepsilon x,21} & \sigma_{22} \end{bmatrix} + o_p(1).$$

Because

(5.40)
$$v_1(1) = y_1 - X_1\theta_{01} - \varepsilon_{(1)}(1)\lambda_{01}$$
$$= y_1 - X_1\theta_{01} - \varepsilon_2(0)\lambda_{01} \quad \text{(using (5.10A))}$$
$$= \varepsilon_1 - [y_2 - X_2\hat{\theta}_2(0)]\lambda_{01}$$
$$= \varepsilon_1 - [\varepsilon_2 - X_2(\hat{\theta}_2(0) - \theta_{02})]\lambda_{01},$$

Then

$$(5.41)\qquad \begin{aligned} T^{-1/2}\hat{X}_1{}'v_1(1) &= T^{-1/2}\hat{X}_1{}'\{\varepsilon_1 - \varepsilon_2\lambda_{01} + X_2[\hat{\theta}_2(0) - \theta_{02}]\lambda_{01}\} \\ &= T^{-1/2}\hat{X}_1{}'\varepsilon_1 - T^{-1/2}\hat{X}_1{}'\varepsilon_2\lambda_{01} \\ &\quad + T^{-1/2}\hat{X}_1{}'X_2[\hat{\theta}_2(0) - \theta_{02}]\lambda_{01} \\ &= A_{1z}A_{zz}^{-1}\cdot T^{-1/2}Z'\varepsilon_1 - A_{1z}A_{zz}^{-1}\cdot(T^{-1/2}Z'\varepsilon_2)\lambda_{01} \\ &\quad + A_{1z}A_{zz}^{-1}A_{z2}\cdot T^{1/2}[\hat{\theta}_2(0) - \theta_{02}]\cdot\lambda_{01} + o_p(1) \end{aligned}$$

using (A5.1)-(A5.2). Combining (A5.3), (5.33) and (5.41), we see that $T^{-1/2}\hat{X}_1{}'v_1(1)$ is asymptotically normal with mean zero.

Similarly

$$(5.42)\qquad \begin{aligned} &T^{-1/2}\varepsilon_{(1)}(1)'v_1(1) \\ &= T^{-1/2}\varepsilon_2(0)'\{\varepsilon_1 - \varepsilon_2\lambda_{01} + X_2[\hat{\theta}_2(0) - \theta_{02}]\lambda_{01}\} \\ &= T^{-1/2}\{\varepsilon_2 - X_2[\hat{\theta}_2(0) - \theta_{02}]\}'\{\varepsilon_1 - \varepsilon_2\lambda_{01} + X_2[\hat{\theta}_2(0) - \theta_{02}]\lambda_{01}\} \\ &= T^{-1/2}\varepsilon_2{}'(\varepsilon_1 - \varepsilon_2\lambda_{01}) - \{T^{1/2}[\hat{\theta}_2(0) - \theta_{02}]\}'[T^{-1}X_2{}'(\varepsilon_1 - \varepsilon_2\lambda_{01})] \\ &\quad + (T^{-1}\varepsilon_2{}'X_2)\cdot T^{1/2}[\hat{\theta}_2(0) - \theta_{02}]\lambda_{01} \\ &\quad - \{T^{1/2}[\hat{\theta}_2(0) - \theta_{02}]\}'(T^{-1}X_2{}'X_2)[\hat{\theta}_2(0) - \theta_{02}]\lambda_{01} \\ &= T^{-1/2}\varepsilon_2{}'(\varepsilon_1 - \varepsilon_2\lambda_{01}) - \{T^{1/2}[\hat{\theta}_2(0) - \theta_{02}]\}'(A_{x\varepsilon,21} - A_{x\varepsilon,22}\lambda_{01}) \\ &\quad + A_{\varepsilon x,22}\cdot T^{1/2}[\hat{\theta}_2(0) - \theta_{02}]\lambda_{01} + o_p(1). \end{aligned}$$

Then combining (A5.4), (5.33) and (5.42), we see that $T^{-1/2}\varepsilon_{(1)}(1)'v_1(1)$ is asymptotically normal with mean zero.

Substituting (5.39), (5.41) and (5.42) into (5.34), we have proved that $T^{1/2}[\hat{\theta}_1(1) - \theta_{01}]$ is asymptotically normal with zero mean. Thus $T^{1/2}[\hat{\theta}_1(1) - \theta_{01}] \to N[0, \Omega_1(1)]$, where $\Omega_1(1)$ is the corresponding asymptotic covariance matrix. (For our present purposes we do not need to evaluate $\Omega_1(1)$.)

The proof that $T^{1/2}[\hat{\theta}_2(1)-\theta_{02}]\rightarrow N[0,\Omega_2(1)]$ is essentially the same as the proof given above. Then, using the inductive method, we can prove that $T^{1/2}[\hat{\theta}_g(n)-\theta_{0g}]\rightarrow N[0,\Omega_g(n)]$ for all g and n.

■

5.4. The Convergence and Asymptotic Efficiency of the Iterated I2SLS Estimators

In this section, we give a convergence result for the iterated I2SLS estimator, and show that it has the same asymptotic efficiency as 3SLS.

Since $[\hat{\theta}_g(n)',\hat{\lambda}_g(n)']'$ is the usual IV estimate of (5.30) using $[\hat{X}_g,\varepsilon_{(g)}(n)]$ as instruments, we have

$$(5.43)\qquad \begin{bmatrix} \hat{X}_g'X_g & \hat{X}_g'\varepsilon_{(g)}(n) \\ \varepsilon_{(g)}(n)'X_g & \varepsilon_{(g)}(n)'\varepsilon_{(g)}(n) \end{bmatrix}\begin{bmatrix} \hat{\theta}_g(n) \\ \hat{\lambda}_g(n) \end{bmatrix}=\begin{bmatrix} \hat{X}_g'y_g \\ \varepsilon_{(g)}(n)'y_g \end{bmatrix}.$$

Because

$$(5.44)\qquad \begin{aligned} y_g &= X_g\theta_{0g}+\varepsilon_g \\ &= X_g\theta_{0g}+\varepsilon_{(g)}\lambda_{0g}+v_g \quad \text{(using (5.7B))} \\ &= [X_g,\varepsilon_{(g)}](\theta_{0g}',\lambda_{0g}')'+v_g, \end{aligned}$$

then

$$(5.45)\qquad \begin{bmatrix} \hat{X}_g'y_g \\ \varepsilon_{(g)}(n)'y_g \end{bmatrix}=\begin{bmatrix} \hat{X}_g' \\ \varepsilon_{(g)}(n)' \end{bmatrix}\{[X_g,\varepsilon_{(g)}]\begin{bmatrix} \theta_{og} \\ \lambda_{0g} \end{bmatrix}+v_g\}$$

$$= \begin{bmatrix} \hat{X}_g{}'X_g & \hat{X}_g{}'\varepsilon_{(g)} \\ \varepsilon_{(g)}(n)'X_g & \varepsilon_{(g)}(n)'\varepsilon_{(g)} \end{bmatrix} \begin{bmatrix} \theta_{0g} \\ \lambda_{0g} \end{bmatrix} + \begin{bmatrix} \hat{X}_g{}'v_g \\ \varepsilon_{(g)}(n)'v_g \end{bmatrix}.$$

Substituting (5.45) into (5.43), we get

(5.46)
$$\begin{bmatrix} \hat{X}_g{}'X_g & \hat{X}_g{}'\varepsilon_{(g)}(n) \\ \varepsilon_{(g)}(n)'X_g & \varepsilon_{(g)}(n)'\varepsilon_{(g)}(n) \end{bmatrix} \begin{bmatrix} \hat{\theta}_g(n) \\ \hat{\lambda}_g(n) \end{bmatrix}$$

$$= \begin{bmatrix} \hat{X}_g{}'X_g & \hat{X}_g{}'\varepsilon_{(g)} \\ \varepsilon_{(g)}(n)'X_g & \varepsilon_{(g)}(n)'\varepsilon_{(g)} \end{bmatrix} \begin{bmatrix} \theta_{0g} \\ \lambda_{0g} \end{bmatrix} + \begin{bmatrix} \hat{X}_g{}'v_g \\ \varepsilon_{(g)}(n)'v_g \end{bmatrix}.$$

In order to examine the convergence of $[\hat{\theta}_g(n)', \hat{\lambda}_g(n)']'$ as $n \to \infty$, we define the following notation:

(5.47A) $$B_0 = \begin{bmatrix} \hat{X}_g{}'X_g & \hat{X}_g{}'\varepsilon_{(g)} \\ \varepsilon_{(g)}{}'X_g & \varepsilon_{(g)}{}'\varepsilon_{(g)} \end{bmatrix}$$

(5.47B) $$B_n = \begin{bmatrix} \hat{X}_g{}'X_g & \hat{X}_g{}'\varepsilon_{(g)}(n) \\ \varepsilon_{(g)}(n)'X_g & \varepsilon_{(g)}(n)'\varepsilon_{(g)}(n) \end{bmatrix}$$

(5.47C) $$C_n = \begin{bmatrix} \hat{X}_g{}'X_g & \hat{X}_g{}'\varepsilon_{(g)} \\ \varepsilon_{(g)}(n)'X_g & \varepsilon_{(g)}(n)'\varepsilon_{(g)} \end{bmatrix}$$

(5.47D) $d_g(n) = \hat{\theta}_g(n) - \theta_{0g}$, for g = 1,2, ..., G and n = 1, 2, ...

(5.47E) $D_g(n) = [X_1 d_1(n), \ldots, X_{g-1} d_{g-1}(n), X_{g+1} d_{g+1}(n-1), \ldots, X_G d_G(n-1)]$ for all n and g

(5.47F) $$\Delta B_n = \begin{bmatrix} 0 & -\hat{X}_g' D_g(n) \\ -D_g(n)' X_g & -\varepsilon_{(g)}' D_g(n) - D_g(n)'\varepsilon_{(g)} + D_g(n)' D_g(n) \end{bmatrix}$$

(5.47G) $$\Delta C_n = \begin{bmatrix} 0 & 0 \\ -D_g(n)' X_g & -D_g(n)'\varepsilon_{(g)} \end{bmatrix}.$$

Then (5.46) can be rewritten as

(5.48) $$B_n \begin{bmatrix} \hat{\theta}_g(n) \\ \hat{\lambda}_g(n) \end{bmatrix} = C_n \begin{bmatrix} \theta_{og} \\ \lambda_{0g} \end{bmatrix} + \begin{bmatrix} \hat{X}_g' v_g \\ \varepsilon_{(g)}(n)' v_g \end{bmatrix}.$$

Because

(5.49) $$\begin{aligned} \varepsilon_g(n) &= y_g - X_g \hat{\theta}_g(n) \\ &= (y_g - X_g \theta_{0g}) - [X_g \hat{\theta}_g(n) - X_g \theta_{0g}] \\ &= \varepsilon_g - X_g [\hat{\theta}_g(n) - \theta_{0g}] \\ &= \varepsilon_g - X_g d_g(n) \end{aligned}$$

using (5.47D), then substituting (5.49) into the definition of $\varepsilon_{(g)}(n)$ in (5.31A), we get

(5.50) $$\begin{aligned} \varepsilon_{(g)}(n) &= [\varepsilon_1(n), \ldots, \varepsilon_{g-1}(n), \varepsilon_{g+1}(n-1), \ldots, \varepsilon_G(n-1)] \\ &= [\varepsilon_1, \ldots, \varepsilon_{g-1}, \varepsilon_{g+1}, \ldots, \varepsilon_G] \\ &\quad - [X_1 d_1(n), \ldots, X_{g-1} d_{g-1}(n), X_{g+1} d_{g+1}(n-1), \ldots, X_G d_G(n-1)] \\ &= \varepsilon_{(g)} - D_g(n) \end{aligned}$$

using the definition of $D_g(n)$ in (5.47E). Substituting (5.50) into (5.47B), we get

(5.51) $$B_n = B_0 + \Delta B_n$$

using the definitions of B_0 in (5.47A) and ΔB_n in (5.47F). Substituting (5.50) into (5.47C), we also get

$$(5.52) \qquad C_n = B_0 + \Delta C_n$$

using the definitions of B_0 in (5.47A) and ΔC_n in (5.47G). Substituting (5.51) and (5.52) into (5.48), we obtain

$$(5.53) \qquad (B_0 + \Delta B_n)\begin{bmatrix} \hat{\theta}_g(n) \\ \hat{\lambda}_g(n) \end{bmatrix} = (B_0 + \Delta C_n)\begin{bmatrix} \theta_{og} \\ \lambda_{0g} \end{bmatrix} + \begin{bmatrix} \hat{X}_g' v_g \\ \varepsilon_{(g)}(n)' v_g \end{bmatrix}.$$

This can also be rewritten as

$$(5.54) \qquad (B_0 + \Delta B_n)\begin{bmatrix} \hat{\theta}_g(n) - \theta_{0g} \\ \hat{\lambda}_g(n) - \lambda_{0g} \end{bmatrix} = (\Delta C_n - \Delta B_n)\begin{bmatrix} \theta_{og} \\ \lambda_{0g} \end{bmatrix} + \begin{bmatrix} \hat{X}_g' v_g \\ \varepsilon_{(g)}(n)' v_g \end{bmatrix},$$

or

$$(5.55) \qquad (\frac{1}{T}B_0 + \frac{1}{T}\Delta B_n)\begin{bmatrix} T^{1/2}[\hat{\theta}_g(n) - \theta_{0g}] \\ T^{1/2}[\hat{\lambda}_g(n) - \lambda_{0g}] \end{bmatrix}$$

$$= T^{-1/2}(\Delta C_n - \Delta B_n)\begin{bmatrix} \theta_{og} \\ \lambda_{0g} \end{bmatrix} + \begin{bmatrix} T^{-1/2}\hat{X}_g' v_g \\ T^{-1/2}\varepsilon_{(g)}(n)' v_g \end{bmatrix}.$$

Premultiplying (5.55) by $(\frac{1}{T}B_o + \frac{1}{T}\Delta B_n)^{-1}$, we arrive at

(5.56)
$$\begin{bmatrix} T^{1/2}[\hat{\theta}_g(n)-\theta_{0g}] \\ T^{1/2}[\hat{\lambda}_g(n)-\lambda_{0g}] \end{bmatrix}$$

$$= (\frac{1}{T}B_o + \frac{1}{T}\Delta B_n)^{-1}\{T^{-1/2}(\Delta C_n - \Delta B_n)\begin{bmatrix} \theta_{0g} \\ \lambda_{0g} \end{bmatrix} + \begin{bmatrix} T^{-1/2}\hat{X}_g{}'v_g \\ T^{-1/2}\varepsilon_{(g)}(n)'v_g \end{bmatrix}\}.$$

Now we wish to show that $\text{plim}_{T\to\infty} T^{-1}\Delta B_n = 0$, and that $T^{-1/2}(\Delta C_n - \Delta B_n)(\theta_{0g}{}', \lambda_{0g}{}')'$, $T^{-1/2}\hat{X}_g{}'v_g$ and $T^{-1/2}\varepsilon_{(g)}(n)'v_g$ each are asymptotically multivariate normal with mean zero.

<u>LEMMA 5.2</u>: For any i, j = 1, 2, ..., G, and n, m = 0, 1, 2, ...,

(1) $\text{plim}_{T\to\infty} T^{-1}\hat{X}_i{}'\varepsilon_{(j)} = 0;$

(2) $\text{plim}_{T\to\infty} T^{-1}X_i{}'X_j d_j(n) = 0;$

(3) $\text{plim}_{T\to\infty} T^{-1}\hat{X}_i{}'X_j d_j(n) = 0;$

(4) $\text{plim}_{T\to\infty} T^{-1}\varepsilon_i{}'X_j d_j(n) = 0;$

(5) $\text{plim}_{T\to\infty} T^{-1}d_i(n)'X_i{}'X_j d_j(n) = 0;$

(6) $\text{plim}_{T\to\infty} T^{-1}X_i{}'D_j(n) = 0;$

(7) $\text{plim}_{T\to\infty} T^{-1}\hat{X}_i{}'D_j(n) = 0;$

(8) $\text{plim}_{T\to\infty} T^{-1}D_i(n)'D_j(m) = 0;$

(9) $\text{plim}_{T\to\infty} T^{-1}\varepsilon_{(i)}{}'D_j(n) = 0.$

Proof: (1) Because

$$\text{plim}\, T^{-1}\hat{X}_i{}'\varepsilon_j = \text{plim}(T^{-1}X_i{}'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z'\varepsilon_j)$$
$$= A_{iz}A_{zz}^{-1}\cdot 0$$
$$= 0$$

using (A5.1)-(A5.2), then

$$\text{plim}\, T^{-1}\hat{X}_i{}'\varepsilon_{(j)} = \text{plim}[T^{-1}\hat{X}_i{}'\varepsilon_1, \ldots, T^{-1}\hat{X}_i{}'\varepsilon_{j-1}, T^{-1}\hat{X}_i{}'\varepsilon_{j+1}, \ldots, T^{-1}\hat{X}_i{}'\varepsilon_G]$$
$$= 0.$$

(2) $\text{plim}\, T^{-1}X_i'X_jd_j(n) = \text{plim}(T^{-1}X_i'X_j)d_j(n) = A_{ij} \cdot 0 = 0$ where the second equality is gotten from the definition of $d_j(n) = \hat{\theta}_j(n) - \theta_{0j}$ in (5.47D) and Lemma 5.1.

(3) $\text{plim}\, T^{-1}\hat{X}_i'X_jd_j(n) = \text{plim}(T^{-1}X_i'Z)(Z'Z)^{-1}(T^{-1}Z'X_j)d_j(n)$

$= A_{iz}A_{zz}^{-1}A_{zj} \cdot 0$ (using (A5.1) and Lemma 5.1)

$= 0$.

(4) $\text{plim}\, T^{-1}\varepsilon_i'X_jd_j(n) = \text{plim}(T^{-1}\varepsilon_i'X_j)d_j(n) = A_{\varepsilon X,ij} \cdot 0 = 0$, where the second equality is gotten from (A5.1) and Lemma 5.1.

(5) $\text{plim}\, T^{-1}d_i(n)'X_i'X_jd_j(n) = \text{plim}\, d_i(n)'(T^{-1}X_i'X_j)d_j(n) = 0 \cdot A_{ij} \cdot 0 = 0$, where the second equality is gotten from (A5.1) and Lemma 5.1.

(6) $\text{plim}\, T^{-1}X_i'D_j(n)$

$= \text{plim}\, T^{-1}X_i'[X_1d_1(n),\ldots,X_{j-1}d_{j-1}(n),X_{j+1}d_{j+1}(n-1),\ldots,X_Gd_G(n-1)]$

(using the definition of $D_j(n)$ in (5.47E))

$= 0$

using part (2) of this lemma.

(7) $\text{plim}\, T^{-1}\hat{X}_i'D_j(n)$

$= \text{plim}\, T^{-1}\hat{X}_i'[X_1d_1(n),\ldots,X_{j-1}d_{j-1}(n),X_{j+1}d_{j+1}(n-1),\ldots,X_Gd_G(n-1)]$

(using the definition of $D_j(n)$ in (5.47E))

$= 0$

using part (3) of this lemma.

(8) $\text{plim}\, T^{-1}D_i(n)'D_j(m)$

$$= \text{plim}T^{-1}\begin{bmatrix} d_1(n)'X_1' \\ \vdots \\ d_{i-1}(n)'X_{i-1}' \\ d_{i+1}(n-1)'X_{i+1}' \\ \vdots \\ d_G(n-1)'X_G' \end{bmatrix}.$$

$$\cdot[X_1 d_1(m),\ldots,X_{j-1}d_{j-1}(m),X_{j+1}d_{j+1}(m-1),\ldots,X_G d_G(m-1)]$$

(using the definition of $D_i(n)$ and $D_j(m)$ in (5.47E))

$$= (\text{p}\lim T^{-1}d_g(n^*)'X_g{}'X_h d_h(m^*))_{(G-1)\times(G-1)}$$

$$= 0$$

using part (5) of this lemma, where $n^* = n$ or n-1, and $m^* = m$ or m-1.

(9) $\quad \text{p}\lim T^{-1}\varepsilon_{(i)}{}'D_j(n)$

$$= \text{p}\lim T^{-1}\begin{bmatrix}\varepsilon_1{}' \\ \vdots \\ \varepsilon_{i-1}{}' \\ \varepsilon_{i+1}{}' \\ \vdots \\ \varepsilon_G{}'\end{bmatrix}[X_1 d_1(n),\ldots,X_{j-1}d_{j-1}(n),X_{j+1}d_{j+1}(n-1),\ldots,X_G d_G(n-1)]$$

$$= (\text{p}\lim T^{-1}\varepsilon_g{}'X_h d_h(n^*))_{(G-1)\times(G-1)}$$

$$= 0$$

using part (4) of this lemma, where $n^* = n$ or n-1.

■

<u>LEMMA 5.3</u>: (1) $\text{p}\lim_{T\to\infty} T^{-1}B_0 = \begin{bmatrix} A_{gz}A_{zz}^{-1}A_{zg} & 0 \\ A^*_{\varepsilon x,gg} & \Sigma_{gg}\end{bmatrix} \equiv B$, where

$A^*_{\varepsilon x,gg} = [A_{\varepsilon x,1g}{}',\ldots,A_{\varepsilon x,(g-1)g}{}',A_{\varepsilon x,(g+1)g}{}',\ldots,A_{\varepsilon x,Gg}{}']'$, and B is nonsingular;

(2) $\text{p}\lim_{T\to\infty} T^{-1}\Delta B_n = 0$.

Proof: (1) We must evaluate the four elements of $T^{-1}B_0$. First,

$$(5.57) \qquad T^{-1}\hat{X}_g{}'X_g = (T^{-1}X_g{}'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z'X_g)$$

$$= A_{gz}A_{zz}^{-1}A_{zg} + o_p(1)$$

using (A5.1)-(A5.2). Second,

(5.58) $$T^{-1}\hat{X}_g{}'\varepsilon_{(g)} = o_p(1)$$

using part (1) of Lemma 5.2. Third,

(5.59) $$T^{-1}\varepsilon_{(g)}{}'X_g = \begin{bmatrix} T^{-1}\varepsilon_1{}'X_g \\ \vdots \\ T^{-1}\varepsilon_{g-1}{}'X_g \\ T^{-1}\varepsilon_{g+1}{}'X_g \\ \vdots \\ T^{-1}\varepsilon_G{}'X_g \end{bmatrix} = \begin{bmatrix} A_{\varepsilon x,1g} \\ \vdots \\ A_{\varepsilon x,(g-1)g} \\ A_{\varepsilon x,(g+1)g} \\ \vdots \\ A_{\varepsilon x,Gg} \end{bmatrix} + o_p(1) = A^{*}_{\varepsilon x,gg} + o_p(1)$$

where the second equality is gotten from (A5.1). Fourth,

(5.60) $$T^{-1}\varepsilon_{(g)}{}'\varepsilon_{(g)} = T^{-1}\begin{bmatrix} \varepsilon_1{}' \\ \vdots \\ \varepsilon_{g-1}{}' \\ \varepsilon_{g+1}{}' \\ \vdots \\ \varepsilon_G{}' \end{bmatrix}[\varepsilon_1,\ldots,\varepsilon_{g-1},\varepsilon_{g+1},\ldots,\varepsilon_G] = \Sigma_{gg} + o_p(1)$$

using (A5.1) and the definition of Σ_{gg} on page 62. So, substituting (5.57)-(5.60) into $T^{-1}B_0$ with B_0 defined in (5.47A), we obtain

(5.61) $$\operatorname{plim} T^{-1}B_0 = \operatorname{plim} T^{-1}\begin{bmatrix} \hat{X}_g{}'X_g & \hat{X}_g{}'\varepsilon_{(g)} \\ \varepsilon_{(g)}{}'X_{(g)} & \varepsilon_{(g)}{}'\varepsilon_{(g)} \end{bmatrix} = \begin{bmatrix} A_{gz}A_{zz}^{-1}A_{zg} & 0 \\ A^{*}_{\varepsilon x,gg} & \Sigma_{gg} \end{bmatrix} = B.$$

But, according to (A5.2), A_{zg} has full column rank, and Σ_{gg} is nonsingular. Thus B is nonsingular.

(2) Using the definition of ΔB_n in (5.47F), we have

(5.62) $\text{p}\lim T^{-1}\Delta B_n$

$$= \text{p}\lim \begin{bmatrix} 0 & -T^{-1}\hat{X}_g{}'D_g(n) \\ -T^{-1}D_g(n)'X_g & -T^{-1}\varepsilon_{(g)}{}'D_g(n) - T^{-1}D_g(n)'\varepsilon_{(g)} + T^{-1}D_g(n)'D_g(n) \end{bmatrix}$$

$$= 0$$

using Lemma 5.2, parts (6)-(9).

■

<u>LEMMA 5.4</u>: For any i, j = 1, 2, ..., G, and n = 1, 2, ..., as the sample size $T \to \infty$,

(1) $T^{-1/2}X_i{}'X_j d_j(n) \to N[0, A_{ij}\Omega_j(n)A_{ji}]$;

(2) $T^{-1/2}\hat{X}_i{}'X_j d_j(n) \to N[0, A_{iz}A_{zz}^{-1}A_{zj}\Omega_j(n)A_{jz}A_{zz}^{-1}A_{zi}]$;

(3) $T^{-1/2}\varepsilon_i{}'X_j d_j(n) \to N[0, A_{\varepsilon x,ij}\Omega_j(n)A_{x\varepsilon,ji}]$.

Proof: (1) $T^{-1/2}X_i{}'X_j d_j(n) = (T^{-1}X_i{}'X_j)[T^{1/2}d_j(n)]$

$$= A_{ij}[T^{1/2}d_j(n)] + o_p(1) \quad \text{(using (A5.1))}$$

$$\to N[0, A_{ij}\Omega_j(n)A_{ji}]$$

using Lemma 5.1.

(2) $T^{-1/2}\hat{X}_i{}'X_j d_j(n) = (T^{-1}X_i{}'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z'X_j)[T^{1/2}d_j(n)]$

$$= A_{iz}A_{zz}^{-1}A_{zj} \cdot T^{1/2}d_j(n) + o_p(1) \quad \text{(using (A5.1))}$$

$$\to N[0, A_{iz}A_{zz}^{-1}A_{zj}\Omega_j(n)A_{jz}A_{zz}^{-1}A_{zi}]$$

using Lemma 5.1.

(3) $T^{-1/2}\varepsilon_i{}'X_j d_j(n) = (T^{-1}\varepsilon_i{}'X_j)[T^{1/2}d_j(n)]$

$$= A_{\varepsilon x,ij} \cdot T^{1/2}d_j(n) + o_p(1) \quad \text{(using (A5.1))}$$

$$\to N[0, A_{\varepsilon x,ij}\Omega_j(n)A_{x\varepsilon,ji}]$$

using Lemma 5.1.

■

LEMMA 5.5: Each element of $T^{-1/2}(\Delta C_n - \Delta B_n)$ converges in distribution to a normal distribution with mean zero as the sample size $T \to \infty$.

Proof: Using the expressions for ΔB_n in (5.47F) and ΔC_n in (5.47G), we get

$$T^{-1/2}(\Delta C_n - \Delta B_n) = \begin{bmatrix} 0 & T^{-1/2}\hat{X}_g{}'D_g(n) \\ 0 & T^{-1/2}\varepsilon_{(g)}{}'D_g(n) - T^{-1/2}D_g(n)'D_g(n) \end{bmatrix}. \tag{5.63}$$

But

$$\begin{aligned} & T^{-1/2}\hat{X}_g{}'D_g(n) \\ & = T^{-1/2}\hat{X}_g{}'[X_1d_1(n),\ldots,X_{g-1}d_{g-1}(n),X_{g+1}d_{g+1}(n-1),\ldots,X_Gd_G(n-1)]. \end{aligned} \tag{5.64}$$

Then using Lemma 5.4, part (2), we see that each column of $T^{-1/2}\hat{X}_g{}'D_g(n)$ converges in distribution to a multivariate normal distribution with mean zero as $T \to \infty$. Similarly,

$$\begin{aligned} & T^{-1/2}\varepsilon_{(g)}{}'D_g(n) \\ & = T^{-1/2}\begin{bmatrix} \varepsilon_1{}' \\ \vdots \\ \varepsilon_{g-1}{}' \\ \varepsilon_{g+1}{}' \\ \vdots \\ \varepsilon_G{}' \end{bmatrix}[X_1d_1(n),\ldots,X_{g-1}d_{g-1}(n),X_{g+1}d_{g+1}(n-1),\ldots,X_Gd_G(n-1)] \\ & = (T^{-1/2}\varepsilon_i{}'X_jd_j(n^*))_{(G-1)\times(G-1)} \end{aligned} \tag{5.65}$$

where i, j = 1, ..., (g-1), (g+1), ..., G, and $n^* = n$ if $j < g$ otherwise $n^* = n - 1$. Then, using Lemma 5.4, part (3), we see that each element of $T^{-1/2}\varepsilon_{(g)}{}'D_g(n)$ is asymptotically normal with mean zero. Finally,

$$(5.66) \qquad T^{-1/2}D_g(n)'D_g(n) = (T^{-1/2}d_i(n_1)'X_i{}'X_j d_j(n_2))_{(G-1)\times(G-1)}$$

using the definition of $D_g(n)$ in (5.47E), where n_1, n_2 = n or n - 1. But, as $T \to \infty$,

$$(5.67) \qquad T^{-1/2}d_i(n_1)'X_i{}'X_j d_j(n_2) = d_i(n_1)'(T^{-1}X_i{}'X_j)[T^{1/2}d_j(n_2)] = o_p(1)$$

using Lemma 5.1 and (A5.1). Substituting (5.67) into (5.66), we obtain

$$(5.68) \qquad T^{-1/2}D_g(n)'D_g(n) = o_p(1)$$

as $T \to \infty$. Therefore, substituting (5.64), (5.65) and (5.68) into (5.63), we prove that each element of $T^{-1/2}(\Delta C_n - \Delta B_n)$ converges in distribution to a normal distribution with mean zero as the sample size $T \to \infty$.

■

<u>LEMMA 5.6</u>: Both $T^{-1/2}\hat{X}_g{}'v_g$ and $T^{-1/2}\varepsilon_{(g)}(n)'v_g$ converge in distribution to multivariate normal distributions with mean zero as the sample size $T \to \infty$.

Proof: Using the definition of v_g in (5.7B), we have

$$(5.69) \qquad \begin{aligned} T^{-1/2}\hat{X}_g{}'v_g &= T^{-1/2}\hat{X}_g{}'[\varepsilon_g - \varepsilon_{(g)}\lambda_{0g}] \\ &= (T^{-1}X_g{}'Z)(T^{-1}Z'Z)^{-1} \cdot T^{-1/2}Z'[\varepsilon_g - \varepsilon_{(g)}\lambda_{0g}] \\ &= -A_{gz}A_{zz}^{-1} \cdot T^{-1/2}Z'\varepsilon\tilde{\lambda}_{0g} + o_p(1), \end{aligned}$$

where we define

$$(5.70)\qquad \tilde{\lambda}_{0g} = (\lambda_{og1},\ldots,\lambda_{og,g-1},-1,\lambda_{og,g+1},\ldots,\lambda_{ogG})'$$

with λ_{0gj} for j = 1, ..., g - 1, g + 1, ..., G defined in (5.7A). (Note that $\tilde{\lambda}_{0g}$ is not an estimate.) But

$$(5.71)\qquad \begin{aligned} T^{-1/2}Z'\varepsilon\tilde{\lambda}_{og} &= vec(T^{-1/2}Z'\varepsilon\tilde{\lambda}_{og}) \\ &= T^{-1/2}(\tilde{\lambda}_{og}'\otimes I_M)vec(Z'\varepsilon) \\ &= T^{-1/2}(\tilde{\lambda}_{og}'\otimes I_M)(I_G\otimes Z')vec(\varepsilon) \\ &= (\tilde{\lambda}_{og}'\otimes I_M)\cdot T^{-1/2}(I_G\otimes Z')vec(\varepsilon) \\ &\rightarrow N(0,\Omega_*) \end{aligned}$$

using (A5.3), where $\Omega_* = (\tilde{\lambda}_{0g}'\otimes I_M)(\Sigma\otimes A_{zz})(\tilde{\lambda}_{0g}\otimes I_M) = (\tilde{\lambda}_{0g}'\Sigma\tilde{\lambda}_{0g})\cdot A_{zz}$. Combining (5.69) and (5.71), we see that

$$(5.72)\qquad T^{-1/2}\hat{X}_g'v_g \rightarrow N(0,\Omega)$$

with $\Omega = A_{gz}A_{zz}^{-1}\cdot\Omega_*\cdot A_{zz}^{-1}A_{zg} = (\tilde{\lambda}_{0g}'\Sigma\tilde{\lambda}_{0g})\cdot(A_{gz}A_{zz}^{-1}A_{zg})$.

Similarly, using the expresion for $\varepsilon_{(g)}(n)$ in (5.50), we have

$$(5.73)\qquad \begin{aligned} T^{-1/2}\varepsilon_{(g)}(n)'v_g &= T^{-1/2}[\varepsilon_{(g)} - D_g(n)]'v_g \\ &= T^{-1/2}\varepsilon_{(g)}'v_g - T^{-1/2}D_g(n)'v_g. \end{aligned}$$

But

$$(5.74)\qquad T^{-1/2}\varepsilon_{(g)}'v_g \rightarrow N(0,\Omega_g)$$

using (A5.4).

$$\text{(5.75)} \qquad T^{-1/2}D_g(n)'v_g = \begin{bmatrix} T^{-1/2}d_1(n)'X_1'v_g \\ \vdots \\ T^{-1/2}d_{g-1}(n)'X_{g-1}'v_g \\ T^{-1/2}d_{g+1}(n-1)'X_{g+1}'v_g \\ \vdots \\ T^{-1/2}d_G(n-1)'X_G'v_g \end{bmatrix}$$

using the definition of $D_g(n)$ in (5.47E). The expression in (5.75) converges to a multivariate normal distribution with mean zero because, for any i and $n^* = n$ or n-1,

$$\begin{aligned} T^{-1/2}d_i(n^*)'X_i'v_g &= [T^{1/2}d_i(n^*)]'\cdot T^{-1}X_i'[\varepsilon_g - \varepsilon_{(g)}\lambda_{0g}] \\ &\quad \text{(using the definition of } v_g \text{ in (5.7B))} \\ &= -\, [T^{1/2}d_i(n^*)]'\cdot T^{-1}X_i'\varepsilon\tilde{\lambda}_{0g} \\ &\quad \text{(using the definition of } \tilde{\lambda}_{0g} \text{ in (5.70))} \\ &= [T^{1/2}d_i(n^*)]'\cdot[A_{x\varepsilon,i1},\ldots,A_{x\varepsilon,iG}]\tilde{\lambda}_{0g} + o_p(1) \end{aligned}$$

using (A5.1). Then, according to Lemma 5.1, $T^{-1/2}d_i(n^*)'X_i'v_g$ converges to a normal distribution with mean zero.

Substituting (5.74) and (5.75) into (5.73), we prove that $T^{-1/2}\varepsilon_{(g)}(n)'v_g$ converges in distribution to a multivariate normal distribution with mean zero.

■

<u>THEOREM 5.1</u>: (i) For any g = 1, 2, ..., G, both $T^{1/2}[\hat{\theta}_g(n) - \theta_{0g}]$ and $T^{1/2}[\hat{\lambda}_g(n) - \lambda_{0g}]$ converge in distribution to multivariate normal distributions with mean zero.

(ii) $T^{1/2}[\hat{\theta}_g(n) - \theta_{0g}] = (A_{gz}A_{zz}^{-1}A_{zg})^{-1}[T^{-1/2}\hat{X}_g'D_g(n)\cdot\lambda_{0g} + T^{-1/2}\hat{X}_g'v_g] + o_p(1)$.

Proof: (i) Applying Lemmas 5.3, 5.5 and 5.6 to (5.56), the result follows immediately.

(ii) Substituting Lemma 5.3 and (5.63) into (5.56), we obtain

$$(5.76)\quad \begin{bmatrix} T^{1/2}[\hat{\theta}_g(n)-\theta_{0g}] \\ T^{1/2}[\hat{\lambda}_g(n)-\lambda_{0g}] \end{bmatrix}$$

$$= \begin{bmatrix} A_{gz}A_{zz}^{-1}A_{zg} & 0 \\ A^{*}_{\varepsilon x,gg} & \Sigma_{gg} \end{bmatrix}^{-1} \left\{ \begin{bmatrix} 0 & T^{-1/2}\hat{X}_g{}'D_g(n) \\ 0 & T^{-1/2}\varepsilon_{(g)}{}'D_g(n)-T^{-1/2}D_g(n)'D_g(n) \end{bmatrix} \begin{bmatrix} \theta_{og} \\ \lambda_{0g} \end{bmatrix} \right.$$

$$\left. + \begin{bmatrix} T^{-1/2}\hat{X}_g{}'v_g \\ T^{-1/2}\varepsilon_{(g)}(n)'v_g \end{bmatrix} \right\} + o_p(1).$$

Then, using the partitioned inverse rule, we obtain

$$(5.77)\quad T^{1/2}[\hat{\theta}_g(n)-\theta_{0g}] = (A_{gz}A_{zz}^{-1}A_{zg})^{-1}[T^{-1/2}\hat{X}_g{}'D_g(n)\cdot\lambda_{0g} + T^{-1/2}\hat{X}_g{}'v_g] + o_p(1).$$

■

We will rewrite (5.77) slightly, as

$$(5.78)\qquad T^{1/2}[\hat{\theta}_g(n)-\theta_{0g}]$$
$$= (T^{-1}\hat{X}_g{}'\hat{X}_g)^{-1}[T^{-1/2}\hat{X}_g{}'D_g(n)\cdot\lambda_{0g} + T^{-1/2}\hat{X}_g{}'v_g] + o_p(1).$$

Premultiplying (5.78) by $(T^{-1}\hat{X}_g{}'\hat{X}_g)$, we obtain

$$(5.79)\qquad (T^{-1}\hat{X}_g{}'\hat{X}_g)\cdot T^{1/2}[\hat{\theta}_g(n)-\theta_{0g}]$$
$$= T^{-1/2}\hat{X}_g{}'D_g(n)\cdot\lambda_{0g} + T^{-1/2}\hat{X}_g{}'v_g + o_p(1)$$

or

$$(5.80)\qquad (T^{-1}\hat{X}_g{}'\hat{X}_g)\cdot T^{1/2}d_g(n) = T^{-1/2}\hat{X}_g{}'D_g(n)\cdot\lambda_{0g} + T^{-1/2}\hat{X}_g{}'v_g + o_p(1)$$

using the definition of $d_g(n)$ in (5.47D). Substituting (5.47E) into (5.80), we obtain

$$\begin{aligned}(5.81)\quad & (T^{-1}\hat{X}_g{}'\hat{X}_g)\cdot T^{1/2}d_g(n)\\ &= T^{-1/2}\hat{X}_g{}'[X_1d_1(n),\ldots,X_{g-1}d_{g-1}(n),X_{g+1}d_{g+1}(n-1),\ldots,X_Gd_G(n-1)]\lambda_{0g}\\ &\quad +T^{-1/2}\hat{X}_g{}'v_g + o_p(1)\\ &= T^{-1/2}\hat{X}_g{}'\{\sum_{j=1}^{g-1}X_jd_j(n)\lambda_{0gj} + \sum_{j=g+1}^{G}X_jd_j(n-1)\lambda_{0gj}\} + T^{-1/2}\hat{X}_g{}'v_g + o_p(1)\end{aligned}$$

using the definition of λ_{0g} in (5.7A). This can also be rewritten as

$$\begin{aligned}(5.82)\quad & (T^{-1}\hat{X}_g{}'\hat{X}_g)\cdot T^{1/2}d_g(n)\\ &= \sum_{j=1}^{g-1}(T^{-1}\hat{X}_g{}'X_j)T^{1/2}d_j(n)\lambda_{0gj} + \sum_{j=g+1}^{G}(T^{-1}\hat{X}_g{}'X_j)T^{1/2}d_j(n-1)\lambda_{0gj}\\ &\quad + T^{-1/2}\hat{X}_g{}'v_g + o_p(1)\\ &= \sum_{j=1}^{g-1}(T^{-1}\hat{X}_g{}'\hat{X}_j)T^{1/2}d_j(n)\lambda_{0gj} + \sum_{j=g+1}^{G}(T^{-1}\hat{X}_g{}'\hat{X}_j)T^{1/2}d_j(n-1)\lambda_{0gj}\\ &\quad + T^{-1/2}\hat{X}_g{}'v_g + o_p(1)\end{aligned}$$

using the fact that $\hat{X}_i{}'X_j = (P_ZX_i)'X_j = (P_ZX_i)'(P_ZX_j) = \hat{X}_i{}'\hat{X}_j$.

Define

$$(5.83A)\quad L = T^{-1}\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0\\ \lambda_{021}\hat{X}_2{}'\hat{X}_1 & 0 & 0 & \cdots & 0 & 0\\ \lambda_{031}\hat{X}_3{}'\hat{X}_1 & \lambda_{032}\hat{X}_3{}'\hat{X}_2 & 0 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ \lambda_{0G1}\hat{X}_G{}'\hat{X}_1 & \lambda_{0G2}\hat{X}_G{}'\hat{X}_2 & \lambda_{0G3}\hat{X}_G{}'\hat{X}_3 & \cdots & \lambda_{0G,G-1}\hat{X}_G{}'\hat{X}_{G-1} & 0\end{bmatrix}$$

(5.83B)

$$U = T^{-1}\begin{bmatrix} 0 & \lambda_{012}\hat{X}_1{}'\hat{X}_2 & \lambda_{013}\hat{X}_1{}'\hat{X}_3 & \cdots & \lambda_{01,G-1}\hat{X}_1{}'\hat{X}_{G-1} & \lambda_{01G}\hat{X}_1{}'\hat{X}_G \\ 0 & 0 & \lambda_{023}\hat{X}_2{}'\hat{X}_3 & \cdots & \lambda_{02,G-1}\hat{X}_2{}'\hat{X}_{G-1} & \lambda_{02G}\hat{X}_2{}'\hat{X}_G \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_{0,G-1,G}\hat{X}_{G-1}{}'\hat{X}_G \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

(5.83C) $$\hat{X} = \begin{bmatrix} \hat{X}_1 & & & \\ & \hat{X}_2 & & \\ & & \ddots & \\ & & & \hat{X}_G \end{bmatrix}, \quad D = T^{-1}\hat{X}'\hat{X}$$

(5.83D) $$d(n) = \begin{bmatrix} d_1(n) \\ \vdots \\ d_G(n) \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ \vdots \\ v_G \end{bmatrix}.$$

With this notation, (5.82) for $g = 1, \ldots, G$ can be expressed in matrix form as

(5.84) $$D \cdot T^{1/2}d(n) = L \cdot T^{1/2}d(n) + U \cdot T^{1/2}d(n-1) + T^{-1/2}\hat{X}'v + o_p(1).$$

Solving for $T^{1/2}d(n)$, we obtain

(5.85) $$T^{1/2}d(n) = (D-L)^{-1}U \cdot T^{1/2}d(n-1) + (D-L)^{-1} \cdot T^{-1/2}\hat{X}'v + o_p(1)$$

for $n = 1, 2, \ldots$.

The iteration procedure defined in (5.85), apart from the $o_p(1)$ term, is just the Gauss-Seidel iteration method (see Varga, 1962). We now wish to show that this iterative process converges to a limit, say d^*, and that $AV(T^{1/2}d^*)$ equals the asymptotic variance of the usual 3SLS estimator of θ_0 in (5.3). In order to prove these results, we first need to establish Lemmas 5.7 and 5.8. Define

(5.86A) $V = (v_1, \ldots, v_G)$ with v_i, i = 1, 2, ..., G defined in (5.7B)

(5.86B) $C = E(T^{-1}\varepsilon' V) \equiv (c_{ij})_{G\times G}$

(5.86C) $\Lambda = (\tilde{\lambda}_{01}, \ldots, \tilde{\lambda}_{0G})$ with $\tilde{\lambda}_{0i}$, i = 1, 2, ..., G defined in (5.70).

LEMMA 5.7: (1) $-V = \varepsilon\Lambda$;

(2) $$C = \begin{bmatrix} c_{11} & & \\ & \ddots & \\ & & c_{GG} \end{bmatrix} \text{ is pd;}$$

(3) $-C = \Sigma\Lambda$;

(4) $L - D + U = T^{-1}\hat{X}'(\Lambda' \otimes I_T)\hat{X}$.

Proof: (1) Using (5.70) and the definitions of ε and $\varepsilon_{(g)}$ in (5.4B) and (5.4D), equation (5.7B) can be rewritten as

(5.87) $$-v_g = \varepsilon\tilde{\lambda}_{0g}.$$

Because (5.87) holds for all g, we can stack the equations together as

(5.88) $$-V = \varepsilon\Lambda.$$

(2) and (3): Because $v_i = \varepsilon_i - L(\varepsilon_i | \varepsilon_{(i)})$, then $E(\varepsilon_{(i)}' v_i) = 0$ for any i. Therefore $c_{ij} = T^{-1}E(\varepsilon_i' v_j) = 0$ for $i \neq j$. Premultiplying (5.88) by $T^{-1}\varepsilon'$, we get

(5.89) $$-T^{-1}\varepsilon'V = T^{-1}\varepsilon'\varepsilon\Lambda.$$

Taking expectations of both sides of (5.89), we obtain

(5.90) $$-C = \Sigma\Lambda$$

using (5.86B) and (A5.1). Equation (5.90) can be rewritten as

(5.91) $$-\Sigma^{-1}C = \Lambda.$$

Denoting $\Sigma^{-1} = (\sigma^{ij})_{G\times G}$, then $\sigma^{ii} > 0$ for all i because $\Sigma^{-1} = (\sigma^{ij})_{G\times G}$ is pd. Comparing the diagonal elements on both sides of (5.91), we get

(5.92) $$-\sigma^{ii}c_{ii} = -1$$

using (5.70), (5.86C) and the diagonality of C. Then $c_{ii} = 1/\sigma^{ii} > 0$ for all i.

(4) Using the definitions of L, D, U, and $\hat{X}$ in (5.83) and Λ in (5.86C), we can easily verify that $L - D + U = T^{-1}\hat{X}'(\Lambda'\otimes I_T)\hat{X}$.

■

<u>LEMMA 5.8</u>: All the eigenvalues of $(D-L)^{-1}U$ equal zero.

Proof:

(5.93) $$\hat{X}'(C\otimes I_T) = \begin{bmatrix} \hat{X}_1' & & \\ & \ddots & \\ & & \hat{X}_G' \end{bmatrix} \cdot \begin{bmatrix} c_{11}I_T & & \\ & \ddots & \\ & & c_{GG}I_T \end{bmatrix}$$

$$= \begin{bmatrix} c_{11}\hat{X}_1' & & \\ & \ddots & \\ & & c_{GG}\hat{X}_G' \end{bmatrix}$$

$$= \begin{bmatrix} c_{11}I_{K_1} & & \\ & \ddots & \\ & & c_{GG}I_{K_G} \end{bmatrix} \cdot \begin{bmatrix} \hat{X}_1' & & \\ & \ddots & \\ & & \hat{X}_G' \end{bmatrix}$$

$$\equiv \tilde{C} \cdot \hat{X}'$$

where K_i is the number of columns in X_i for i = 1, 2, ..., G. According to Lemma 5.7, part (2), $c_{ii} > 0$ for all i. So we can define

(5.94) $$C_* = \begin{bmatrix} c_{11}^{1/2}I_{K_1} & & \\ & \ddots & \\ & & c_{GG}^{1/2}I_{K_G} \end{bmatrix}.$$

Then

(5.95) $$\tilde{C} = C_*^2 .$$

Using Lemma 5.7, part (4),

(5.96)
$$\begin{aligned} L - D + U &= T^{-1}\hat{X}'(\Lambda' \otimes I_T)\hat{X} \\ &= T^{-1}\hat{X}'((-\Sigma^{-1}C)' \otimes I_T)\hat{X} \quad \text{(using Lemma 5.7, part (3))} \\ &= -T^{-1}\hat{X}'(C \otimes I_T)(\Sigma^{-1} \otimes I_T)\hat{X} \\ &= -T^{-1}\tilde{C}\hat{X}'(\Sigma^{-1} \otimes I_T)\hat{X} \quad \text{(using (5.93))} \\ &= -T^{-1}C_*^2\hat{X}'(\Sigma^{-1} \otimes I_T)\hat{X} \end{aligned}$$

using (5.95). Then

$$(5.97) \qquad C_*^{-1}(D-L-U)C_* = C_*[T^{-1}\hat{X}'(\Sigma^{-1}\otimes I_T)\hat{X}]C_* \equiv M$$

using (9.96). Because both C_* and $T^{-1}\hat{X}'(\Sigma^{-1}\otimes I_T)\hat{X}$ are pd, then the matrix M defined in (5.97) is also pd. Therefore there exists a nonsingular matrix, say P, such that

$$(5.98) \qquad P^{-1}MP = I_K$$

where $K = \sum_{i=1}^{G} K_i$.

Suppose x is any eigenvalue of the matrix $(D-L)^{-1}U$, then it satisfies

$$(5.99) \qquad \left|xI_K - (D-L)^{-1}U\right| = 0$$

where $|A| = \det(A)$. Using the facts that $|A\cdot B| = |A|\cdot|B|$ and (D - L) is nonsingular, (5.99) is equivalent to

$$(5.100) \qquad \left|x(D-L)-U\right| = 0,$$

which is also equivalent to

$$(5.101) \qquad \left|C_*^{-1}[x(D-L)-U]C_*\right| = 0$$

because $|C_*| \neq 0$. Substituting (5.97) into (5.101), we get

$$(5.102) \qquad \left|xM + (x-1)C_*^{-1}UC_*\right| = 0.$$

Because $|P| \neq 0$, (5.102) is also equivalent to

(5.103) $$\left|P^{-1}[xM + (x-1)C_*^{-1}UC_*]P\right| = 0.$$

Substituting (5.98) into (103), we obtain

(5.104) $$\left|xI_K + (x-1)(C_*P)^{-1}U(C_*P)\right| = 0.$$

Using the fact that all the eigenvalues of a strict upper triangular matrix are zero, we conlude that all the eigenvalues of U are zero, since U is a strict upper triangular matrx. Because $(C_*P)^{-1}U(C_*P)$ conjugates with U, then all the eigenvalues of $(C_*P)^{-1}U(C_*P)$ are zero. Next we use the facts that if an $H \times H$ matrix Q has eigenvalues ρ_1, ..., ρ_H, then (1) for any scalar α, the eigenvalues of matrix αQ are $\alpha\rho_1$, ..., $\alpha\rho_H$; and (2) for any scalar α, the eigenvalues of matrix $(\alpha I_H + Q)$ are $(\alpha + \rho_1)$, ..., $(\alpha + \rho_H)$. From this we can conclude that all the eigenvalues of the matrix $[xI_K + (x-1)(C_*P)^{-1}U(C_*P)]$ are equal to x. Then (5.104) is equivalent to

(5.105) $$x^K = 0$$

because the determinant of a matrix equals the product of its eigenvalues. Solving for x, we get x = 0 with multiplicity K. Therefore we have proved that all the eigenvalues of the matrix $(D-L)^{-1}U$ are equal to zero.

■

Define d^* to be the limit of the iterative process

(5.106) $$T^{1/2}d(n) = (D-L)^{-1}U \cdot T^{1/2}d(n-1) + (D-L)^{-1} \cdot T^{-1/2}\hat{X}'v,$$

which is the same as (5.85) except for an $o_p(1)$ term. Because all of the eigenvalues of $(D-L)^{-1}U$ equal zero, d^* exists and the iterative process (5.106) reaches d^* in no more than K iterations. Furthermore, since (5.85) and (5.106) differ only by an $o_p(1)$ term, the probability that the process (5.85) has a limit (in n) approaches one as $T\to\infty$; and the limit of this iterated I2SLS estimator has the same asymptotic distribution as d^*. We now proceed to show that d^* (and hence the iterated I2SLS estimator) has the same asymptotic distribution as the 3SLS estimator.

<u>THEOREM 5.2</u>: $T^{1/2}d^* \to N(0,W)$, with $W=[A'(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}$, where $A=\mathrm{diag}(A_{z1},\ldots,A_{zG})$.

Proof: Since d^* is the limit of the process (5.106), it satisfies

$$(5.107)\qquad T^{1/2}d^* = (D-L)^{-1}U\cdot T^{1/2}d^* + (D-L)^{-1}\cdot T^{-1/2}\hat{X}'v.$$

Solving for $T^{1/2}d^*$, we get

$$(5.108)\qquad \begin{aligned} T^{1/2}d^* &= -(L+U-D)^{-1}\cdot T^{-1/2}\hat{X}'v \\ &= -[T^{-1}\hat{X}'(\Lambda'\otimes I_T)\hat{X}]^{-1}\cdot T^{-1/2}\hat{X}'v, \end{aligned}$$

using Lemma 5.7, part (4). But

$$(5.109)\qquad \begin{aligned} T^{-1}\hat{X}'(\Lambda'\otimes I_T)\hat{X} &= T^{-1}\begin{bmatrix}\hat{X}_1' & & \\ & \ddots & \\ & & \hat{X}_G'\end{bmatrix}(\Lambda'\otimes I_T)\begin{bmatrix}\hat{X}_1 & & \\ & \ddots & \\ & & \hat{X}_G\end{bmatrix} \\ &= T^{-1}\begin{bmatrix}X_1' & & \\ & \ddots & \\ & & X_G'\end{bmatrix}(\Lambda'\otimes P_Z)\begin{bmatrix}X_1 & & \\ & \ddots & \\ & & X_G\end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} T^{-1}X_1{}'Z & & \\ & \ddots & \\ & & T^{-1}X_G{}'Z \end{bmatrix} [\Lambda'\otimes(T^{-1}Z'Z)^{-1}]\cdot$$

$$\cdot\begin{bmatrix} T^{-1}Z'X_1 & & \\ & \ddots & \\ & & T^{-1}Z'X_G \end{bmatrix}$$

$$= A'(\Lambda'\otimes A_{zz}^{-1})A + o_p(1) \quad \text{(using (A5.1))}$$

$$= -A'(C\Sigma^{-1}\otimes A_{zz}^{-1})A + o_p(1)$$

(using Lemma 5.7, part (3) and the diagonality of C)

$$= -A'(C\otimes I_M)(\Sigma^{-1}\otimes A_{zz}^{-1})A + o_p(1).$$

Similarly

(5.110) $$T^{-1/2}\hat{X}'v = T^{-1/2}\begin{bmatrix} \hat{X}_1{}' & & \\ & \ddots & \\ & & \hat{X}_G{}' \end{bmatrix}\begin{bmatrix} v_1 \\ \vdots \\ v_G \end{bmatrix}$$

$$= \begin{bmatrix} T^{-1/2}\hat{X}_1{}'v_1 \\ \vdots \\ T^{-1/2}\hat{X}_G{}'v_G \end{bmatrix}$$

$$= \begin{bmatrix} (T^{-1}X_1{}'Z)(T^{-1}Z'Z)^{-1}\cdot T^{-1/2}Z'v_1 \\ \vdots \\ (T^{-1}X_G{}'Z)(T^{-1}Z'Z)^{-1}\cdot T^{-1/2}Z'v_G \end{bmatrix}$$

$$= \begin{bmatrix} A_{1z}A_{zz}^{-1} \cdot T^{-1/2}Z'v_1 \\ \vdots \\ A_{Gz}A_{zz}^{-1} \cdot T^{-1/2}Z'v_G \end{bmatrix} + o_p(1) \quad \text{(using (A5.1))}$$

$$= A'(I_G \otimes A_{zz}^{-1}) \cdot T^{-1/2}(I_G \otimes Z')\text{vec}(V) + o_p(1)$$

(using the definition of v in (5.83D) and V in (5.86A))

$$= -A'(I_G \otimes A_{zz}^{-1}) \cdot T^{-1/2}(I_G \otimes Z')\text{vec}(\varepsilon\Lambda) + o_p(1)$$

(using Lemma 5.7, part (1))

$$= -A'(I_G \otimes A_{zz}^{-1}) \cdot T^{-1/2}(I_G \otimes Z')(\Lambda' \otimes I_T)\text{vec}(\varepsilon) + o_p(1)$$

$$= -A'(I_G \otimes A_{zz}^{-1})(\Lambda' \otimes I_M) \cdot T^{-1/2}(I_G \otimes Z')\text{vec}(\varepsilon) + o_p(1)$$

$$\rightarrow N(0, W_1)$$

using (A5.3), where

(5.111) $$W_1 = A'(I_G \otimes A_{zz}^{-1})(\Lambda' \otimes I_M) \cdot (\Sigma \otimes A_{zz}) \cdot [A'(I_G \otimes A_{zz}^{-1})(\Lambda' \otimes I_M)]'$$

$$= A'(\Lambda'\Sigma\Lambda \otimes A_{zz}^{-1})A$$

$$= A'[(-\Sigma^{-1}C)'\Sigma(-\Sigma^{-1}C) \otimes A_{zz}^{-1}]A \quad \text{(using Lemma 5.7, part (3))}$$

$$= A'(C\Sigma^{-1}C \otimes A_{zz}^{-1})A$$

$$= A'(C \otimes I_M)(\Sigma^{-1} \otimes A_{zz}^{-1})(C \otimes I_M)A.$$

Combining (5.108)-(5.111), we obtain

(5.112) $$T^{1/2}d^* = -[T^{-1}\hat{X}'(\Lambda' \otimes I_T)\hat{X}]^{-1} \cdot T^{-1/2}\hat{X}'v$$

$$= [A'(C \otimes I_M)(\Sigma^{-1} \otimes A_{zz}^{-1})A]^{-1} \cdot T^{-1/2}\hat{X}'v + o_p(1)$$

$$\rightarrow N(0, W^*)$$

with

(5.113) $$W^* = [A'(C \otimes I_M)(\Sigma^{-1} \otimes A_{zz}^{-1})A]^{-1} \cdot W_1 \cdot$$

$$\cdot\{[A'(C\otimes I_M)(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}\}'$$

$$= [A'(C\otimes I_M)(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}\cdot$$

$$\cdot A'(C\otimes I_M)(\Sigma^{-1}\otimes A_{zz}^{-1})(C\otimes I_M)A\cdot$$

$$\cdot\{[A'(C\otimes I_M)(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}\}'$$

$$= [A'(C\otimes I_M)(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}\cdot$$

$$\cdot A'(C\otimes I_M)(\Sigma^{-1}\otimes A_{zz}^{-1})(C\otimes I_M)A\cdot$$

$$\cdot[A'(\Sigma^{-1}\otimes A_{zz}^{-1})(C\otimes I_M)A]^{-1}.$$

But

$$(5.114)\qquad (C\otimes I_M)A = \begin{bmatrix} c_{11}I_M & & \\ & \ddots & \\ & & c_{GG}I_M \end{bmatrix}\begin{bmatrix} A_{z1} & & \\ & \ddots & \\ & & A_{ZG} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11}A_{z1} & & \\ & \ddots & \\ & & c_{GG}A_{ZG} \end{bmatrix}$$

$$= \begin{bmatrix} A_{z1} & & \\ & \ddots & \\ & & A_{ZG} \end{bmatrix}\begin{bmatrix} c_{11}I_{K_1} & & \\ & \ddots & \\ & & c_{GG}I_{K_G} \end{bmatrix}$$

$$\equiv A\cdot\tilde{C}$$

where K_i = the number of columns in X_i, i = 1, 2, ..., G. Substituting (5.114) into (5.113), we get

$$(5.115)\quad W^* = [(A\tilde{C})'(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}\cdot(A\tilde{C})'(\Sigma^{-1}\otimes A_{zz}^{-1})A\tilde{C}\cdot[A'(\Sigma^{-1}\otimes A_{zz}^{-1})A\tilde{C}]^{-1}$$

$$= [A'(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}\tilde{C}^{-1}\cdot\tilde{C}A'(\Sigma^{-1}\otimes A_{zz}^{-1})A\tilde{C}\cdot\tilde{C}^{-1}[A'(\Sigma^{-1}\otimes A_{zz}^{-1})A]^{-1}$$

$$= [\mathbf{A}'(\Sigma^{-1} \otimes \mathbf{A}_{zz}^{-1})\mathbf{A}]^{-1}$$

$$= \mathbf{W}.$$

■

Because $\mathbf{W} = [\mathbf{A}'(\Sigma^{-1} \otimes \mathbf{A}_{zz}^{-1})\mathbf{A}]^{-1}$ is just the asymptotic variance of the the usual 3SLS estimator of θ_0 in (5.3), theorem 5.2 tells us that the iterated I2SLS estimator defined in section 5.3 of this chapter does indeed have the same asymptotic efficiency as the usual 3SLS estimator.

CHAPTER 6

CONCLUSION

In this dissertation we have shown how to improve on standard GMM estimators, given observable extra information. For linear models, and for certain kinds of nonlinear models, the additional information consists of variables that are uncorrelated with the instruments but correlated with the error(s) of the equation(s) being estimated. We believe that these results are empirically relevant, notably in the estimation of rational expectations models. An obvious further research question is the size of the efficiency gain that can be obtained in actual empirical work. Here the relevant issue is the strength of the correlation between the forecast errors in related series.

We have also considered the case that the extra moment conditions involve parameters that need to be estimated, and we discussed the 3SLS problem in detail. More generally, we could consider GMM based on the moment conditions

$$(6.1) \qquad 0 = E[\phi(y_t^*,\theta_0)] = E\begin{bmatrix}\phi_1(y_t^*,\theta_{01},\theta_{02})\\ \phi_2(y_t^*,\theta_{01},\theta_{02})\end{bmatrix},$$

and using the weighting matrix C^{-1}, where

$$(6.2) \qquad C = \lim_{T\to\infty} E[T\cdot\phi_T(\theta_{01},\theta_{02})\phi_T(\theta_{01},\theta_{02})'] \equiv \begin{bmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{bmatrix}$$

and where $\phi_T(\theta_1,\theta_2) = \begin{bmatrix}\phi_{T1}(\theta_1,\theta_2)\\ \phi_{T2}(\theta_1,\theta_2)\end{bmatrix} = T^{-1}\sum_{t=1}^{T}\phi(y_t^*,\theta_1,\theta_2)$. Let $\tilde\theta = (\tilde\theta_1',\tilde\theta_2')'$ be the corresponding GMM estimates. An alternative to this (standard) GMM treatment is to consider an iterative procedure, and this is feasible if ϕ_1 identifies θ_{01} with θ_{02} given, and

ϕ_2 identifies θ_{02} with θ_{01} given. Then, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are any initial estimates, we can consider the GMM problems

(6.3A) $$\bar{\theta}_1 = \arg\min_{\theta_1} \{\phi_{1|2}(\theta_1,\hat{\theta}_2)'C^{11}\phi_{1|2}(\theta_1,\hat{\theta}_2)\}$$

(6.3B) $$\bar{\theta}_2 = \arg\min_{\theta_2} \{\phi_{2|1}(\hat{\theta}_1,\theta_2)'C^{22}\phi_{2|1}(\hat{\theta}_1,\theta_2)\}$$

where

(6.4A) $$\phi_{1|2}(\theta_1,\hat{\theta}_2) = [\phi_{T1}(\theta_1,\hat{\theta}_2) - C_{12}C_{22}^{-1}\phi_{T2}(\hat{\theta}_1,\hat{\theta}_2)]$$

(6.4B) $$\phi_{2|1}(\hat{\theta}_1,\theta_2) = [\phi_{T2}(\hat{\theta}_1,\theta_2) - C_{21}C_{11}^{-1}\phi_{T1}(\hat{\theta}_1,\hat{\theta}_2)]$$

(6.4C) $$C^{11} = (C_{11} - C_{12}C_{22}^{-1}C_{21})^{-1}$$

(6.4D) $$C^{22} = (C_{22} - C_{21}C_{11}^{-1}C_{12})^{-1}.$$

This yields new estimates $\bar{\theta}_1$, $\bar{\theta}_2$, and the iterative process can be continued in an obvious fashion. We conjecture that the limit of this iterative process is asymptotically equivalent to the GMM estimate $\tilde{\theta}$. This result, if true, is a considerable generalization of our results of Chapter 5 on 3SLS and iterated improved 2SLS.

REFERENCES

REFERENCES

[1] Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, Mass.: Harvard University Press.

[2] Gallant, A.R. and H. White (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford: Basil Blackwell.

[3] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[4] Imbens, G.W. (1992), "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling," *Econometrica*, 60, 1187-1214.

[5] Imbens, G.W. (1993), "A New Approach to Generalized Method of Moments Estimation," unpublished manuscript, Harvard University.

[6] Imbens, G.W. and T. Lancaster (1994), "Combining Micro and Macro Data in Microeconometric Models," *Review of Economic Studies*, 61, 655-680.

[7] Schmidt, P. (1986), "A Curious IV Result," unpublished manuscript, Michigan State University.

[8] Schmidt, P. (1988), "Estimation of A Fixed-Effect Cobb-Douglas System Using Panel Data," *Journal of Econometrics*, 37, 361-380.

[9] Telser, L.G. (1964), "Iterative Estimation of a Set of Linear Regression Equations," *American Statistical Association Journal*, September, 845-862.

[10] Varga, R.S. (1962), *Matrix Iterative Analysis*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

[11] Wooldridge, J.M. (1993), "Efficient Estimation with Orthogonal Regressors," *Econometric Theory*, 9, 687.