







This is to certify that the  
dissertation entitled

ESTIMATION WITH PANEL DATA

presented by

KYUNG SO IM

has been accepted towards fulfillment  
of the requirements for

Ph.D. degree in Economics

  
Major professor

Date March 31, 1994

**LIBRARY**  
**Michigan State**  
**University**

**PLACE IN RETURN BOX to remove this checkout from your record.**  
**TO AVOID FINES return on or before date due.**

DATE DUE	DATE DUE	DATE DUE
<del>245</del> JUL 25 2007		
<del>245</del> SEP 26 1996		
<del>48</del> FEB 11 1997		
<del>245</del> APR 30 2007		
<del>245</del> SEP 03 2008		
<del>245</del> JUL 2 2007 051010		

MSU Is An Affirmative Action/Equal Opportunity Institution

c:\circ\datsdue.pm3-p.1

**ESTIMATION WITH PANEL DATA**

By

Kyung So Im

**A DISSERTATION**

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

**DOCTOR OF PHILOSOPHY**

Department of Economics

1994



## **ABSTRACT**

### **ESTIMATION WITH PANEL DATA**

By

Kyung So Im

This dissertation studies standard panel data models with repeated observations for large cross sections. In Chapter 2, we compare the 3SLS estimator and the generalized IV estimator, and derive some equivalence results. Also, we obtain some redundancy conditions for models where the regressors are strictly exogenous. Block diagonality of the optimal GMM weighting matrix turns out to be crucial for some instruments to be superfluous. Also, we propose some GMM estimators that are computationally simple and asymptotically no less efficient than GLS.

Chapter 3 covers weakly exogenous models. If the instruments are weakly exogenous and the errors are serially correlated, it seems that the currently used moment conditions lead to inconsistent estimators in general. The source of the serial correlations appears crucial to determine the set of orthogonality conditions. Also, we suggest some reduced lists of instruments in several useful models that would produce nearly efficient estimators.

In Chapter 4, we derive asymptotic variances of estimators when the moment conditions from covariance restrictions are used. As nonlinear optimization is not necessary to estimate these variances, this result, in

practice, would motivate people to use covariance restrictions more frequently. We also detail when the moment conditions from covariance restrictions are redundant in several popular models. An interesting result is that the instrumental variables can be useful even when they are not correlated with the regressors. We also argue that the moment conditions from covariance restrictions are useful always unless the GLS efficiency is reached.

To the memory of my parents



## **ACKNOWLEDGMENTS**

Many people have assisted me in getting through the graduate program. Although there are too many names to mention them all, I am happy to have a chance to express my gratitude.

Were it not for the aid of Professor Jeffrey Wooldridge, the dissertation committee chair, I would still be idling my time away. He provided careful advice on every aspects of my thesis with exceeding intelligence and consistent patience. I was a demanding student, but he always split his tight schedule and found time for listening to my questions. Especially, his careful correction on the final draft of Chapter 2 is appreciated deeply.

I am grateful to the other two committee members, Professors Peter Schmidt and Ching-Fan Chung, for their time and efforts. The detailed and critical comments by Professor Schmidt on the early drafts of Chapter 2 and 4 improved the quality of this thesis substantially, and will be invaluable lessons through my career.

The wonderful lectures from many faculty and the assistance of the departmental staff are appreciated. I must thank Professor Richard Baillie for his careful reading and comments on Chapter 2 and 4, and Professor Paul Chen for his advice and encouragement.

It was my fortune to have known Dongin Lee, my four year officemate. We talked on whatever occurred to us. Without him, this course should have been much more arduous. Joonsoo Lee's guidance was also quite helpful.

My greatest luck in graduate school was meeting and marrying my wife, Eunhee Kuh. She sacrificed her own study and brought a lovely girl and a healthy boy into this wonderful world. Words cannot be found to express my thanks to her. My baby brother Kyung Tae and big sister Hye Kyung have provided thorough support throughout the course. And I am also indebted to my other brothers Kyung Rae, Kyung Ho and Kyung Hun for their encouragement.

Thank you all who prayed for me.

## TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION .....	1
CHAPTER 2. ESTIMATION USING PANEL DATA UNDER STRICT EXOGENEITY	
1. INTRODUCTION .....	14
2. 3SLS, GIV, AND REDUNDANCY CONDITIONS	
2.1. PRELIMINARIES .....	18
2.2. EFFICIENCY COMPARISON OF 3SLS AND GIV .....	20
2.3. NUNERICAL EQUIVALENCE OF 3SLS AND GIV .....	23
2.4. ALGEBRAIC REDUNDANCY OF INSTRUMENTS IN 3SLS ..	26
3. MODEL WHERE THE REGRESSORS ARE UNCORRELATED WITH THE ERRORS	
3.1. UNRESTRICTED COVARIANCE MATRIX .....	26
3.2. DIAGONAL COVARIANCE MATRIX .....	28
3.3. RANDOM EFFECTS STRUCTURE .....	29
3.4. A GENERALIZATION OF THE RANDOM EFFECTS ASSUMPTION .....	34
4. MODEL WHERE THE REGRESSORS ARE CORRELATED WITH THE TIME CONSTANT ERROR COMPONENTS .....	37
4.1. A "FIXED EFFECTS" TYPE MODEL .....	38
4.2. HAUSMAN AND TAYLOR MODEL .....	41
4.3. HT MODEL WITH SERIALY CORRELATED TIME-VARYING ERRORS .....	44
5. CONCLUSION .....	46
APPENDIX 1 .....	48
APPENDIX 2 .....	52
CHAPTER 3. ESTIMATION USING PANEL DATA UNDER WEAK EXOGENEITY	
1. INTRODUCTION .....	56
2. SERIAL CORRELATION AND CONSISTENCY OF ESTIMATORS .	60
2.1. MOMENT CONDITIONS UNDER WEAK EXOGENEITY .....	61
2.2. SERIAL CORRELATION AND MOMENT CONDITIONS ....	63
3. ESTIMATION WITH THE BMS ASSUMPTION .....	70



4. NEARLY EFFICIENT ESTIMATION .....	76
5. CONCLUSION .....	82
CHAPTER 4. INFORMATION FROM COVARIANCE RESTRICTIONS IN PANEL DATA MODELS	
1. INTRODUCTION .....	84
2. PRELIMINARIES	
2.1. REDUNDANCY CONDITIONS FOR MOMENT RESTRICTIONS	87
2.2. SCALAR COVARIANCE AND THE ASYMPTOTIC VARIANCE OF GMM .....	89
2.3. RANDOM EFFECTS COVARIANCE AND THE ASYMPTOTIC VARIANCE OF GMM .....	95
3. STRICTLY EXOGENOUS MODELS .....	99
3.1. GENERAL RESULTS ON NONREDUNDANCY UNDER IDEAL CONDITIONS .....	99
3.2. STRICTLY EXOGENOUS MODEL: SCALAR COVARIANCE ..	101
3.3. STRICTLY EXOGENOUS MODEL: RANDOM EFFECTS COVARIANCE .....	106
3.4. STRICTLY EXOGENOUS MODEL: FIXED EFFECTS TYPE .	107
4. WEAKLY EXOGENOUS MODELS .....	111
4.1. WEAKLY EXOGENOUS MODEL: DIAGONAL COVARIANCE ..	111
4.2. WEAKLY EXOGENOUS MODEL: RANDOM EFFECTS COVARIANCE .....	115
5. CONCLUSION .....	121
CHAPTER 5. CONCLUDING REMARKS .....	123
LIST OF REFERENCES .....	126

## CHAPTER ONE

### INTRODUCTION

This dissertation deals with linear panel data models with repeated time observations for large cross sections. A basic model is

$$(1) \quad y_{it} = x_{it}\beta + u_{it}, \quad t = 1, \dots, T,$$

where  $x_{it}$  is a  $1 \times k$  vector and  $\beta$  is a  $k \times 1$  parameter vector of primary interest. Let  $x_i = (x_{i1}', \dots, x_{iT}')'$  with  $y_i$  and  $u_i$  similarly defined. Allowing time-constant unobserved individual effects, that may in many instances bias the estimators obtained from single cross section data, we write

$$(2) \quad u_{it} = \phi_i + \epsilon_{it}, \quad t = 1, \dots, T.$$

Thus,  $\phi_i$  is the time-constant error component and  $\epsilon_{it}$  is the idiosyncratic error. We assume there is a  $T \times h$  matrix of instrumental variables  $w_i$ ; these instruments are suggested by various assumptions. We are interested in the case where  $\phi_i$  are treated as random, and not as fixed parameters to be estimated. We do consider the case where  $\phi_i$  is correlated with some or all of the regressors; for many applications this is an important feature.

This chapter provides a summary of the main results contained in the subsequent chapters, and links them to

previous studies. The following chapters are essentially independent of each other, and can be read separately. In each chapter we define the relevant notation. Whenever we refer to theorems, equations or assumptions contained in other chapters, we will specify the chapter number.

References are gathered at the end of the thesis.

In Chapter 2, we are primarily concerned with estimation in models where the regressors are strictly exogenous to the idiosyncratic errors:

$$(3) \quad E(x_i \otimes \epsilon_i) = 0.$$

But, before dealing with specific models, we compare 3SLS and generalized IV (GIV). This comparison appears in Bowden and Turkington (1984, p. 72) and White (1984, pp. 83-105; 1986). But, a general result has not been established yet.

We assume

$$(4) \quad E(u_i u_i' | w_i) = E(u_i u_i') \equiv \Sigma.$$

The 3SLS and GIV estimators are defined as

$$\hat{\beta}_{3SLS} = [X'W(W'\Omega W)^{-1}W'X]^{-1}X'W(W'\Omega W)^{-1}W'Y,$$

and

$$\hat{\beta}_{GIV} = [X'\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}X]^{-1}X'\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}Y,$$

where  $(Y, X, W)$  is the data matrix stacking  $(y_i, x_i, w_i)$ ,  $i = 1, \dots, N$ , and  $\Omega = I_N \otimes \Sigma$ . 3SLS and GIV utilize the instruments  $w_i$  and  $\Sigma^{-1}w_i$ , respectively. We show that there in general is



no dominance one over the other between these two, and they are numerically identical when the time periods use common instruments or when  $\Sigma$  is diagonal. Thus, the well-known equivalence result between OLS and GLS in the SUR model follows as a corollary of our result when  $w_i = x_i$ .

We then turn to the specific models and provide several reduced lists of instrumental variables that lead to the fully efficient estimators under several different assumptions. Asymptotically, there is no reason to reduce the set of instruments since GMM never loses asymptotic efficiency by adding orthogonality conditions. However, GMM based on restricted instruments is not only computationally simpler but could have better finite sample properties.

The unobserved effects model is standard if we add the assumption of the random effects covariance matrix

$$(5) \quad \Sigma = \sigma_{\epsilon}^2 I_T + \sigma_{\phi}^2 e_T e_T',$$

where  $E(\epsilon_{it}^2) = \sigma_{\epsilon}^2$ ,  $t = 1, \dots, T$ ,  $E(\phi_i^2) = \sigma_{\phi}^2$ ,  $I_T$  is the  $T \times T$  identity matrix, and  $e_T$  is the  $T \times 1$  vector of ones. If  $x_i$  and  $\phi_i$  are not correlated with each other, the model is the popular random effects model, and the random effects estimator (GLS) is the most efficient. We show that 3SLS utilizing the instrumental variables (PX, QX) is GLS, and propose GMM using (PX, QX) as the instruments, where PX and QX denote the  $NT \times k$  meaned and demeaned matrices of the data matrix X. If the assumption (4) does not hold - for

example, in the presence of heteroskedasticity and/or serial correlation in the  $\epsilon_{it}$  conditioned on  $w_i$  - GMM using the instruments (PX,QX) is generally more efficient than the random effects estimator.

Hausman and Taylor (1981) (HT hereafter) allowed for some of  $x_i$  to be correlated with  $\phi_i$ , and showed that how the coefficients on the time-constant variables are identified. Subsequently, Amemiya and MaCurdy (1986) and Breusch, Mizon and Schmidt (1989) (BMS henceforth) developed more efficient estimators under some additional assumptions. We argue that the optimal weighting matrix  $E(w_i' u_i u_i' w_i)$  needs to be block diagonal to generate the fully efficient GMM with reduced lists of the instrumental variables. This unified theme provides intuition behind the redundancy results established by BMS, Ahn and Schmidt (1992) and many of the theorems in this thesis including the previous result that 3SLS using (PX,QX) is GLS in the random effects model. If  $\Sigma$  is of the random effects form, then  $\Sigma$  can be expressed as  $aP_T + bQ_T$ , for some scalars  $a$  and  $b$ . Thus if  $w_i$  can be decomposed into  $(P_T w_{1i}, Q_T w_{2i})$ , where  $w_{1i}$  is generally constructed from the regressors that are not correlated with  $\phi_i$  (an important exception is the instrument set suggested by BMS) and  $w_{2i}$  usually is based on all of the time varying regressors, and provided assumption (4) holds, the optimal weighting matrix become block diagonal simply because  $P_T Q_T = 0$ . We show a redundancy result through an example when  $w_{2i} = L \otimes x_i^0$ , which

is standard, where  $L$  is the  $T \times (T-1)$  differencing matrix and  $x_i^0 = (x_{i1}, \dots, x_{iT})$ . Then  $W_2(W_2'W_2)^{-1}W_2'X = QX$  and the resulting estimators are the same whether  $W_2$  or  $QX$  are used as instruments.  $W_2$  contains  $T(T-1)k$  instruments, but only the  $k$  instruments  $QX$  are useful and the rest are redundant.

This also explains the well-known lemma showing GLS in the random effects model is a linear combination of the within and the between estimators, and generalizes this lemma to the case when the optimal weighting matrix is block diagonal. We show that  $\Sigma = aP_T + bQ_T \Rightarrow P_T'\Sigma Q_T = 0$ , but the converse is not true.  $P_T'\Sigma Q_T = 0$  is sufficient to make the optimal weighting matrix block diagonal, provided  $w_i = (P_T'w_{1i}, Q_T'w_{2i})$  and assumption (4) holds. Another important case when the optimal weighting matrix become block diagonal is when

$$(6) \quad \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_T^2),$$

that is, when there are no time-constant unobserved effects and the errors are serially uncorrelated. Under the assumptions (4) and (6),  $E(w_i'u_i u_i'w_i)$  is block diagonal with the  $t$ -th block  $\sigma_t^2 E(w_{it}'w_{it})$ , where  $w_{it}$  is the instruments for the  $t$ -th period equation. If  $w_{it} \supset x_{it}$  for  $t = 1, \dots, T$ , 3SLS utilizing the instruments  $\text{diag}(x_{i1}, \dots, x_{iT})$  is GLS. This covariance matrix is especially relevant for the rational expectations models where  $\phi_i$  does not present and the errors are necessarily serially uncorrelated.



Assumption (5) is now almost standard in the panel data literature. But, in general, there are neither a priori grounds nor any technical reasons that justify this form of  $\Sigma$ . Therefore, we consider the case when the idiosyncratic errors are serially correlated. Ahn and Schmidt (1991) showed that 3SLS using all the instruments  $I_i \otimes x_i^0$  is GLS. We provide a simpler proof than theirs, and reconsider the HT model allowing for the idiosyncratic errors to be serially correlated in an arbitrary manner. Some of the instrumental variables turn out to be redundant, but the number is smaller than under (5). Also, we show that GIV is not consistent unless the equi-correlation assumption of BMS holds, but when the BMS assumption holds, GIV can reduce the number of instrumental variables substantially.

We also consider the model when all of the regressors are correlated with  $\phi_i$ . If the idiosyncratic errors are arbitrarily correlated, the model is very similar to the fixed effects model with arbitrary intertemporal covariance considered by Kiefer (1980). We show that the several estimators, including the Kiefer's estimator (GLS in demeaned equation using a generalized inverse of  $Q_T \Sigma Q_T$ ), GLS in the differenced equations, and GLS in the demeaned equations after deleting any one equation, are numerically identical. Thereby, generalized inverting in this case is an unnecessary complication.

In Chapter 3, we study the models where the regressors

are weakly exogenous to the idiosyncratic errors. Thus in place of (3) we have the assumption

$$(7) \quad E(x_{it}'\epsilon_{is}) = 0, \quad s \geq t.$$

Dynamic models and rational expectations models are the leading examples of weakly exogenous but not strictly exogenous models. However, weakly exogenous models would be suitable in broader applications. We are primarily concerned with the consistency of estimators when the errors are serially correlated, and with the consistency of the usual standard errors of the 3SLS estimators when certain moment conditions are used. Also, some nearly efficient estimators based on some reduced lists of instruments are proposed

We ask a basic question whether the moment conditions in (7) are valid when the idiosyncratic errors are serially correlated. In dynamic models, it now is well known that no moment conditions exist between the lagged dependent variables and the disturbances if the idiosyncratic errors are arbitrarily serially correlated. We show that a similar relation exists between the general weakly exogenous regressors and the errors unless the time-varying errors contain two components; the serially correlated components to which the regressors are strictly exogenous and the serially uncorrelated components to which the regressors are weakly exogenous.

Keane and Runkle (1992) proposed 2SLS upon the forward filtered equations when the errors are serially correlated, adapting a suggestion by Hayashi and Sims (1983) in a pure time series context. Schmidt, Ahn and Wyhowski (1992) (SAW henceforth), in a comment on Keane and Runkle, provided the maximal sets of the instrumental variables in several weakly exogenous models, and showed that the Keane and Runkle estimator is numerically identical to 3SLS when all the instruments are used. Hayashi and Sims (1983) and SAW indicated that eliminating the serial correlations by forward filtering is justified only when "the serial correlations in the errors are independent of the current and lagged values of the instrumental variables". But, if the Keane and Runkle estimator is inconsistent, so is 3SLS since they are the same. Thereby, the requirement for vindicating forward filtering noted by SAW and Hayashi and Sims indeed is needed for the moment conditions in (7) to be valid.

Wooldridge (1993) showed that the usual standard errors for the nonlinear 3SLS estimators in hedonic pricing models are not consistent, and derived a condition for the usual 3SLS standard errors to be consistent. Ahn (1990) obtained a similar result in the dynamic model when certain instruments are used. If BMS's equi-correlation assumption that  $E(x_{it}'\phi_i)$  are the same over  $t$  holds, we have  $(T-1)k$  additional instruments. We show that the usual 3SLS

standard errors are not consistent if these instruments are used. Applying BMS's condition to dynamic models, we have the condition that  $E(y_{it}'\phi_i)$  are the same over  $t$ , which is implied by the stationarity of  $\{(y_{it}\phi_i):t=0,\dots,T\}$  suggested by Arellano and Bover (1990). This condition implies  $T-1$  instruments, and the usual 3SLS standard errors are not consistent if these instruments are used. In fact, the result obtained by Ahn (1990) is based on the instrumental variables obtained from the moment conditions  $E[(y_{it}-\alpha y_{it-1})\phi_i]$  being the same over  $t$ , which is weaker than BMS condition (or Arellano and Bover), but the structures of the instruments from these conditions are quite similar. Thus, Ahn's result is closely related to ours.

As we argued above, all the instrumental variables are useful unless the optimal weighting matrix is block diagonal in general. When the instruments are weakly exogenous the diagonal would be the only structure of  $\Sigma$  that makes the optimal weighting matrices block diagonal. Thus, any attempts to find some reduced lists of instruments that produce the fully efficient estimators may not be fruitful. But, it is practically useful to find some reduced set of instruments. In many applications, the weakly and the strictly exogenous regressors exist together in a model. A leading example is the dynamic model with strictly exogenous regressors considered by many. Letting  $X_1$  and  $X_2$  be weakly and strictly exogenous regressors, respectively, we suggest

instruments  $(R, \Omega^{-1}X_2)$ , where  $R$  includes all the instruments between  $X_1$  and the errors.

In Chapter 4, we study the moment conditions from covariance restrictions, which are essentially nonlinear in the parameters. Covariance restrictions are rarely used in practice in standard models, perhaps because people are reluctant to utilize a priori restrictions that will cause inconsistency of estimators when they are false. Another important reason would be the computational burden of numerical optimization. But, if covariance restrictions bring non-trivial efficiency gains, computational burden is secondary. We show how to consistently estimate the asymptotic variances of the nonlinear GMM estimators that use the moment conditions from covariance restrictions without numerical optimization. If the efficiency gains by adding the moment conditions from covariance restrictions is non-trivial, then it is worth doing numerical optimization. Testing whether the covariance restrictions are valid is straightforward, so we can get around possible inconsistency problem. This result applies to the general simultaneous equations models as well as to the panel data models.

Next, we find when the moment conditions from covariance restrictions are redundant. We consider the scalar and the random effects covariance matrices in strictly and weakly exogenous models. In the strictly exogenous model with scalar covariance matrix, so where OLS

is efficient under standard set of assumptions, it turns out that the moment conditions from covariance restrictions are useful unless certain third moment conditions of the errors are met. In this case, the nonlinear GMM estimator is equivalent to the linear GMM estimator using the instruments made of residuals from initial consistent estimators, and these instruments are not correlated with the regressors. Therefore, what we find is that the instruments can be useful even when they are not correlated with the regressors. This deviates from convention. In fact, the efficiency gains in this case follow from the correlations between the instruments and the error square sequence  $\{u_{it}^2\}$ . In another words, the additional instrumental variables from covariance restrictions necessarily cause heteroskedasticity to be useful. This relates to the results obtained by Cragg (1983) and Chamberlain (1982) that additional instruments (other than the regressors) can be useful in standard regression models where all the regressors are valid instruments under heteroskedasticity of unknown form. The asymptotic variance of the estimator when we use instruments  $w_i$  is (we consider here a single equation case for simplicity),

$$(8) \quad [E(x_i' w_i) E(u_i^2 w_i' w_i)^{-1} E(w_i' x_i)]^{-1}.$$

Let  $w_i = (x_i, z_i)$ , where  $z_i$  is the set of additional instruments, and suppose  $E(z_i' x_i) = 0$  (this is not necessary



but simplify the algebra), then (8) becomes

$$E(x_i' x_i)^{-1} \{ E(u_i^2 x_i' x_i) - E(u_i^2 x_i' z_i) [E(u_i^2 z_i' z_i)]^{-1} E(u_i^2 z_i' x_i) \} E(x_i' x_i)^{-1},$$

which is strictly smaller than OLS variance as long as  $E(u_i^2 x_i' z_i) \neq 0$ . Thus,  $z_i$  contributes by explaining  $u_i^2$ .

Another general result we obtain in Chapter 4 is that the information from covariance restrictions is useful if the GLS efficiency is not reached. Along with the above arguments, it becomes obvious that the GLS efficiency is necessary but not sufficient for the moment conditions from covariance restrictions to be redundant.

In the Hausman and Taylor (1981) model, GLS is not consistent. But, the regression is separated into the two orthogonal spaces as we argued above, and the GLS efficiency is reached in the deviation space. As was conjectured by Ahn and Schmidt (1992), the instrumental variables from the random effects covariance restrictions are in the deviation space, and they are redundant if certain higher conditions on the error are satisfied.

In weakly exogenous models, GLS is not consistent unless  $\Sigma$  is diagonal. It seems that covariance restrictions are useful if  $\Sigma$  is not diagonal, and they can be redundant if  $\Sigma$  is diagonal. We show these through several examples that quite often appear in applications. If there present unobserved effects  $\phi_i$  in weakly exogenous model, the moment conditions from covariance are useful because  $\Sigma$  is not

diagonal. This holds whether  $\phi_i$  is correlated with the regressors or not.

## **CHAPTER 2**

### **ESTIMATION WITH PANEL DATA UNDER STRICT EXOGENEITY**

#### **1. INTRODUCTION**

This chapter has two purposes. The first is to establish equivalences between certain three stage least squares (3SLS) and generalized instrumental variables (GIV) estimators in several panel data models with strictly exogenous regressors. The second purpose is to find minimal sets of nonredundant instruments for 3SLS under various assumptions that have been used in the panel data literature. Extensions of the standard assumptions are also considered.

The 3SLS estimator considered in this paper appears in Amemiya (1977, equation 5.4), Hausman, Newey and Taylor (1987), and Schmidt (1990, equation 5), and is the generalized method of moments estimator (GMM) under standard assumptions. The GIV estimator has been considered by White (1984, pp. 85-105; 1986), Bowden and Turkington (1984, pp. 68-72), Schmidt (1990), and many others.

In the models we consider, there will always be an estimator that is asymptotically no less efficient than 3SLS and GIV, namely the GMM estimator using all orthogonality conditions and an unrestricted weighting matrix. Thus, in

terms of achieving asymptotic efficiency there is really no need to weed out redundant orthogonality conditions. As a practical matter, though, it is very useful to know whether, under certain assumptions, the list of nonredundant instruments is shorter than the list of all possible instruments. For example, in a panel data model with 10 strictly exogenous time-varying regressors and six time periods, the total number of instruments is 360. This can cause computational problems for GMM, especially when the cross section dimension is small.

Even if there is no computational issue with GMM, there might be good statistical reasons for using estimators based on fewer orthogonality conditions. As an illustration, consider a result obtained in section 3.3. There, it is shown that a 3SLS estimator based on a reduced instrument list is, under the usual random effects assumptions, equivalent to the random effects (GLS) estimator. This 3SLS estimator is based on many fewer instruments than an unstructured GMM analysis would be. This result implies that, if we then compute the GMM estimator with optimal weighting matrix using the restricted 3SLS instruments, we obtain an estimator no less efficient than random effects. In addition, if the random effects assumption fails, then the estimator is generally more efficient than random effects. And while this GMM estimator based on restricted instruments is generally less efficient than the full GMM

estimator if the random effects assumptions fail, it has computational advantages and could very well have better finite sample properties. Thus, the redundancy conditions do have practical applications.

Throughout this chapter (and in later chapters) we focus on general unobserved effects panel data models, where the time constant unobserved effect may or may not be correlated with some or all of the observed regressors. We cover models with serially uncorrelated idiosyncratic errors as well as models where the idiosyncratic errors are allowed to be serially correlated with time-varying variances. Such a setup captures the flavor of both random effects and fixed effects-type specifications. In a random effects framework, the key assumption is that the observed regressors are uncorrelated with the unobserved effects. For many fixed effects applications, the key feature is that the unobservable effects can be correlated with some or all of the regressors. We consider both cases in what follows, and this is sufficient for the vast majority of applications. The strict fixed effects framework, which assumes that the unobserved effects are constants that differ across individual, is not treated here. See Hsiao (1986, pp. 41-49) for a discussion of the conceptual issues underlying the fixed versus random effects dichotomy.

Section 2 contains some general results concerning the equivalence between 3SLS and GIV. These are extensions of

the well-known equivalences between OLS and GLS under certain conditions in the seemingly unrelated regressions model, and of the equivalence between 3SLS and 2SLS in simultaneous equations models.

Section 3 studies panel data models where the regressors are uncorrelated with the composite errors. In a general model the equivalence of the 3SLS estimator using all orthogonality conditions and GLS is established, giving a different proof of a result of Ahn and Schmidt (1991). A new result showing the equivalence of a 3SLS estimator with reduced instrument set and GLS is presented in section 3.3.

Section 4 turns to models where the time-constant unobserved effects are potentially correlated with some or all of the regressors. Several estimators are shown to be identical for estimating the parameters in the unobserved effects analog of Kiefer's (1980) fixed effects model. We also study the Hausman and Taylor (1981) (HT hereafter) model under more general assumptions. Hausman and Taylor showed how the coefficients on the time constant regressors could be identified when the time constant regressors are correlated with the unobserved effects in a model with serially independent idiosyncratic errors. Efficient estimation in the HT model was considered further by Breusch, Mizon and Schmidt (1989) (BMS hereafter) under some additional assumptions. These results are extended to allow for arbitrary serial correlations in the idiosyncratic

errors.

## 2. 3SLS, GIV, AND REDUNDANCY CONDITIONS

### 2.1. Preliminaries

Consider a linear panel data model,

$$y_i = x_i\beta + u_i, \quad (2.1)$$

where  $y_i = (y_{i1}, \dots, y_{iT})'$  and  $x_i = (x_{i1}', \dots, x_{iT}')'$  of dimension  $T \times 1$  and  $T \times k$ , respectively.  $\{(y_i, x_i) : i=1, \dots, N\}$  is an i.i.d. random sequence.

Throughout the paper, for any  $T \times p$  matrix  $m_i$ ,  $M \equiv (m_1', \dots, m_N')'$  of dimension  $NT \times p$ , where  $N$  is the number of observations. Thus, the matrix  $M$  is the stacked matrix of  $m_i$  for  $i = 1, \dots, N$  with the  $i$ -th block  $m_i$ . In the sample,  $Y = X\beta + U$ .

Most of the results we discuss in this chapter deal with algebraic equivalences of various estimators and have nothing to do with statistical properties such as consistency and asymptotic normality. Nevertheless, one would probably not use the estimators unless certain assumptions are satisfied. It is useful to set out some assumptions that typically underlie method of moments-type estimators. Because we are studying both 3SLS and GIV we make assumptions that traditionally underlie application of



these methods.

To consistently estimate  $\beta$  appearing in (2.1) we assume that there is a set of  $T \times h$  instruments,  $w_i$ , that are appropriately orthogonal to  $u_i$ . A fairly standard set of assumptions is

ASSUMPTION 2.1:

$$(a) \ E(w_i' u_i) = 0.$$

$$(b) \ E(w_i' \Sigma^{-1} u_i) = 0, \text{ where } \Sigma \equiv E(u_i u_i') \text{ is nonsingular.}$$

ASSUMPTION 2.2:

$$(a) \ E(w_i' x_i) \text{ has full column rank and } E(w_i' \Sigma w_i) \text{ is positive definite.}$$

$$(b) \ E(w_i' \Sigma^{-1} x_i) \text{ has full column rank and } E(w_i' \Sigma w_i) \text{ is positive definite.}$$

ASSUMPTION 2.3:

$$(a) \ E(w_i' u_i u_i' w_i) = E(w_i' \Sigma w_i).$$

$$(b) \ E(w_i' \Sigma^{-1} u_i u_i' \Sigma^{-1} w_i) = E(w_i' \Sigma^{-1} w_i).$$

Assumption 2.1(a) and 2.2(a) ensure that the 3SLS estimator is consistent under standard regularity conditions. As a practical matter, Assumption 2.2(a) implies that the 3SLS estimator exists with probability approaching one (as the sample size grows); for what follows, we just assume the estimator exists for any sample. Assumption 2.3(a) is the weakest assumption that guarantees that the usual formula for the asymptotic variance of 3SLS is valid. Assumptions 2.1(b) and 2.2(b) imply consistency of the GIV; Assumption

2.3(b) implies that its asymptotic variance matrix is of a relatively simple form. Note that a sufficient condition for both parts of Assumption 2.3 is  $E(u_i u_i' | w_i) = E(u_i u_i') = \Sigma$ .

When we study 3SLS in later sections, we typically have in mind assumptions such as Assumption 2.1(a), 2.2(a), and 2.3(a). The key will be to find instruments  $w_i$  that satisfy these conditions under more primitive assumptions about the models at hand. Assumption 2.1(a) is critical and dictates the choice of  $w_i$ . Assumption 2.2(a) can be viewed as a regularity condition. Assumption 2.3(a) cannot be guaranteed a priori, but it is useful as a starting point.

In practice, one needs a consistent estimator of  $\Sigma$  to perform 3SLS or GIV. Nothing is lost in the following analysis by assuming  $\Sigma$  is known because it is consistently estimable in the models we deal with.

## 2.2. Efficiency Comparison of 3SLS and GIV

Given the data matrices  $X$ ,  $W$ , and  $Y$ , and defining  $\Omega \equiv I_N \otimes \Sigma$ , the 3SLS estimator is defined to be

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

Equivalently, this is the GMM estimator based on the orthogonality condition  $E(W'U)=0$  using weighting matrix  $(W'\Omega W)^{-1}$ . The GIV estimator first transforms (2.1) to spherical disturbances by premultiplying by  $\Sigma^{-1/2}$ , and then

uses  $\Sigma^{-1/2}w_i$  as the instruments. This gives

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

Under Assumptions 2.1-2.3, we have the following asymptotic variances:

$$Avar \sqrt{N}(\hat{\beta}_{3SLS} - \beta) = [E(x_i' w_i) \{E(w_i' \Sigma w_i)\}^{-1} E(w_i' x_i)]^{-1},$$

and

$$Avar \sqrt{N}(\hat{\beta}_{GIV} - \beta) = [E(x_i' \Sigma^{-1/2} w_i) \{E(w_i' \Sigma^{-1/2} w_i)\}^{-1} E(w_i' \Sigma^{-1/2} x_i)]^{-1}.$$

Rather than compare the asymptotic variances, it is easier to work with the estimates of the asymptotic variances that these formulas imply, which are

$$\widehat{Var}(\hat{\beta}_{3SLS}) = [X' W (W' \Omega W)^{-1} W' X]^{-1} = (X' \Omega^{-1/2} P_{(\Omega^{1/2} W)} \Omega^{-1/2} X)^{-1}, \quad (2.2a)$$

and

$$\widehat{Var}(\hat{\beta}_{GIV}) = [X' \Omega^{-1} W (W' \Omega^{-1} W)^{-1} W' \Omega^{-1} X]^{-1} = (X' \Omega^{-1/2} P_{(\Omega^{-1/2} W)} \Omega^{-1/2} X)^{-1}, \quad (2.2b)$$

where  $P_{(\cdot)}$  denotes the projection onto the columns of  $(\cdot)$ . Efficiency comparison of the two estimators is ranking the two idempotent matrices of the same rank  $P_{(\Omega^{1/2} W)}$  and  $P_{(\Omega^{-1/2} W)}$ , which is not possible without further information on  $X$ ,  $W$ , and  $\Omega$ . The problem is equally stated as finding the optimal  $\delta$  which maximizes  $P_{(\Omega^\delta W)}$ , which does not seem to be possible in general.

An important well-known special case is when  $W = X$ , in which case  $\hat{\beta}_{GIV}$  is the GLS estimator and  $\hat{\beta}_{3SLS}$  is the OLS estimator. However, when  $W \neq X$ , general dominance of one

estimator over the other has not been established.

Bowden and Turkington (1984, p. 72) argued that there would be no clear dominance of one estimator over the other without further information on  $X$  and  $W$  when  $X$  and  $W$  are of the same order. White (1984, pp. 83-105; 1986) showed that the GIV estimator is no less efficient than the 3SLS estimator if  $\Omega^{-1/2}W$  is the optimal set of instruments in the transformed equation multiplied by  $\Omega^{-1/2}$ . White's proof is based on the fact that the GIV estimator is the 2SLS estimator on the transformed equation and the 2SLS estimator is the most efficient when the covariance matrix is scalar and the optimal instruments are used. However, as we can see in (2.2), the 3SLS estimator is also a 2SLS estimator in the transformed equation with the scalar covariance matrix. The difference is in the instruments to be used; GIV uses  $\Omega^{-1/2}W$ , whereas 3SLS uses  $\Omega^{1/2}W$ . It is not clear which instrument set is optimal without further information on  $X$ ,  $W$  and  $\Omega$ .

Before turning to algebraic equivalence results it should be noted that the efficiency issue is unambiguous if we strengthen the assumptions as in Chamberlain (1987). He shows that if (i)  $E(u_i|w_i) = 0$  and (ii)  $E(u_i u_i' | w_i) = \Sigma$ , then the most efficient estimator that ignores second moment information is the 3SLS estimator using the instruments  $\Sigma^{-1}E(x_i|w_i)$ . Unfortunately, the condition  $E(u_i|w_i) = 0$  is too strong for several of the panel data applications we have in

mind.

### 2.3. Numerical Equivalence of 3SLS and GIV

We now turn to the algebraic equivalence of 3SLS and GIV. Therefore, Assumptions 2.1-2.3 are not needed; we only need for the estimators to exist along with assumptions about  $W$  or  $\Sigma$ , to be given below.

As is well known, for nonsingular  $\Omega$ , OLS = GLS iff there exists nonsingular  $R$  such that  $\Omega^{-1}X = XR$ . Essentially the same relationship holds between the 3SLS and the GIV estimators for given instruments  $W$ . From a GMM viewpoint, the estimators are invariant to any nonsingular transformation of the orthogonality conditions as was pointed out by Schmidt, Ahn and Wyhowski (1992) (SAW hereafter).

**THEOREM 2.1:** In model (2.1), if there exists nonsingular  $B$  such that  $\Omega^{-1}W = WB$ , then  $\hat{\beta}_{3SLS} = \hat{\beta}_{GIV}$ .

**PROOF:**  $X'\Omega^{-1}W(W'\Omega^{-1}W)^{-1}W'\Omega^{-1}X = X'WB(B'W'\Omega WB)^{-1}B'W'X$   
 $= X'W(W'\Omega W)^{-1}W'X. \blacksquare$

The most widely known special case of Theorem 2.1 is the SUR model with either common regressors or diagonal  $\Sigma$ , in which case OLS = GLS. We now extend these results to show 3SLS and GIV are equivalent under analogous assumptions.

We first consider the common instrument case, that is, where the same set of instruments is used for all  $t$ . This is given by the following assumption.

ASSUMPTION 2.4:  $w_i = I_T \otimes w_i^0$ , where  $w_i^0$  is a  $1 \times q$  vector.

As we will see, Assumption 2.4 is applicable to many panel data models with strictly exogenous regressors since the regressors in each time period are orthogonal to errors in all time periods.

THEOREM 2.2: In model (2.1), when common instruments are used, that is  $w_i = I_T \otimes w_i^0$ , the 3SLS and the GIV estimators are the same.

PROOF; From Theorem 2.1, it is sufficient to show that

$$\Sigma^{-1} w_i = \Sigma^{-1} (I_T \otimes w_i^0) = \Sigma^{-1} \otimes w_i^0 = (I_T \otimes w_i^0) (\Sigma^{-1} \otimes I_T) = w_i B. \blacksquare$$

Just as with Theorem 2.1, no statistical assumptions are imposed. Since the proof is based on each observation, technically the results holds in the samples of size no smaller than  $h$ . It is worth emphasizing that the common instruments  $w_i^0$  do not have to be of the form  $(w_{i1}, \dots, w_{iT})$ . Theorem 2.2 holds no matter what  $w_i^0$  is.

Theorem 2.2 still holds after we replace  $\Sigma^\delta$  for  $\Sigma^{-1}$  for any scalar  $\delta$ . Hence, there are infinite sets of instruments that generate the same estimator. The same result has been provided by SAW (1992) when  $Fw_i$  is used as instruments, where  $F$  is the forward filtering matrix; this corresponds to

the case  $\delta = -1/2$ .

A wide class of models implies common instruments. Standard panel data models where the instruments are strictly exogenous to errors, standard simultaneous models, and the usual SUR models are examples. Thus, the GIV estimators are the same as the 3SLS estimators in these models.

We now turn to the case of diagonal  $\Sigma$ , where the instruments are essentially unrestricted.

ASSUMPTION 2.5:  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$ .

THEOREM 2.3: For any  $1 \times h_t$  vectors  $w_{it}$ , define  $w_i = \text{diag}(w_{i1}, \dots, w_{iT})$ . Then, under Assumption 2.5, 3SLS and GIV are the same.

PROOF:  $\Sigma^{-1}w_i = \text{diag}(\sigma_1^{-2}w_{i1}, \dots, \sigma_T^{-2}w_{iT}) = \text{diag}(w_{i1}\sigma_1^{-2}, \dots, w_{iT}\sigma_T^{-2}) \equiv w_i B$ , where,  $B \equiv \text{diag}(I_{g_1} \otimes \sigma_1^{-2}, I_{g_2} \otimes \sigma_2^{-2}, \dots, I_{g_T} \otimes \sigma_T^{-2})$  of dimension  $g \times g$ , and  $I_{g_t}$  is the identity matrix of dimension  $g_t$ . ■

It turns out that the theorems in this section can be used to derive some of the results for the specific panel data models we turn to next. In cases where the comparison is between 3SLS using one set of instruments and GIV using another set, direct arguments are easier.

## 2.4. Algebraic Redundancy of Instruments in 3SLS

In section 4 we use a general result on redundancy of instruments for 3SLS. The following is the algebraic equivalence analog of White (1984, Proposition 4.50).

**THEOREM 2.4:** Let  $\hat{\beta} = [X'W_1(W_1'\Omega W_1)^{-1}W_1'X]^{-1}X'W_1(W_1'\Omega W_1)^{-1}W_1'Y$  and  $\tilde{\beta} = [X'W(W'\Omega W)^{-1}W'X]^{-1}X'W(W'\Omega W)^{-1}W'Y$ , where  $W = (W_1, W_2)$ . Then  $\hat{\beta} = \tilde{\beta}$  if  $W_2'X = W_2'\Omega W_1(W_1'\Omega W_1)^{-1}W_1'X$ .

**PROOF:** Appendix 2. ■

Similarly, the two GIV estimators using instruments  $\Omega^{-1}W_1$  and  $\Omega^{-1}(W_1, W_2)$  are numerically identical if

$$W_2'\Omega^{-1}X = W_2'\Omega^{-1}W_1(W_1'\Omega^{-1}W_1)^{-1}W_1'\Omega^{-1}X.$$

## 3. **MODEL WHERE THE REGRESSORS ARE UNCORRELATED WITH THE ERRORS**

### 3.1. Unrestricted Covariance Matrix

We now consider model (2.1) under the assumption that each element of  $x_i$  is orthogonal to each element of  $u_i$ ; thus, we have in mind that

$$E(x_i \otimes u_i) = 0. \quad (3.1)$$

Under (3.1), for each  $t$  the instruments can be taken to be



all nonredundant elements of

$$x_i^0 \equiv (x_{i1}, \dots, x_{iT}). \quad (3.2)$$

Thus, we want to analyze the 3SLS estimator under the following assumption.

**ASSUMPTION 3.1:**  $E(w_i' u_i) = 0$ , where  $w_i = I_T \otimes w_i^0$ , and  $w_i^0$  contains all nonredundant elements of  $x_i^0$ .

If there are no time constant elements in  $x_{it}$  then  $w_i^0 = x_i^0$ . Typically,  $w_i^0$  will have fewer than  $Tk$  elements since  $x_{it}$  often contains at least a constant, if not other time constant variables. Any time constant variables only appear once in  $w_i^0$ .

In this subsection we place no restrictions on the variance matrix  $\Sigma$ , which puts us exactly in the situation of Theorem 2.2. Ahn and Schmidt (1991) showed that the 3SLS estimator using all of the instruments  $I_T \otimes x_i^0$  is the GLS estimator. We can restate their finding with a simpler proof.

**THEOREM 3.1:** Under Assumption 3.1, the 3SLS estimator is the GLS estimator.

**PROOF:** It follows immediately from Theorem 2.2 that 3SLS = GIV using the same set of instruments. But the GIV estimator using instruments  $I_T \otimes w_i^0$  is the GLS estimator since  $P_{(X^0 \otimes \Sigma)} X = P_{(X^0 \otimes I_T)} X = X$ . ■

### 3.2. Diagonal Covariance Matrix

If the errors  $u_{it}$  are serially uncorrelated over time, we have Assumption 2.5. Let  $x_i^* = \text{diag}(x_{i1}, \dots, x_{iT})$ .

**THEOREM 3.2:** Under Assumption 2.5, the 3SLS estimator using instruments  $x_i^*$  is the GLS estimator.

**PROOF;** 3SLS=GIV from Theorem 2.3. Let  $\tilde{x}_i^* = \Sigma^{-1/2}x_i^*$  and  $\tilde{x}_i = \Sigma^{-1/2}x_i$ . Since  $\Sigma$  is diagonal,  $\tilde{x}_i^* = \text{diag}(\sigma_1^{-1}x_{i1}, \dots, \sigma_T^{-1}x_{iT})$ . Thus,  $P_{(\tilde{X}^*)}\tilde{X} = \tilde{X}$ . The result follows immediately. ■

The definition of  $x_i^*$  leaves the instruments for each equation entirely unrestricted. In fact, because  $E(x_i^*u_i) = 0$  is sufficient for consistency of 3SLS under Assumption 2.5, the strict exogeneity condition (3.1) is not needed for consistency. Of course we are only proving algebraic equivalence results here anyway.

Recalling the conclusion of Theorem 3.1, Theorem 3.2 is seen to be a redundancy result. Theorem 3.1 showed that, without any restrictions on  $\Sigma$ , the 3SLS estimator using  $w_i = I_T \otimes w_i^0$ , where  $w_i^0$  contains all nonredundant elements of  $x_i^0$ , equals the GLS estimator. Together, Theorem 3.1 and Theorem 3.2 show that, under Assumption 2.5, 3SLS using instruments  $w_i$  is the same as 3SLS using instruments  $x_i^*$ . Without Assumption 2.5 this redundancy does not necessarily hold.

In the SUR model where  $\beta$ 's are different across the equations, the regressor itself is  $x_i^*$ . OLS = GLS if  $\Sigma$  is

diagonal. But in the panel data models with the same  $\beta$  across time periods, GLS is strictly more efficient than OLS. Of course, if  $\sigma_t^2 = \sigma^2$  for  $t = 1, \dots, T$ , OLS = GLS.

### 3.3. Random Effects Structure

In many panel data applications  $\Sigma$  is entirely unrestricted as in section 3.1. Further, it is essentially never diagonal in unobserved components models. We now turn to the popular random effects model.

To study the random effects setup we need to introduce some notation similar to that used by BMS (1989). The instruments  $I_T \otimes x_i^0$  are equivalent to the instruments  $(e_T, L) \otimes x_i^0$ , since all of the columns in  $(e_T, L)$  are linear combinations of the columns in  $I_T$  and both are of the same column rank, where  $e_T$  denotes the  $T \times 1$  vector of ones and  $L$  denotes the  $T \times (T-1)$  differencing matrix

$$L = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & -1 \end{bmatrix}.$$

The instruments  $e_T \otimes x_i^0$  and  $L \otimes x_i^0$  are in the space spanned by  $e_T$  and  $L$ , and of dimension  $T \times Tk$  and  $T \times T(T-1)k$ , respectively.

For the same reason, the instruments  $x_i^0$  are equivalent to  $(\bar{x}_i, \tilde{x}_i)$ , where  $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$  and  $\tilde{x}_i = (x_{i1} - \bar{x}_i, \dots, x_{iT-1} - \bar{x}_i)$ .

That is,  $x_i^0$  and  $(\bar{x}_i, \tilde{x}_i)$  preserve the same information. Since  $x_{iT} - \bar{x}_i$  is the negative sum of the rest of the terms  $(x_{i1} - \bar{x}_i, \dots, x_{iT-1} - \bar{x}_i)$ ,  $x_{iT} - \bar{x}_i$  is trimmed away to avoid the singularity problem without losing any information. Let  $\tilde{x}_{it} = x_{it} - \bar{x}_i$ , then  $\tilde{x}_i = (\tilde{x}_{i1}, \dots, \tilde{x}_{iT-1})$ . The dimensions of  $\bar{x}_i$  and  $\tilde{x}_i$  are  $1 \times k$  and  $1 \times (T-1)k$ .

In summary, the instruments  $I_T \otimes x_i^0$  are equivalent to;  $(L, e_T) \otimes x_i^0$  or  $I_T \otimes (\bar{x}_i, \tilde{x}_i)$  or  $(L, e_T) \otimes (\bar{x}_i, \tilde{x}_i)$ . In the sample,  $(X^0, \tilde{X}, \bar{X}) \otimes L$  and  $(X^0, \tilde{X}, \bar{X}) \otimes e_T$  stand for the stacked instruments  $L \otimes (x_i^0, \tilde{x}_i, \bar{x}_i)$  and  $e_T \otimes (x_i^0, \tilde{x}_i, \bar{x}_i)$ , respectively, where

$$X^0 = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NT} \end{bmatrix} \quad \tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1T-1} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2T-1} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \tilde{x}_{N1} & \tilde{x}_{N2} & \cdots & \tilde{x}_{NT-1} \end{bmatrix} \quad \bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \vdots \\ \bar{x}_N \end{bmatrix}.$$

In the standard random effects model each  $u_{it}$  can be written as

$$u_{it} = \phi_i + \epsilon_{it}, \quad t = 1, \dots, T,$$

where  $\phi_i$  is the time-constant unobserved effect and the  $\epsilon_{it}$  are the idiosyncratic errors. We assume that  $\phi_i$  and  $\epsilon_{it}$  have zero means, are uncorrelated for all  $t$ , and that  $\{\epsilon_{it}: t=1, \dots, T\}$  is an uncorrelated sequence with constant variance  $\sigma_\epsilon^2$ . The variance of  $\phi_i$  is  $\sigma_\phi^2$ . These assumptions lead to a well known form for  $\Sigma$ .

ASSUMPTION 3.2:  $\Sigma = \sigma_\epsilon^2 I_T + \sigma_\phi^2 e_T e_T'$ .

In applying random effects it is assumed that  $E(x_i' \phi_i) = 0$  and  $E(x_i \otimes \epsilon_i) = 0$ , so that the strict exogeneity condition (3.1) holds. Thus, the set of potential instruments is exactly as in section 3.1: the nonredundant elements of  $x_i^0$  can be used as instruments for each time period  $t$ . Section 3.2 showed how the number of instruments can be reduced when  $\Sigma$  is diagonal. The next result shows that one can get by with many fewer instruments in the random effects model. The proof is much simplified by writing  $\Sigma$  under Assumption 3.2 as  $\Sigma = P_T + bQ_T$ , where  $P_T = e_T(e_T' e_T)^{-1} e_T' = \frac{1}{T} e_T e_T'$ ,  $Q_T = L(L' L)^{-1} L' = I_T - P_T$ , and  $b$  is a positive scalar. To see how to do this, note that

$$\Sigma = \sigma_\epsilon^2 I_T + \sigma_\phi^2 e_T e_T' = \sigma_\epsilon^2 I_T + T\sigma_\phi^2 P_T = (\sigma_\epsilon^2 + T\sigma_\phi^2) P_T + \sigma_\epsilon^2 Q_T \equiv P_T + bQ_T.$$

$\sigma_\epsilon^2 + T\sigma_\phi^2$  (sum of each column in  $\Sigma$ ) is assumed to be one without loss of generality, and  $b \equiv \sigma_\epsilon^2$ . Note further that  $(P_T + bQ_T)^{-1} = P_T + \frac{1}{b}Q_T$ , that holds since two projection matrices  $P_T$  and  $Q_T$  span the two orthogonal bases  $e_T$  and  $L$ . Let  $P = I_N \otimes P_T$ ,  $Q = I_N \otimes Q_T$ .

THEOREM 3.3: Under Assumption 3.2, the 3SLS estimator using instruments  $(P_T x_i, Q_T x_i)$  is the GLS estimator.

PROOF:

$$\begin{aligned} & X' (PX, QX) \left( \begin{bmatrix} X' P \\ X' Q \end{bmatrix} (P+bQ) (PX, QX) \right)^{-1} \begin{bmatrix} X' P \\ X' Q \end{bmatrix} X \\ &= X' PX + \frac{1}{b} X' QX = X' (P+bQ)^{-1} X. \blacksquare \end{aligned}$$

Another way to say this is that 3SLS using all the instruments is the same as 3SLS using  $(P_T x_i, Q_T x_i)$ .

Since both 3SLS and GLS need a consistent estimator of  $\Sigma$  and GLS is computationally simpler, one might think 3SLS is not so useful in practice. But in fact Theorem 3.3 is quite useful. Recall that 3SLS is a GMM estimator using a restrictive weighting matrix. Under Assumption 3.2, 3SLS is asymptotically equivalent to GMM using instruments  $(P_T x_i, Q_T x_i)$  which is robust in the presence of the conditional heteroskedasticity and/or the conditional serial correlation (White, 1980, 1982; Hansen, 1982). If Assumption 2.3(a) is violated, for example, if there exists conditional heteroskedasticity or serial correlation, the inference based upon the GLS estimator is not valid. And while the robust variance estimator of the GLS estimator can be reported as

$$\widehat{\text{Var}}(\hat{\beta}_{\text{GLS}}) = (X' \Omega^{-1} X)^{-1} \left( \sum_{i=1}^N x_i' \Sigma^{-1} \hat{u}_i \hat{u}_i' \Sigma^{-1} x_i \right) (X' \Omega^{-1} X)^{-1}$$

(Wooldridge, 1992), this is not even necessarily smaller than that of the OLS estimator if A2.3 is violated. On the other hand, the GMM estimator using the instruments  $(P_T x_i, Q_T x_i)$  is the most efficient (among the estimators based on these instruments), and Theorem 3.3 tells us it is no less efficient than GLS, and it is more efficient than GLS if Assumption 3.2 fails. Further, the GMM estimator using the instruments  $(P_T x_i, Q_T x_i)$  is no less efficient than the OLS

estimator whether Assumption 2.3(a) holds or not, since the instruments  $(P_T x_i, Q_T x_i)$  are equivalent to the instruments  $(x_i, P_T x_i)$ , and the additional instruments  $P_T x_i$  are not redundant upon the instruments  $x_i$ . This GMM estimator has a lot fewer instruments than the GMM estimator based on all orthogonality conditions.

The representation  $\Sigma = P_T + bQ_T$  has many other uses as well. For example, it leads to a straightforward proof that the GLS estimator can be written as a convex combination of the between and within estimators. Let  $\hat{\beta}_b$  and  $\hat{\beta}_w$  be the between and the within estimators, then

$$\begin{aligned}\hat{\beta}_{GLS} &= [X' (P + \frac{1}{b}Q) X]^{-1} X' (P + \frac{1}{b}Q) Y \\ &= [X' (P + \frac{1}{b}Q) X]^{-1} X' P Y + \frac{1}{b} [X' (P + \frac{1}{b}Q) X]^{-1} X' Q Y \\ &= (X' P X + \frac{1}{b} X' Q X)^{-1} X' P X \hat{\beta}_b + (X' P X + \frac{1}{b} X' Q X)^{-1} \frac{1}{b} X' Q X \hat{\beta}_w.\end{aligned}$$

Theorems 3.2 and 3.3 suggest that the minimal set of instruments depends on the structure of error covariance. It appears that the minimal set of instruments depends on the block diagonality of the optimal weighting matrix  $E(w_i' \Sigma w_i)$ . In the model with a diagonal error covariance, the optimal weighting matrix is a block diagonal with the  $t$ -th block  $\sigma_t^2 E(x_i^0' x_i^0)$ , for which only  $x_{it}$  is non-redundant since  $\sigma_t^2 E(x_i^0' x_i^0)^{-1}$  meets the regressor  $x_{it}$  only.

Recall that the instruments  $I_T \otimes x_i^0$  are equivalent to  $(e_T \otimes x_i^0, L \otimes x_i^0)$ . In the random effect model, the optimal weighting matrix becomes block diagonal between the two

blocks  $e_i' e_i \otimes E(x_i^0 x_i^0)$  and  $bL' L \otimes E(x_i^0 x_i^0)$ . In the sample, the first block corresponds to the regression

$$[X' P_{(X^0 \otimes I_T)} (I_N \otimes P_T) X]^{-1} X' P_{(X^0 \otimes I_T)} (I_N \otimes P_T) Y.$$

Thus, replacing  $PX$  for  $X^0 \otimes e_i$  produces the same result. For the same reason, nothing differs whether we use  $QX$  or  $X^0 \otimes L$  for the regression in the space spanned by  $L$ .

It would be worth noting that if  $\Sigma$  is diagonal the optimal weighting matrix becomes block diagonal even when the instruments are weakly exogenous to the errors, but when  $\Sigma = P_T + bQ_T$ , the optimal weighting matrix is not block diagonal upon the weakly exogenous instruments. This would be the reason why Ahn and Schmidt (1992) get the result that all of the instruments are not redundant in the dynamic panel data model with the random effect covariance structure. In dynamic model, the instruments corresponding to the lagged dependent variables are weakly exogenous. Details are in Chapter 3.

We end section 3 with another model where the optimal weighting matrix is block diagonal.

### 3.4. A Generalization of the Random Effects Assumption

Note that  $\Sigma = P_T + bQ_T$  is sufficient but not necessary for the block diagonality of the optimal weighting matrix upon the instruments  $(e_i \otimes x_i^0, L \otimes x_i^0)$ .  $\Sigma = P_T + bQ_T \Rightarrow P_T \Sigma Q_T = 0$ ,



but not vice versa (Lemma A1.2 in Appendix 1). Even though it seems unlikely in applications for  $\Sigma$  to satisfy  $P_T \Sigma Q_T = 0$  but not be of the form  $P_T + b Q_T$ , theoretically it is worth looking into this case in greater details. The properties of  $\Sigma$  which satisfies  $P_T \Sigma Q_T = 0$  are collected in Appendix 1.

Because the nonredundant set of instruments depends in different ways on time-constant and time-varying regressors, we now explicitly separate the two. Write

$$y_{it} = x_{it}\beta + z_i\gamma + u_{it}, \quad t = 1, \dots, T, \quad (3.3)$$

where  $x_{it}$  is  $1 \times k$  and  $z_i$  is  $1 \times g$ ; note that  $z_i$  can include a constant. Note that  $u_{it}$  need not be separated into a time constant and time varying errors for stating the results of this section. For consistent estimation of  $\beta$  and  $\gamma$  by, say, GLS, in addition to (3.1) we would now need the condition

$$E(z_i \otimes u_i) = 0.$$

Interestingly, if  $\Sigma$  satisfies certain conditions, the 3SLS estimator with a reduced set of instruments is the GLS estimator. The condition on  $\Sigma$  is formally stated as

ASSUMPTION 3.3:  $P_T \Sigma Q_T = 0$ .

As we mentioned earlier, if  $\Sigma$  is of the random effects form then it satisfies Assumption 3.3, but the converse is not true.

**THEOREM 3.4:** In model (3.1) under Assumption 3.3, 3SLS using the instruments  $(e_t \otimes \bar{x}_i, e_t \otimes z_i, L \otimes \bar{x}_i)$  is GLS.

**PROOF:** Appendix 2. ■

Theorem 3.4 shows that the  $(T-1)(2k+g)$  instruments  $(L \otimes \bar{x}_i, L \otimes z_i, e_t \otimes \bar{x}_i)$  are redundant when  $P_T \Sigma Q_T = 0$ . Redundancy of  $e_t \otimes \bar{x}_i$  is rather obvious. The regression is separated into the two orthogonal spaces and the error covariance is idempotent in the space spanned by  $e_t$ . On the other hand, the intuition behind why the instruments  $L \otimes (\bar{x}_i, z_i)$  are redundant is not that obvious. Note that  $\bar{x}_i$ , the time constant component of time-varying instruments  $x_i^o$ , behaves just like the time-constant instruments  $z_i$ .

In the previous subsection, if the error covariance is of the random effects structure then the GLS estimator is a convex combination of the between and the within estimators. Similarly, when the error covariance satisfies Assumption 3.3, the GLS estimator is a convex combination of the between estimator and the GLS estimator on the differenced data. To show this, let  $A = X'(I_N \otimes \Sigma^{-1})X$ , and note that the GLS estimator on the differenced data is

$$\bar{\beta}_{GLS} = [X'(I_N \otimes L(L'\Sigma L)^{-1}L')X]^{-1}X'(I_N \otimes L(L'\Sigma L)^{-1}L')Y, \text{ and } P_T \Sigma^{-1} P_T = \frac{1}{a} P_T \text{ (Lemma A1.5 in Appendix 1). Then,}$$

$$\begin{aligned} \hat{\beta}_{GLS} &= A^{-1}X'(I_N \otimes \Sigma^{-1})Y = A^{-1}[X'(I_N \otimes P_T \Sigma^{-1} P_T)Y + X'(I_N \otimes Q_T \Sigma^{-1} Q_T)Y] \\ &= A^{-1} \frac{1}{a} X'PY + A^{-1}X'(I_N \otimes L(L'\Sigma L)^{-1}L')Y \\ &= A^{-1} \frac{1}{a} (X'PX) \hat{\beta}_b + A^{-1}[X'(I_N \otimes L(L'\Sigma L)^{-1}L')X] \bar{\beta}_{GLS}. \end{aligned}$$

#### 4. MODEL WHERE THE REGRESSORS ARE CORRELATED WITH THE TIME CONSTANT ERROR COMPONENTS

In this section we consider two models where the time-constant unobserved effect may be correlated with some or all of the regressors. In the first model all regressors are time-varying and possibly correlated with the unobserved effect. If in addition the idiosyncratic errors are assumed to be serially uncorrelated with time-constant variance, this effectively corresponds to the traditional fixed effects model. When the variance-covariance matrix of the idiosyncratic disturbances is unrestricted we get the unobserved effects analog of Kiefer's (1980) fixed effect model. In the general case we derive the equivalence of several estimators that are suggested by the structure of the model.

In sections 4.2 and 4.3 we study the HT model, where some regressors are assumed to be orthogonal to the unobserved effect. The original HT model assumed i.i.d. idiosyncratic errors. We cover this case in section 4.2 and in section 4.3 derive new redundancy results when the variance-covariance matrix of the idiosyncratic errors is unrestricted.

#### 4.1. A "Fixed Effects" Type Model

The model can be written as

$$y_{it} = x_{it}\beta + \phi_i + \epsilon_{it} = x_{it}\beta + u_{it}, \quad t = 1, \dots, T. \quad (4.1)$$

Now the orthogonality condition underlying the analysis is

$$E(x_i \otimes \epsilon_i) = 0, \quad (4.2)$$

which is a strict exogeneity condition but allows  $x_{it}$  and  $\phi_i$  to be arbitrarily correlated. This arbitrary relationship between  $x_i$  and  $\phi_i$  gives (4.1) a fixed effects flavor.

Under (4.2), only coefficients on time-varying regressors are identified; thus, for this subsection,  $x_{it}$  contains only time-varying regressors.

Under (4.2) the valid instruments for estimating  $\beta$  are given by

$$w_i = L \otimes x_i^0,$$

where recall that  $x_i^0 \equiv (x_{i1}, \dots, x_{iT})$  is a  $1 \times Tk$  row vector and  $L$  is the  $T \times (T-1)$  differencing matrix (see section 3.3). Thus,  $w_i$  is  $T \times (T-1)k$ . Not surprisingly, a reduced set of instruments is available under standard assumptions.

Under Assumption 3.2, that is  $\Sigma = \sigma_\epsilon^2 I_T + \sigma_\phi^2 e_T e_T'$ , the within (or "fixed effects") estimator is known to be efficient (provided Assumption 2.3(a) holds with  $w_i = L \otimes x_i^0$ ). Thus, it seems natural that other efficient estimators would

be the same as fixed effects.

**THEOREM 4.1:** In model (4.1) under Assumption 3.2, the 3SLS estimators using the instruments  $L\otimes x_i^0$  and  $Q_T x_i$  and the GLS estimators on the demeaned and differenced data are the within estimator.

**PROOF:** The two 3SLS estimators using the instrumental variables  $L\otimes x_i^0$  and  $Q_T x_i$  are the within estimators since

$$X' (X^0 \otimes L) [X^0' X^0 \otimes L(P_T + bQ_T)]^{-1} (X^0 \otimes L)' X = \frac{1}{b} X' Q X.$$

And the GLS estimators on the demeaned and on the differenced data are the within estimator because

$$x_i' Q_T Q_T^* Q_T x_i = x_i' Q_T x_i = x_i' L(L'L)^{-1} L' x_i,$$

where  $Q_T^*$  denotes the generalized inverse of  $Q_T$ . ■

If we allow  $E(\epsilon_i \epsilon_i')$  to have an unrestricted form, then  $\Sigma = E(u_i u_i')$  is also unrestricted and this effectively gives the setup of Kiefer (1980). We can apply Theorem 2.4 to show that some instruments used in 3SLS are redundant.

**THEOREM 4.2:** In model (4.1), the two 3SLS estimators using the instruments  $(L\otimes \bar{x}_i, L\otimes \bar{x}_i)$  and  $L\otimes \bar{x}_i$  are numerically identical. Thus,  $L\otimes \bar{x}_i$  are redundant.

**PROOF:** From Theorem 2.4, it is sufficient to show that

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

But, the RHS is  $(\bar{X}' \otimes L' \Sigma) [P_{(\bar{X})} \otimes L(L' \Sigma L)^{-1} L'] X$

$$= (\bar{X}' \otimes L' \Sigma) (I_N \otimes L(L' \Sigma L)^{-1} L') (P_{(\bar{X})} \otimes Q_T) X = (\bar{X} \otimes L)' X. \quad \text{The last}$$

equality holds from Lemma A2.1 in Appendix 2. ■

Theorem 4.2 shows that among the  $T(T-1)k$  instruments  $L\bar{\otimes}x_i^0$ , the  $(T-1)k$  instruments  $L\bar{\otimes}\bar{x}_i$  are redundant. As before, this result can be used to construct a GMM estimator that is no less efficient than 3SLS when Assumptions 2.1(a), 2.2(a), and 2.3(a) hold with  $w_i = L\bar{\otimes}x_i^0$ ; if, in particular, Assumption 2.3(a) should fail, this GMM estimator is more efficient than 3SLS, and it adds no more orthogonality conditions.

Other estimators under these assumptions also suggest themselves. One can apply GLS on the first differenced or demeaned equations. Kiefer (1980) proposed GLS using the demeaned data using  $(Q_T \Sigma Q_T)^+$ , the generalized inverse of error covariance on the demeaned data. It seems clear that no information is lost by deleting any one equation in the demeaned data, since any one equation is the negative sum of the rest of the equations. Also demeaning and differencing preserve the same information. There are several estimators that are numerically identical. Let

$\hat{\beta}_{3SLS}$ ; 3SLS estimator using the instruments  $L\bar{\otimes}\bar{x}_i$  in the original data,

$\hat{\beta}_{KF}$  ; Kiefer's estimator,

$\hat{\beta}_{DM}$  ; GLS estimator in the demeaned equation after deleting any one equation,

$\bar{\beta}_{DF}$  ; GLS estimator in the differenced equation,

$\bar{\beta}_{3SLS}$ ; 3SLS estimator in the differenced equation using all

of the instruments

$\bar{\beta}_{GIV}$  ; GIV estimator in the differenced equation using all of the instruments.

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

**PROOF:** Appendix 2. ■

#### 4.2. Hausman and Taylor Model

The HT model is the model (3.3) where  $x_i = (x_{1i}, x_{2i})$  and  $z_i = (z_{1i}, z_{2i})$ . Thus,

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + (e_i \otimes z_{1i})\gamma_1 + (e_i \otimes z_{2i})\gamma_2 + \phi_i e_i + \epsilon_i. \quad (4.3)$$

The dimensions of  $x_{1it}$ ,  $x_{2it}$ ,  $z_{1i}$  and  $z_{2i}$  are  $1 \times k_1$ ,  $1 \times k_2$ ,  $1 \times g_1$  and  $1 \times g_2$ , where  $k = k_1 + k_2$  and  $g = g_1 + g_2$ .  $\beta = (\beta_1', \beta_2')'$  and  $\gamma = (\gamma_1', \gamma_2')'$ . Assumptions 2.1 - 2.3 and Assumption 3.2 that the error covariance is that of the random effect model are assumed in the HT model. The distinctive feature of the HT model lies in the assumptions

$$E(x_{1i} \otimes u_i) = 0, \quad (4.4)$$

$$E(x_{2i} \otimes \epsilon_i) = 0, \quad (4.5)$$

$$E(z_{1i} \otimes u_i) = 0, \quad (4.6)$$

$$E(z_{2i} \otimes \epsilon_i) = 0, \quad (4.7)$$

$$E(x_{2it}' \phi_i) \text{ is the same for } t = 1, \dots, T. \quad (4.8)$$

The conditions (4.4)-(4.8) determine the instruments

available in the model. (4.4) implies the  $T^2k_1$  instruments  $I_T \otimes x_{1i}^0$ . (4.5) implies the  $T(T-1)k_2$  instruments  $L \otimes x_{2i}^0$ . (4.6) implies the  $Tg_1$  instruments  $I_T \otimes z_{1i}$ , and (4.7) implies the  $(T-1)g_2$  instruments  $L \otimes z_{2i}$ . (4.8) adds the  $(T-1)k_2$  instruments  $e_T \otimes \tilde{x}_{2i}$ . The condition (4.8) and the additional instruments  $e_T \otimes \tilde{x}_{2i}$  were proposed by BMS. Together, we have  $[T^2k_1 + (T^2-1)k_2 + Tg_1 + (T-1)g_2]$  instruments  $w_i = (I_T \otimes x_{1i}^0, L \otimes x_{2i}^0, I_T \otimes z_{1i}, L \otimes z_{2i}, e_T \otimes \tilde{x}_{2i})$ , which are equally represented by  $[L \otimes (x_{1i}^0, x_{2i}^0, z_{1i}, z_{2i}), e_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})]$ .

For the model to be identified, the number of instruments in the space spanned by  $e_T$  should not be smaller than the number of the time constant regressors, that is,  $Tk_1 + (T-1)k_2 + g_1 \geq g_1 + g_2$  should hold.

Under the random effects covariance structure in Assumption 3.2, BMS and Ahn and Schmidt (1992) showed that the minimal instruments needed for the most efficient 3SLS estimator are  $[Q_T x_i, e_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})]$ . All the instruments in the space spanned by  $e_T$  are not redundant, but only the  $k$  instruments  $Q_T x_i$  are not redundant among the  $T(T-1)k + (T-1)g$  instruments  $L \otimes (x_i^0, z_i)$  in the space spanned by  $L$ . Since  $\Sigma = P_T + bQ_T$  and all the instruments belong exclusively either to the space spanned by  $e_T$  or to the space spanned by  $L$ , the regression is separated into the two orthogonal spaces. It is entirely valid to find the minimal set of instruments in each space separately. In the space spanned by  $L$ , the error covariance  $Q_T \Sigma Q_T = bQ_T$  is a scalar idempotent, thus it is not



surprising that the instruments  $Q_T x_i$  are sufficient to reach the GLS efficiency. In the space spanned by  $e_T$ , the error covariance is idempotent but  $e_T \otimes (\bar{x}_{1i}, z_{2i})$  are correlated with  $\phi_i$ , thus it fits our intuition that all the instruments in the space spanned by  $e_T$  are not redundant.

**THEOREM 4.4:** In model (4.3) under Assumption 3.2, the 3SLS estimator using the instruments  $[Q_T x_i, e_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})]$  is a convex combination of the within estimator and the 2SLS estimator using the instruments  $e_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})$ .

**PROOF:** Let  $d_i = (x_{1i}^0, \tilde{x}_{2i}, z_{1i})$ . In the sample the instruments are  $(QX, D \otimes e_T)$ . Let  $R = (X, Z \otimes e_T)$ , the regressors. Note that  $QR = QX$ . Let

$$\begin{aligned} A &= R' (QX, D \otimes e_T) \begin{bmatrix} bX'QX & 0 \\ 0 & D'D \otimes e_T' e_T \end{bmatrix}^{-1} \begin{bmatrix} X'QX \\ (D \otimes e_T)' R \end{bmatrix} = \\ &= \frac{1}{b} X'QX + R' P_{(D \otimes e_T)} R. \quad \text{Then,} \\ \hat{\beta}_{3SLS} &= A^{-1} \left( \frac{1}{b} X'QY + R' P_{(D \otimes e_T)} Y \right) = \frac{1}{b} A^{-1} X'QX \hat{\beta}_w + A^{-1} R' P_{(D \otimes e_T)} R \hat{\beta}_{2SLS}. \quad \blacksquare \end{aligned}$$

**THEOREM 4.5:** In model (4.4) under Assumption 3.2, the 3SLS estimator using the instruments  $(Q_T x_i, e_T \otimes d_i)$  is the GIV estimator using the same set of instruments.

**PROOF:** By Theorem 2.1, it is sufficient to show that

$$(P_T + \frac{1}{b} Q_T) (Q_T x_i, e_T \otimes d_i) = (\frac{1}{b} Q_T x_i, e_T \otimes d_i) = (Q_T x_i, e_T \otimes d_i) B,$$

where  $B$  is the  $Tk_1 \times Tk_1$  diagonal matrix with  $\frac{1}{b}$ 's in the first  $k_1$  diagonal and 1's in the last  $(T-1)k_1$  diagonal.  $\blacksquare$

Thus, GIV does not reduce the number of instruments for the

original HT model.

#### 4.3. HT Model with Serially Correlated Time-Varying Errors

We now study the HT model without Assumption 3.2; in fact, the error covariance  $\Sigma$  is entirely unrestricted. We get the following redundancy result for 3SLS.

**THEOREM 4.6:** In model (4.3), the two 3SLS estimators using the instruments  $[I_T \otimes (x_{1i}^0, \bar{x}_{2i}, z_{1i}), L \otimes (\bar{x}_{2i}, z_{2i})]$  and  $I_T \otimes (x_{1i}^0, \bar{x}_{2i}, z_{1i})$  are the same. Thus  $L \otimes (\bar{x}_{2i}, z_{2i})$  are redundant.

**PROOF:** Appendix 2. ■

The instruments  $L \otimes (\bar{x}_{2i}, z_{2i})$  are redundant whereas  $L \otimes (\bar{x}_{1i}, z_{1i})$  are not. The difference is that the instruments  $e_T \otimes (\bar{x}_{1i}, z_{1i})$  are included in the instrument set while  $e_T \otimes (\bar{x}_{2i}, z_{2i})$  are not. It is intuitively reasonable that the redundancy of the two sets of the instrumental variables  $\bar{x}_i$  and  $z_i$ , namely the instruments from the time constant component of time varying regressors and the instruments from the time constant regressors, show the same pattern. A similar result was provided in Theorem 3.5, but there the covariance structure satisfying  $P_T \Sigma Q_T = 0$  is critical.

The BMS condition (4.8) is crucial for the GIV estimators to consistent. Without (4.8),  $L \otimes (x_{2i}^0, z_{2i})$  are not valid for GIV since

$$E[(L\otimes x_{2i}^0)' \Sigma^{-1} u_i] = E[(L\otimes x_{2i}^0)' \Sigma^{-1} \phi_i] = L' \Sigma^{-1} e_T \otimes E(x_{2i}^0 \phi_i),$$

which is not zero unless  $P_T \Sigma Q_T = 0$ . The same proof shows that  $L\otimes z_{2i}$  are not valid instruments.

The instruments  $L\otimes(\bar{x}_{2i}, z_{2i})$  which are redundant for 3SLS are not valid for GIV even with the BMS condition (4.8), since

$$E[(L\otimes z_{2i})' \Sigma^{-1} u_i] = E[(L\otimes z_{2i})' \Sigma^{-1} \phi_i e_T] = L' \Sigma^{-1} e_T \otimes E(z_{2i} \phi_i),$$

which is not zero unless  $P_T \Sigma Q_T = 0$ . The same procedure shows invalidity of the instruments  $L\otimes \bar{x}_{2i}$ . Thus the instruments available for GIV are  $I_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})$ , which are no more than the non-redundant instruments for the 3SLS estimator. Here,  $(x_{1i}^0, \tilde{x}_{2i}, z_{1i})$  are the common instruments.

**THEOREM 4.7:** In the model (4.3), the 3SLS estimator using the instruments  $I_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})$  is the GIV estimator using the instruments  $[Q_T x_i, e_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})]$ .

**PROOF:** Appendix 2. ■

Thus, GIV reduces the number of instruments in the space spanned by  $L$  from  $L\otimes(\tilde{x}_i, z_{1i})$  to  $Q_T x_i$ , say,  $(T-1)^2 k + (T-1)g_1$  to  $k$ , but does not reduce the number of instruments in the space spanned by  $e_T$ .

The relationship between the two GIV estimators using the instruments  $I_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})$  and  $[Q_T x_i, e_T \otimes (x_{1i}^0, \tilde{x}_{2i}, z_{1i})]$  is similar to the relationship between the two 3SLS estimators

using the same set of instruments in the HT model of the previous subsection, where error covariance is of random effects form.

If there are no  $x_{1i}$ ,  $z_i$  in the model we are dealing with, the model is the fixed-effects type model of section 4.1 with unrestricted covariance matrix but with the BMS assumption. Then, the instruments  $(e_i \otimes \tilde{x}_{2i})$  are not redundant, and there are more efficient estimators than the GLS estimator in the differenced equation or Kiefer's estimator.

## 5. CONCLUSION

For the two estimators, GIV and 3SLS, it has been shown that there is no clear dominance of one over the other in general but the two are the same numerically in many popular panel data models. We have also derived new redundancy results for 3SLS estimator for a variety of panel data models with strictly exogenous regressors.

The result that block diagonality of the optimal weighting matrix plays the central role in determining the nonredundant set of instrumental variables provides a sound intuition behind the findings in Amemiya and MaCurdy (1986), BMS, Ahn and Schmidt (1992), and most of the findings in this paper. The idea also is applied to other models. For example, construction of the GMM estimator which is robust

to conditional heteroskedasticity and no less efficient than the GLS estimator is straightforward in the usual group data models.

When the optimal weighting matrix is not block diagonal, a large number of instruments is not redundant for 3SLS. Even if the GIV estimators are not robust, GIV reduces the number of relevant instruments under an unrestricted error covariance structure. GIV would therefore be useful in practice.

As was shown in Schmidt (1990), in general GIV produces inconsistent estimators when the instruments are only weakly exogenous to the errors. This raises many interesting questions both for the consistency of the estimators and for the consistency of the standard errors. We study weakly exogenous panel data models in Chapter 3.

Throughout this paper, we only consider orthogonality conditions between the regressors and the errors. An interesting question is whether the orthogonality conditions from the error covariance matrix in each of the models are useful or not. Chapter 4 gives the conditions when the information from the error covariance matrix is useful in several popular models.

## APPENDIX 1

**LEMMA A1.1:**  $P_T \Sigma Q_T = 0$  iff the element sums of each row (column) are the same for all rows (columns) of  $\Sigma$ .

**PROOF:**  $P_T \Sigma Q_T = 0 \Leftrightarrow \Sigma = P_T \Sigma P_T + Q_T \Sigma Q_T$ , since  $\Sigma = (P_T + Q_T) \Sigma (P_T + Q_T)$ . The element sum of each row (column) in  $Q_T \Sigma Q_T$  is 0 and all of the elements in  $P_T \Sigma P_T$  are the same. Hence, necessity follows. If the element sums of each column are the same, all of the elements in  $P_T \Sigma$  are the same. Thus  $P_T \Sigma = P_T \Sigma P_T$ , which implies  $P_T \Sigma Q_T = 0$ . Hence, sufficiency follows. ■

**LEMMA A1.2:** For non-negative scalar  $a$  and  $b$ ,  $\Sigma = aP_T + bQ_T \Rightarrow P_T \Sigma Q_T = 0$ , but not vice versa.

**PROOF:** Direct substitution proves that  $\Sigma = aP_T + bQ_T \Rightarrow P_T \Sigma Q_T = 0$ . A counter-example which satisfies  $P_T \Sigma Q_T = 0$  but not of the form  $\Sigma = aP_T + bQ_T$  is sufficient for the proof. Suppose

$$\Sigma = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -1 & 5 \end{bmatrix}.$$

$\Sigma$  is symmetric positive definite and the sum of each row is 4, but  $\Sigma$  is not of the form  $aP_T + bQ_T$ , because not all of its diagonal elements are equal and not all of its off diagonal elements are equal. ■

LEMMA A1.3:  $\Sigma = aP_T + bQ_T \Leftrightarrow P_T \Sigma Q_T = 0$  when  $T = 2$ .

PROOF: Equality of the off diagonal terms is guaranteed from the symmetry of  $\Sigma$  and the equality of diagonal terms is enforced from Lemma A1.1. Thus  $\Sigma$  is of the form  $P_T + bQ_T$ . ■

LEMMA A1.4: If  $P_T \Sigma Q_T = 0$  and either all of the off diagonals in  $\Sigma$  are equal or all the diagonal terms in  $\Sigma$  are the same, then  $\Sigma = aP_T + bQ_T$ .

PROOF: If the off diagonal terms are equal, all of the diagonal terms should be the same each other for  $P_T \Sigma Q_T = 0$  to hold from Lemma A1.1. Thus the two statements  $\Sigma = aP_T + bQ_T$  and  $P_T \Sigma Q_T = 0$  are equivalent when the off diagonal terms of  $\Sigma$  are the same. To prove the statement that the equal diagonal elements of  $\Sigma$  and  $P_T \Sigma Q_T = 0$  implies  $\Sigma = aP_T + bQ_T$ , mathematical induction is used. When  $T = 3$ , imposing the equality of the diagonal terms and from the symmetry of  $\Sigma$ ,  $\Sigma$  is expressed as

$$\Sigma = \begin{bmatrix} \sigma & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma \end{bmatrix}.$$

From Lemma A1.1,  $\sigma_{12} + \sigma_{13} = \sigma_{12} + \sigma_{23} = \sigma_{13} + \sigma_{23}$ , hence,  $\sigma_{12} = \sigma_{13} = \sigma_{23}$ . Suppose the off diagonals are the same. If the diagonal terms are the same for  $T = t > 3$ , then for  $T = t+1$ , the  $(t+1)$ -th off diagonal terms are forced to be the same (Lemma A1.1). ■

**LEMMA A1.5:** If  $P_T \Sigma Q_T = 0$ , then  $P_T \Sigma P_T = a P_T$ ,  $P_T \Sigma^{-1} P_T = \frac{1}{a} P_T$ , and  $e_T (e_T' \Sigma e_T)^{-1} e_T' = P_T \Sigma^{-1} P_T$ , where  $a$  is the sum of each column of  $\Sigma$ .

**PROOF:**  $P_T \Sigma P_T = e_T (e_T' e_T)^{-1} e_T' \Sigma e_T (e_T' e_T)^{-1} e_T' = \frac{s}{T} P_T$ , where  $s = e_T' \Sigma e_T$  is the sum of all elements of  $\Sigma$ , thus  $\frac{s}{T} = a$ . Hence, the first result follows.  $(P_T \Sigma^{-1} P_T) (P_T \Sigma P_T) = P_T$ . Thus,  $(P_T \Sigma^{-1} P_T) a P_T = P_T$  and  $P_T \Sigma^{-1} P_T = \frac{1}{a} P_T$ . The third result follows trivially. ■

**LEMMA A1.6:** If  $P_T \Sigma Q_T = 0$ , then  $Q_T \Sigma^{-1} Q_T = L(L' \Sigma L)^{-1} L'$ .

**PROOF:** It is sufficient to show that  $Q_T \Sigma^{-1} Q_T \Sigma = L(L' \Sigma L)^{-1} L' \Sigma$ . Note that  $Q_T \Sigma^{-1} Q_T \Sigma = Q_T \Sigma^{-1} \Sigma Q_T = Q_T$ , since  $P_T \Sigma Q_T = 0 \Rightarrow Q_T \Sigma = Q_T \Sigma Q_T = \Sigma Q_T$ . Thus, what we need to show is that  $L(L' \Sigma L)^{-1} L' \Sigma = Q_T$ . But,  $L(L' \Sigma L)^{-1} L' \Sigma (P_T + Q_T) = L(L' \Sigma L)^{-1} L' \Sigma Q_T$ , since  $P_T \Sigma Q_T = 0 \Leftrightarrow L' \Sigma e_T = 0$  and  $L(L' \Sigma L)^{-1} L' \Sigma Q_T = L(L' \Sigma L)^{-1} L' \Sigma L (L' L)^{-1} L' = Q_T$ . ■

**LEMMA A1.7:**  $P_T \Sigma Q_T = 0 \Leftrightarrow P_T \Sigma^{-1} Q_T = 0$ .

**PROOF:**  $P_T \Sigma Q_T = 0 \Leftrightarrow \Sigma = P_T \Sigma P_T + Q_T \Sigma Q_T \Leftrightarrow P_T \Sigma = P_T \Sigma P_T$  and  $Q_T \Sigma = Q_T \Sigma Q_T$ . Post-multiplying by  $\Sigma^{-1}$ ,  $P_T = P_T \Sigma P_T \Sigma^{-1} = P_T \Sigma P_T \Sigma^{-1} P_T = \Sigma P_T \Sigma^{-1} P_T$  and pre-multiplying by  $\Sigma^{-1}$ , we have  $\Sigma^{-1} P_T = P_T \Sigma^{-1} P_T$ , which is the condition we are looking for. Exactly the same procedure shows that  $\Sigma^{-1} Q_T = Q_T \Sigma^{-1} Q_T$ . Given sufficiency, necessity is obvious. ■



In fact,  $\Sigma^n = (P_T \Sigma P_T + Q_T \Sigma Q_T)^n = (P_T \Sigma^n P_T + Q_T \Sigma^n Q_T)$  for any integer  $n$ , which implies  $\Sigma^{-1} = (P_T \Sigma^{-1} P_T + Q_T \Sigma^{-1} Q_T)$  if  $P_T \Sigma Q_T = 0$ .

## APPENDIX 2

LEMMA A2.1:  $(P_{(\tilde{X})} \otimes Q_T)X = (P_{(\tilde{X})} \otimes Q_T)QX = QX$ .

PROOF: Note that when  $k = 1$ ,  $X = \text{vec}(X^{o'})$  and  $\tilde{X} = X^{o'}Q_T$ .

Hence,  $(P_{(\tilde{X})} \otimes Q_T)X = (P_{(\tilde{X})} \otimes Q_T)\text{vec}(X^{o'}) = \text{vec}(Q_T X^{o'} P_{(\tilde{X})}) = \text{vec}(Q_T X^{o'}) = QX$ , which is valid when  $k > 1$  by applying this argument to each of the regressors separately. In fact,  $\tilde{X}$  are the first  $(T-1)k$  columns of  $X^{o'}Q_T$ , but  $P_{(\tilde{X})}X^{o'}Q_T = X^{o'}Q_T$ , simply because the projection of  $Q_T$  after deleting any one column of  $Q_T$  is still  $Q_T$ . ■

PROOF of Theorem 2.4:  $\hat{\beta} = \tilde{\beta}$  if  $X'W_1(W_1'\Omega W_1)^{-1}W_1'X$   
 $= X'W(W'\Omega W)^{-1}W'X$ . But,

$$\begin{aligned} X'W(W'\Omega W)^{-1}W'X &= (X'W_2 \ X'W_1) \begin{bmatrix} W_2'\Omega W_2 & W_2'\Omega W_1 \\ W_1'\Omega W_2 & W_1'\Omega W_1 \end{bmatrix}^{-1} \begin{bmatrix} W_2'X \\ W_1'X \end{bmatrix} \\ &= X'W_2 D^{-1} W_2'X - X'W_2 D^{-1} W_2'\Omega W_1 (W_1'\Omega W_1)^{-1} W_1'X - X'W_1 (W_1'\Omega W_1)^{-1} W_1'\Omega W_2 D^{-1} W_2'X \\ &\quad + X'W_1 (W_1'\Omega W_1)^{-1} W_1'X + X'W_1 (W_1'\Omega W_1)^{-1} W_1'\Omega W_2 D^{-1} W_2'\Omega W_1 (W_1'\Omega W_1)^{-1} W_1'X, \end{aligned}$$

from the partitioned inverse lemma. Thus the condition is

$$X'W_2 D^{-1} W_2'X - X'W_2 D^{-1} W_2'\Omega W_1 (W_1'\Omega W_1)^{-1} W_1'X - X'W_1 (W_1'\Omega W_1)^{-1} W_1'\Omega W_2 D^{-1} W_2'X$$

$$+ X'W_1 (W_1'\Omega W_1)^{-1} W_1'\Omega W_2 D^{-1} W_2'\Omega W_1 (W_1'\Omega W_1)^{-1} W_1'X = 0, \text{ which is } A'D^{-1}A,$$

where  $A = W_2'X - W_2'\Omega W_1 (W_1'\Omega W_1)^{-1} W_1'X$  and  $D = W_2'\Omega W_2 -$   
 $W_2'\Omega W_1 (W_1'\Omega W_1)^{-1} W_1'\Omega W_2$ , a nonsingular positive definite. Thus  
 $A'D^{-1}A = 0$ , iff  $A = 0$ . ■

PROOF of Theorem 3.4: Let  $H=(\bar{X},Z)$ . It is sufficient to show that

$$\begin{aligned} & (\bar{X}\otimes L, H\otimes e_T) \left[ \begin{array}{c} \bar{X}'\bar{X}\otimes L'\Sigma L \\ H'H\otimes e_T'\Sigma e_T \end{array} \right]^{-1} \left[ \begin{array}{c} (\bar{X}\otimes L)' \\ (H\otimes e_T)' \end{array} \right] (X, Z\otimes e_T) \\ &= (I_N\otimes\Sigma^{-1})(X, Z\otimes e_T). \end{aligned}$$

The LHS is

$$\begin{aligned} & [P_{(\bar{X})}\otimes L(L'\Sigma L)^{-1}L'](X, Z\otimes e_T) + [P_{(H)}\otimes e_T(e_T'\Sigma e_T)^{-1}e_T'](X, Z\otimes e_T) \\ &= [I_N\otimes L(L'\Sigma L)^{-1}L']X + [I_N\otimes e_T(e_T'\Sigma e_T)^{-1}e_T'](X, Z\otimes e_T) = (I_N\otimes Q_T\Sigma^{-1}Q_T)X \\ &+ (I_N\otimes P_T\Sigma^{-1}P_T)(X, Z\otimes e_T) = (I_N\otimes\Sigma^{-1})(X, Z\otimes e_T). \end{aligned}$$

The first equality follows from Lemma A2.1 and the second equality follows from Lemmas A1.5 and A1.6. ■

PROOF of Theorem 4.3:  $\bar{\beta}_{GLS} = \bar{\beta}_{3SLS} = \bar{\beta}_{GIV}$  in the differenced equation from Theorem 2.1 and Theorem 3.1.  $\bar{\beta}_{GLS} = \hat{\beta}_{3SLS}$  since

$$\begin{aligned} & X'(\bar{X}\otimes L)(\bar{X}'\bar{X}\otimes L'\Sigma L)^{-1}(\bar{X}\otimes L)'X = X'[P_{(\bar{X})}\otimes L(L'\Sigma L)^{-1}L']X \\ &= X'[I_N\otimes L(L'\Sigma L)^{-1}L'] [P_{(\bar{X})}\otimes Q_T]X = X'(I_N\otimes L)(I_N\otimes L'\Sigma L)^{-1}(I_N\otimes L)'X. \end{aligned}$$

The last equality follows from Lemma A2.1. To show  $\hat{\beta}_{KF} = \hat{\beta}_{DF}$ , it is sufficient to show that  $L(L'\Sigma L)^{-1}L' = (Q_T\Sigma Q_T)^+$ .

But,  $Q_T\Sigma Q_T L(L'\Sigma L)^{-1}L'Q_T\Sigma Q_T = Q_T\Sigma Q_T$ ,  $L(L'\Sigma L)^{-1}L'Q_T\Sigma Q_T L(L'\Sigma L)^{-1}L' = L(L'\Sigma L)^{-1}L'$  and  $Q_T\Sigma Q_T L(L'\Sigma L)^{-1}L' = L(L'\Sigma L)^{-1}L'Q_T\Sigma Q_T = Q_T$ . Thus  $L(L'\Sigma L)^{-1}L'$  is the unique generalized inverse of  $Q_T\Sigma Q_T$

(Theil, 1971, pp 269). To prove  $\hat{\beta}_{KF} = \hat{\beta}_{DM}$ . Let

$$Q_T x_i = \begin{bmatrix} \tilde{x}_i \\ x_i^\ell \end{bmatrix}, \text{ and } (Q_T\Sigma Q_T)^+ = \begin{bmatrix} s^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

where  $\tilde{x}_i$  denotes the first  $T-1$  rows and  $x_i^\ell$  denotes the last row of  $Q_T x_i$ . Hence, we are looking at the case when the

last row of the demeaned data is deleted. Then,

$$x_i' Q_T (Q_T \Sigma Q_T)^{-1} Q_T x_i = [\tilde{x}_i', x_i^{\ell'}] \begin{bmatrix} s^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_i \\ x_i^{\ell} \end{bmatrix} = \tilde{x}_i' s^{-1} \tilde{x}_i.$$

Deleting any one other row instead of the last row of the demeaned data makes no difference. ■

PROOF of Theorem 4.6: Let  $R = (X, Z \otimes e_T)$ ,  $H = (\bar{X}_2, Z_2)$  and  $G = (X_1^0, \bar{X}_2, Z_1)$ . Note  $(H \otimes L)' R = (H \otimes L)' X$ . From Theorem 2.4 it is sufficient to show that

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

The LHS is  $(H \otimes L)' (P_{(G)} \otimes I_T)' R = (H \otimes I_T)' (P_{(G)} \otimes L)' X = (H \otimes L)' X$  (Lemma A2.1). ■

PROOF of Theorem 4.7: Let  $R = (X, Z \otimes e_T)$  and  $G = (X_1^0, \bar{X}_2, Z_1)$ .

For the 3SLS estimator using the instruments  $(G \otimes I_T)$ ,

$$\begin{aligned} & (G \otimes I_T) [(G \otimes I_T)' (I_N \otimes \Sigma) (G \otimes I_T)]^{-1} (G \otimes I_T)' (X, Z \otimes e_T) \\ &= (P_{(G)} \otimes \Sigma^{-1}) (X, Z \otimes e_T). \end{aligned}$$

And, for the GIV estimator using the instruments  $(QX, G \otimes e_T)$

$$\begin{aligned} & (I_N \otimes \Sigma^{-1}) (QX, G \otimes e_T) \left( \begin{bmatrix} X' Q \\ (G \otimes e_T)' \end{bmatrix} (I_N \otimes \Sigma^{-1}) (QX, G \otimes e_T) \right)^{-1} \begin{bmatrix} X' Q \\ (G \otimes e_T)' \end{bmatrix} (I_N \otimes \Sigma^{-1}) (X, Z \otimes e_T) \\ &= [ (I_N \otimes \Sigma^{-1} Q_T) X D^{-1} X' (I_N \otimes Q_T \Sigma^{-1}) \\ &\quad - (I_N \otimes \Sigma^{-1} Q_T) X D^{-1} X' (P_{(G)} \otimes Q_T \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1}) \\ &\quad - (P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} Q_T) X D^{-1} X' (I_N \otimes Q_T \Sigma^{-1}) \\ &\quad + P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} \\ &\quad + (P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} Q_T) X D^{-1} X' (P_{(G)} \otimes Q_T \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1}) ] \\ &\quad \cdot (X, Z \otimes e_T), \end{aligned}$$

where  $D = X' (I_N \otimes Q_T \Sigma^{-1} Q_T) X - X' [P_{(G)} \otimes Q_T \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} Q_T] X$ .

Let  $A = [X' (I_N \otimes Q_T \Sigma^{-1}) - X' (P_{(G)} \otimes Q_T \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1})] (X, Z \otimes e_T)$ .

$$A = [X' (I_N \otimes Q_T \Sigma^{-1} Q_T) - X' (P_{(G)} \otimes Q_T \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} Q_T)] (X, Z \otimes e_T) \\ + [X' (I_N \otimes Q_T \Sigma^{-1} P_T) - X' (P_{(G)} \otimes Q_T \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} P_T)] (X, Z \otimes e_T) = D.$$

Thus, the 1st and the 2nd terms add up to  $(I_N \otimes \Sigma^{-1} Q_T) X$  and the 3rd and the 5th terms add up to  $-(P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} Q_T) X$ .

Together, we have  $(I_N \otimes \Sigma^{-1} Q_T) X - (P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} Q_T) X +$

$[P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1}] (X, Z \otimes e_T)$ . For the regressor  $Z \otimes e_T$ ,

$[P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1}] (Z \otimes e_T) = (P_{(G)} \otimes \Sigma^{-1}) (Z \otimes e_T)$ , and for  $X$ ,

$$[I_N \otimes \Sigma^{-1} Q_T - P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1} Q_T + P_{(G)} \otimes \Sigma^{-1} e_T (e_T' \Sigma^{-1} e_T)^{-1} e_T' \Sigma^{-1}] X \\ = (I_N \otimes \Sigma^{-1} Q_T) X + (P_{(G)} \otimes \Sigma^{-1} P_T) X = (P_{(G)} \otimes \Sigma^{-1}) X. \blacksquare$$

## CHAPTER 3

### ESTIMATION USING PANEL DATA UNDER WEAK EXOGENEITY

#### 1. INTRODUCTION

In this chapter we study linear panel data models where the regressors are only weakly exogenous. The primary concern is with the consistency of estimators when the errors are serially correlated, and with the consistency of the usual standard errors of 3SLS (appropriately defined, see Chapter 2) estimators when certain instruments are used. We also discuss how to construct some reduced lists of instrumental variables that would lead nearly efficient estimators.

A leading example of the weakly exogenous model is the dynamic model with lagged dependent variables. Anderson and Hsiao (1981), Bargava and Sargan (1983), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bover (1990), Arellano and Bond (1991), Ahn (1990), and Ahn and Schmidt have studied efficient estimation in dynamic models. The rational expectations model is another important example. Using panel data to test the rational expectations hypothesis has lead to renewed interest in studying weak exogeneity in panel data models (Zeldes, 1989; Kean and Runkle, 1990; Runkle, 1991). Generally, there is a growing realization

both in time series and panel data contexts that many regressors in general models would be only weakly exogenous to the errors.

However, only a few study exist that deal with the general weak exogeneity (Keane and Runkle (1992) and comments). An important feature of weak exogeneity is that different instruments are available for each period so that GLS transformation, in general, will bring the inconsistency of the resulting estimators (Schmidt, 1990). An important exceptional case is when  $\Sigma$  is diagonal (Chapter 2, Theorem 2.3). When  $\Sigma$  is diagonal, the redundancy result of Theorem 3.2 of Chapter 2 also applies to weakly exogenous case. Consequently, a general result that GMM using all the moment conditions is the best specially has a force in weakly exogeneous case with non-diagonal covariance matrix (Ahn and Schmidt, 1992; comments on Keane and Runkle, 1992).

Schmidt, Ahn, and Wyhowski (1992) (SAW henceforth) provide the lists of instrumental variables for each of several weakly exogenous models. The structure of the weakly exogenous instrumental variables provided by SAW is quite similar to the structure of the instruments for lagged dependent variables in dynamic models. It now is well-known that no moment conditions between the lagged dependent variables and the disturbances exist unless the covariance matrix is somehow restricted. We ask a basic question about whether the a priori population moment conditions that are

currently suggested in weakly exogenous models are valid when the idiosyncratic errors are serially correlated. Generally, it is likely that there is a link between covariance restrictions and orthogonality conditions as is in dynamic models. We investigate these in the next section.

If the variances of the disturbances conditioned on the instrumental variables is the same as unconditional variances, the usual 3SLS standard errors are consistent in general. Wooldridge (1993) showed that the usual standard errors from nonlinear 3SLS in hedonic pricing models are not consistent even when the conditional variance of the errors are constants. A similar result was obtained by Ahn (1990) in dynamic panel data models. SAW suggested that the equi-correlation assumption of Breusch, Mizon, and Schmidt (1989) (BMS hereafter) can hold in weakly exogenous models. We show, in section 3, that the usual 3SLS standard errors are not consistent if the instrumental variables from the BMS assumption are used under any plausible assumptions for weakly exogenous models. Also, we link this to the result obtained by Ahn (1990) in dynamic models.

Keane and Runkle (1992) proposed 2SLS after forward filtering the equations in weakly exogenous panel data models when the errors are serially correlated, adapting a suggestion by Hayashi and Sims (1983) in a pure time series context. But, it was shown by SAW that the Keane and Runkle



estimator is numerically identical to 3SLS when all of the instruments are used, and thereby forward filtering is an unnecessary complication. However, it still remains an interesting question that how the forward filtering can reduce the list of instruments, since, in many instances, using all of the instruments is not even feasible. Keane and Runkle provided evidence through an example that forward filtering can bring non-trivial efficiency gains, though this does not generalize to other cases. The arguments of Chapter 2 comparing 3SLS and GIV apply to this case. The main idea of Keane and Runkle is that forward filtering whitens the errors and applying instruments (without transformation) to the forward filtered equations would result in better estimators. However, 3SLS also is 2SLS on the forward filtered equations using the instruments  $\Omega^{1/2}W$ . Therefore, what Keane and Runkle suggest is using  $W$  instead of  $\Omega^{1/2}W$  on the forward filtered equations. To compare these two, we need to compare  $P_{(W)}$  and  $P_{(\Omega^{1/2}W)}$ , where  $P_{(\cdot)}$  denotes the projection onto the columns of  $(\cdot)$ . This comparison, in general, is not entirely clear. See Chapter 2 for more details.

3SLS and the Keane and Runkle estimators, in the presence of heteroskedasticity, are less efficient than GMM that uses the same instruments. The comparison between the two GMM estimators using instruments  $W$  and  $\Omega^{1/2}W$  is even more ambiguous.

However, in finite samples, GMM estimators based on a huge number of instrumental variables might not have desirable properties. For example, in finite samples, the standard errors can grow by adding instruments, and this would happen more likely when the number of instruments becomes closer to the number of observations. Thus, it is practically useful to find some reduced lists of instruments that generate estimators with desirable properties. In section 4, we suggest some weighted sums of the given long lists of instrumental variables, which do not generate fully efficient estimators but would lead to nearly efficient estimators. However, we should note that the usefulness of applying these reduced lists of instrumental variables remains to be seen. Section 5 concludes.

## **2. SERIAL CORRELATION AND CONSISTENCY OF ESTIMATORS**

This section shows that the moment conditions in weakly exogenous models are restricted in general by structures of covariance matrices. Before doing this, we review previous results on the moment conditions for weakly exogenous regressors, in particular those provided by SAW.

### 2.1. Moment Conditions under Weak Exogeneity

The model we consider is

$$y_{it} = x_{it}\beta + u_{it}, \quad t = 1, \dots, T. \quad (2.1)$$

Let  $y_i = (y_{i1}, \dots, y_{iT})'$ ,  $x_i = (x_{i1}', \dots, x_{iT}')'$  and  $u_i = (u_{i1}, \dots, u_{iT})'$  of dimensions  $T \times 1$ ,  $T \times k$  and  $T \times 1$ , respectively. (2.1) is equally expressed as  $y_i = x_i\beta + u_i$ .  $\{(y_i, x_i) : i=1, \dots, N\}$  is an i.i.d. sequence and the fourth moments of  $(y_i, x_i)$  exist. Let  $\Sigma = E(u_i u_i')$ .

If there are no unobserved individual effects that are correlated with the regressors, weak exogeneity is defined by an assumption

**ASSUMPTION 2.1:**  $E(x_{it}' u_{is}) = 0, \quad 1 \leq t \leq s \leq T.$

This implies  $\frac{1}{2}T(T+1)k$  instruments  $\text{diag}(x_{i1}^0, x_{i2}^0, \dots, x_{iT}^0)$ , where  $x_{it}^0 = (x_{i1}, \dots, x_{it})$ ,  $t = 1, \dots, T$ . Note that we did not give any restrictions on  $\Sigma$ . Also, Assumption 2.1 allows for unobserved effects that are uncorrelated with the regressors.

By introducing unobserved fixed effects that are correlated with the regressors, the errors are composites of the time-constant and the time-idiosyncratic components, so  $u_{it} = \phi_i + \epsilon_{it}$ ,  $t = 1, \dots, T$ . The assumption that corresponds to the weak exogeneity and the presence of explicit unobserved effects is

ASSUMPTION 2.2:  $E(x_{it}'\epsilon_{is}) = 0, 1 \leq t \leq s \leq T$ .

Under Assumption 2.2,  $\phi_i$  is allowed to be arbitrarily correlated with the regressors.

Under Assumption 2.2, the coefficients on the time-constant variables are not identified. Thus we simply assume that all the regressors are time-varying. It is usual, under Assumptions 2.2, to estimate the parameters after differencing. In the differenced equations, the errors are  $\Delta u_{it} \equiv u_{it} - u_{it+1} = \epsilon_{it} - \epsilon_{it+1}$ ,  $t = 1, \dots, T-1$ , and the orthogonality conditions are

$$E(x_{it}'\Delta u_{is}) = 0, t \leq s = 1, \dots, T-1. \quad (2.2)$$

Thus, we have the instrumental variables  $\text{diag}(x_{i1}^0, \dots, x_{iT-1}^0)$  of dimension  $T \times \frac{1}{2}T(T-1)k$  in the differenced equations.

Let  $w_i = \text{diag}(x_{i1}^0, x_{i2}^0, \dots, x_{iT-1}^0)$ . Applying  $w_i$  to the differenced equations amounts to applying the instrumental variables  $Lw_i$  to the original equations before differencing (SAW), where  $L$  is the  $T \times (T-1)$  differencing matrix (for the definition of  $L$ , see SAW or Chapter 2). While applying instruments  $Lw_i$  to the original equations and applying  $w_i$  to the differenced equations yield the numerically identical estimators, using  $Lw_i$  in the original equations has several important advantages, both for the identification of coefficients on the time-constant variables if they exist and for the efficiency of all of the estimators, whenever we have some instrumental variables that are not in the space

spanned by  $L$ .

## 2.2. Serial Correlation and Moment Conditions

It is usually assumed that  $\{\epsilon_{it}: t=1, \dots, T\}$  is an uncorrelated sequence with a constant variance and is not correlated with  $\{\phi_i\}$ . In this case  $\Sigma$  takes the random effects covariance structure

$$\Sigma = \sigma_\epsilon^2 I_T + \sigma_\phi^2 e_T e_T', \quad (2.3)$$

where  $I_T$  is the  $T \times T$  identity matrix and  $e_T$  is the  $T \times 1$  vector of ones. But, there is no reason to think (2.3) holds universally. In GMM, the restriction (2.3) makes no difference unless we utilize the moment conditions implied by (2.3), namely, the moment conditions from covariance restrictions.  $\Sigma$  is estimable in the models we deal with and imposing restrictions like (2.3) without testing can be too limiting.

In dynamic models, the set of the instrumental variables corresponding to the lagged dependent variables relies heavily on whether the idiosyncratic errors are serially correlated (Ahn and Schmidt, 1992; Arellano and Bond, 1991). Nevertheless, serial correlation of the idiosyncratic errors in weakly exogenous models have often been presumed to have nothing to do with the set of the instrumental variables (Runkle (1991) is an exception; he

expressed the concern that the usual instrumental variables might not be valid if the time-varying errors are serially correlated).

We will consider whether instruments  $Lw_i$  (or Assumption 2.2 that generally is based on apriority) are valid when the idiosyncratic errors are serially correlated in the model with fixed effects. Although we will study the case when the regressors are correlated with the unobserved individual effects, the results equally apply to the model with no fixed effects.

To this end, though it is unnecessary to give any parametric restrictions on serial correlations, for simplicity, we do impose them. We consider two examples that are simple and of particular interest in practice: AR(1) and MA(1). For the AR(1) case, suppose

$$\epsilon_{it} = \psi \epsilon_{it-1} + \zeta_{it}, \text{ for some non-zero constant } \psi, \quad (2.4a)$$

$$E(\zeta_{it} x_{it}^0) = 0, \quad (2.4b)$$

$$E(\zeta_{it} \phi_i) = E(\zeta_{it} \epsilon_{it-1}^0) = 0, \quad (2.4c)$$

for  $t = 1, \dots, T$ . Under weak exogeneity, the regressors are correlated with the lagged errors so that  $x_{it}$  is correlated with  $\epsilon_{it-j}$  for  $j > 0$ , thus

$$E(x_{it}^1 \epsilon_{it-j}) \neq 0, \quad j > 0. \quad (2.5)$$

We now examine whether moment conditions of Assumption 2.2, given (2.4) and (2.5), are valid. But,

$$E(x_{it}' \epsilon_{it}) = \psi E(x_{it}' \epsilon_{it-1}) + E(x_{it}' \zeta_{it}) = \psi E(x_{it}' \epsilon_{it-1}) \neq 0.$$

And the same course shows that  $E(x_{it}' \epsilon_{it+j}) \neq 0$ ,  $j > 0$ . Thus given (2.4), Assumption 2.2 is at odds with condition (2.5), as is in dynamic models.

Condition (2.4), in fact, implies a set of orthogonality conditions that are not linear in parameters. Ignoring (2.4c), covariance restrictions, (2.4b) implies

$$E[(\Delta u_{it} - \psi \Delta u_{it-1}) x_{it}^0] = 0, \quad t = 2, \dots, T-1,$$

since  $(u_{it} - u_{it+1}) - \psi(u_{it-1} - u_{it}) = \zeta_{it} - \zeta_{it+1}$ . These are  $[\frac{1}{2}T(T-1)-1]k$  moment conditions so that, to compare with  $Lw_i$ , the number of moment conditions reduced by the AR(1) serial correlations is  $k$ .

We next consider the case when the idiosyncratic errors follow the MA(1), so when we have

$$\epsilon_{it} = \eta_{it} - \rho \eta_{it-1}, \quad (2.6a)$$

$$E(\eta_{it} x_{it}^0) = 0, \quad (2.6b)$$

$$E(\eta_{it} \eta_{it-1}^0) = E(\eta_{it} \phi_i) = 0, \quad (2.6c)$$

for  $t = 1, \dots, T$ . Then,

$$E(x_{it}' \Delta u_{it}) = E[x_{it}' (\eta_{it} - \rho \eta_{it-1} - \eta_{it+1} + \rho \eta_{it})] = -E(x_{it}' \eta_{it-1}) \neq 0.$$

Thus, if we use the instruments  $Lw_i$ , the obtained estimator will be inconsistent in this case.

However, from (2.6b) we have the  $\frac{1}{2}(T-1)(T-2)$  moment conditions

$$E(x_{it}^0 \Delta u_{it+j}) = 0, \quad j \geq 1. \quad (2.7)$$

These are linear in  $\beta$ , and we have the instruments

$$\begin{bmatrix} 0 \\ x_{i1}^0 \\ -x_{i1}^0 & x_{i2}^0 \\ & -x_{i2}^0 & \ddots \\ & & & x_{iT-2}^0 \\ & & & & -x_{iT-2}^0 \end{bmatrix}.$$

Comparing this with  $Lw_i$ , the number of instruments shrunk by the MA(1) serial correlation is  $(T-1)k$ . The instruments that becomes invalid is  $L[\text{diag}(x_{i1}, \dots, x_{iT-1})]$ , which corresponds to the statement in (2.7) that  $E(x_{it}^0 \eta_{it-1}) \neq 0$ .

However, it is interesting to note that the condition  $E(x_{it}^0 \eta_{it-1}) = 0$  can be valid, giving an alternative explanation for the serial correlation. As appears in rational expectations literature, time lags until the shocks are observed by individuals can cause serial correlation. If this is the case, it is quite possible that the one period lagged errors are uncorrelated with the current regressors though the errors follow MA(1), thus the moment condition  $E(x_{it}^0 \eta_{it-1}) = 0$  can be plausible. Therefore, allowing for time lags until shock is observed, it would not be necessary to reduce the set of instruments.

We showed that the set of instrumental variables in weakly exogenous model is closely connected to the structure



of  $\Sigma$  through a couple of examples. However, we can give a different interpretation for serial correlation of the idiosyncratic errors. Suppose that serial correlation is caused by some omitted variables that are uncorrelated with the regressors of all periods (like time constant-errors in the random effects model) and that the error components to which the regressors are weakly exogenous are serially uncorrelated. Then  $Lw_i$  is valid and  $\Sigma$  is unrestricted. In this case, the errors should be of the form

$$u_{it} = \phi_i + \epsilon_{it} = \phi_i + \epsilon_{it}^s + \epsilon_{it}^w, \quad t = 1, \dots, T, \quad (2.11)$$

where the regressors are strictly exogenous to the serially correlated errors  $\epsilon_{it}^s$  and weakly exogenous to the intertemporally uncorrelated errors  $\epsilon_{it}^w$ . This distinguishes general weakly exogenous models from dynamic models. In dynamic models, there can not exist the error components to which the regressors are strictly exogenous. A familiar example is the unobserved individual effects  $\phi_i$ . The correlation between  $\phi_i$  and the lagged dependent variables are guaranteed in dynamic models, but it is not necessarily the case in general weakly exogenous models.

SAW and Hayashi and Sims (1983) pointed out that eliminating serial correlations by forward filtering requires a similar situation. We quote SAW (p. 11),

"Forward filtering requires that the serial correlations in the errors do not depend on the values current and lagged

values of the instruments."

As was noted earlier, SAW showed that the Keane and Runkle estimator, the 2SLS estimator based on forward filtered equations, is numerically identical to 3SLS if all the instruments  $Lw_i$  are used. Thus, whenever the Keane and Runkle estimator is inconsistent, so is 3SLS. The requirement for vindicating forward filtering noted by SAW and Hayashi and Sims indeed is needed for currently utilized instrumental variables to be valid.

Finally, we note on the relationship between the moment conditions for lagged dependent variables and the structure of  $\Sigma$  in the dynamic model considered by Ahn and Schmidt and many others. For simplicity, let  $x_{it} = y_{it-1}$ ,  $t = 1, \dots, T$ . We start with the assumption that

$$E(\epsilon_{it} y_{it-1}^0) = 0, \quad t = 1, \dots, T, \quad (2.12)$$

which implies that  $\{\epsilon_{it}: t=1, \dots, T\}$  is serially uncorrelated. We do not impose the homoskedasticity restriction of  $\{\epsilon_{it}: t=1, \dots, T\}$  so that  $E(\epsilon_{it}^2) \neq E(\epsilon_{is}^2)$ ,  $t \neq s$ . Assumption (2.12) alone implies the set of instruments  $Lw_i$  that are usually used in dynamic models, but, as was noted by Ahn and Schmidt, we need an additional assumption

$$E(\epsilon_{it} \phi_i) \text{ are the same, } t = 1, \dots, T, \quad (2.13)$$

in order to have the restricted covariance matrix where all of the off-diagonal elements are the same. The T-2

additional moment conditions suggested by Ahn and Schmidt,

$$E(u_{it}u_{it+1} - u_{it+1}u_{it+2}) = 0, \quad t = 1, \dots, T-2, \quad (2.14)$$

along with the instruments  $Lw_i$ , encompass all of the moment conditions implied by conditions (2.12) and (2.13). Thus, the Hausman test (1978) and the GMM test (Hansen, 1982), given the instruments  $Lw_i$ , of testing the validity of the moment conditions (2.14) essentially test whether the condition (2.13) holds.

When (2.13) is violated, the covariance matrix still will be restricted as long as  $\{\epsilon_{it}: t=1, \dots, T\}$  is serially uncorrelated. The off-diagonal part of  $\Sigma$  has  $\frac{1}{2}T(T-1)$  possibly distinctive elements, but these are composed of the  $T+1$  elements  $E(\phi_i \epsilon_{it})$ ,  $t = 1, \dots, T$ , and  $\sigma_\phi^2$ , and so there should be the  $\frac{1}{2}T(T-3)-1$  restrictions. But, these add no useful moment conditions given the instruments  $Lw_i$ , since  $Lw_i$  stands for the  $\frac{1}{2}(T-1)(T-2)$  moment conditions from covariance restrictions when condition (2.13) holds, and  $\frac{1}{2}(T-1)(T-2) > \frac{1}{2}T(T-3)-1$ .

### 3. ESTIMATION USING THE BMS ASSUMPTION

Wooldridge (1993) showed that the usual standard errors based on N3SLS in hedonic pricing models are not consistent, and he derived a condition for the usual 3SLS standard errors to be valid. Ahn (1990) obtained a similar result in

dynamic panel data models when certain moment conditions from covariance restrictions are used. These results essentially show that some moment conditions necessarily cause heteroskedasticity in the models considered by Wooldridge (1993) and Ahn (1990).

It was suggested by SAW that the equi-correlation assumption of BMS can hold in weakly exogenous models with unobserved effects, and then  $(T-1)k$  instruments are added. This section shows that the usual 3SLS standard errors are not consistent if these instruments are used. Also, we apply this result to the dynamic case and link it to the result obtained by Ahn (1990). We assume that the covariance matrix is of the random effects form to avoid the consistency arguments of the previous section.

For the usual 3SLS standard errors to be consistent when the instruments  $Lw_i$  are used, we need the assumption of no heteroskedasticity

ASSUMPTION 3.1:  $E(w_i' L' u_i u_i' L w_i) = E(w_i' L' \Sigma L w_i)$ .

Wooldridge (1993, Example 5.2) showed that Assumption 3.1 is satisfied under the assumptions

$$E(\epsilon_{it} | x_{it}^0, \epsilon_{it-1}^0, \phi_i) = 0 \quad (3.1a)$$

and

$$E(\epsilon_{it}^2 | x_{it}^0, \epsilon_{it-1}^0, \phi_i) = \sigma_\epsilon^2, \quad (3.1b)$$

that are plausible and used quite often in the rational

expectations models. To see this, note that  $E(w_i' L' u_i u_i' L w_i)$  has elements  $E(x_{it}^0 \Delta u_{it} \Delta u_{is} x_{is}^0) = E[x_{it}^0 (\epsilon_{it} - \epsilon_{it+1}) (\epsilon_{is} - \epsilon_{is+1}) x_{is}^0]$ ,  $t, s = 1, \dots, T-1$ , and so the result follows immediately by the law of iterated expectations.

The assumption suggested by BMS is

ASSUMPTION 3.2:  $E(x_{it}' \phi_i)$  is the same,  $t = 1, \dots, T$ ,

which, given Assumption 2.2, is expressed as

$$E(x_{it} u_{it} - x_{it+1} u_{it+1}) = 0, \quad t = 1, \dots, T-1. \quad (3.2)$$

It is convenient to see the orthogonality conditions implied by this through the moment matrix

$$E \begin{bmatrix} x_{i1}' u_{i1} & \cdots & x_{i1}' u_{iT} \\ \vdots & & \vdots \\ x_{iT}' u_{i1} & \cdots & x_{iT}' u_{iT} \end{bmatrix}. \quad (3.3)$$

Under Assumptions 2.2 and 3.2, all of the elements in the upper triangular of the moment matrix are the same, while Assumption 2.2 only implies that the elements in each row of the upper triangular of (3.3) are the same. Thus, Assumption 3.2 adds the  $(T-1)k$  moment conditions that the upper triangular elements of (3.3) are the same across the rows. The moment conditions of (3.2) describes these and we have instrumental variables

$$h_i = \begin{bmatrix} x_{i1} \\ -x_{i2} & x_{i2} \\ & -x_{i3} & \cdot \\ & & \cdot & x_{iT-1} \\ & & & -x_{iT} \end{bmatrix}.$$

A notable distinction between  $h_i$  and  $Lw_i$  is that  $h_i$  is not in the space spanned by  $L$  and it is useful for identification of the parameters on the time-constant variables.

For the usual 3SLS standard errors to be consistent when the instruments  $h_i$  are used, the conditions

$$E(h_i' u_i u_i' h_i) = E(h_i' \Sigma h_i) \quad (3.4a)$$

and

$$E(w_i' L' u_i u_i' h_i) = E(w_i' L' \Sigma h_i) \quad (3.4b)$$

should be satisfied. Now, we show that condition (3.4) can not hold under any plausible assumptions for weakly exogenous models with fixed effects. To this end, we will show that the equalities for the first  $k \times k$  blocks of (3.2a) and (3.2b) do not hold. The first  $k \times k$  block of  $E(h_i' \Sigma h_i) = E[h_i' (\sigma_\epsilon^2 I_T + \sigma_\phi^2 e_i e_i') h_i]$  is

$$\sigma_\epsilon^2 E(x_{i1}' x_{i1} + x_{i2}' x_{i2}) + \sigma_\phi^2 E[(x_{i1} - x_{i2})' (x_{i1} - x_{i2})] \quad (3.5a)$$

And for  $E(h_i' u_i u_i' h_i)$ , we have

$$\begin{aligned}
& E(x_{i1}' u_{i1}^2 x_{i1} - x_{i1}' u_{i1} u_{i2} x_{i2} - x_{i2}' u_{i2} u_{i1} x_{i1} + x_{i2}' u_{i2}^2 x_{i2}) \\
& = E[\epsilon_{i1}^2 x_{i1}' x_{i1} - \epsilon_{i1} \epsilon_{i2} (x_{i1}' x_{i2} + x_{i2}' x_{i1}) + \epsilon_{i2}^2 x_{i2}' x_{i2}] \\
& \quad + E[\phi_i^2 (x_{i1} - x_{i2})' (x_{i1} - x_{i2})] \tag{3.5b} \\
& \quad + E[2\phi_i \epsilon_{i1} x_{i1}' x_{i1} - \phi_i (\epsilon_{i1} + \epsilon_{i2}) (x_{i1}' x_{i2} + x_{i2}' x_{i1}) + 2\phi_i \epsilon_{i2} x_{i2}' x_{i2}]
\end{aligned}$$

For the equality between (3.5a) and (3.5b) to hold, three conditions should be met:

- (i)  $E[\epsilon_{i1} \epsilon_{i2} (x_{i1}' x_{i2} + x_{i2}' x_{i1})]$  in the first term of (3.5b) is zero,
- (ii)  $E(\phi_i^2 \Delta x_{i1}' \Delta x_{i1}) = \sigma_\phi^2 E(\Delta x_{i1}' \Delta x_{i1})$  for equality of the second terms in (3.5a) and (3.5b),
- (iii) the last term of (3.5b) is zero.

Condition (i) holds under Assumption 3.1. Condition (ii) holds if we are willing to assume

$$E(\phi_i^2 | \Delta x_{i1}, \dots, \Delta x_{iT-1}) = \sigma_\phi^2, \tag{3.6}$$

which is a strong assumption, but still is plausible along with Assumption 3.2. Condition (iii) is a different matter. From the assumptions in (3.1), the last term of (3.5b) becomes

$$E[\phi_i (\epsilon_{i1} + \epsilon_{i2}) (x_{i1}' x_{i2} + x_{i2}' x_{i1})] = E[\phi_i \epsilon_{i1} (x_{i1}' x_{i2} + x_{i2}' x_{i1})],$$

which, however, never becomes zero under any assumptions that are plausible for weakly exogenous models with fixed effects.

Next we compare the first  $k \times k$  block of  $E(w_i' L' u_i u_i' h_i)$

and  $E(w_i' L' \Sigma h_i)$ . For  $E(w_i' L' \Sigma h_i)$ , we have

$$E[w_i' L' (\sigma_\epsilon^2 I_T + \sigma_\phi^2 e_T e_T') h_i] = \sigma_\epsilon^2 E(w_i' L' h_i) = \sigma_\epsilon^2 E(x_{i1}' x_{i1} + x_{i1}' x_{i2}),$$

and from the assumption (3.1),  $E(w_i' L' u_i u_i' h_i)$  becomes

$$\sigma_\epsilon^2 E(x_{i1}' x_{i1} + x_{i1}' x_{i2}) + E(\phi_i \epsilon_{i1} x_{i1}' x_{i2}),$$

the second term of which is not zero in weakly exogenous cases with fixed effects. Therefore, the usual 3SLS standard errors are not consistent if the instruments  $h_i$  are used.

We now apply this result to dynamic models and compare it with the result obtained by Ahn (1990). For simplicity, we focus on a simple AR(1) dynamic model with no exogenous regressors, so  $x_{it} = y_{it-1}$ ,  $t = 1, \dots, T$ . The covariance matrix is assumed to be of the random effects form, and we keep the assumption (3.1) of conditional moment conditions. The BMS assumption, in this case, tells that

$$E(y_{it} \phi_i) \text{ is the same for } t = 0, \dots, T, \quad (3.7)$$

which is implied by the stationarity of  $\{(y_{it} \phi_i) : t=0, \dots, T\}$ , the assumption suggested by Arellano and Bover (1990). The moment conditions are

$$E(y_{it} u_{it+1} - y_{it+1} u_{it+2}) = 0, \quad t = 0, \dots, T-2, \quad (3.8)$$

and we have the set of the instruments  $(Lw_i, h_i)$ . Note that all of the elements in the upper triangular of the moment



matrix (3.3) are the same, and so the number of moment conditions is  $\frac{1}{2}T(T+1)-1$ . This covers the moment conditions from the random effects covariance restrictions (except for those from the restriction of equal diagonals in  $\Sigma$ ). Nothing essentially differs from the previous model, and the usual standard errors from 3SLS are not consistent if  $h_i$  is used as instruments.

Without the stationarity of  $\{(y_{it}\phi_i):t=0,\dots,T\}$ , there are  $\frac{1}{2}T(T-1)-1$  moment conditions from the equal off-diagonal restriction of  $\Sigma$ , and  $T-1$  conditions that  $E(y_{i0}\Delta u_{it}) = 0$ ,  $t = 1, \dots, T-1$ . Together, we have  $\frac{1}{2}T(T+1)-2$  conditions, one less than the case under the stationarity of  $\{(y_{it}\phi_i):t=0,\dots,T\}$ . These comprise  $\frac{1}{2}T(T-1)$  moment conditions  $E(w_i' L' u_i) = 0$ , and  $T-2$  conditions (2.14) suggested by Ahn and Schmidt. The conditions in (2.14) are essentially nonlinear in parameters. Ahn and Schmidt showed that, given that the diagonals of  $\Sigma$  are equal, these additional moment conditions are represented as linear in parameters like

$$E(y_{it}\Delta u_{it} - y_{it+1}\Delta u_{it+1}) = 0, \quad t = 1, \dots, T-2. \quad (3.9)$$

These are, in fact, linear combinations of the conditions in (2.14) with the moment conditions from the restrictions of equal diagonals, and with  $E(w_i' L' u_i) = 0$ . Ahn (1990) showed that the usual 3SLS standard errors are not consistent when the instrumental variables from (3.9) are used, which is quite closely related to the result we obtained. The

structure of instruments from (3.9) is quite similar to  $h_i$ .

The conventional N3LS methods do not generalize to implement the moment conditions from covariance restrictions. Thereby, there is no point to compare between GMM and N3SLS when the moment condition (2.14) are used.

#### 4. NEARLY EFFICIENT ESTIMATION

GMM using all of the moment conditions leads to the fully efficient estimator in large samples. In many panel data sets (e.g. when we construct the data from PSID), there is trade-off between  $N$  and  $T$  in applications; as  $T$  increases, the size of cross-section  $N$  shrinks. Further, as  $T$  increases, the number of moment conditions grows by  $T^2$ . Thereby, when  $T$  is relatively large (like 6 - 10), situations where it is not even feasible to use all of the moment conditions could arise, and finding some shorter lists of instruments is of practical importance.

In this section, we propose several reduced lists of instruments that would lead to nearly efficient estimators. However, we cannot provide "how near", since the efficiencies of the resulting estimators depend on too many factors to be sorted out clearly. The estimators we propose are intended to serve as only possible estimators among many, and they need to be compared to other possible estimators in practice. The arguments apply equally to the

estimator proposed by Keane and Runkle (1992).

In many cases some of the regressors in weakly exogenous models are, in fact, strictly exogenous. A leading example is the dynamic panel data model with additional strictly exogenous regressors considered by many. Suppose  $x_{it} = (x_{1it}, x_{2it})$  for  $t = 1, \dots, T$  and  $\beta = (\beta_1', \beta_2')'$  in model (2.1), where  $x_{1it}$  is weakly exogenous and  $x_{2it}$  are strictly exogenous to the errors. The dimensions of  $x_{1it}$  and  $x_{2it}$  are  $1 \times k_1$  and  $1 \times k_2$ , where  $k = k_1 + k_2$ . We first consider a simple model where all the regressors are not correlated with the time-constant error  $\phi_i$ . We have

ASSUMPTION 4.1:  $E(x_{1it}' u_{is}) = 0, 1 \leq t \leq s \leq T$ .

ASSUMPTION 4.2:  $E(x_{2i} \otimes u_i) = 0$ .

The set of instrumental variables implied by Assumption 4.1 is  $w_{1i} = \text{diag}(x_{1i1}^0, x_{1i2}^0, \dots, x_{1iT}^0)$ . Assumption 4.2 implies the instruments  $w_{2i} = (I_T \otimes x_{2iT}^0)$ . Let  $w_i = (w_{1i}, w_{2i})$  of the column dimension  $\frac{1}{2}T(T+1)k_1 + T^2k_2$ . We further assume that there presents no conditional heteroskedasticity, so

ASSUMPTION 4.3:  $E(w_i' u_i u_i' w_i) = E(w_i' \Sigma w_i)$ .

Under Assumptions 4.1 - 4.3 and if  $\Sigma$  is diagonal so that if there are no unobserved individual effects and the errors are intertemporally uncorrelated, the reduced set of instruments  $\text{diag}(x_{i1}, x_{i2}, \dots, x_{iT})$  generate the fully efficient estimator (Chapter 2). No result exists that finds some

reduced set of instruments that lead to the fully efficient estimator when  $\Sigma$  is not diagonal (Ahn and Schmidt, Chapter 2, Chapter 4). Here we allow momentarily that  $\Sigma$  to be unrestricted, but Assumptions 4.1 - 4.3 hold. These assumptions rules out time constant unobservables correlated with the regressors, but the result obtained under these assumptions will be generalized to more practical models.

For the rest of the paper, for any  $T \times p$  matrix  $m_i$ ,  $M \equiv (m_1', \dots, m_N')'$  of dimension  $NT \times p$ . Thus, the matrix  $M$  is the stacked matrix of  $m_i$  for  $i = 1, \dots, N$  with the  $i$ -th block  $m_i$ . Let  $\Omega = I_N \otimes \Sigma$ . Then,  $W = (W_1, W_2)$ , note that  $W_2$  is equally represented by  $X_2^0 \otimes I_T$  (Chapter 2).

If all the regressors are strictly exogenous, the GLS estimator is fully efficient. Nevertheless,  $E(x_{1i}' \Sigma^{-1} u_i) \neq 0$ , and  $\Sigma^{-1} x_{1i}$  are not valid instruments. However,  $\Sigma^{-1} x_{2i}$  are still valid. It would be natural to consider the property of the 3SLS estimator using the instruments  $z_i = (w_{1i}, \Sigma^{-1} x_{2i})$ , the column dimension of which is  $\frac{1}{2}T(T+1)k_1 + k_2$ . Estimators are defined as

$$\tilde{\beta} = [X'W(W'\Omega W)^{-1}W'X]^{-1}X'W(W'\Omega W)^{-1}W'Y$$

and

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

We will compare the variances of  $\tilde{\beta}$  and  $\hat{\beta}$ . Let  $P_{(\cdot)}$  denote the projection onto the columns of  $(\cdot)$ . Then,  $P_{(X_2^0 \otimes I_T)} X_2 = X_2$ , and  $P_{(X_2^0 \otimes \Sigma^{1/2})} = P_{(X_2^0 \otimes \Sigma^{-1/2})} = P_{(X_2^0 \otimes I_T)}$ . Thus,

we have

$$W(W'\Omega W)^{-1}W'X_2 = \Omega^{-1/2}P_{(\Omega^{1/2}Z_1, X_2'\Sigma^{1/2})}\Omega^{-1/2}X_2 = \Omega^{-1}X_2, \quad (4.2a)$$

and

$$Z(Z'\Omega Z)^{-1}Z'X_2 = \Omega^{-1}X_2 \quad (4.2b)$$

since  $\Omega^{1/2}Z(Z'\Omega Z)^{-1}Z'\Omega^{1/2}\Omega^{-1/2}X_2 = P_{(\Omega^{1/2}W_1, \Omega^{-1/2}X_2)}\Omega^{-1/2}X_2 = \Omega^{-1/2}X_2$ .

From (4.2), it follows that

$$[X'W(W'\Omega W)^{-1}W'X]^{-1} = \begin{bmatrix} X_1'W(W'\Omega W)^{-1}W'X_1 & X_1'\Omega^{-1}X_2 \\ X_2'\Omega^{-1}X_1 & X_2'\Omega^{-1}X_2 \end{bmatrix}^{-1}, \quad (4.3a)$$

and

$$[X'Z(Z'\Omega Z)^{-1}Z'X]^{-1} = \begin{bmatrix} X_1'Z(Z'\Omega Z)^{-1}Z'X_1 & X_1'\Omega^{-1}X_2 \\ X_2'\Omega^{-1}X_1 & X_2'\Omega^{-1}X_2 \end{bmatrix}^{-1}. \quad (4.3b)$$

The difference between these two arises from the difference between  $X_1'W(W'\Omega W)^{-1}W'X_1$  and  $X_1'Z(Z'\Omega Z)^{-1}Z'X_1$ . We know that  $\tilde{\beta}$  is more efficient than  $\hat{\beta}$ . However, the difference in efficiency might not be substantial, since both  $W$  and  $Z$  include  $W_1$  which provides direct information for  $X_1$ . This result depends on Assumption 4.3 of no conditional heteroskedasticity, and asymptotically the GMM estimator using the instruments  $Z$  could be dominated by GMM using simpler instruments  $(W, X_2)$  in the presence of heteroskedasticity. This is the reason why the instruments  $Z$  is limited to serve as only a choice to be compared with many other constructable set of instruments.

If the covariance matrix is of the random effects form

and the unobserved effects are not correlated with the regressors, it is better to use  $(W_1, PX_2, QX_2)$  than  $(W_1, \Omega^{-1}X_2)$  as instruments whether there is heteroskedasticity or not, since  $\Omega^{-1}X_2 = aPX_2 + bQX_2$  and we never worse off by using  $(PX_2, QX_2)$  instead of  $aPX_2 + bQX_2$ , where  $P = I_N \otimes \frac{1}{T} e_1 e_1'$  and  $Q = I_{NT} - P$ . See Chapter 2 for the result of  $\Omega^{-1} = aP + bQ$ . In this case,  $(PX_2, QX_2)$  explains  $X_1$  better than  $aPX_2 + bQX_2$ . We omit complicated algebra comparing the performance of two sets of instrumental variables  $(W_1, PX_2, QX_2)$  and  $(W_1, \Omega^{-1}X_2)$  since it is intuitively clear.

In the case when the unobserved individual effects are correlated with all of the regressors and the covariance matrix is of the random effects form, we have the instruments  $r_i = L \cdot \text{diag}(x_{1i1}^0, x_{1i2}^0, \dots, x_{1iT-1}^0)$  and  $L \otimes x_{2i}^0$ . In this case, the reduced set of instruments  $(R, QX_2)$  would produce a nearly efficient estimator, since  $P_{(R, QX_2)} X_2 = P_{(R, X_2^0 \otimes L)} X_2 = QX_2$ . The same algebra compares the variances of estimator from these instruments, and produces a result like (4.3).

The above arguments also apply to more general models where only some of the regressors are correlated with the fixed effects. We study the dynamic version of the Hausman and Taylor model considered by Ahn and Schmidt:

$$Y_{it} = \alpha Y_{it-1} + x_{1it} \beta_1 + x_{2it} \beta_2 + z_{1i} \gamma_1 + z_{2i} \gamma_2 + \phi_i + \epsilon_{it}, \quad (4.4)$$

for  $t = 1, \dots, T$ . All the regressors but the lagged dependent variable are strictly exogenous to the

idiosyncratic errors, and only  $(x_{2it}, z_{2i})$  is correlated with unobserved fixed effects  $\phi_i$ . The covariance matrix takes the random effects form. Then, we have the set of instruments  $\Gamma = [R_1, (X_1, Z_1 \otimes e_T) \otimes I_T, (X_2, Z_2 \otimes e_T) \otimes L]$ , where  $R_1$  includes all the instruments between the lagged dependent variables and the disturbances. Note that  $(Z_1, Z_2) \otimes e_T$  is the time-constant regressors. The reduced lists of instrumental variables  $F = (R_1, QX, PX_1, Z_1 \otimes e_T)$  would produce a nearly efficient estimator, since  $P_{(\Gamma)}(X_1, Z_1 \otimes e_T) = P_{(F)}(X_1, Z_1 \otimes e_T) = (X_1, Z_1 \otimes e_T)$  and  $P_{(\Gamma)}(X_2, Z_2) = P_{(F)}(X_2, Z_2)$ . Hence, the reasoning is the same as the previous cases. Note that  $(QX, X_1 \otimes e_T, Z_1 \otimes e_T)$  is the reduced lists of instruments in the static Hausman and Taylor model, that produces the fully efficient estimator when there is no heteroskedasticity. For more details, see Ahn and Schmidt (1992).

## 5. CONCLUSION

We showed that the moment conditions that are currently utilized in weakly exogenous models may not be valid in some cases if the idiosyncratic errors are serially correlated. Were the serial correlations of the idiosyncratic errors detected from an initial stage estimator of  $\beta$ , testing exogeneity of the instruments would be constructive. Difficulty arises when  $\Sigma$  is entirely unrestricted. Then, identification of  $\beta$  becomes a serious problem unless there

are a sufficient number of strictly exogenous instrumental variables. However, testing the structure of  $\Sigma$  would lead to some nice specification tests that are not doable in single equation models, which we leave for further study.

Though we prove that GMM should be used when we estimate using the BMS assumption, the reason why there necessarily arises heteroskedasticity problem is not so clear. But, the condition derived by Wooldridge (1993) provides a partial answer for this. He essentially shows that weak exogeneity is minimally required for the usual 3SLS standard errors to be consistent. Any instruments  $w_{it}$  for the  $t$ -th equation that satisfy  $E(\epsilon_{it}, \epsilon_{it+1}, \dots, \epsilon_{iT} | w_t) = 0$  do not raise this problem, but the instruments from the BMS assumption and the instruments suggested by Ahn (1990) necessarily relate to disturbances across equations and thereby violate the Wooldridge's condition.



## CHAPTER 4

### INFORMATION FROM COVARIANCE RESTRICTIONS IN PANEL DATA MODELS

#### 1. INTRODUCTION

In this chapter, we study the orthogonality conditions from covariance restrictions. The main purpose is to find whether the covariance restrictions are useful for more efficient estimation in several panel data models. We focus on the restrictions from scalar and random effects covariance matrices, but the results can be extended to more general restrictions. Also, we derive the asymptotic variances of generalized method of moment (GMM) estimators that use the moment conditions from covariance restrictions.

Covariance restrictions have largely been studied in the context of simultaneous equations models, and identification has been the major concern. For efficiency, Rothenberg and Leenders (1964) showed that the exploitation of covariance restrictions lowers the Cramer-Rao bound in standard simultaneous equations models when the errors are normally distributed. Hausman, Newey and Taylor (1987) proposed augmented 3SLS as a handy way to realize the efficiency gains from covariance restrictions.

For panel data models, previous studies have focused on simple dynamic models. No results exist for static models

like the Hausman and Taylor (1981) model (HT henceforth) and general weakly exogenous models. The random effects covariance structure, which has become an almost standard assumption in panel data analysis, implies a set of orthogonality conditions. Instrumental variables for lagged dependent variables considered by Anderson and Hsiao (1982), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991) and Ahn and Schmidt are based on the orthogonality conditions from the random effects covariance matrix.

Covariance restrictions are rarely used in practice except for dynamic panel data models and triangular simultaneous equations models, where covariance restrictions are crucial for identification. This may be due to reluctance to utilize a priori restrictions that will cause inconsistency of estimators when they are false. Another important reason would be computational burden of numerical optimization, which is required in general to realize the efficiency gains from covariance restrictions. But, if covariance restrictions bring non-trivial efficiency gains, the computational burden is secondary. Therefore, it would be nice to have an easy way (without getting nonlinear estimators) to approach the possible efficiency gain when moment conditions from covariance restrictions are added. We show how to consistently estimate the asymptotic variances of the nonlinear GMM estimators that incorporate covariance restrictions without numerical optimization. By

comparing the two variance estimates of estimators (with and without covariance restrictions) we can see the possible efficiency gains when we add the moment conditions from covariance restrictions. If the efficiency gain is non-trivial, it would be worth doing numerical optimization. Once we get a nonlinear GMM estimator, it is straightforward to apply Hausman test (Hausman, 1978) or GMM test (Hansen, 1982) to test whether the covariance restrictions used are valid. Thus, it is not hard to get around the possible inconsistency problem of estimators from using false restrictions.

In section 2, we study a general model and give some preliminary results used for the chapter. The asymptotic variances of GMM estimators that use the orthogonality conditions from covariance restrictions will be derived, and it will be shown how they are consistently estimated without numerical optimization. Also, we provide the conditions when the linear GMM estimators using the residuals as instrumental variables are asymptotically identical to the nonlinear GMM.

In sections 3 and 4 we study covariance restrictions in specific models and derive conditions when the moment conditions from covariance restrictions are redundant. Section 3 deals with models where the regressors are strictly exogenous to the time-varying errors. It turns out that certain moment conditions from covariance restrictions

are useful unless some third moment conditions - essentially symmetry conditions - of the errors are met. We cover three models: the model with scalar covariance matrix, the random effects model, and a fixed effects type model where the unobserved individual effect is correlated with the regressors.

Section 4 studies weakly exogenous models. We argue that the orthogonality conditions from covariance restrictions can be redundant when the covariance matrix is diagonal, but whenever the covariance matrix is not diagonal, they are essentially always useful. Section 5 concludes.

## 2. PRELIMINARIES

### 2.1. Redundancy Conditions for Moment Restrictions

We study a linear panel data model

$$y_{it} = x_{it}\beta + u_{it}, \quad t = 1, \dots, T, \quad (2.1)$$

where  $\{(y_{it}, x_{it}) : i = 1, \dots, N\}$  is an i.i.d. random sequence.

Let  $y_i = (y_{i1}, \dots, y_{iT})'$ ,  $x_i = (x_{i1}', \dots, x_{iT}')'$ , and  $u_i = (u_{i1}, \dots, u_{iT})'$  of dimensions  $T \times 1$ ,  $T \times k$  and  $T \times 1$ , respectively. (2.1)

is equally expressed as  $y_i = x_i\beta + u_i$ . Let  $\Sigma \equiv E(u_i u_i')$ , a  $T \times T$  nonsingular matrix. We assume that the 4-th moments of  $(y_i, x_i)$  exists. Throughout the chapter, for any  $T \times p$  matrix

$m_i$ ,  $M = (m_1', \dots, m_N')'$  of dimension  $NT \times p$ , where  $N$  is the number of cross-section observations.

There is a set of  $T \times h$  observable instrumental variables  $w_i$ , that satisfy

ASSUMPTION 2.1:  $E(w_i' u_i) = 0$ .

ASSUMPTION 2.2:  $E(w_i' x_i)$  has full column rank and  $E(w_i' w_i)$  is positive definite.

Assumption 2.2 is a regularity condition that ensures identification, and it is assumed for the rest of the paper without being stated further.

Throughout the paper, we let  $E[g_{1i}(\beta)] = 0$  denote an initial set of moment conditions and  $E[g_{2i}(\beta)] = 0$  be additional moment conditions from covariance restrictions. Therefore, our major concern is whether the additional moment conditions  $E[g_{2i}(\beta)] = 0$  are redundant, given the conditions  $E[g_{1i}(\beta)] = 0$ .

Let  $g_i(\beta) = [g_{1i}(\beta)', g_{2i}(\beta)']'$ . If we use the orthogonality conditions  $E[g_i(\beta)] = 0$ , GMM solves the problem

$$\min_{\beta} Q_N(\beta) = \left[ \frac{1}{N} \sum_{i=1}^N g_i(\beta) \right]' \hat{\Lambda}_N^{-1} \left[ \frac{1}{N} \sum_{i=1}^N g_i(\beta) \right].$$

It is well known that the best choice of weighting matrix  $\hat{\Lambda}_N$  is a consistent estimator of  $\Lambda \equiv E[g_i(\beta)g_i(\beta)']$ , so we take

$$\hat{\Lambda}_N = \frac{1}{N} \sum_{i=1}^N g_i(\hat{\beta})g_i(\hat{\beta})',$$

where  $\hat{\beta}$  is a consistent estimator of  $\beta$  (e.g. Hansen, 1982).

Then

$$\text{Avar}/N(\tilde{\beta}_{\text{GMM}} - \beta) = [D' \Lambda^{-1} D]^{-1}, \quad (2.2)$$

where  $D \equiv E[\frac{\partial g_i(\beta)}{\partial \beta'}]$ . Let  $\Lambda_{j\ell} = E[(g_{ji}(\beta)g_{\ell i}(\beta))']$ ,  $j, \ell = 1, 2$ , and  $D = (D_1', D_2')'$ , where  $D_j \equiv E[\frac{\partial g_{ji}(\beta)}{\partial \beta'}]$ ,  $j = 1, 2$ . Then, the asymptotic variance of the GMM estimator that uses only  $E[g_{1i}(\beta)] = 0$  is

$$\text{Avar}/N(\tilde{\beta}_{\text{GMM}} - \beta) = [D_1' \Lambda_{11}^{-1} D_1]^{-1}. \quad (2.3)$$

From (2.2) and (2.3) it is seen that interest centers on the difference between  $D' \Lambda^{-1} D$  and  $D_1' \Lambda_{11}^{-1} D_1$ . The former is no smaller than the latter since GMM never becomes worse asymptotically by adding orthogonality conditions. Thus, the information from covariance restrictions is useful unless  $D' \Lambda^{-1} D = D_1' \Lambda_{11}^{-1} D_1$ . Schmidt (1991) shows that this equality holds if and only if

$$D_2 = \Lambda_{21} \Lambda_{11}^{-1} D_1. \quad (2.4)$$

We will use this condition at several points in the remainder of the chapter.

## 2.2. Scalar Covariance and the Asymptotic Variance of GMM

We begin by considering the moment conditions from the scalar covariance matrix.

ASSUMPTION 2.3:  $\Sigma = \sigma^2 I_T$ .

Assumption 2.3 tells that the off-diagonals are zero and the diagonals are the same in  $\Sigma$ . The number of off-diagonal elements in  $\Sigma$  is  $T(T-1)$ , but due to the symmetry of  $\Sigma$ , the upper triangular of  $\Sigma$  is a duplicate of the lower triangular. Thus, the condition of zero off-diagonals implies the  $\frac{1}{2}T(T-1)$  orthogonality conditions

$$E(u_{is}u_{it}) = 0, s > t = 1, \dots, T-1. \quad (2.5)$$

The moment conditions (2.5) can be expressed as  $E(b_{1i}'u_i) = 0$  or  $E(b_{2i}'u_i) = 0$ , where

$$b_{1i} = \begin{bmatrix} u_{i2} & \dots & u_{iT} & & & \\ & & u_{i3} & \dots & u_{iT} & \\ & & & & \dots & \\ & & & & & u_{iT} \\ 0 & \dots & & & & 0 \end{bmatrix},$$

and

$$b_{2i} = \begin{bmatrix} 0 & \dots & & & 0 \\ u_{i1} & & 0 & \dots & 0 \\ & u_{i1} & u_{i2} & & \\ & & \cdot & \cdot & \\ & & & \cdot & \dots \\ & & & & u_{iT-1} \end{bmatrix}.$$

The dimension of both  $b_{1i}$  and  $b_{2i}$  is  $T \times \frac{1}{2}T(T-1)$ .

The condition of equal diagonals implies the  $T-1$  moment conditions

$$E(u_{it}^2 - u_{it+1}^2) = 0, \quad t = 1, \dots, T-1. \quad (2.6)$$

We express (2.6) as  $E(c_i' u_i) = 0$ , where

$$c_i = \begin{bmatrix} u_{i1} \\ -u_{i2} & u_{i2} & & \\ & -u_{i3} & \cdot & \\ & & \cdot & \cdot & \\ & & & \cdot & u_{iT-1} \\ & & & & -u_{iT} \end{bmatrix}.$$

The moment conditions in (2.5) and (2.6) contain different information, and will be considered separately.

We derive the asymptotic variance of the GMM estimator expressed in terms of  $x_i$ ,  $w_i$ ,  $b_{1i}$ ,  $b_{2i}$  and  $c_i$ . First,

$$D_1 = -E(w_i' x_i). \quad (2.7a)$$

For the moment conditions in (2.5), or  $E(b_{1i}' u_i) = 0$ , the elements of  $D_2$  are

$$-E(u_{is} x_{it} + u_{it} x_{is}), \quad s > t = 1, \dots, T-1.$$

Thus, it follows that  $D_2$  can be written as

$$D_2 = -E[(b_{1i} + b_{2i})' x_i]. \quad (2.7b)$$

For the moment conditions  $E(c_i' u_i) = 0$ , we have the elements of  $D_2$

$$-2E(u_{it} x_{it} - u_{it+1} x_{it+1}), \quad t = 1, \dots, T-1,$$

so

$$D_2 = -2E(c_i' x_i). \quad (2.7c)$$



Consider first the estimator that uses the moment conditions  $E(w_i' u_i) = 0$  and  $E(b_{1i}' u_i) = 0$ . Then from (2.2),

$$\text{Avar}/N(\tilde{\beta}_{\text{GMM1}} - \beta) = \left[ E[x_i' w_i, x_i' (b_{1i} + b_{2i})] E \begin{bmatrix} w_i' u_i u_i' w_i & w_i' u_i u_i' b_{1i} \\ b_{1i}' u_i u_i' w_i & b_{1i}' u_i u_i' b_{1i} \end{bmatrix}^{-1} E \begin{bmatrix} w_i' x_i \\ (b_{1i} + b_{2i})' x_i \end{bmatrix} \right]^{-1} \quad (2.8a)$$

Similarly for the GMM estimator that uses the moment conditions  $E(w_i' u_i) = 0$  and  $E(c_i' u_i) = 0$ , we have

$$\text{Avar}/N(\tilde{\beta}_{\text{GMM2}} - \beta) = \left[ E[x_i' w_i, 2x_i' c_i] E \begin{bmatrix} w_i' u_i u_i' w_i & w_i' u_i u_i' c_i \\ c_i' u_i u_i' w_i & c_i' u_i u_i' c_i \end{bmatrix}^{-1} E \begin{bmatrix} w_i' x_i \\ 2c_i' x_i \end{bmatrix} \right]^{-1} \quad (2.8b)$$

The equations (2.8a) and (2.8b) are useful in practice. They are consistently estimated using residuals  $\hat{u}_{it}$  in place of disturbances  $u_{it}$ ,  $t = 1, \dots, T$ , where  $\hat{u}_{it}$  is based on a consistent estimator of  $\beta$  from the initial instruments  $w_i$ . For a proof of consistency, See White (1984, pp. 135-138). Define  $\hat{b}_{1i}$ ,  $\hat{b}_{2i}$  and  $\hat{c}_i$  to be  $b_{1i}$ ,  $b_{2i}$  and  $c_i$  after replacing  $\hat{u}_{it}$  for  $u_{it}$  for  $t = 1, \dots, T$ , and let  $w_i^* = (w_i, \hat{b}_{1i})$ . Then, the ratio between the corresponding diagonal elements of the two estimators (standard errors) of the asymptotic variances

$$[X' W \left( \sum_{i=1}^N w_i' \hat{u}_i \hat{u}_i' w_i \right)^{-1} W' X]^{-1}$$

and

$$[X' (W, \hat{B}_1 + \hat{B}_2) \left( \sum_{i=1}^N w_i^* \hat{u}_i \hat{u}_i' w_i^* \right)^{-1} (\hat{B}_1 + \hat{B}_2, W)' X]^{-1}$$

will provide guidance about whether it is worth trying to

use the moment conditions from zero off-diagonal covariance restrictions through numerical optimization. Similar arguments apply for the equal diagonal restrictions.

Were they available,  $b_{1i}$ ,  $b_{2i}$  and  $c_i$  could serve themselves as instrumental variables. In the equation (2.8a), it is not hard to see that if  $E(b_{2i}'x_i) = 0$  the asymptotic variance of the nonlinear GMM estimator becomes the asymptotic variance of the linear GMM estimator using  $b_{1i}$  as instruments. Similarly, if  $E(b_{1i}'x_i) = 0$ , the linear GMM estimator using  $b_{2i}$  as instrumental variables is asymptotically identical to the nonlinear GMM estimator. Also equation (2.8b) shows that when  $E(c_i'x_i) = 0$ , the nonlinear GMM estimator is the same asymptotically as the linear GMM estimator using  $c_i$  as instrumental variables. It is interesting to ask what will happen if we use  $\hat{b}_{1i}$ ,  $\hat{b}_{2i}$  and  $\hat{c}_i$  as instruments instead of  $b_{1i}$ ,  $b_{2i}$  and  $c_i$ . As shortly will be shown, there is an interesting correspondence. If  $E(b_{2i}'x_i) = 0$ , there asymptotically is no difference between using  $b_{1i}$  and  $\hat{b}_{1i}$  as instruments. Similarly, if  $E(b_{1i}'x_i) = 0$  and  $E(c_i'x_i) = 0$ , we lose nothing by doing linear GMM using instruments  $\hat{b}_{2i}$  and  $\hat{c}_i$ , respectively.

We now verify these assertions. If we use  $\hat{b}_{1i}$ ,  $\hat{b}_{2i}$  or  $\hat{c}_i$  as instruments the resulting estimators are consistent since

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \hat{u}_{is} u_{it} = E(u_{is} u_{it}), \quad s, t = 1, \dots, T.$$

For the limiting distributions of the resulting estimators not to be affected by replacing  $b_{1i}$  by  $\hat{b}_{1i}$ , the two random variables  $\frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{u}_{is} u_{it}$  and  $\frac{1}{\sqrt{N}} \sum_{i=1}^N u_{is} u_{it}$ , for  $s > t$ , should have the same limiting distribution. But,  $\hat{u}_{is} = u_{is} - x_{is}(\hat{\beta} - \beta)$  and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{u}_{is} u_{it} = \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{is} u_{it} - \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{it} x_{is} \sqrt{N}(\hat{\beta} - \beta).$$

Because  $\sqrt{N}(\hat{\beta} - \beta) = o_p(1)$ , the limiting distributions of the two GMM estimators using  $\hat{b}_{1i}$  and using  $b_{1i}$  are the same provided

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N u_{it} x_{is} = E(u_{it} x_{is}) = 0, \quad s > t,$$

so when  $E(b_{2i}' x_i) = 0$ . Similarly, if  $E(b_{1i}' x_i) = 0$ , replacing  $b_{2i}$  by  $\hat{b}_{2i}$  do not affect the limiting distribution of the estimators. For  $\hat{c}_i$ , since

$$\begin{aligned} & \frac{1}{\sqrt{N}} \sum_{i=1}^N (\hat{u}_{it} u_{it} - \hat{u}_{it+1} u_{it+1}) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N (u_{it}^2 - u_{it+1}^2) - \frac{1}{\sqrt{N}} \sum_{i=1}^N (u_{it} x_{it} - u_{it+1} x_{it+1}) \sqrt{N}(\hat{\beta} - \beta), \end{aligned}$$

if  $E(u_{it} x_{it}) = 0$ ,  $t = 1, \dots, T$ ,  $c_i$  can be replaced by  $\hat{c}_i$  without affecting the limiting distribution of the estimators. In summary, the linear GMM estimators using  $\hat{b}_{1i}$ ,  $\hat{b}_{2i}$  and  $\hat{c}_i$  as instrumental variables are asymptotically identical to the nonlinear GMM estimators if

$$E(b_{2i}' x_i) = 0, \quad E(b_{1i}' x_i) = 0 \quad \text{and} \quad E(c_i' x_i) = 0, \quad (2.9)$$

respectively. Further, if (2.9) holds, it is valid to replace  $y_{it}$  for  $u_{it}$ ,  $t = 1, \dots, T$  in  $b_{1i}$ ,  $b_{2i}$  and  $c_i$ .

Note that the condition (2.9) holds for each column in  $b_{1i}$ ,  $b_{2i}$  and  $c_i$ , and thus should apply to any covariance restrictions that are subsets of the scalar covariance restriction.

If  $\hat{b}_{1i}$ ,  $\hat{b}_{2i}$  and  $\hat{c}_i$  are used as instruments when (2.9) does not hold, the usual standard errors are not consistent and adjustment terms should be added (for a general treatment, see Newey and McFadden, 1993, Section 6).

We end this sub-section by stating the general redundancy condition (2.4) more explicitly for the current problem. The additional moment conditions  $E(b_{1i}'u_i) = 0$  are redundant, given the initial moment conditions  $E(w_i'u_i) = 0$ , iff

$$E[(b_{1i} + b_{2i})'x_i] = E(b_{1i}'u_i u_i' w_i) [E(w_i' u_i u_i' w_i)]^{-1} E(w_i' x_i). \quad (2.10a)$$

In the remainder of this chapter we are interested in checking this condition for various choices of  $w_i$ . Now consider using the orthogonality conditions from (2.6). Given the initial moment conditions  $E(w_i'u_i) = 0$ , the orthogonality conditions  $E(c_i'u_i) = 0$  are redundant iff

$$2E(c_i'x_i) = E(c_i'u_i u_i' w_i) [E(w_i' u_i u_i' w_i)]^{-1} E(w_i' x_i). \quad (2.10b)$$

### 2.3. Random Effects Covariance and the Asymptotic Variance of GMM

The random effects error structure allows the time-constant error component, thus in model (2.1)

$$u_{it} = \phi_i + \epsilon_{it}, \quad t = 1, \dots, T. \quad (2.11)$$

where  $\phi_i$  is the time-constant error.

Instead of Assumption 2.3 that  $\Sigma$  is a scalar covariance matrix, we assume the random effects covariance:

ASSUMPTION 2.4:  $\Sigma = \sigma_\epsilon^2 I_T + \sigma_\phi^2 \mathbf{e}_T \mathbf{e}_T'$ .

where  $\sigma_\phi^2 = E(\phi_i^2)$ ,  $\sigma_\epsilon^2 = E(\epsilon_{it}^2)$ ,  $t = 1, \dots, T$ , and  $I_T$  and  $\mathbf{e}_T$  are  $T \times T$  identity matrix and  $T \times 1$  vector of ones, respectively.

Assumption 2.4 implies that the idiosyncratic errors  $(\epsilon_{i1}, \dots, \epsilon_{iT})$  are uncorrelated with each other, are uncorrelated with  $\phi_i$ , and have constant variance.

The difference between the random effects covariance matrix and the scalar covariance matrix is that the off-diagonal elements of the random effects are non-zero but the same each other. We have the  $\frac{1}{2}T(T-1)-1$  moment conditions from the restriction of equal off-diagonals, one less than for the case of scalar covariance, and the restriction of equal diagonals adds the  $T-1$  moment conditions. Together, we have the  $\frac{1}{2}T(T+1)-2$  moment conditions, which should be the case since we have only two parameters  $\sigma_\epsilon^2$  and  $\sigma_\phi^2$  among the

$\frac{1}{2}T(T+1)$  possibly distinct elements of  $\Sigma$ .

The condition of equal off diagonals of  $\Sigma$  is

$$\begin{aligned} & \mathbf{E}(\mathbf{u}_{i2}\mathbf{u}_{i1}) \\ &= \mathbf{E}(\mathbf{u}_{i3}\mathbf{u}_{i1}) = \mathbf{E}(\mathbf{u}_{i3}\mathbf{u}_{i2}) \\ & \quad \dots \\ &= \mathbf{E}(\mathbf{u}_{iT}\mathbf{u}_{i1}) = \mathbf{E}(\mathbf{u}_{iT}\mathbf{u}_{i2}) = \dots = \mathbf{E}(\mathbf{u}_{iT}\mathbf{u}_{iT-2}) = \mathbf{E}(\mathbf{u}_{iT}\mathbf{u}_{iT-1}) \end{aligned} \quad (2.12)$$

$$h_{2i} = \left[ \begin{array}{cccc|cccc} 0 & & & & \Delta u_{i2} & \Delta u_{i3} & \cdots & \Delta u_{iT-1} \\ 0 & & & & 0 & 0 & \cdots & 0 \\ \Delta u_{i1} & & & & & & & \\ & \Delta u_{i1} & \Delta u_{i2} & & & & & \\ & & & \ddots & & & & \\ & & & & \Delta u_{iT-2} & \cdots & \Delta u_{iT-1} & \\ & & & & 0 & \cdots & 0 & \end{array} \right],$$

where  $\Delta u_{it} \equiv u_{it} - u_{it+1}$ . The left and right blocks of  $h_{1i}$  and  $h_{2i}$  correspond to the moment conditions in (2.13a) and (2.13b), respectively.

To derive the asymptotic variance of the nonlinear GMM estimator, we follow the same path as we did in the last sub-section. Since

$$\frac{\partial u_{is} \Delta u_{it}}{\partial \beta'} = -(u_{is} \Delta x_{it} + x_{is} \Delta u_{it}),$$

it follows that

$$\frac{\partial h_{1i}' u_i}{\partial \beta'} = -(h_{1i} + h_{2i})' x_i,$$

and we have

$$D_2 = -E[(h_{1i} + h_{2i})' x_i].$$

The first derivatives of the moment conditions from the random effects covariance matrix is quite similar to that from the scalar covariance, which is the case because the moment conditions from the random effects covariance are some linear combination of those from the scalar covariance. Thus, the results we obtained for the scalar covariance restriction equally applies to the random effects covariance

restriction. If we use the orthogonality conditions  $E(w_i' u_i) = 0$  and  $E(h_{1i}' u_i) = 0$ , then

$$\text{Avar} \sqrt{N}(\tilde{\beta}_{\text{GMM}} - \beta) = \left[ E[x_i' w_i, x_i' (h_{1i} + h_{2i})] E \begin{bmatrix} w_i' u_i u_i' w_i & w_i' u_i u_i' h_{1i} \\ h_{1i}' u_i u_i' w_i & h_{1i}' u_i u_i' h_{1i} \end{bmatrix}^{-1} E \begin{bmatrix} w_i' x_i \\ (h_{1i} + h_{2i})' x_i \end{bmatrix} \right]^{-1} \quad (2.14)$$

The linear GMM estimator using the instrumentals  $\hat{h}_{1i}$  ( $\hat{h}_{2i}$ ) is asymptotically identical to the nonlinear GMM if  $E(h_{2i}' x_i) = 0$  ( $E(h_{1i}' x_i) = 0$ ).

Applying the redundancy condition (2.4), the moment conditions from the restriction of equal off-diagonals are redundant iff

$$E[(h_{1i} + h_{2i})' x_i] = E(h_{1i}' u_i u_i' w_i) [E(w_i' u_i u_i' w_i)]^{-1} E(w_i' x_i), \quad (2.15)$$

which is an analogy of the condition (2.10a)

### 3. STRICTLY EXOGENOUS MODELS

In this section, we find the conditions when the moment conditions from covariance restrictions are redundant in the models where the regressors are strictly exogenous to the time-varying errors. We study the scalar and the random effects covariance matrices. Before considering redundancy, we present a theorem that provides intuition for our later discussion.



### 3.1. General Results on Nonredundancy under Ideal Conditions

In the model  $y_i = x_i\beta + u_i$ , we assume:

ASSUMPTION 3.1:  $E(u_i|x_i) = 0$ ,

ASSUMPTION 3.2:  $E(u_i u_i' | x_i) = \sigma^2 I_T$ .

Assumptions 3.1 and 3.2 are "ideal" conditions. OLS is BLUE under these assumptions (along with nonsingularity of  $X'X$  matrix). Chamberlain (1987) showed that, ignoring the moment conditions from (3.1b) below, if all the instrumental variables  $w_i$  that include  $x_i$  satisfy

$$E(u_i | w_i) = 0, \quad (3.1a)$$

$$E(u_i u_i' | w_i) = \sigma^2 I_T, \quad (3.1b)$$

the optimal set of instruments is  $E(x_i | w_i) = x_i$ , and OLS is the most efficient. Condition (3.1) is stronger than Assumptions 3.1 and 3.2, and Chamberlain's result allows that there would be nonredundant instrumental variables other than  $x_i$  under Assumptions 3.1 and 3.2. We write it down more explicitly.

THEOREM 3.1: In model (2.1) under Assumptions 3.1 - 3.2, suppose there are instrumental variables  $a_i$  of dimension  $T \times q$  such that (i)  $E(a_i' u_i) = 0$  and (ii)  $E(a_i' u_i u_i' x_i) = \sigma^2 E(a_i' x_i)$ . Then GMM using the instruments  $(x_i, a_i)$  is more efficient than OLS.

PROOF: It is sufficient to show that  $a_i$  is not redundant. From (2.5), given the initial instruments  $x_i$ ,  $a_i$  are redundant iff

$$E(a_i'x_i) = E(a_i'u_iu_i'x_i) [E(x_i'u_iu_i'x_i)]^{-1} E(x_i'x_i), \quad (3.2)$$

which holds iff  $E(a_i'u_iu_i'x_i) = \sigma^2 E(a_i'x_i)$ . ■

Theorem 3.1 holds even when the errors are normally distributed and when the regressors are independent of the errors, but applies only to large samples. Generally GMM should be used to realize the efficiency gain from the additional instruments  $a_i$ .

The idea underlying Theorem 3.1 is suggested in Cragg (1983) and Chamberlain (1982). They showed that there can exist nonredundant instrumental variables in addition to the regressors in the presense of conditional heteroskedasticity of unknown form, even when all the regressors are valid instruments. Cragg's estimator is a GMM estimator with more instrumental variables that are correlated with the conditional error covariance. The efficiency gain in Chamberlain's optimal minimum distance estimator has a similar interpretation.

Note that Theorem 3.1 shows that, to be useful, the additional instrumental variables  $a_i$  do not have to be correlated with the regressors  $x_i$ . Instruments that are uncorrelated with the regressors appear frequently in the models we study subsequently.

### 3.2. Strictly Exogenous Model: Scalar Covariance

Assumption 3.1 is stronger than needed. The weakest assumption with strictly exogenous regressors is

ASSUMPTION 3.3:  $E(w_i' u_i) = 0$ , where  $w_i = I_T \otimes x_i^0$  and  $x_i^0 = (x_{i1}, \dots, x_{iT})$ .

The choice of instruments in Assumption 3.3 simply means that  $x_{it}$  is uncorrelated with  $u_{is}$ , all  $t, s = 1, \dots, T$ . In this section, we find when the moment conditions from the scalar covariance of Assumption 2.3 are redundant, given the initial instruments  $I_T \otimes x_i^0$ .

Under Assumption 3.3,  $E(b_{1i}' x_i) = E(b_{2i}' x_i) = E(c_i' x_i) = 0$ , so from (2.9) the linear GMM estimators using either  $\hat{b}_{1i}$  or  $\hat{b}_{2i}$  and  $\hat{c}_i$  as instruments has the same limiting distribution as the nonlinear GMM. Thus, we treat  $b_{1i}$ ,  $b_{2i}$  and  $c_i$  as being available. Applying the condition (2.10a) to see if  $b_{1i}$  is redundant or not, given the initial instruments  $I_T \otimes x_i^0$ , we get

$$E[b_{1i}' u_i u_i' (I_T \otimes x_i^0)] [E(u_i u_i' \otimes x_i^0' x_i^0)]^{-1} E[(I_T \otimes x_i^0)' x_i] = 0. \quad (3.3)$$

If there exists no conditional heteroskedasticity, thus if the assumption

ASSUMPTION 3.4:  $E(w_i' u_i u_i' w_i) = E(w_i' \Sigma w_i)$

holds, then combined with the scalar covariance assumption  $\Sigma$

$= \sigma^2 I_T$ , the weighting matrix which is in the middle of the LHS of the equation (3.3) becomes  $\sigma^2 I_T \otimes E(x_i^0 x_i^0)$ . Thus, OLS is efficient and only  $k$  instruments  $x_i$  are useful among  $T^2 k$  instruments  $I_T \otimes x_i^0$  since  $P_{(X^0 \otimes I_T)} X = X$ . Otherwise all the instruments  $I_T \otimes x_i^0$  are useful (Chapter 2). Under Assumption 3.4, the equation (3.3) becomes

$$E(b_{1i}' u_i u_i' x_i) = 0. \quad (3.4a)$$

This is no more than the redundancy condition of  $b_{1i}$  on the initial instruments  $x_i$ .  $E(b_{1i}' u_i u_i' x_i)$  contains the elements

$$\sum_{\tau=1}^T E(u_{is} u_{it} u_{i\tau} x_{i\tau}), \quad s > t = 1, \dots, T-1.$$

A sufficient condition for (3.4a) is

$$E(u_{is} u_{it}^2 x_{it}) = E(u_{is} u_{it}^2 x_{is}) = E(u_{is} u_{it} u_{i\tau} x_{i\tau}) = 0, \quad s \neq t \neq \tau. \quad (3.4b)$$

There are other situations where (3.4a) holds, but they are not very intuitive.

If Assumption 3.4 is violated, condition (3.3) holds if

$$E[b_{1i}' u_i u_i' (I_T \otimes x_i^0)] = 0. \quad (3.5a)$$

(3.5a) is stronger than (3.3), but it would be very unusual if (3.3) holds without (3.5a). Because  $E[b_{1i}' u_i u_i' (I_T \otimes x_i^0)]$  has elements

$$E(u_{is} u_{it} u_{i\tau} x_i^0), \quad s > t = 1, \dots, T-1 \text{ and } \tau = 1, \dots, T,$$

condition (3.3) holds if

$$E(u_{is}^2 u_{it} x_i^0) = 0 \text{ and } E(u_{is} u_{it} u_{ir} x_i^0) = 0, s \neq t \neq r. \quad (3.5b)$$

Condition (3.5b) is stronger than (3.4b), but as long as the strong exogeneity assumption holds the two conditions are quite similar. A sufficient condition that ensures (3.5b) is  $E(u_{it} u_{is} | x_i^0, u_{ir}) = 0$ ,  $r \neq s, t$ , (including  $s = t$ ), which is met if the errors are independent over time. This condition particularly rules out ARCH presentation in panel data context. A constructive way to understand what the condition (3.5b) represents is that for  $b_{1i}$  to be redundant, there should be no conditional heteroskedasticity when  $b_{1i}$  is used as instruments. Otherwise, though being uncorrelated with the regressors,  $b_{1i}$  becomes useful by explaining the second moments of errors. Thus, the reason why  $b_{1i}$  can be useful is the same as why additional instrumental variables  $a_i$  of the previous subsection can be useful even when they are uncorrelated with the regressors.

We follow the same procedure to find the redundancy condition for  $c_i$ . From (2.10b), when Assumption 3.4 holds  $c_i$  is redundant iff  $E(c_i' u_i u_i' x_i) = 0$ , or equivalently

$$\sum_{t=1}^T E(u_{is}^2 - u_{is+1}^2) u_{it} x_{it} = 0, s = 1, \dots, T-1. \quad (3.6a)$$

Given the condition that  $b_{1i}$  is redundant, this condition becomes

$$E(u_{it}^3 x_{it} - u_{it+1}^3 x_{it+1}) = 0, t = 1, \dots, T-1. \quad (3.6b)$$

A sufficient condition for (3.6b) is

$$E(u_{it}^3 x_{it}) = 0, \quad t = 1, \dots, T, \quad (3.6c)$$

that demands the symmetry of the error distribution as a minimum.

If Assumption 3.4 does not hold,  $c_i$  is redundant if

$$E[c_i' u_i u_i' (I_T \otimes x_i^0)] = 0, \quad (3.7a)$$

or,

$$E[(u_{it}^2 - u_{it+1}^2) u_{i\tau} x_i^0] = 0, \quad t = 1, \dots, T-1, \quad \tau = 1, \dots, T. \quad (3.7b)$$

Given (3.5a), condition (3.7b) becomes

$$E(u_{it}^3 x_i^0) = 0, \quad t = 1, \dots, T-1. \quad (3.7c)$$

It is interesting to note that the conditions (3.5) and (3.7) are usually assumed in the literature concerning about covariance restrictions in simultaneous equations models, namely:

ASSUMPTION 3.5:  $E[(u_i u_i' \otimes u_i) w_i] = 0$ .

Examples are Hausman, Newey and Taylor (1987) and Arellano (1989). In particular, standard errors from augmented 3SLS estimator proposed by Hausman, Newey and Taylor are not consistent when Assumption 3.5 is violated (Section 4 of their paper). To show why Assumption 3.5 ensures the conditions (3.5) and (3.7), we define selection matrices  $S_j$  of dimension  $[\frac{1}{2}T(T+1)-1] \times T^2$  such that  $\Sigma = \sigma^2 I_T \otimes S_j \text{vec}(\Sigma) = 0$

(Magnus and Neudecker, 1980). Thus, the matrix  $S_j$  selects the elements in  $\text{vec}(\Sigma)$ . In our model,  $S_j[\text{vec}(u_i u_i')] = (b_{ji}, c_i)' u_i$ , for  $j = 1, 2$ . Since  $\text{vec}(u_i u_i') = (I_T \otimes u_i) u_i$ , the conditions (3.5a) and (3.7a) are equally expressed as  $S_j E[(I_T \otimes u_i) u_i u_i' (I_T \otimes x_i^0)] = S_j E[(u_i u_i' \otimes u_i) (I_T \otimes x_i^0)] = 0$  from Assumption 3.5. Thus, Assumption 3.5 is sufficient to ensure the moment conditions from covariance restrictions are redundant when the regressors are strictly exogenous.

Allowing for individual effects would be a primary reason why people use panel data models, and the model we studied in this section rarely appears in panel data applications. We now turn to more widely applicable models.

### 3.3. Strictly Exogenous Model: Random Effects Covariance

In this section, we will consider the covariance restrictions in the popular random effects model. Thus, the errors are composites of time-constant  $\phi_i$  and time-varying  $\epsilon_{it}$ , and all the regressors are exogenous to  $\phi_i$  as well as to  $\epsilon_{it}$ ,  $t = 1, \dots, T$ . We have the initial instruments  $I_T \otimes x_i^0$  and the set of moment conditions from the random effects covariance is a subset of the moment conditions from the scalar covariance. Thus, the moment conditions from the random effects covariance matrix are redundant as long as the higher conditional moment conditions of the errors such as Assumption 3.5 are satisfied.

We focus on the higher moment conditions on the time-constant error  $\phi_i$  that usually is thought to be caused by omitting some unobserved variables that are invariant over the time periods in question.

The additional moment conditions are  $E(h_{1i}'u_i) = 0$  or  $E(h_{2i}'u_i) = 0$ , and  $E(c_i'u_i) = 0$ . Since  $E(h_{1i}'x_i) = E(h_{2i}'x_i) = E(c_i'x_i) = 0$ , we handle  $h_{1i}$ ,  $h_{2i}$  and  $c_i$  like they are available. From the condition (2.15),  $h_{1i}$  is redundant if

$$E[h_{1i}'u_i u_i' (I_T \otimes x_i^0)] = 0. \quad (3.9a)$$

and which is equally stated as,

$$(i) E(u_{is} \Delta u_{it} u_{ir}' x_i^0) = 0, \text{ and } (ii) E(u_{i1} \Delta u_{i\ell} u_{ir}' x_i^0) = 0, \\ s-1 > t = 1, \dots, T-1, \tau = 1, \dots, T, \text{ and } \ell = 2, \dots, T-1. \quad (3.9b)$$

The first and the second conditions of (3.9b) correspond to the left and the right blocks of  $h_{1i}$ . Due to homogeneity of the moment conditions, not much is lost by examining only the (1,1) element of  $E[h_{1i}'u_i u_i' (I_T \otimes x_i^0)]$ , so the condition  $E(u_{i3} \Delta u_{i1} u_{i1}' x_i^0) = 0$ . Then, we have

$$E\{[\phi_i^2(\epsilon_{i1} - \epsilon_{i2}) + \phi_i(\epsilon_{i1}^2 - \epsilon_{i1}\epsilon_{i2} + \epsilon_{i1}\epsilon_{i3} - \epsilon_{i2}\epsilon_{i3}) + (\epsilon_{i1}^2\epsilon_{i3} - \epsilon_{i1}\epsilon_{i2}\epsilon_{i3})]x_i^0\} = 0. \quad (3.9c)$$

The condition (3.9c) is met if (i)  $E(\epsilon_i | x_i, \phi_i) = 0$ , (ii)  $E(\epsilon_{it}\epsilon_{is} | x_i, \phi_i, \epsilon_{ir}) = E(\epsilon_{it}\epsilon_{is})$ ,  $s, t \neq r$  (including  $s = t$ ).

For the moment conditions from equal diagonals, we apply the condition (3.7b) that  $E[(u_{it}^2 - u_{it+1}^2)u_{ir}' x_i^0] = 0$ ,  $t = 1, \dots, T-1$ , and  $r = 1, \dots, T$ . We consider the case when  $t = r$



= 1, then the condition becomes

$$E\{[2(\epsilon_{i1}-\epsilon_{i2})\phi_i^2 + (3\epsilon_{i1}^2-\epsilon_{i1}^2-2\epsilon_{i1}\epsilon_{i2})\phi_i + \epsilon_{i1}^3 - \epsilon_{i1}\epsilon_{i2}^2]x_i^0\} = 0. \quad (3.10)$$

Equality of this equation holds if we add the condition

$E(\epsilon_{i1}^3 x_i^0) = 0$  to (3.9). One notable thing is that the condition (3.7c) that  $E(u_{it}^3 x_i^0) = 0$  does not necessarily apply for this case. Note that in (3.10)  $\phi_i^3$  is differenced away.

#### 3.4. Strictly Exogenous Model: Fixed Effects Type

We allow for arbitrary correlations between the regressors and the time-constant error  $\phi_i$ . The usual assumption is

ASSUMPTION 3.6:  $E(x_i \otimes \epsilon_i) = 0$ .

Note that we still are working on the random effects covariance matrix. Assumption 3.6 implies a set of instrumental variables  $L \otimes x_i^0$ , where  $L$  is the  $T \times (T-1)$  differencing operator (Chapter 2; Ahn and Schmidt). We assume that there are no time-constant variables in  $x_{it}$  to ensure that  $\beta$  and  $\Sigma$  are identified.

Ahn and Schmidt showed that the moment conditions from the random effects covariance restrictions are redundant under Assumption 3.6. Their reasoning is that the moment conditions from the random effects covariance restrictions add information only through the regression corresponding to

the instruments that are in the space spanned by  $L$  where the regression reaches the GLS efficiency. Their finding is plausible under certain set of assumptions, that we will detail.

$E(h_{2i}'x_i) = 0$  under Assumption 3.6, thus the linear GMM using  $\hat{h}_{1i}$  as instruments is asymptotically identical to the nonlinear GMM, but using the instruments  $\hat{h}_{2i}$  would in general lead to less efficient estimator than the nonlinear GMM since  $E(h_{1i}'x_i) \neq 0$ . Thus,  $h_{1i}$  will be considered as being available. Note that  $h_{1i}$  is in the space spanned  $L$  in the sense that  $Q_T h_{1i} = h_{1i}$ , where  $Q_T = L(L'L)^{-1}L'$ .

If Assumption 3.4 of no conditional heteroskedasticity is met, so that if we have,

$$E[(L \otimes x_i^0)' u_i u_i' (L \otimes x_i^0)] = \sigma_e^2 L' L \otimes E(x_i^0 x_i^0)',$$

the only relevant instrumental variables are  $Q_T x_i$ , and OLS on the demeaned equations is efficient. It does not matter whether we use  $Q_T x_i$  or  $L \otimes x_i^0$  as the initial instruments since  $P_{(X^0 \otimes L)} X = QX$ , where  $Q = I_N \otimes Q_T$ . Applying the redundancy condition (2.15), we have

$$\sigma_e^2 E(h_{1i}' x_i) = E(h_{1i}' u_i u_i' Q_T x_i). \quad (3.11)$$

To simplify this condition, we will consider an example when  $T = 3$ . Then,  $h_{1i}^L = (u_{i3} \ -u_{i3} \ 0)'$  and  $h_{1i}^R = (0 \ u_{i1} \ -u_{i1})'$ , that are pretty much of the same sort in a sense that one is redundant if the other is. We will consider  $h_{1i}^L$  only. Since

$$E(h_{i1}^1, x_i) = E[\phi_i(x_{i1} - x_{i2})]$$

and

$$E(h_{i1}^1, u_i u_i' Q_T x_i) = E[(\epsilon_{i3} + \phi_i)(\epsilon_{i1} - \epsilon_{i2}) \sum_{r=1}^3 \epsilon_{ir}(x_{ir} - \bar{x}_i)],$$

the equality in (3.11) holds under

**ASSUMPTION 3.7:**  $E(\epsilon_{it}\epsilon_{is} | x_i, \phi_i, \epsilon_{ir}) = E(\epsilon_{it}\epsilon_{is}), s, t \neq r.$

Note that Assumption 3.7 includes the case when  $t = s$ .

For the restriction of equal diagonals in  $\Sigma$ , we have the orthogonality conditions  $E(c_i' u_i) = 0$ , and the redundancy condition becomes

$$2\sigma_\epsilon^2 E(c_i' x_i) = E(c_i' u_i u_i' Q_T x_i). \quad (3.12)$$

We will consider the simplest case when  $T = 2$ . Then, the LHS of (3.12) becomes  $2\sigma_\epsilon^2 E[\phi_i(x_{i1} - x_{i2})]$ . For the RHS of (3.11), we have

$$E[(\epsilon_{i1}^2 - \epsilon_{i2}^2) \sum_{r=1}^2 \epsilon_{ir}(x_{ir} - \bar{x}_i)] + E[2\phi_i(\epsilon_{i1} - \epsilon_{i2}) \sum_{r=1}^2 \epsilon_{ir}(x_{ir} - \bar{x}_i)],$$

$t = 1, \dots, T.$

For both terms to be the same, it generally requires

**ASSUMPTION 3.8:**  $E(\epsilon_{it}^3 | x_i) = 0, t = 1, \dots, T,$

as well as Assumption 3.7.

Recall that Assumption 3.7 and 3.8 are quite similar to the conditions we derived for the random effects model.

They also are quite similar to the redundancy conditions in

the equations (3.5) and (3.7) of the moment conditions from the scalar covariance matrix if we consider  $\phi_i$  as a regressor. Allowing for correlation between the individual effects and the regressors does not alter appreciably the redundancy condition for the moment conditions from the random effects covariance restrictions. The intuition provided by Ahn and Schmidt (1992) is plausible, but it is interesting to note that there are more efficient estimators than GLS when certain conditional third moments conditions on errors are violated.

Throughout this section we have studied redundancy conditions of covariance restrictions in the models where the GLS efficiency is reached. In the following, we turn to the models where the regressors are weakly exogenous to the errors.

#### 4. WEAKLY EXOGENOUS MODELS

As we noted in Chapter 3, dynamic models and the rational expectations models are typical weakly exogenous models. There is a growing concern that many regressors in standard panel data models would be only weakly exogenous to the time-varying errors. We will not work on dynamic models explicitly, since there is a large body of previous work and much of which studied covariance restrictions. For listings of references, see Chapter 3. We study the models under the

assumptions that are usual in the rational expectations models. But, our results apply to standard models where some of regressors are weakly exogenous and also to general dynamic models. There, in general, is no point to argue whether covariance restrictions are useful in dynamic models, because covariance restrictions coincide with the instruments for the lagged dependent regressor. But, there at least is one model (probably the only model) that draws our interest, which we study first.

#### 4.1. Weakly Exogenous Model: Diagonal Covariance

We first study the weakly exogenous panel data model with sequential conditional moment restrictions of the type in Chamberlain (1992), but with no individual effects. The model is a typical rational expectations model that appears in panel data applications. The diagonal covariance matrix rarely appears in standard panel data models. Nevertheless, it frequently is assumed in the rational expectations models as the hypothesis itself implies. Further, many tests failed to reject the null of no individual effects (Keane and Runkle, 1990; 1992; Runkle, 1991). Other important models that can have the diagonal covariance matrix are dynamic models. Most of the previous studies on dynamic models concerned the random effects covariance. However, it has been observed that allowing for rich dynamics diminishes

the importance of individual effects (e.g. Holtz-Eakin, 1988).

We continue to consider model (2.1), and assume

ASSUMPTION 4.1:  $E(u_{it} | x_{it}^0, u_{it-1}^0) = 0$ ,  $t = 1, \dots, T$ ,

where  $x_{it}^0 = (x_{i1}, \dots, x_{it})$  and  $u_{it}^0 = (u_{i1}, \dots, u_{it})$ ,  $t = 1, \dots, T$ .

Assumption 4.1 implies many instrumental variables.

Utilizing every moment condition is not feasible. We restrict our attention to the second moment conditions of  $(x_i, u_i)$ , thus we have

$$E(u_{it} x_{it}^0) = 0, \quad t = 1, \dots, T, \quad (4.1)$$

and

$$E(u_{it} u_{is}) = 0, \quad s \neq t. \quad (4.2)$$

(4.1) implies the set of instruments that appears in panel data literatures (e.g. Schmidt, Ahn and Wyhowski, 1992), and (4.2) is the covariance restrictions that the off-diagonals of  $\Sigma$  are zero. It is usual to assume

ASSUMPTION 4.2:  $E(u_{it}^2 | x_{it}^0, u_{it-1}^0) = \sigma_t^2$ ,  $t = 1, \dots, T$ .

This excludes conditional heteroskedasticity. Under Assumptions 4.1 and 4.2, GMM using the instrumental variables  $x_i^* = \text{diag}(x_{i1}, \dots, x_{iT})$  is asymptotically identical to GLS, and no other instruments are useful ignoring the higher moment conditions on the errors (Chapter 2). That is to say, among  $tk$  instruments for the  $t$ -th period equation,

only  $k$  instruments  $x_{it}$  are useful and others are redundant, and any other functions of  $x_i$  are also redundant.

We find whether the moment conditions (4.2) are redundant under Assumptions 4.1 and 4.2, given the initial instruments  $x_i^*$ . Note that  $E(b_{1i}'x_i) = 0$ , but  $E(b_{2i}'x_i) \neq 0$ . Thus, the linear GMM using the instruments  $\hat{b}_{2i}$  is identical to the nonlinear GMM asymptotically. Thus, we treat  $b_{2i}$  like it is known.

**THEOREM 4.1:** In model (2.1) under Assumptions 4.1 - 4.2, the orthogonality conditions in (4.2) are redundant, given the initial instrumental variables  $x_i^*$ .

**PROOF:** We apply the redundancy condition (2.10a), which becomes

$$E(b_{2i}'x_i) = E(b_{2i}'u_i u_i' x_i^*) [E(x_i^* u_i u_i' x_i^*)]^{-1} E(x_i^* x_i) \quad (4.3)$$

$$E(b_{2i}'u_i u_i' x_i^*) = E(u_{it} u_{is} u_{ir} x_{ir}^*), \quad s > t = 1, \dots, T-1, \quad r = 1, \dots, T.$$

When  $r = s$ ,  $E(u_{it} u_{is} u_{ir} x_{ir}^*) = E(u_{it} u_{is}^2 x_{is}) = E[u_{it} E(u_{is}^2 | u_{it}, x_{is}) x_{is}] = \sigma_s^2 E(u_{it} x_{is}) \neq 0$ . When  $r = t$ ,  $E(u_{it} u_{is} u_{ir} x_{ir}^*) = E(u_{it}^2 u_{is} x_{it}) = E[u_{it}^2 E(u_{is} | u_{it}, x_{it}) x_{it}] = 0$  from Assumption 4.1. For  $r \neq s, t$ ,  $E(u_{it} u_{is} u_{ir} x_{ir}^*) = 0$ . Thus, non-zero elements in  $E(b_{2i}'u_i u_i' x_i^*)$  are  $\sigma_s^2 E(u_{it} x_{is})$  for  $s > t$ . For simplicity, we show this when  $T = 3$ . Then,

$$E(b_{2i}'u_i u_i' x_i^*) = E \begin{bmatrix} 0 & \sigma_2^2 u_{i1} x_{i2} & 0 \\ 0 & 0 & \sigma_3^2 u_{i1} x_{i3} \\ 0 & 0 & \sigma_3^2 u_{i2} x_{i3} \end{bmatrix},$$

and  $[E(x_i^* u_i u_i' x_i^*)]^{-1} E(x_i^* x_i) = (\sigma_1^{-2} \sigma_2^{-2} \sigma_3^{-2})' \otimes I_k$ . Thus, the RHS

of the condition (4.3) becomes  $E(u_{it}x_{is})$  for  $t < s = 2, 3$ , which is no more than the LHS of the condition (4.3). The argument is more tedious for general  $T$ . ■

Even if the regressors are only weakly exogenous, GLS is consistent since  $\Sigma$  is diagonal, and efficient under Assumption 4.2. This result also applies to dynamic models with one or more lagged dependent variables. As an example, suppose a simple AR(1) dynamic model where  $y_{it-1}$  is the only regressor. Then, GLS is equivalent to 3SLS using the instruments  $\text{diag}(y_{i0}, \dots, y_{iT-1})$ . The moment conditions to be used are  $E(y_{t-1}u_t) = 0$ ,  $t = 1, \dots, T$ , which are equivalent to  $E(u_{it-1}u_{it}) = 0$ ,  $t = 1, \dots, T-1$ , and  $E(y_{i0}u_{i1}) = 0$ . Thus, among  $\frac{1}{2}T(T-1)$  zero off-diagonal restrictions only  $(T-1)$  restrictions that correspond to the second moments between regressors and errors are useful and rest are redundant. Similar arguments apply when more than one lagged dependent variable appear as regressors.

Theorem 4.1 depends heavily on the diagonality of  $\Sigma$ . If  $\Sigma$  is not diagonal and the regressors are weakly exogenous, GLS is not consistent (Schmidt, 1990), and all the covariance restrictions would be useful in general. We study a special case of this in the next section.



#### 4.2. Weakly Exogenous Model: Random Effects Covariance

We now combine the individual effects and the weak exogeneity of the regressors to the idiosyncratic errors. From Ahn and Schmidt, and section 3.3, we know that the moment conditions from the random effects error covariance are useful unless GLS efficiency is reached in the space spanned by  $L$ . In another words, for the orthogonality conditions from the random effect covariance to be redundant, GLS in the differenced equations should be at least consistent or equivalently the instruments  $L\otimes x_i^0$  should be valid in the original equations before differencing. In this sense, the model with a diagonal covariance matrix we studied in the last sub-section is an exception, where the instruments  $L\otimes x_i^0$  are not valid, but the GLS efficiency (not in the space spanned by  $L$ ) is reached because the covariance matrix is diagonal and so the optimal weighting matrix becomes block diagonal.

We argue, throughout this section, that whenever the GLS efficiency is not reached, the moment conditions from covariance restrictions are useful. There probably is a nice and simple proof for this statement, but we could not provide it. Thus, we only provide a heuristic discussion through a couple of examples that look useful in applications.

The model we deal with in this section is (2.1) with

the random effects error structure. Once we allow for the time-constant error  $\phi_i$ , Assumption 4.1 is not plausible. Instead, we assume

ASSUMPTION 4.3:  $E(\epsilon_{it} | x_{it}^0, \epsilon_{it-1}^0, \phi_i) = 0, t = 1, \dots, T.$

Assumption 4.3 is standard in the rational expectation models that allow for arbitrary correlations between the regressors and the time-constant unobservable  $\phi_i$ .

Like we did in the previous subsection, we will consider only the orthogonality conditions from the second moments of  $(x_i, u_i)$ . Then, we get a set of instruments

$$w_i = \begin{bmatrix} x_{i1}^0 \\ -x_{i1}^0 & x_{i2}^0 \\ & -x_{i2}^0 & \ddots \\ & & & x_{iT-1}^0 \\ & & & & -x_{iT-1}^0 \end{bmatrix}.$$

For more details, see Chapter 3 or Schmidt, Ahn and Wyhowsky (1992). Also, Assumption 4.3 implies that the off-diagonals of  $\Sigma$  are the same each other.

We add no conditional heteroskedasticity assumption

ASSUMPTION 4.4:  $E(\epsilon_{it}^2 | x_{it}^0, \epsilon_{it-1}^0, \phi_i) = \sigma_\epsilon^2, t = 1, \dots, T.$

This assumption is stronger than usual. It generally is allowed that  $E(\epsilon_{it}^2) \neq E(\epsilon_{is}^2), t \neq s$ . Assuming that they are the same, along with Assumption 4.3, leads to the random

effects covariance matrix. But, as will shortly become clear, the restriction of equal diagonals of  $\Sigma$  does not alter the results we obtain. The moment conditions from equal diagonals are not our concern anyway; our interest centers on whether the equal off-diagonal restriction of covariance matrix are useful under Assumption 4.3 and 4.4.

The orthogonality conditions from equal off-diagonals of the random effects covariance matrix are  $E(h_{1i}'u_i) = 0$ . To simplify our discussion, we will consider when  $T = 3$ . We focus on the moment condition  $E(h_{1i}'u_i) = 0$ , and find whether it is redundant. We lose no generality from this simplification, since the moment conditions  $E(h_{1i}'u_i) = 0$  and  $E(h_{2i}'u_i) = 0$  represent the same sort of moment conditions of equal off-diagonals of covariance matrix. The moment condition  $E(h_{1i}'u_i) = 0$ , given the initial instruments  $w_i$ , is redundant iff

$$E[(h_{1i}' + h_{2i}')x_i] = E(h_{1i}'u_i u_i' w_i) [E(w_i' u_i u_i' w_i)]^{-1} E(w_i' x_i). \quad (4.4)$$

Recall that  $h_{1i}' = (u_{i3} \ -u_{i3} \ 0)'$  and  $h_{2i}' = (0 \ 0 \ \Delta u_{i1})'$ . Thus, the LHS becomes

$$E[\phi_i(x_{i1} - x_{i2})] + (\epsilon_{i1} - \epsilon_{i2})x_{i3}.$$

Straightforward algebra using the matrix inverse lemma and the identity  $E(x_{i1}^0 x_{i2}^0) [E(x_{i2}^0 x_{i2}^0)]^{-1} E(x_{i2}^0 x_{i1}^0) = E(x_{i1}^0 x_{i1}^0)$  shows that the RHS of the condition (4.4) is

$$E(\phi_i x_{i1}) - \frac{1}{2}E(\phi_i x_{i2}) - \frac{3}{2}E(\phi_i x_{i1}) [E(x_{i1}' x_{i1})]^{-1} E(x_{i1}' x_{i2}) \\ + \frac{1}{2}E(\phi_i x_{i1}) [E(x_{i1}' x_{i1})]^{-1} E(x_{i1}' x_{i3}) + \frac{1}{2}E(\phi_i x_{i2}^0) [E(x_{i2}^0' x_{i2}^0)]^{-1} E(x_{i2}^0' x_{i3})$$

Equality between two terms do not hold unless  $E(\phi_i x_i) = 0$ , and  $E(\epsilon_{i1} x_{i3}) = E(\epsilon_{i2} x_{i3}) = 0$ , that make both terms zero.

Thus, the moment condition  $E(h_{i1}^L u_i) = 0$  is not redundant under Assumption 4.3 and 4.4.

Nonlinear optimization is necessary in general to implement the moment condition  $E(h_{i1}^L u_i) = 0$ , since both  $E(h_{i1}^L x_i)$  and  $E(h_{i2}^L x_i)$  are non-zero. However,  $E(h_{i2}^R x_i) = 0$  so that the linear GMM estimator using  $\hat{h}_{i1}^R$  as instrumental variables is asymptotically equivalent to the nonlinear GMM that uses the same set of moment conditions. Note that we construct the  $\frac{1}{2}T(T-1)-1$  instruments  $h_{i1}^L$  (or  $h_{i2}^L$ ) first by equalizing the elements in each row of (2.13a), and then the  $T-1$  instruments  $h_{i1}^R$  (or  $h_{i2}^R$ ) are constructed by equalizing the elements in columns of (2.13a). But, it is not hard to see that the column dimensions of  $h_{i1}^L$  and  $h_{i1}^R$  (or  $h_{i2}^L$  and  $h_{i2}^R$ ) are reversed if we construct the orthogonality conditions by equalizing all the elements in each column of (2.13a) first. Thus,  $\frac{1}{2}T(T-1)-1$  moment conditions can be implemented without numerical optimization.

One would still be doubtful that efficiency gains from the random effects covariance restrictions would come from allowing correlations between the regressors and the time-constant error  $\phi_i$  rather than from covariance structure. To

provide a firm idea that covariance structure plays an important role for redundancy of the moment conditions from covariance restrictions, we consider one more model under a quite strong set of assumptions, which would not be much of practical use.

We add assumptions

ASSUMPTION 4.5:  $E(\phi_i | x_i, \epsilon_i) = 0$ ,

ASSUMPTION 4.6:  $E(\phi_i^2 | x_i, \epsilon_i) = \sigma_\phi^2$ .

Assumption 4.5 is like the random effects assumption in strictly exogenous model, Assumption 4.6 is an assumption of no conditional heteroskedasticity. Obviously, Assumptions 4.5 and 4.6 exclude dynamic models. Now, the only difference between the model under Assumption 4.3 - 4.6 that we will consider and the model under 4.1 - 4.2 that we considered in the last subsection is in the covariance structure. The initial set of instruments is the same, that is  $w_i = \text{diag}(x_{i1}^0, x_{i2}^0, \dots, x_{iT}^0)$ .

For simplicity, we again consider the case when  $T = 3$  and the moment condition  $E(h_{i1}^l u_i) = 0$ . The condition (4.4) is the redundancy condition, given the initial instruments  $w_i = \text{diag}(x_{i1}^0, x_{i2}^0, \dots, x_{iT}^0)$ . The LHS of (4.4) is

$$E[(h_{i1}^l + h_{i2}^l)' x_i] = E[(\epsilon_{i1} - \epsilon_{i2}) x_{i3}].$$

For the RHS of (4.4), we have

$$E(h_i' u_i u_i' w_i) = \begin{bmatrix} 0 & 0 & (\sigma_\epsilon^2 + \sigma_\phi^2) E\{(\epsilon_{i1} - \epsilon_{i2}) x_{i3}^0\} \end{bmatrix},$$

$$E(w_i' x_i) = E(x_{i1}^0 x_{i1} \quad x_{i2}^0 x_{i2} \quad x_{i3}^0 x_{i3})'$$

and

$$\begin{aligned} E(w_i' u_i u_i' w_i) &= \sigma_\epsilon^2 E(w_i' w_i) + \sigma_\phi^2 E(w_i' e_T e_T' w_i) \\ &= E \begin{bmatrix} (\sigma_\epsilon^2 + \sigma_\phi^2) x_{i1}^0 x_{i1} & \sigma_\phi^2 x_{i1}^0 x_{i2}^0 & \sigma_\phi^2 x_{i1}^0 x_{i3}^0 \\ (\sigma_\epsilon^2 + \sigma_\phi^2) x_{i2}^0 x_{i2} & \sigma_\phi^2 x_{i2}^0 x_{i3}^0 & \\ (\sigma_\epsilon^2 + \sigma_\phi^2) x_{i3}^0 x_{i3} & & \end{bmatrix}. \end{aligned}$$

Though it is onerous to invert  $E(w_i' u_i u_i' w_i)$ , it is not hard to see that equality in (4.4) holds when the off-blocks of  $E(w_i' u_i u_i' w_i)$  are zero, as  $P_{(X_3^0)} X_t = X_t$ ,  $t = 1, 2, 3$ , the case we considered in the last section. Thus, covariance restrictions become useful by allowing the time-constant error.

Note again that the only difference between this and the model we studied in the last section is in the structure of the covariance matrix. Given weak exogeneity of the regressors, appearance of the time-constant error breaks the block diagonality of the optimal weighting matrix and make GLS inconsistent.

In many of the rational expectations models, MA(1) serial correlation of the errors has been detected (e.g., Keane and Runkle, 1990; 1992; Runkle, 1991). As we discussed in Chapter 3, we do not have to shrink the set of instruments if the serial correlation is caused by the time lag for observing past shocks. Then, the set of instruments in those models and in the model we are dealing with are the

same except those from covariance restrictions. However, GLS anyway is not consistent under MA(1) error structure. Thus, we conjecture that covariance restrictions are useful in those models. Also, numerical optimization will not be necessary to realize the efficiency gains from covariance restrictions in those models.

## 5. CONCLUSION

GMM provides a new aspect of instrumental variables that, to be useful, they do not have to be correlated with the regressors as long as they are correlated with the error squared sequence  $\{u_i^2\}$ . It is interesting to ask how much we can improve estimators from using those instruments. Finding that kind of instrumental variables outside of the models we are interested in would be unusual, but as sections 3 and 4 show, residuals generated from initial consistent estimators could play that role. And generally they are useful when error distribution is not symmetric.

From section 4, we know that diagonality of covariance matrix is crucial for GLS to be consistent and efficient when the regressors are weakly exogenous. However, there are models, though not of practical importance, where GLS is consistent but covariance restrictions are always useful. Suppose a model where the regressors are only currently uncorrelated with the errors (known as contemporaneous

uncorrelated model) and  $\Sigma$  is diagonal, then GLS is consistent. But, conditional heteroskedasticity is guaranteed in this case, and more instruments generally are useful if exist, and covariance restrictions also are useful. This arguments are directly related with Wooldridge (1993) and Chapter 3.

We only considered redundancy of the moment conditions from the second moments of errors and it proves that the conditional third moment conditions of errors are crucial. If covariance restrictions are not useful since the conditional third moment conditions of the errors are met. Then, those conditional third moment conditions of the errors in turn become a new set of moment conditions, and they, to be redundant, would require certain set of the conditional fourth moment conditions of the errors, and so on. Though we do not pursue the redundancy of higher moments in this paper, we conjecture that unless the conditional errors are from a normal distribution, higher moment conditions would matter at some point. For an example, consider a moment condition  $E(\epsilon_{it}^3) = 0$ ,  $t = 1, \dots, T$ . It is not hard to show that these moment conditions are not redundant unless  $E(\epsilon_{it}^4) = 3\sigma_{\epsilon_t}^4$ , which holds when the errors are from a normal distribution.



## CHAPTER FIVE

### CONCLUDING REMARKS

Finding some additional moment conditions and the conditions under which certain moment conditions are superfluous in panel data models has been an important branch of research; Hausman and Taylor (1981), Amemiya and MaCurdy (1986), Breusch, Mizon and Schmidt (1989), Anderson and Hsiao (1981), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bover (1990), Schmidt, Ahn and Wyhowski (1992), and Ahn and Schmidt are examples of contributions. The results in this thesis unify and extend results in several of these papers. One consequence of the analysis is the emergence of some new estimators that either exploit redundancy results or new useful orthogonality conditions.

Another important line of research is the specification test. In most of applications, people presume that the covariance matrix takes the random effects form, and many existing tests that are suitable for the panel data framework focus on testing whether the time-constant unobserved effects are correlated with the explanatory variables (Chamberlain, 1982; Holtz-Eakin, 1986; Jakubson, 1991). However, as we discussed in Chapter 3, in weakly exogenous models it is highly probable that the moment conditions depend on the structure of the covariance matrix,

and also as we showed in Chapter 2, in strictly exogenous model the redundancy of moment conditions hinges heavily on the structure of the covariance matrix. Thus, testing the structure of covariance matrices will be quite useful and necessary in many cases. While the over-identification test (Sargan, 1958; Hansen, 1982) and the Hausman test (Hausman, 1978) are directly applicable for testing the covariance structure, these tests need the estimators using the moment conditions from covariance restrictions, and therefore, in general, will involve numerical optimization.

There might be simpler ways of testing the structure of covariance matrix. Arellano and Bond (1991) devised a test statistic for testing the null  $\sigma_{st} = 0$ ,  $s \neq t$ , in dynamic models, where  $\sigma_{st}$  is the  $(s,t)$  element of the covariance matrix of the differenced errors. This direct test can be generalized to general weakly exogenous, and also to strictly exogenous models. There also will be many more test statistics to be devised. For an example, it would be nice to have a simple test statistic that jointly tests the null that the covariance matrix is of the random effects form.

In addition, these direct test of covariance structure, combined with the Hausman test or with the GMM test, would lead to even nicer results. For an example, suppose we have a conflict result in the rational expectations model considered in Chapter 3; the null of the MA(1) serial

correlation of the time-varying errors cannot be rejected from the direct test, but the Hausman test cannot reject the hypothesis that the instruments are valid, that are supposed to be invalid in the presence of the MA(1) serial correlation. Then, this result will lead to the conclusion that the serial correlation is due to the time lag until the shock is observed. We leave these topics for future works.

## **LIST OF REFERENCES**

## REFERENCES

- Ahn, S.C. (1990), "Three Essays on Share Contracts Labor Supply, and The Estimation of Models for Dynamic Panel Data," unpublished Ph.D. dissertation, Michigan State University.
- Ahn, S.C. and P. Schmidt (1991), "Generalized Least Squares Estimation and Specification Test for Panel Data Models," unpublished manuscript.
- Ahn, S.C. and P. Schmidt (1992), "Efficient Estimation of Models for Panel Data," Journal of Econometrics, forthcoming.
- Ahn, S.C. and P. Schmidt (1993), "A Separability Result for GMM Estimation, with Application to GLS Prediction and Conditional Moment Tests," Econometric Review, forthcoming
- Amemiya, T. (1977), "The Maximum Likelihood and The Nonlinear Three Stage Least Square Estimator in The General Nonlinear Simultaneous Equation Model," Econometrica, 45, 955-968.
- Amemiya, T. (1985), Advanced Econometrics, Basil Blackwell.
- Amemiya, T. and T.E. MaCurdy (1986), "Instrumental-Variable Estimation of An Error-components Model," Econometrica, 54, 869-880.
- Anderson, T. and C. Hsiao (1981), "Estimation of Dynamic Model with Error Components," Journal of the American Statistic Association, 76, 598-606.
- Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and Application to Employment Equation," Review of Economic Studies, 58, 277-297.
- Arellano, M. and O. Bover (1990), "Another Look at The Instrumental Variable Estimation of Error-Components Models," Review of Economic Studies, forthcoming.
- Bhargava, A., and J.D. Sargan (1983), "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," Econometrica, 51, 1635-1659

- Bowden, R.J. and D.A. Turkington (1984), Instrumental Variables, New York, Cambridge University Press.
- Breusch, T.S., G.E. Mizon and P. Schmidt (1989), "Efficient Estimation Using Panel Data," Econometrica, 57, 695-701.
- Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," Journal of Econometrics, 18, 5-18.
- Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," Journal of Econometrics, 34, 305-334.
- Chamberlain, G. (1992a), "Comment: Sequential Moment Restrictions in Panel Data," Journal of Business and Economic Statistics, 10, 20-26.
- Chamberlain, G. (1992b), "Efficiency Bound for Semiparametric Regression," Econometrica, 60, 567-596.
- Cragg, J.G. (1983), "More Efficient Estimation in The Presence of Heteroskedasticity of Unknown Form," Econometrica, 51, 751-763.
- Hansen, L.P. (1982), "Large Sample Properties of Generalized Methods of Moments Estimators," Econometrica, 50, 1029-1054.
- Hausman, J.A. (1978), "Specification Tests in Econometrics," Econometrica, 46, 1251-1272.
- Hausman, J.A. and W.E. Taylor (1981), "Panel Data and Unobservable Individual Effects," Econometrica, 49, 1377-1398.
- Hausman, J.A., W.K. Newey and W.E. Taylor (1987), "Efficient Estimation of Simultaneous Equation Models with Covariance Restrictions," Econometrica, 55, 849-874.
- Hayashi, F., and C. Sims (1983), "Nearly Efficient Estimation of Time Series Models with Predetermined but Not Exogenous Instruments," Econometrica, 51, 783-792.
- Holtz-Eakin, D.W. (1986), "Testing for Individual Effects in Autoregressive Models," Journal of Econometrics, 39, 297-307.
- Holtz-Eakin, D., W. Newey and H.S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," Econometrica, 56, 1371-1396.

- Hsiao, C. (1986), Analysis of Panel Data, New York, Cambridge University Press.
- Jakubson, G. (1991), "Estimation and Testing of the Union Wage Effect Using Panel Data," Review of Economic Studies, 58, 971-991.
- Keane, M.P. and D.E. Runkle (1990), "Testing The Rationality of Price Forecasts: New Evidence form Panel Data," American Economic Review, 80, 714-735.
- Keane, M.P. and D.E. Runkle (1992), "On The Estimation of Panel Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous," Journal of Business and Economic Statistics, 10, 1-9.
- Kiefer, N.M. (1980), "Estimation of Fixed Effects Models for Time Series of Cross Sections with Arbitrary Intertemporal Covariance," Journal of Econometrics, 14, 195-202.
- Newey, W.K. and D. McFadden (1993), "Estimation in Large Samples," Handbook of Econometrics, Vol.4, forthcoming.
- Rothenberg, T.J. and C.T. Leenders (1964): "Efficient Estimation of Simultaneous Equations Systems," Econometrica, 32, 57-76.
- Runkle, D.E. (1991), "Liquidity Constraints and The Permanent Income Hypothesis: Evidence from Panel Data," Journal of Monetary Economics, 97, 73-98.
- Sargan, J.D. (1958), "The Estimation of Economic Relations Using Instrumental Variables," Econometrica, 26, 393-415.
- Schmidt, P. (1990), "Three-Stage Least Squares with Different Instruments for Different Equations," Journal of Econometrics, 43, 389-394.
- Schmidt, P. (1990), Lecture notes.
- Schmidt, P., S.C. Ahn and D. Wyhowski (1992), "Comment," Journal of Business and Economic Statistics, 10, 10-14.
- Theil, H. (1971), Principles of Econometrics, John Wiley & Sons.
- White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and A Direct Test for Heteroskedasticity," Econometrica, 48, 817-838.

White, H. (1982), "Instrumental Variables Regression with Independent Observations," Econometrica, 50, 483-499.

White, H. (1984), Asymptotic Theory for Econometricians, Orlando, Academic Press.

White, H. (1986), "Instrumental Variables Analogs of Generalized Least Squares Estimators," Advances in Statistical Analysis and Statistical Computing, 1, 173-227.

Wooldridge, J.M. (1992), "System Estimation Procedures," Lecture notes.

Wooldridge, J.M. (1993), "Estimating Systems of Equations with Different Instruments for Different Equations," unpublished manuscript.



MICHIGAN STATE UNIV. LIBRARIES



31293010515538