This is to certify that the

dissertation entitled

INSTRUMENTAL-VARIABLE ESTIMATION OF
A PANEL DATA MODEL

presented by

Donald J. Wyhowski

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Peter Schmidt
Major professor

Date July 15, 1988

INSTRUMENTAL-VARIABLE ESTIMATION OF A PANEL DATA MODEL

By

Donald J. Wyhowski

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Economics

1988

This

regressic

develcps

under dif

explanatc

My t

extended

containin

or all of

small T.

allow the

both type

Seco

Taylor to

system;

regressor

the error

sets prop

(1986), a

Thi

estimator

individua

ABSTRACT

INSTRUMENTAL-VARIABLE ESTIMATION OF
A PANEL DATA MODEL

By

Donald J. Wyhowski

This dissertation involves the estimation of a linear regression model in the presence of panel data. My research develops appropriate econometric techniques for such models, under differing assumptions about the correlation between the explanatory variables and the (unobserved) effects.

My three major contributions are: First, I have extended the analysis of Hausman and Taylor (1981) to a model containing individual and time effects correlated with some or all of the regressors, under the assumption of large N and small T. I consider random individual and time effects, and allow the regressors to be correlated or not with either or both types of effects.

Second, I have extended the analysis of Hausman and Taylor to a single equation in a simultaneous equations system; that is, to a regression model in which some of the regressors are correlated with the random noise component of the error. I propose 2SLS estimators based on instrument sets proposed by Hausman and Taylor, Amemiya and MaCurdy (1986), and Breusch, Mizon, and Schmidt (1987).

Third, the dissertation proposes full-information (3SLS) estimators for a simultaneous equations system with random individual effects correlated with some or all of the

exogenous
the usual
are corre
estimator
the exoge
also cons
be correl
others, s
equation

exogenous variables. These estimators are shown to reduce to the usual fixed-effects treatment if all exogenous variables are correlated with the effects, and to reduce to an estimator previously proposed by Baltagi (1981) if none of the exogenous variables are correlated with the effects. I also consider the case in which some exogenous variables may be correlated with the effects in some equations but not in others, so that the available instrument set varies from equation to equation.

I wo

contribu

Committe

suggesti

contribu

would li

read thi

comments

ACKNOWLEDGMENTS

I would like to express my sincere graditude to the contributors to this dissertation. Peter Schmidt, the Committee Chairman, offered encouragement and helpful suggestions at all stages of my research and without his contribution this dissertation would not be possible. I would like to thank the other members of the committee who read this dissertation in its entirety and provided useful comments.

Chapter

1. In

2. In

2.

2.

2.

2.

2.5

2.6

2.7

2.8

3. Indi

3.1

3.2

TABLE OF CONTENTS

Chapter Page

1. Introduction 1

2. Individual Effects but no Time Effects . . . 8

2.1 Introduction 8

2.2 Geometry 10

2.3 Fixed Effects 12

2.4 Random Effects not Correlated with the Regressors 13

2.4.1 Within Estimation 15

2.4.2 Generalized Least Squares Estimation 15

2.4.3 Weighted Least Squares Estimation . 17

2.5 Random Effects Correlated with the Regressors 23

2.6 Variance Estimation when the Random Effects are not Correlated with the Regressors . 28

2.6.1 Counting Rules for Identification . 30

2.6.2 Estimation of q and r 31

2.7 Variance Estimation when the Random Effects are Correlated with the Regressors . . 34

2.7.1 Necessary Conditions for the Existence of d_{IV} 37

2.7.2 Consistency of d_{IV} 37

2.7.3 A Consistent Estimate of r 38

2.8 Conclusions 45

3. Individual and Time Effects 47

3.1 Introduction 47

3.2 Geometry 49

4.

3.3 Fixed Effects 52

3.4 Random Effects, not Correlated with Regressors 54

3.4.1 Within Estimation 56
3.4.2 Generalized Least Squares Estimation 56
3.4.3 Weighted Least Squares Estimation . 58

3.5 Random Effects, Correlated with Regressors 63

3.5.1 Weighted Least Squares Estimation . 65
3.5.2 Counting Rules for Identification . 68

3.6 Random Effects when T is Fixed 71

3.6.1 Random Effects not Correlated with Regressors 71
3.6.2 Random Effects Correlated with the Regressors 76
3.6.3 An Alternative Approach to Estimation 80

3.7 Variance Estimation when the Random Effects are not Correlated with the Regressors . 93

3.8 Variance Estimation when the Random Effects are Correlated with the Regressors . . 101

3.8.1 Necessary Conditions for the Existence of d_{IV} and c_{IV} 105
3.8.2 Consistency of d_{IV} and c_{IV} 105
3.8.3 Consistent Estimates of q and r . . 107

3.9 Conclusions 121

4. Simultaneous Equations with Effects 122

4.1 Introduction 122

4.2 Single Equation Estimation 126

4.2.1 Two-Stage Least Squares 130
4.2.2 An Orthogonality Derivation of the 2SLS Estimator 132
4.2.3 Baltagi's Error-Component Two-Stage Least Squares Estimator 135

4.3 System Estimation 136

4.3.1 Three-Stage Least Squares 138
4.3.2 Instrumental Variables Estimation . 139

5.

Bibliogr

4.3.3 Special Cases 144

4.4 3SLS with Different Instruments 147

4.5 Conclusions 160

5. Conclusion 161

Bibliography 164

CHAPTER 1

Introduction

In this thesis, we consider the estimation of a linear regression model using panel data. Following the usual practice in the literature, we assume that this data consists of T time-series observations on each of N individuals. Models using panel data present the possibility that some of the explanatory variables could be constant over either of the two indices (T or N) and that these variables could be unobservable. Such unobservable time-invariant and individual-invariant variables are called individual and time effects, respectively. Our research will develop appropriate econometric techniques for panel data models, under differing assumptions about the correlation between the explanatory variables and the (unobserved) effects.

It is commonly argued (e.g., Theil (1972), p. 104) that the stochastic disturbance in the usual regression model reflects the joint influence of the variables not included in the model. In the case of panel data, the individual effects would represent the influences of those neglected variables which are time-invariant, and similiarly the time effects would represent the influences of those neglected variables which are

to a mode

these neg

regressor

treatment

the corre

The

models wi

time effe

effects t

correlated

then is t

the regres

variables.

effects as

This direc

the error

(1981). A

be random

between th

papers by

(1986), an

the case i

error comp

and have p

estimators

MaCurdy, n

which are indivual-invariant. Clearly, least squares applied to a model including either type of effects will be biased if these neglected variables are correlated with the included regressors, and we therefore will distinguish different treatments of the model which vary according to the nature of the correlation between the regressors and the effects.

The literature on panel data has covered separately models with individual effects and models with individual and time effects. One strand of the literature has assumed the effects to be fixed, or, more or less equivalently, to be correlated with all the regressors. The point of the model then is to remove the potential bias caused by correlation of the regressors with omitted time or individual-invariant variables. A second strand of literature has viewed the effects as being random and uncorrelated with the regressors. This direction of thought includes the textbook treatment of the error component model as well as the work of Baltagi (1981). A third direction of thought assumes the effects to be random but allows for the possibility of correlation between the effects and some of the regressors. Recent papers by Hausman and Taylor (1981), Amemiya and MaCurdy (1986), and Breusch, Mizon and Schmidt (1987) have considered the case in which the indivdual effects (the time-invariant error component) are correlated with explanatory variables, and have proposed different instrumental variables estimators. However, with the exception of Amemiya and MaCurdy, none of these papers considers the case in which some of

sense of

error, a

Amemiya

estimati

ours are

The

analysis

correlat

effects

the HT,

there ar

error ha

of the f

and e_{it}

earlier

the case

so that

In this

BMS to t

sets of

interest

assumpti

is done

individu

so that

as both

some of the explanatory variables are endogenous (in the sense of being correlated with the noise component of the error, as well as the individual effect). Furthermore, Amemiya and MaCurdy consider only limited information (2SLS) estimation, and their model is restrictive in some ways that ours are not.

The first concern of this dissertation is to extend the analysis for the case when effects are allowed to be correlated with some of the regressors to the case when time effects are present as well as individual effects. That is, the HT, AM, and BMS articles all consider a model in which there are individual effects but no time effects, so that the error has only two components. That is, their error term is of the form $s_{it} = u_i + e_{it}$ where u_i is the individual effect and e_{it} is the random noise. As pointed out above, the earlier literature on panel data also considered prominently the case in which the error also contains a time effect, v_t, so that the error, $s_{it} = u_i + e_{it}$, contains three components. In this dissertation, we extend the results of HT, AM, and BMS to the three component case. This results in different sets of allowable instruments than they use, and to some interesting results on how many and what kind of exogeniety assumptions must be made to estimate the model. The analysis is done mostly under the assumption that both the number of individuals (N) and the number of time periods (T) is large, so that asymptotic properties of the estimators are derived as both N and T approach infinity. However, we will include

a separat
small, th
leads to
The
of simult
equations
individua
point of
earlier b
variables
with the
natural e
and it ca
every str
leads to
types: <u>er</u>
individua
noise but
<u>doubly</u> <u>ex</u>
and noise
generaliz
stage lea
Thir
the singl
(3SLS) es
effects e
correlate

a separate treatment of the case in which N is large but T is small, the common assumption in the two-component case. This leads to some novel estimators in the three-component case.

The second concern of this dissertation is the problem of simultaneity. We consider the usual simultaneous equations model, but with panel data and with an unobservable individual effect in every structural equation. The basic point of view in this thesis, motivated by an argument given earlier by Breusch, Mizon and Schmidt (1985), is that all variables correlated with the noise should also be correlated with the individual effects, but not conversely. This is a natural extension of the point of view in Hausman and Taylor, and it can be justified by consideration of a system in which every structural equation contains individual effects. It leads to a classification of exogenous variables into three types: endogenous, meaning correlated with noise and individual effects; singly exogenous, meaning correlated with noise but possibly correlated with individual effects; and doubly exogenous meaning uncorrelated with individual effects and noise. Several estimators are derived, which are natural generalizations both of the HT estimators and the usual two-stage least squares estimator.

Third, this dissertation generalizes the estimators from the single-equation literature just cited to full information (3SLS) estimators. These estimators reduce to the fixed effects estimators when all exogenous variables are correlated with the effects, and they reduce to previous

estimator
exogenous
addition,
are corre
equations

The
we survey
is used i
approach
data. Th
models wi
topic of
model, in
parameter
which the
uncorrela
which inc
potential
the probl
noise and
necessary
estimator

In c
considere
individua
method of
model, an

estimators for the random effects model when none of the exogenous variables are correlated with the effects. In addition, we discuss the case in which different variables are correlated with the effects in different structural equations.

The plan of this thesis is as follows. In chapter two, we survey the existing literature, review the geometry which is used in our subsequent analysis, and introduce a new approach for the analysis of regression models with panel data. This approach proves to be useful in the analysis of models with both individual effects and time effects, the topic of chapter three. We then consider the fixed effects model, in which the individual effects are treated as fixed parameters to be estimated; the random effects model, in which the individual effects are treated as random and uncorrelated with the regressors; and the model of HT, in which individual effects are treated as random but potentially correlated with the regressors. We also consider the problem of consistent estimation of the variances of the noise and the individual effects. Such estimates are necessary to implement the generalized least squares estimators considered above.

In chapter three, we extend the linear regression model considered in the previous chapter to include unobservable individual-invariant time effects, and we then apply the HT method of instrumental variables estimation to this extended model, and derive the subsequent estimator. The analysis of

the regres
using the
consider t
time effec
and we cor
individual
uncorrelat
version of
are treate
regressors
are charac
but only r
previous t
fixed. Fi
estimation
effects, a
to impleme
considered

In ch
equations
individual
consider a
of the exp
individual
variables
informatio
and discus

the regression models considered in this chapter is done using the approach introduced in chapter two. We then consider the fixed effects model, in which the individual and time effects are treated as fixed parameters to be estimated; and we consider the random effects model, in which both the individual and time effects are treated as random and uncorrelated with the regressors. We consider an extended version of HT, in which both the individual and time effects are treated as random but potentially correlated with the regressors. Since many currently available panel data sets are characterized by having many cross-sectional observations but only relatively few time periods, we then consider the previous two models for the case when N is large but T is fixed. Finally, we consider the problem of consistent estimation of the variances of the noise, the individual effects, and the time effects. Such estimates are necessary to implement the feasible weighted least squares estimator considered.

In chapter four, we consider the usual simultaneous equations model, but with panel data and with an unobservable individual effect in each structural equation. We then consider a natural extension of the HT model by allowing some of the explanatory variables to be correlated with the individual effects. We apply the HT method of instrumental variables estimation, derive the subsequent limited information (2SLS) and full information (3SLS) estimators, and discuss their relative efficiency. In addition, we provide a s

equations w

the notatio

interesting

equations m

correlated

instruments

In cha

suggestions

provide a survey of the current literature on simultaneous equations with effects and translate previous estimators into the notation of this thesis. Finally, we consider an interesting problem which arises for the linear simultaneous equations model with effects when some of the variables are correlated with the individual effects; namely, the instruments need not be the same for every equation.

In chapter five, we summarize our results and make suggestions for future directions of research.

CHAPTER 2

Individual Effects but no Time Effects

2.1 Introduction

In this chapter, we consider the estimation of a linear regression model using panel data. Following the usual practice in the literature, we assume that this data consists of T time-series observations on each of N individuals; we distinguish regressors which vary over time and individuals from those which vary over individuals but are time-invariant; and we assume the presence of unobservable, time-invariant individual effects as well as the usual statistical noise. In chapter 3, we will extend this model to include unobservable time effects.

We write the model to be considered in this chapter as

$$(2.1.1)\quad y_{it} = X_{it}B + Z_iD + u_i + e_{it},\ i = 1,..,N;\ t = 1,..,T$$

where y_{it} is the dependent variable, X_{it} is a vector (of dimension 1 x g) of time-varying explanatory variables, Z_i is a vector (of dimension 1 x k) of time-invariant explanatory variables, and B and D are vectors of parameters to be estimated. The errors e_{it} are iid with mean zero and variance σ_e

and various

all cases t

It is

the stochas

reflects th

model. In

(our u_i) wo

variables w

(of y on X

are correla

will distin

according

individual

The p

we review

analyses.

various as

effects mod

fixed param

to remove t

regressors

2.4 we cons

individual

with the re

problem of

concern. I

variance σ_e^2. The individual effects u_i are unobservable, and various assumptions about them will be made. However, in all cases they will be treated as time-invariant.

It is commonly argued (e.g., Theil (1972), p. 104)) that the stochastic disturbance in the usual regression model reflects the joint influence of variables not included in the model. In the case of panel data, the individual effects (our u_i) would represent the influences of those neglected variables which are time-invariant. Clearly, least squares (of y on X and Z) will be biased if these neglected variables are correlated with the included regressors, and we therefore will distinguish different treatments of the model which vary according to the nature of the correlation between the individual effects and the regressors.

The plan of this chapter is as follows. In section 2.2 we review the geometry which is used in our subsequent analyses. We then consider the estimation of the model under various assumptions. In section 2.3 we consider the fixed effects model, in which the individual effects are treated as fixed parameters to be estimated. The point of this model is to remove the potential bias caused by correlation of the regressors with omitted time-invariant variables. In section 2.4 we consider the random effects model, in which the individual effects are treated as random and uncorrelated with the regressors. Under these assumptions there is no problem of bias, and efficiency of estimation is our central concern. In section 2.5 we consider the model of Hausman and

Taylor (19

random but

Finally, in

consistent

individual

implement t

in section

This ch

However, it

it introduce

models with

useful in th

and time effe

2.2 <u>Geometry</u>

A useful

remainder of

(2.1.1) can be

equations

(2.2.2) $(y_{it}$

(2.2.3)

where $i = 1,..$

$(1/T)\sum_{t=1}^{T} X_{it}$, an

expresses the c

time, while eq

its deviations

Taylor (1981), in which the individual effects are treated as random but potentially correlated with the regressors. Finally, in section 2.6 and 2.7 we consider the problem of consistent estimation of the variances of the noise and the individual effects. Such estimates are necessary to implement the generalized least squares estimators considered in section 2.4 and 2.5.

This chapter does not contain any new estimators. However, it provides a survey of the existing literature, and it introduces a new approach for the analysis of regression models with panel data. This approach will prove to be useful in the analysis of models with both individual effects and time effects, as we will see in chapter 3.

2.2 Geometry

A useful fact, and one to be used throughout the remainder of this chapter, is that the original equation (2.1.1) can be equivalently written as the two orthogonal equations

$$(2.2.2) \qquad (y_{it} - y_{i.}) = (X_{it} - X_{i.})B + (e_{it} - e_{i.})$$

$$(2.2.3) \qquad y_{i.} = X_{i.}B + Z_iD + u_i + e_{i.},$$

where $i = 1,..,N$; $t = 1,...,T$; $y_{i.} = (1/T)\sum_{t=1}^{T} y_{it}$, $X_{i.} = (1/T)\sum_{t=1}^{T} X_{it}$, and $e_{i.} = (1/T)\sum_{t=1}^{T} e_{it}$. Equation (2.2.3) expresses the data in terms of its individual averages over time, while equation (2.2.2) expresses the data in terms of its deviations around the mean for each individual.

(2.2.

where

and Z

respec

Taylor

individ

are (NT

each blo

To

above, w

(2.2.5)

where jr

(T x 1).

of the ind

observatio

observation

respective

elements of

(2.2.6) (P

respectively

Since Z

time-series o

elements of t

Writing equation (2.1.1) in matrix form, we have

$$(2.2.4) \quad y = XB + ZD + u + e$$

where y, u, and e denote (NT x 1) dimensioned vectors; and X and Z denote (NT x g) and (NT x k) dimensioned matrices, respectively. Following the convention of Hausman and Taylor (1981), the observations are ordered first by individual and then by time, so that u and each column of Z are (NT x 1) dimensioned vectors consisting of N blocks, with each block containing T identical entries.

To achieve the same decomposition as was accomplished above, we define the two orthogonal projections

$$(2.2.5) \quad P = (I_N \otimes j_T j_T^T/T) \text{ and } Q = I_{NT} - P$$

where $j_T = (1,...,1)^T$ is a vector of ones, having dimension (T x 1). The transformation P determines the means for each of the individual groups and repeats each of these N observations T times. The transformation Q transforms each observation into the difference between itself and its respective individual group mean. Explicitly, the (i,t) elements of Py and Qy can be written as

$$(2.2.6) \quad (Py)_{it} = y_{i.} \quad \text{and} \quad (Qy)_{it} = y_{it} - y_{i.},$$

respectively.

Since Z contains variables that are constant across all time-series observations for a given individual, QZ = 0. The elements of the columns of Z are, on the other hand, unaffec

Analogo

i.e. Q

can now

equati

(2.2.7

(2.2.8

2.3 <u>Fi</u>

In

regress

effects

approac

and the

estimat

(2.3.1)

The est

uses or

sometim

regress

covaria

as eith

all wel

pp. 329

A

not pos

unaffected by the transformation P; that is, PZ = Z. Analogous results hold true for the individual effects u; i.e. Qu = 0 and Pu = u. Thus, the original equation (2.2.4) can now be written equivalently as the two orthogonal equations

$$(2.2.7) \quad Qy = QXB + Qe$$

$$(2.2.8) \quad Py = PXB + ZD + u + Pe$$

2.3 Fixed Effects

In this section, we discuss the estimation of the linear regression equation (2.2.4) when the individual-specific effects are treated as fixed constants. The standard approach is to use individual dummy variables as regressors, and then to apply least squares. This yields the following estimator for B:

$$(2.3.1) \quad b_W = (X^T QX)^{-1} X^T Qy.$$

The estimator b_W is the familiar within-group estimator; it uses only the variation within each group. This estimator is sometimes called the covariance estimator since the regression just described is in fact the usual analysis of covariance. The estimator is unbiased, and it is consistent as either N or T (or both) approaches infinity. These are all well-known results; for example, see Judge et al. (1985, pp. 329).

A problem with this estimation procedure is that it is not possible to obtain estimates of the coefficients of the time-inv

is perfe

equivale

to devia

containe

coeffici

(2.3.2)

and thes

as T ap

time-in

interpr

An

to defi

(2.2.2)

least s

ignorin

Us

(2.2.2)

computa

equatio

stateme

it igno

2.4 Ran

In

a linea

u_1) are

time-invariant regressors (Z). Any time-invariant regressor is perfectly collinear with the individual dummy variables; equivalently, it is removed by the transformation of the data to deviations from individual means. If the original model contained no time-invariant regressors, the estimated coefficients of the individual dummy variables are

$$(2.3.2) \quad u_w = Py - PXb_w ,$$

and these estimates of the individual effects are consistent as T approaches infinity. If the original model contained time-invariant regressors, then u_w defined above is interpreted as an estimate of (ZD + u) rather than of just u.

An equivalent derivation of the within estimator b_w is to define it as the least squares estimator in equation (2.2.2), ignoring (2.2.3). Similiarly, the estimator u_w is least squares applied to (2.2.3), after setting B = b_w, and ignoring the time-invariant variables Z.

Using only one part of equation (2.1.1), namely equation (2.2.2), when estimating B has the advantage of being computationally more convenient than estimating the whole of equation (2.1.1). This approach also makes explicit the statement that b_w ignores the between-group variation; i.e., it ignores the cross-sectional variation in equation (2.2.3).

2.4 Random Effects not Correlated with the Regressors

In the previous section, we discussed the estimation of a linear regression model when the individual effects (the u_i) are treated as fixed constants. In this section, we treat th

the erro

The N in

from som

viewed a

We a

zero and

uncorrela

(2.4.1)

The

(2.4.2) v

The varian

components

of the err

and e_{it} ar

this model

variance-c

The pr

term result

cross-secti

cross-secti

explicit if

vector (s_{i1},

the matrix

treat the individual effects similiarly to the way we treat the error term e_{it}; we assume the u_i to be random variables. The N individuals are now to be interpreted as being drawn from some larger population, and so the effects u_i can be viewed as a random sample from some distribution.

We assume specifically that the u_i are iid with mean zero and variance σ_u^2. We also assume that X and Z are uncorrelated with u. The model is written as

$$(2.4.1)\quad y_{it} = X_{it}B + Z_iD + u_i + e_{it}$$

$$= X_{it}B + Z_iD + s_{it} \quad i = 1,..,N;\ t = 1,...,T$$

The variance of y_{it}, conditional on X_{it} and Z_i, is

$$(2.4.2)\quad var(y_{it}) = var(s_{it}) = \sigma_e^2 + \sigma_u^2.$$

The variances σ_e^2 and σ_u^2 are sometimes called variance components; each is itself a variance as well as a component of the error variance, $var(s_{it})$. Similiarly, the errors u_i and e_{it} are sometimes called error components. Therefore, this model is often referred to as an error-components or variance-components model.

The presence of the random effects u_i in the disturbance term results in correlation among the errors of the same cross-sectional unit, although the errors from different cross-sectional units are independent. This can be made explicit if we let s_i denote the (T x 1) dimensioned error vector $(s_{i1},...,s_{iT})^T$. The covariance matrix of s_i is then the matrix

(2.4.3)

where jτ =

correlatio

over time

2.4.1 Wit

The

whether t

variables

squares a

effects d

are treat

is still

Hsiao (19

effects a

2.4.2 Ge

As

periods b

errors i

(2.4.4)

are autoc

use the g

and Taylo

(2.4.5)

$$(2.4.3)\quad \mathrm{Cov}(\, s_i \,) \equiv S_i = \sigma_e^2 I_T + \sigma_u^2 (j_T j_T^T)$$

where $j_T = (1,\ldots,1)^T$ is a (T x 1) vector of 1's. Thus, this correlation of the errors at the individual level is constant over time and is identical for all individuals.

2.4.1 Within Estimation

The within-group estimator can be used regardless of whether the u_i's are viewed as fixed constants or as random variables. The within estimator of B can be viewed as least squares applied to equation (2.2.2), and the individual effects do not appear in this equation. So, whether the u_i are treated as nonstochastic or stochastic, the estimator b_W is still unbiased and consistent. However, as pointed out by Hsiao (1986), the Within estimator is inefficient when the effects are random and uncorrelated with the regressors.

2.4.2 Generalized Least Squares Estimation

As was shown above, since the s_{it} in different time periods but for the same individual both contain u_i, the errors in the equation

$$\begin{aligned}(2.4.4)\quad y_{it} &= X_{it}B + Z_iD + u_i + e_{it}\\ &= X_{it}B + Z_iD + s_{it} \qquad i = 1,..,N;\quad t = 1,\ldots,T\end{aligned}$$

are autocorrelated. Efficient estimation requires that we use the generalized least squares method. Following Hausman and Taylor (1981), we write

$$(2.4.5)\quad S \equiv \mathrm{Cov}(\, s \,) = \mathrm{Cov}(\, u + e \,) = \sigma_e^2 I_{TN} + T\sigma_u^2 P$$

where σ_e

both idem

factor of

(2.4.6)

where c^2

equation

(2.4.7)

where

assume

squares

(2.4.8)

Eq

squares

Taylor

(2.4.9

so tha

transf

differ

(2.4.1

and th

that c

where σ_e^2 = var(e_{it}) and σ_u^2 = var(u_i). Since P and Q are both idempotent and orthogonal, it follows that, up to a factor of proportionality,

$$(2.4.6) \quad S^{-1} = Q + c^2 P$$

where $c^2 = [\ \sigma_e^2/(\ \sigma_e^2 + T\sigma_u^2\ )\ ]$. Now, if we rewrite equation (2.2.4) as

$$(2.4.7) \quad y = XB + ZD + u + e$$
$$= RA + s$$

where R = (X, Z) and A = (B^T, D^T $)^T$, and if we assume that σ_e^2 and σ_u^2 are known, the generalized least squares estimator of A from equation (2.4.7) is simply

$$(2.4.8) \quad a_{GLS} = (R^T S^{-1} R)^{-1} R^T S^{-1} y.$$

Equivalently, the GLS estimator is ordinary least squares of ($S^{-1/2}y$) on ($S^{-1/2}R$). Again following Hausman and Taylor (1981), we can note that

$$(2.4.9) \quad S^{-1/2} = Q + cP = I_{NT} - (1-c)P$$

so that $S^{-1/2}y = y - (1-c)Py$ (and similiarly for R). This transformation is what Hausman and Taylor call "(1-c) differences." For example,

$$(2.4.10) \quad (S^{-1/2}y)_{it} = y_{it} - (1-c)y_{i}.$$

and this differs from the within transformation to the extent that $c \neq 0$.

2.4.3 W

As

squares

from th

equatio

(2.4.11

(2.4.12

Fi

(1978b),

thesis.

the cova

satisfy

Lemma (2

conditic

idempote

= My ar

form b

definite

is the m

class.

The

s and c

involves

idempote

2.4.3 Weighted Least Squares Estimation

As an alternative approach to the generalized least squares estimator of A, consider the equations which result from the decomposition of equation (2.4.7). These orthogonal equations can be written as

(2.4.11) $Qy = QRA + Qs$

(2.4.12) $Py = PRA + Ps$

First, we consider the following lemma, due to Mundlak (1978b), and which is to be used throughout much of this thesis. It concerns the need to correct for the failure of the covariance matrix associated with the disturbance term to satisfy the ideal conditions.

Lemma (2.1): Suppose $y = XB + s$ satisfies the ideal conditions except that $Cov(y) = Cov(s) = S$. Let M be an idempotent matrix other than the identity matrix, and let $y^* = My$ and $X^* = MX$. Consider the class of estimators of the form $b_* = (X^{*T}H^{-1}X^*)^{-1}X^{*T}H^{-1}y^*$, where H is any positive definite matrix. Then the estimator $b_0 = (X^{*T}X^*)^{-1}X^{*T}y^*$ is the minimum variance unbiased estimator of B within this class.

The point of the Lemma is as follows. We have $y = XB + s$ and $cov(s) = S$. The best (GLS) estimator of B certainly involves S. However, if we transform the equation by an idempotent matrix M:

(2.4.13)

the best
transfor
This is
dealing
P and Q.

We r
errors in

(2.4.14)

and

(2.4.15)

respective
these two
times an i
the same b
the weights
follows fro
system so w
within the
the paramet
these equat
We will ref
system of or

(2.4.16) (1

(2.4.13) $(My) = (MX)B + (MS)$,

the best (minimum variance unbiased) estimator of B from this transformed equation is just OLS, which does not depend on S. This is relevant in the present context because we are dealing with equations transformed by the idempotent matrices P and Q.

We note that the covariance matrices associated with the errors in (2.4.11) and (2.4.12) may be written as

(2.4.14) $\mathrm{Cov}(\ Qs\ ) \ = \ QSQ \ = \ qQ$

and

(2.4.15) $\mathrm{Cov}(\ Ps\ ) \ = \ PSP \ = \ rP$,

respectively, where $q = \sigma_e^2$ and $r = \sigma_e^2 + T\sigma_u^2$. Each of these two covariance matrices is of the form of a constant times an idempotent matrix. These two constants may be made the same by multiplying equations (2.4.11) and (2.4.12) by the weights (1/q) and (1/r), respectively. Moreover, it follows from Lemma (2.1) that least squares applied to the system so weighted yields the best minimum unbiased estimator within the class containing all least square estimators of the parameter vector A from any further transformation of these equations or, in fact, the original equation (2.4.7). We will refer to the least squares estimator of A from the system of orthogonal equations

(2.4.16) $(1/q)Qy \ = \ (1/q)QRA + (1/q)Qs$

(2.4.17)

as the we

may be wr

(2.4.18)

The (

transforma

correlatio

matrix of

space. Si

sum is the

been reduc

equations,

equations

original e

unbiased e

to the gen

equation.

Theorem (2

equal to t

Proof:

The g

equation y

a_{cis} = (R

= (R

(2.4.17) $(1/r)Py = (1/r)PRA + (1/r)Ps$

as the weighted least squares estimator of A. This estimator may be written as

(2.4.18) $a_{WLS} = (R^TQR/q + R^TPR/r)^{-1}(R^TQ/q + R^TP/r)y.$

The decomposition of the original equation by the transformations, P and Q, has the effect of isolating the correlations found in the non-block diagonal covariance matrix of its error vector, S, to the particular orthogonal space. Since these transformations are orthogonal, and their sum is the identity matrix, equation (2.4.7) is said to have been reduced by the pair (Q, P) into the two orthogonal equations, (2.4.11) and (2.4.12). Since this pair of equations contains exactly the same information as the original equation, we would expect that the minimum variance unbiased estimator from the two equations would be equivalent to the generalized least squares estimator from the original equation. This result is stated in the following theorem.

Theorem (2.2): The weighted least squares estimator, a_{WLS}, is equal to the generalized least squares estimator, a_{GLS}.

Proof:

The generalized least squares estimator of A from the equation y = RA + s where Cov(s) = S, can be written as

$$a_{GLS} = (R^TS^{-1}R)^{-1}(R^TS^{-1}y)$$

$$= (R^T\{ Q + c^2P \}R)^{-1}(R^T\{ Q + c^2P \}y)$$

=

Theref

N

the be

= (R^T

in the

estimat

(2.4.11

groups.

squares

of the w

estimato

variatio

the equa

and r =

and (2.4.

The

estimatio

estimator

For t

equations

equation (

as

$$= (R^T\{ Q + (q/r)P \}R)^{-1}(R^T\{ Q + (q/r)P \}y)$$

since $c^2 = (q/r)$

$$= (R^T\{ (1/q)Q + (1/r)P \}R)^{-1}(R^T\{ (1/q)Q + (1/r)P \}y)$$

$$= (R^TQR/q + R^TPR/s)^{-1}(R^TQ/q + R^TP/r)y$$

Therefore, $a_{WLS} = a_{GLS}$. Q.E.D.

Now least squares applied to equation (2.4.12) is called the between-group estimator; explicitly, it is $a_B = (R^TPR)^{-1}R^TPy$. It utilizes the cross-sectional variation in the individual means. Recall that the within-group estimator can be viewed as least squares applied to equation (2.4.11); it utilizes the variation within the individual groups. As Maddala (1971) has shown, the generalized least squares estimator can be viewed as an efficient combination of the within-group estimator and the between-group estimator. The optimal weights for the two different sets of variation are the constants being used to normalize each of the equations; i.e. the reciprocal of the variances $q = \sigma_e^2$ and $r = \sigma_e^2 + T\sigma_u^2$ for the respective equations, (2.4.11) and (2.4.12).

The following two theorems concern alternative estimation procedures which yield the weighted least squares estimator defined above.

For the first such derivation, consider again the equations resulting from the decomposition of the original equation (2.4.7). These orthogonal equations can be written as

(2

wh

and

Let

(2.4

so S.

the e

any i

there

is

(2.4.2

Applyi

genera

again a

formall

<u>Theorem</u>

from eq

estimato

$$(2.4.19) \qquad y_* = R_* A + s_*$$

where

$$y_* = \begin{bmatrix} Py \\ Qy \end{bmatrix}, \qquad R_* = \begin{bmatrix} PR \\ QR \end{bmatrix},$$

and

$$s_* = \begin{bmatrix} P(e + u) \\ Q(e) \end{bmatrix}.$$

Let

$$(2.4.20) \qquad S_* \equiv Cov(s_*) = \begin{bmatrix} qP & 0 \\ 0 & rQ \end{bmatrix},$$

so S_* denotes the singular covariance matrix associated with the error term of the above system. It is well known that any idempotent matrix is its own generalized inverse, and therefore the generalized inverse of the singular matrix, S_*, is

$$(2.4.21) \qquad S_*^{+} = \begin{bmatrix} (1/q)P & 0 \\ 0 & (1/r)Q \end{bmatrix}$$

Applying generalized least squares to (2.4.19), using the generalized inverse of the error covariance matrix, we arrive again at the weighted least squares estimator; this is stated formally in the following theorem.

<u>Theorem (2.3)</u>: The generalized least squares estimator of A from equation (2.4.19) equals the weighted least squares estimator of A.

Proof:

The

equation

a_{GLS*} =

=

Thus, a_{WL}

A se

Battese (

consider

using the

q^* = (q

equation c

(2.4.22)

Thus, usin

covariance

theorem.

Theorem (2.

(2.4.22) is

A.

Proof:

Now the

written as

Proof:

The generalized least squares estimator of A from the equation (2.4.19) can be written as

$$a_{GLS*} = (R_*^T S_*^+ R_*)^{-1} R_*^T S_*^+ y_*$$
$$= (R^T QR/q + R^T PR/r)^{-1} (R^T Q/q + R^T P/r)y$$

Thus, $a_{WLS} = a_{GLS*}$ Q.E.D.

A second derivation follows the lines of Fuller and Battese (1974). We note that Cov(s) $\equiv$ S = qQ + rP, and we consider the transformation of the original equation (2.4.7), using the matrix $S^{-1/2}$, where $S^{-1/2} = (1/q^*)Q + (1/r^*)P$, $q^* = (q)^{1/2}$ and $r^* = (r)^{1/2}$. The transformed equation can be written as

$$(2.4.22) \quad S^{-1/2}y = S^{-1/2}RA + S^{-1/2}(s)$$
$$= S^{-1/2}RA + S^{-1/2}(e + u)$$

Thus, using the Fuller and Battese expression for the covariance of the error term, s, we have the following theorem.

Theorem (2.4): The least squares estimator of A from equation (2.4.22) is equal to the weighted least squares estimator of A.

Proof:

Now the decomposition of equation (2.4.22) can be written as

(2.4.23

(2.4.24

The leas

$a_* = ($

time

$= ($

t

$= (R$

As shown

estimator

$a_{WLS} = (R$

Thus, a_{WLS}

2.5 Random

In som

may be reas

unobservabl

correlated

variables.

random effec

variables, t

(2.4.23) $QS^{-1/2}y = QS^{-1/2}RA + QS^{-1/2}(e)$

(2.4.24) $PS^{-1/2}y = PS^{-1/2}RA + PS^{-1/2}(e + u)$

The least squares estimator of A from this system is

$$a_* = (R^T S^{-1/2}QS^{-1/2}R + R^T S^{-1/2}PS^{-1/2}R)^{-1}$$
$$\text{times} \quad (R^T S^{-1/2}QS^{-1/2} + R^T S^{-1/2}PS^{-1/2})y$$

$$= (R^T(Q/q^* + P/r^*)Q(Q/q^* + P/r^*)R$$
$$+ R^T(Q/q^* + P/r^*)P(Q/q^* + P/r^*)R)^{-1}$$
$$\text{times} \quad (R^T(Q/q^* + P/r^*)Q(Q/q^* + P/r^*)$$
$$+ R^T(Q/q^* + P/r^*)P(Q/q^* + P/r^*))y$$

$$= (R^T QR/q + R^T PR/r)^{-1}(R^T Q/q + R^T P/r)y$$

As shown in equation (2.4.18), the weighted least squares estimator of A is written as

$$a_{WLS} = (R^T QR/q + R^T PR/r)^{-1}(R^T Q/q + R^T P/r)y$$

Thus, $a_{WLS} = a_*$. Q.E.D.

2.5 Random Effects Correlated with the Regressors

In some applications of the error-component model, there may be reasons to believe that the individual-specific unobservable effects found in the error term may, in fact, be correlated with some or all of the included explanatory variables. If we take the view suggested earlier, that the random effects represent omitted individual-specific variables, this correlation would seem inevitable. When there is c

explanatory

estimator i

(1978a) tak

always pres

rejects the

the within

However

effects are

consider in

(1981), in

regressors.

introduce s

(2.5.1) y_i

where X_{1it}

time-varyin

($1 \times k_1$) di

variables,

both errors

time-varyin

dimensioned

Z_{2i}, are bot

uncorrelated

component, e

well as inde

The mat

there is correlation between the random effects and the explanatory variables, the generalized least squares estimator is biased and inconsistent. Indeed, Mundlak (1978a) takes the extreme view that such correlation is always present in the error-component model, and therefore rejects the generalized least squares estimator in favor of the within estimator.

However, Mundlak (1978a) considers the case in which the effects are correlated with <u>all</u> of the regressors. We consider instead the case treated by Hausman and Taylor (1981), in which the effects are correlated with <u>some</u> of the regressors. To consider this case, we first need to introduce some notation. Consider the equation

$$(2.5.1) \quad y_{it} = (X_{1it}, X_{2it})B + (Z_{1i}, Z_{2i})D + u_i + e_{it}$$

where X_{1it} represents the $(1 \times g_1)$ dimensioned vector of time-varying explanatory variables and Z_{1i} represents the $(1 \times k_1)$ dimensioned vector of time-invariant explanatory variables, both of which are assumed to be uncorrelated with both errors, u_i and e_{it}. The $(1 \times g_2)$ dimensioned vector of time-varying explanatory variables, X_{2it}, and the $(1 \times k_2)$ dimensioned vector of time-invariant explanatory variables, Z_{2i}, are both assumed to be correlated with u_i but uncorrelated with e_{it}. As before, both the random noise component, e_{it}, and the individual effects, u_i, are i.i.d. as well as independent of one another.

The matrix form of equation (2.5.1) can be written as

(2.5.2)

where y,

and Z is

Now

tradition

regressors

In the pr

interestin

estimator.

based on a

present in

First

(Cov(s))-1

scalar cov

(2.5.3) S

Second

(Q, X_1, Z_1

efficient i

(2.5.3). I

the column

when M has

estimator o

(2.5.4) aH r

$$(2.5.2) \quad y = (X_1, X_2)B + (Z_1, Z_2)D + u + e$$
$$= RA + s$$

where y, u, and e are (NT x 1); X is (NT x g), $g = g_1 + g_2$; and Z is (NT x k), $k = k_1 + k_2$.

Now the method of instrumental variables has traditionally been viewed as the response to the problem of regressors correlated with the equation's disturbance term. In the present context, Hausman and Taylor (1981) propose an interesting variation to the usual instrumental variables estimator. Unlike the usual approach, their estimator is based on a set of instruments made up of regressors already present in the equation being estimated.

First, they multiply equation (2.5.2) by $S^{-1/2} \equiv (Cov(s))^{-1/2}$ to transform the error term so that it has a scalar covariance matrix. The transformed equation is simply

$$(2.5.3) \quad S^{-1/2}y = S^{-1/2}(X_1, X_2)B + S^{-1/2}(Z_1, Z_2)D + S^{-1/2}s$$
$$= S^{-1/2}RA + S^{-1/2}s$$

Second, they use as their instruments the set H = (Q, X_1, Z_1), and derive what they consider to be the efficient instrumental variables estimator of A from equation (2.5.3). If we define for any matrix M the projection onto the column space of M as P[M] (so that $P[M] = M(M^TM)^{-1}M^T$ when M has full column rank), the Hausman and Taylor estimator of A can be written as

$$(2.5.4) \quad a_{HT} = (R^TS^{-1/2}P[H]S^{-1/2}R)^{-1}(R^TS^{-1/2}P[H]S^{-1/2}y)$$

The l

H is not

following

Lemma (2.

However, w

estimator,

understand

somewhat m

equation (

equations

(2.5.5) Q

(2.5.6) P

Since

errors and

can readil

variables

uncorrelat

(2.5.6) on

of orthogo

(2.5.7)

(2.5.8) P

where P_1 =

The c

The Hausman-Taylor instrument set is cumbersome because H is not of full column rank. We can evaluate P[H] using the following Lemma:

Lemma (2.5): $P[H] = Q + P[(PX_1, Z_1)]$

However, while this solves the problem of calculating the estimator, it is not very satifactory in helping us to understand why the estimator is efficient. Perhaps a somewhat more intuitive approach to the estimation of equation (2.5.2) is to decompose it into the two orthogonal equations

$$(2.5.5) \quad Qy = (QX_1, QX_2)B + Qe$$

$$(2.5.6) \quad Py = (PX_1, PX_2)B + (Z_1, Z_2)D + P(e + u)$$

Since $Qu = 0$, there is no problem of correlation between errors and regressors in (2.5.5). Furthermore, (PX_1, Z_1) can readily be seen to be the largest available set of variables in equation (2.5.6) which have been assumed to be uncorrelated with the random effects. Projecting equation (2.5.6) onto the column space of (PX_1, Z_1), we have the set of orthogonal equations

$$(2.5.7) \quad Qy = QRA + Q(e + u)$$

$$(2.5.8) \quad P_1Py = P_1PRA + P_1P(e + u)$$

where $P_1 = P[(PX_1, Z_1)]$.

The covariance matrix associated with the errors in equations

(2.5.9)

and

(2.5.10)

respectiv

matrices

matrix. T

attempt at

equation w

estimator.

squares es

becomes

(2.5.11)

Using

(2.5.12)

We ha

Theorem (2

weighted l

(2.5.8).

equations (2.5.7) and (2.5.8) can be written as

$$(2.5.9) \quad \mathrm{Cov}(\ Q(\ e\ )\ ) = qQ$$

and

$$(2.5.10) \quad \mathrm{Cov}(\ P[(PX_1,\ Z_1)]P(\ e + u\ )\ ) = rP[(PX_1,\ Z_1)].$$

respectively. We note that each of these two covariance matrices has the form of a constant times an idempotent matrix. Thus, Lemma (2.1) would imply that any further attempt at diagonalizing the covariance matrices in either equation would not improve the efficiency of the resulting estimator. Using the weights q and r, the weighted least squares estimator of A from equations (2.5.7) and (2.5.8) becomes

$$(2.5.11) \quad a^*_{IV} = \{\ R^T(1/q)QR + R^T(1/r)P[(X_1,\ Z_1)]R\ \}^{-1} \text{ times } \{\ R^T(1/q)Qy + R^T(1/r)P[(PX_1,\ Z_1)]y\ \}$$

Using Lemma (2.5), a_{HT} can be rewritten as

$$(2.5.12) \quad a_{HT} = (\ R^T(\ (1/q)Q + (1/r)P[(\ PX_1,\ Z_1\ )]\ )R\ )^{-1} \text{ times } (\ R^T(\ (1/q)Q + (1/r)P[(PX_1,\ Z_1)]\ )y$$

We have now proved the following theorem.

<u>Theorem (2.6)</u>: The Hausman and Taylor estimator of A equals weighted least squares applied to equations (2.5.7) and (2.5.8).

2.6 <u>Va</u>

<u>Corre</u>

Wh

procedu

compone

not the

therefo

compone

example

Un

(1973)

estimat

distrib

known v

large N

the est

to zero

estimato

other ha

covarian

feasible

definite

suggests

squares

estimato

Eff

and its s

2.6 Variance Estimation when the Random Effects are not Correlated with the Regressors

When discussing the generalized least squares estimation procedure, we have implicitly assumed that the variance components, σ_e^2 and σ_u^2, were known. In practice, this is not the case; the variance components are usually unknown and therefore must be estimated. When estimates of the variance components are used in place of the actual values, we have an example of feasible generalized least squares.

Under mild regularity conditions, Fuller and Battese (1973) have shown that the feasible generalized least squares estimator is consistent and has the same asymptotic distribution as the generalized least squares estimators with known variance components. This result holds true for either large N or large T. Swamy and Mehta (1979) caution that, if the estimator of σ_u^2 is unreliable, say because σ_u is close to zero or N is small, the feasible generalized least squares estimator may also be unreliable. Taylor (1980), on the other hand, has shown that the difference between the covariance matrices of the Within estimator and of the feasible generalized least squares estimator is nonnegative definite for even moderate sizes of either N or T. This suggests that, in practice, the feasible generalized least squares estimator may be more efficient than the Within estimator.

Efficiency in the estimation of the variance components and its subsequent effect on the efficiency of the feasible genera

Amemiy

(1973)

estima

in eff

I

(2.4.1

(2.6.1

where

(2.6.2

where

k.

If

instead

procedu

estimat

estimat

$= \sigma_1^2$ +

groundw

Swamy (

estimat

additio

of the

We

(2.6.1)

generalized least squares estimator has been discussed by Amemiya (1971). Similiarly, papers by Maddala and Mount (1973) and Taylor (1980) have shown that using more efficient estimates of the variance components need not lead to a gain in efficiency of the estimates.

In the following discussion, we rewrite equation (2.4.11) and (2.4.12) as

(2.6.1) $$Qy = R_1 A_1 + Qs$$

where $R_1 = (\ QX\ )$, $A_1 = B$, and $rank(\ R_1\ ) = g$; and

(2.6.2) $$Py = R_2 A_2 + Ps$$

where $R_2 = (\ PX,\ Z\ )$, $A_2 = (\ B^T,\ D^T\ )^T$, and $rank(\ R_2\ ) = g + k$.

If feasible weighted least squares is to be implemented instead of the equivalent feasible generalized least squares procedure, the weights q and r are the parameters we need estimate. One approach to estimating these weights is to estimate $q = \sigma_0^2$ using residuals from equation (2.6.1) and $r = \sigma_1^2 + T\sigma_2^2$ using residuals from equation (2.6.2). The groundwork for such an approach is laid by Maddala (1971), Swamy (1971), and Arora (1973). We now proceed to show that estimators so defined are both unbiased and consistent. In addition, we find the necessary conditions for identification of the model.

We define the sum of squared residuals from equation (2.6.1) as

(2.6.3)

where th

squares

namely

(2.6.4)

And we d

(2.6.2)

(2.6.5)

where th

equation

(2.6.6)

2.6.1 Co

To

identifi

equation

Thus, th

model ar

(2.6.7)

Sinc

necessary

more succ

(2.6.8)

(2.6.3) $SSE_1 = (Qy - R_1 a_1)^T (Qy - R_1 a_1)$

where the residuals have been computed using the least squares estimates of the coefficients in equation (2.6.1), namely

(2.6.4) $a_1 = (R_1^T R_1)^{-1} R_1^T y.$

And we define the sum of squared residuals from equation (2.6.2) as

(2.6.5) $SSE_2 = (Py - R_2 a_2)^T (Py - R_2 a_2)$

where the least squares estimates of the coefficients in equation (2.6.2) are given as

(2.6.6) $a_2 = (R_2^T R_2)^{-1} R_2^T y.$

2.6.1 Counting Rules for Identification

To insure that the parameters in the model are identified requires that the parameters in each of the two equations, (2.6.1) and (2.6.2), separately be identified. Thus, the necessary conditions for the identification of the model are that

(2.6.7) $g + k \leq N$ and $g \leq N(T - 1).$

Since the second condition follows from the first, the necessary condition for identification of the model can be more succinctly written as

(2.6.8) $g + k \leq N.$

2.6.2 Est

Theorem (

s_2^2 = SS

q and s_2^2

Proof:

Let P

the regre

$R_1(R_1^T R_1$

and $P_1 P$

Firs

Residual

Then we

SSE_1 = (

=

=

Taking t

Exp{ SSE

2.6.2 Estimation of q and r

Theorem (2.7): Let $s_1^2 = SSE_1/[\, N(T-1) - g \,]$, and let $s_2^2 = SSE_2/[\, N - g - k \,]$. Then s_1^2 is an unbiased estimator q and s_2^2 is an unbiased estimator of r.

Proof:

Let P_1 represent the projection onto the column space of the regressors in equation (2.6.1); i.e. $P_1 = P[R_1] = R_1(R_1^T R_1)^{-1} R_1^T$. Then $QP_1 = P_1 Q = P_1$, $P_1 R_1 = R_1$, $P_1^T = P_1$, and $P_1 P = 0$.

First we write the residual from equation (2.6.1) as

$$
\begin{aligned}
Residual_1 &= (Qy - QP_1 y) = R_1 A_1 + Qs - P_1 Qy \\
&= R_1 A_1 + Qs - P_1 R_1 A_1 - P_1 Qs \\
&= R_1 A_1 - R_1 A_1 + Qs - P_1 s \\
&= (Q - P_1)s
\end{aligned}
$$

Then we form the expression

$$
\begin{aligned}
SSE_1 &= (Qy - QP_1 y)^T (Qy - QP_1 y) = y^T (Q - QP_1)^T (Q - QP_1) y \\
&= s^T (Q - QP_1)^T (Q - QP_1) s = s^T (Q - P_1 Q - QP_1 + P_1 QP_1) s \\
&= s^T (Q - P_1 - P_1 + P_1) s = s^T (Q - P_1) s
\end{aligned}
$$

Taking the expectation of the SSE_1, we write

$$
\begin{aligned}
Exp\{ SSE_1 \} &= Exp\{ s^T (Q - P_1) s \} \\
&= Exp\{ trace\{ s^T (Q - P_1) s \} \} \\
&= Exp\{ trace\{ (Q - P_1) s s^T \} \}
\end{aligned}
$$

since trace(AB) = trace(BA) if both AB, BA defined and square.

Thus,

N

space

= R_2(P

P_2, ar

R

Resid

Then w

SSE_2 =

=

=

Taking

Exp{ SS

$$= \text{trace}\{ (Q - P_1)E\{ ss^T \} \}$$

$$= \text{trace}\{ (Q - P_1)\{ qQ + rP \} \}$$

since $E\{ ss^T \} \equiv S = qQ + rP$;

$$= (q)\text{trace}\{ (Q - P_1) \} \quad \text{since } P_1 P = 0.$$

$$= (q)\text{rank}(Q - P_1)$$

since trace(A) = rank(A) if A is idempotent.

$$= (q)\{\text{rank}(Q) - \text{rank}(R_1)\}$$

Thus, $\text{Exp}\{ s_1^2 \} = q$.

Now, let P_2 represent the projection onto the column space of the regressors in equation (2.4.2); i.e. $P_2 = P[R_2] = R_2(R_2^T R_2)^{-1} R_2^T$. Then $PP_2 = P_2 P = P_2$, $P_2 R_2 = R_2$, $P_2^T = P_2$, and $P_2 Q = 0$.

First we write the residual from equation (2.6.2) as

$$\begin{aligned} \text{Residual}_2 &= (Py - PP_2 y) = R_2 A_2 + Ps - P_2 Py \\ &= R_2 A_2 + Ps - P_2 R_2 A_2 - P_2 Ps \\ &= R_2 A_2 - R_2 A_2 + Ps - P_2 s \\ &= (P - P_2)s \end{aligned}$$

Then we form the expression

$$\begin{aligned} SSE_2 &= (Py - PP_2 y)^T (Py - PP_2 y) = y^T (P - PP_2)(P - PP_2)y \\ &= s^T (P - PP_2)(P - PP_2)s = s^T (P - P_2 P - PP_2 + P_2 PP_2)s \\ &= s^T (P - P_2 - P_2 + P_2)s = s^T (P - P_2)s \end{aligned}$$

Taking the expectation of the SSE we write

$$\begin{aligned} \text{Exp}\{ SSE_2 \} &= \text{Exp}\{ s^T (P - P_2)s \} \\ &= \text{Exp}\{ \text{trace}\{ s^T (P - P_2)s \} \} \end{aligned}$$

Theore

s_2^2 =

estima

estima

Proof:

plim s_1

Th

= $[s^T R$

$R_1^T s/N($

N ->

The

results

idempote

$= \text{Exp}\{ \text{trace}\{ (P - P_2)ss^T \} \}$

since trace(AB) = trace(BA) if both AB and BA are defined and square.

$= \text{trace}\{ (P - P_2)E\{ ss^T \} \}$

$= \text{trace}\{ (P - P_2)\{ qQ + rP \} \}$

since $E\{ ss^T \} = qQ + rP$.

$= (r)\text{trace}\{ (P - P_2) \}$ since $P_2 Q = 0$.

$= (r)\text{rank}(P - P_2)$

since trace(A) = rank(A) if A idempotent.

$= (r)\{\text{rank}(P) - \text{rank}(R_2)\}$ Q.E.D.

Theorem (2.8): Let $s_1{}^2 = SSE_1/\{ N(T - 1) - g \}$ and $s_2{}^2 = SSE_2/\{ N - g - k \}$. Then $s_1{}^2$ is a consistent estimator of q as N or T gets large, and $s_2{}^2$ is a consistent estimator of $r = \sigma_e{}^2 + T\sigma_u{}^2$ as N gets large.

Proof:

$\text{plim } s_1{}^2 = \text{plim } SSE_1/\{ \text{rank}(Q) - \text{rank}(R_1) \}$

$= \text{plim } SSE_1/N(T - 1) = \text{plim } s^T(Q - P_1)s/N(T - 1)$

$= \text{plim } s^T Qs/N(T - 1) - \text{plim } s^T P_1 s/N(T - 1)$

The last term is zero since $s^T P_1 s/N(T - 1)$ $= [s^T R_1/N(T - 1)][R_1{}^T R_1/N(T - 1)]^{-1} R_1{}^T s/N(T - 1)$ and $R_1{}^T s/N(T - 1) \rightarrow 0$ as $N(T - 1) \rightarrow \infty$ (as either $N \rightarrow \infty$ or $T \rightarrow \infty$).

The first term equals $\sigma_e{}^2$ because, using standard results (e.g. Rao (1973, p 185)) on the distribution of idempotent quadratic forms in normals, $s^T Qs$ is distributed as

$\sigma_e^2\chi^2_{N(T}$

plim s_2^2

The

= [$s^T R_2$

N -> o

The

standard

distribu

is distr

2.7 <u>Varia</u>

<u>Correla</u>

So f

feasible

consider

regressor

again we

and σ_u^2,

Taylor in

weighted

σ_e^2 based

is still

r = σ_e^2 +

consistent

squares ap

$\sigma_e^2 \chi^2_{N(T-1)}$.

$$\text{plim } s_2^2 = \text{plim } SSE_2/\{ \text{rank}(P) - \text{rank}(R_2) \}$$
$$= \text{plim } SSE_2/N = \text{plim } s^T(P - P_2)s/N$$
$$= \text{plim } s^T Ps/N - \text{plim } s^T P_2/N$$

The last term is zero since $s^T P_2 s/N = [s^T R_2/N][R_2^T R_2/N]^{-1} R_2^T s/N$ and $R_2^T s/N \to 0$ as $N \to \infty$.

The last term equals $r = \sigma_e^2 + T\sigma_u^2$ because, using standard results (e.g. Rao (1973, p 185)) on the distribution of idempotent quadratic forms in normals, $s^T Qs$ is distributed as $r\chi^2_N$. Q.E.D.

2.7 Variance Estimation when the Random Effects are Correlated with the Regressors

So far we have considered variance estimation for the feasible weighted least squares estimator only. We now consider the model of section 2.5, in which some of the regressors are correlated with the individual effects. Once again we will need to estimate the variance components σ_e^2 and σ_u^2, since they are needed to implement the Hausman and Taylor instrumental variables estimator (or the equivalent weighted instrumental variables estimator). The estimate of σ_e^2 based on the within residuals, discussed in section 2.6, is still consistent in this model. However, the estimate of $r = \sigma_e^2 + T\sigma_u^2$ which was discussed in section 2.6 is not consistent, since it was based on the residuals from least squares applied to (2.6.2), and this least squares estimator is incon

We

finding

consiste

consiste

is the w

estimate

give a r

assumpti

Assumpti

(i)

(ii)

(iii)

(iv)

(v)

Eve

assumed c

a consist

the distu

the probl

reduced t

following

an estima

Lemma (2.

(2.7.1)

is inconsistent when regressors and effects are correlated.

We therefore turn our attention to the problem of finding consistent estimates of B and D. Then, using these consistent estimates of $A_2 = (B^T , D^T)^T$, we derive a consistent estimate of r. The background for this approach is the work of Hausman and Taylor (1981), who suggest the estimate of r which we discuss here. However, they do not give a rigorous proof that it is consistent. The following assumptions will be made.

<u>Assumption (2.9)</u>: Let $H = [PX_1 , Z_1]$. Then we assume that

(i) plim $X^T Qe/N = 0$ as N -> oo.

(ii) plim $H^T P(e + U)/N = 0$ as N -> oo.

(iii) plim $(X^T QX)/N$ is finite and nonsingular as N -> oo.

(iv) plim $(H^T Z)/N$ is finite as N -> oo.

(v) plim $(H^T X)/N$ is finite as N -> oo.

Even after the introduction of X_2 and Z_2 - regressors assumed correlated with the effects - the within estimator is a consistent estimator of B; no correlation exists between the disturbance and the regressors in equation (2.5.5). So the problem of finding a consistent estimator of A_2 is reduced to finding a consistent estimator of D. The following regression equation will be used in deriving such an estimator.

<u>Lemma (2.10)</u>: Let $d^* = P(y - Xb_w)$. Then

(2.7.1) $d^* = ZD + (P - PX(X^T QX)^{-1} X^T Q)s.$

<u>**Proof**</u>:

$$\begin{aligned} d^* &= P(y - Xb_3) = Py - PXb_3 = Py - PX(X^TQX)^{-1}X^TQy \\ &= P(XB + ZD + s) - PX(X^TQX)^{-1}X^TQ(XB + ZD + s) \\ &= P(XB + ZD + s) - PX(X^TQX)^{-1}X^T(QXB + Qs) \\ &= P(XB + ZD + s) - PX(X^TQX)^{-1}X^TQXB + PX(X^TQX)^{-1}X^TQs \\ &= PXB + PZD + Ps - PXB + PX(X^TQX)^{-1}X^TQs \\ &= ZD + (P - PX(X^TQX)^{-1}X^TQ)s \qquad \text{Q.E.D.} \end{aligned}$$

Since part of Z is correlated with the error term, least squares applied to equation (2.7.1) does not yield a consistent estimator of D. But, using $H = (PX_1, Z_1)$ as a set of instruments, the instrumental variable estimator of D is defined as

(2.7.2) $$d_{IV} = (X^TP[H]X)^{-1}X^TP[H]d^*$$

where $P[H] = H(H^TH)^{-1}H^T$.

It is interesting to note that using $d^{**} = (y - Xb_W)$ instead of $d^* = P(y - Xb_W)$ would not increase the efficiency of the estimator, d_{IV}. Indeed, since $P[H]P = PP[H] = P[H]$, $Z_1^TQ = 0$, and the first order conditions (i.e. the "normal equations") defining b_W imply that $X_1^TQ(y - Xb_W) = 0$,

(2.7.3) $$H^TP(y - Xb_W) = H^T(y - Xb_W);$$

thus the <u>same</u> estimator would result if we used d^{**} in place of d^*.

Given the estimator d_{IV}, the question is whether this

estimator is, indeed, a consistent estimator of D. But first, we consider the conditions necessary to assure that d_{IV} does exist.

2.7.1 Necessary Conditions for the Existence of d_{IV}

A necessary condition for the existence of d_{IV} is that the rank of H be at least as large as the rank of Z; that is, there must be at least as many instruments as regressors. This requires $g_1 + k_1 \geq k$, or $g_1 \geq k_2$, as noted by Hausman and Taylor (1981). Intuitively, PX_1 is serving as instruments for Z_2, and so there must be at least as many variables in X_1 as in Z_2.

2.7.2 Consistency of d_{IV}

Lemma (2.11): Given assumption (2.9),

(1) plim $Z^T P_1 P(e + u)/N = 0$ as $N \to \infty$

(2) plim $Z^T P_1 Z/N$ is finite and non-singualar as $N \to \infty$

(3) plim $Z^T P_1 X/N$ is finite as $N \to \infty$

Lemma (2.11) can be easily proved by noting that $P_1 = P[H] = H(H^T H)^{-1} H^T$, where $H = P(X_1, Z_1)$.

Theorem (2.12): The instrumental variable estimator d_{IV} is a consistent estimator of D as $N \to \infty$.

Proof:

First, rewrite d_{IV} as

$$d_{IV} = (Z^T P_1 Z)^{-1} Z^T P_1 d^*$$

2

e

r

te

es

$$= (Z^T P_1 Z)^{-1} Z^T P_1 (ZD + (P - PX(X^T QX)^{-1} X^T Q)s)$$

$$= (Z^T P_1 Z)^{-1} Z^T P_1 ZD + (Z^T P_1 Z)^{-1} Z^T P_1 (P - PX(X^T QX)^{-1} X^T Q)s)$$

$$= D + (Z^T P_1 Z)^{-1} Z^T P_1 Ps - PX(X^T QX)^{-1} X^T Qs$$

$$= D + (Z^T P_1 Z/N)^{-1} \{Z^T P_1 Ps/N\}$$

$$- (Z^T P_1 Z/N)^{-1} \{Z^T P_1 PX/N\} (X^T QX/N)^{-1} \{X^T Qs/N\}$$

By assumption

$$\text{plim } X^T Q(e + u)/N = 0 \quad \text{as} \quad N \to \infty.$$

and

$$\text{plim } (X^T QX)/N \text{ is finite and nonsingular as } N \to \infty.$$

Using Lemma (2.11), it follows that

$$\text{plim } Z^T P_1 P(e + u)/N = 0 \quad \text{as} \quad N \to \infty,$$

$$\text{plim } (Z^T P_1 Z)/N \text{ is finite and nonsingular as } N \to \infty,$$

and

$$\text{plim } (Z^T P_1 PX)/N \text{ is finite as } N \to \infty.$$

Thus,

$$\text{plim } d_{IV} = D + \{\text{finite}\}\{ 0 \}$$

$$- \{ \text{finite} \}\{ \text{finite} \}\{ \text{finite} \}\{ 0 \}$$

$$= D \quad \text{as} \quad N \to \infty.$$ Q.E.D.

2.7.3 <u>A Consistent Estimate of r</u>

Using as a consistent estimate of $A_2 = (B^T , D^T)^T$ the estimators b_W and d_{IV}, we will now form a vector of residuals. We will then show that the sum of the squared terms of this residual vector, divided by N, is a consistent estimator of $r = \sigma_e^2 + T\sigma_u^2$.

Lemma (2.13): Let Residual = $Py - PXb_W - PZd_{IV}$. Then

$$\text{Residual} = P(e + u) - PX(X^TQX)^{-1}X^TQe - PZ(Z^TP_1Z)^{-1}Z^TP_1P(e + u) + PZ(Z^TP_1Z)^{-1}Z^TP_1PX(X^TQX)^{-1}X^TQe$$

Proof:

$$\begin{aligned}
\text{Residual} &= Py - PXb_W - PZd_{IV} \\
&= Py - PX(X^TQX)^{-1}X^TQy - PZ(Z^TP_1Z)^{-1}Z^TP_1d^* \\
&= P\{XB + ZD + WC + s\} \\
&\quad - PX(X^TQX)^{-1}X^TQ\{XB + ZD + WC + s\} \\
&\quad - PZ(Z^TP_1Z)^{-1}Z^TP_1\{ZD \\
&\quad + (P - PX(X^TQX)^{-1}X^TQ)(e + u)\} \\
&= PXB + ZD + P(e + u) \\
&\quad - PX(X^TQX)^{-1}X^TQXB - PX(X^TQX)^{-1}X^TQe \\
&\quad - PZ(Z^TP_1Z)^{-1}Z^TP_1ZD \\
&\quad - PZ(Z^TP_1Z)^{-1}Z^TP_1P(e + u) \\
&\quad + PZ(Z^TP_1Z)^{-1}Z^TP_1PX(X^TQX)^{-1}X^TQe \\
&= PXB + ZD + P(e + u) \\
&\quad - PXB - PX(X^TQX)^{-1}X^TQ(e + u) \\
&\quad - PZD \\
&\quad - PZ(Z^TP_1Z)^{-1}Z^TP_1P(e + u) \\
&\quad + PZ(Z^TP_1Z)^{-1}Z^TP_1PX(X^TQX)^{-1}X^TQe \\
&= P(e + u) - PX(X^TQX)^{-1}X^TQe \\
&\quad - PZ(Z^TP_1Z)^{-1}Z^TP_1P(e + u) \\
&\quad + PZ(Z^TP_1Z)^{-1}Z^TP_1PX(X^TQX)^{-1}X^TQe
\end{aligned}$$

Q.E.D.

We now define a consistent estimator for r. Define SSE^* as the sum of squared residuals defined in Lemma (2.13). Our

estimator is just SSE^*/N.

(2.7.4) $SSE^* = (\text{Residual})^T(\text{Residual})$.

Theorem (2.14): $\text{plim } SSE^*/N = r \equiv \sigma_e^2 + T\sigma_u^2$

Proof:

First, SSE^* can be written as

$$
\begin{aligned}
SSE^* &= (\text{Residual})^T(\text{Residual}) \\
&= (e + u)^T P(e + u) \\
&\quad - (e + u)^T PX(X^T QX)^{-1}X^T Qe \\
&\quad - (e + u)^T PZ(Z^T P_1 Z)^{-1}Z^T P_1 P(e + u) \\
&\quad + (e + u)^T PZ(Z^T P_1 Z)^{-1}Z^T P_1 PX(X^T QX)^{-1}X^T Qe \\
&\quad - e^T QX(X^T QX)^{-1}X^T P(e + u) \\
&\quad + e^T QX(X^T QX)^{-1}X^T PX(X^T QX)^{-1}X^T Qe \\
&\quad + e^T QX(X^T QX)^{-1}X^T PZ(Z^T P_1 Z)^{-1}Z^T P_1 P(e + u) \\
&\quad - e^T QX(X^T QX)^{-1}X^T PZ(Z^T P_1 Z)^{-1}Z^T P_1 PX(X^T QX)^{-1}X^T Qe \\
&\quad - (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T P(e + u) \\
&\quad + (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T PX(X^T QX)^{-1}X^T Qe \\
&\quad + (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T PZ(Z^T P_1 Z)^{-1}Z^T P_1 P(e + u) \\
&\quad - (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T PZ(Z^T P_1 Z)^{-1}Z^T P_1 PX(X^T QX)^{-1}X^T Qe \\
&\quad + e^T QX(X^T QX)^{-1}X^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T P(e + u) \\
&\quad - e^T QX(X^T QX)^{-1}X^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T PX(X^T QX)^{-1}X^T Qe \\
&\quad - e^T QX(X^T QX)^{-1}X^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T PZ(Z^T P_1 Z)^{-1}Z^T P_1 P(e + u) \\
&\quad + e^T QX(X^T QX)^{-1}X^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T PZ(Z^T P_1 Z)^{-1} \\
&\qquad \text{times } Z^T P_1 PX(X^T QX)^{-1}X^T Qe
\end{aligned}
$$

Now, from the above expression, taking the probability limit of SSE^* as N gets large is equivalent to taking the probability limit of the sum of sixteen different terms. Evaluation of these sixteen terms shows that the first term has a probability limit equal to r and that the remaining fifteen terms each have a probability limit equal to zero with all limits being taken as N -> oo. These probability limits are evaluated below.

1) $\text{plim } (e + u)^T P(e + u)/N$

$$= \text{plim } e^T Pe/N + \text{plim } u^T Pu/N$$

Consider these term by term. First,

$$e^T Pe/N = T\sum_{i=1}^{N} e_{i.}^2/N.$$

Each term $e_{i.}^2$ has a mean of σ_e^2/T, and the terms are independent. Therefore,

$$e^T Pe/N \to T\sigma_e^2/T = \sigma_e^2 \quad \text{as } N \to \infty.$$

Second,

$$u^T Pu/N = T\sum_{i=1}^{N} u_i^2/N \to T\sigma_u^2 \quad \text{as } N \to \infty.$$

Third,

$$e^T Pu/N = T\sum_{i=1}^{N} e_{i.}u_i/N \to 0 \quad \text{as } N \to \infty$$

because e and u are uncorrelated. Therefore,

$$(e + u)^T P(e + u)/N \to \sigma_e^2 + T\sigma_u^2 \quad \text{as } N \to \infty.$$

2) $\text{plim } (e + u)^T PX(W^T QW)^{-1} W^T Qe/N$

$$= \text{plim } \{(e + u)^T PW/N\}(W^T QW/N)^{-1}\{W^T Qe/N\}$$

5

6

7)

$= \text{plim } \{(e + u)^T PW/N\} \text{ plim } (W^T QW/N)^{-1} \text{ plim } \{W^T Qe/N\}$

$= 0 \quad \text{as} \quad N \to \infty.$

3) $\text{plim } (e + u)^T PZ(Z^T P_1 Z)^{-1} Z^T P_1 P(e + u)/N$

$= \text{plim } \{(e + u)^T PZ/N\}(Z^T P_1 Z/N)^{-1}\{Z^T P_1 P(e + u)/N\}$

$= \text{plim } \{(e + u)^T PZ/N\} \text{ plim } (Z^T P_1 Z/N)^{-1}$

$\text{times plim } \{Z^T P_1 P(e + u)/N\}$

$= 0 \quad \text{as} \quad N \to \infty.$

4) $\text{plim } (e + u)^T PZ(Z^T P_1 Z)^{-1} Z^T P_1 PW(W^T QW)^{-1} W^T Qe/N$

$= \text{plim } \{(e + u)^T PZ/N\}(Z^T P_1 Z/N)^{-1}$

$\text{times } \{Z^T P_1 PW/N\}(W^T QW/N)^{-1}\{W^T Qe/N\}$

$= \text{plim } \{(e + u)^T PZ/N\} \text{ plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T P_1 PW/N\}$

$\text{times plim } (W^T QW/N)^{-1} \text{ plim } \{W^T Qe/N\}$

$= 0 \quad \text{as} \quad N \to \infty.$

5) $\text{plim } e^T QW(W^T QW)^{-1} W^T P(e + u)/N$

$= \text{plim } \{e^T QW/N\}(W^T QW/N)^{-1}\{W^T P(e + u)/N\}$

$= \text{plim } \{e^T QW/N\} \text{ plim } (W^T QW/N)^{-1} \text{ plim } \{W^T P(e + u)/N\}$

$= 0 \quad \text{as} \quad N \to \infty.$

6) $\text{plim } e^T QW(W^T QW)^{-1} W^T PW(W^T QW)^{-1} W^T Qe/N$

$= \text{plim } \{e^T QW/N\}(W^T QW/N)^{-1}\{W^T PW/N\}(W^T QW/N)^{-1}\{W^T Qe/N\}$

$= \text{plim } \{e^T QW/N\} \text{ plim } (W^T QW/N)^{-1} \text{ plim } \{W^T PW/N\}$

$\text{times plim } (W^T QW/N)^{-1} \text{ plim } \{W^T Qe/N\}$

$= 0 \quad \text{as} \quad N \to \infty.$

7) $\text{plim } e^T QW(W^T QW)^{-1} W^T PZ(Z^T P_1 Z)^{-1} Z^T P_1 P(e + u)/N$

$= \text{plim } \{e^T QW/N\}(W^T QW/N)^{-1}\{W^T PZ/N\}$

$\text{times } (Z^T P_1 Z/N)^{-1}\{Z^T P_1 P(e + u)/N\}$

$= \text{plim}\ \{e^T QW/N\}\ \text{plim}\ (W^T QW/N)^{-1}\ \text{plim}\ \{W^T PZ/N\}$

times $\text{plim}\ (Z^T P_1 Z/N)^{-1}\ \text{plim}\ \{Z^T P_1 P(e + u)/N\}$

$= 0$ as $N \to \infty$.

8) $\text{plim}\ e^T QW(W^T QW)^{-1} W^T PZ(Z^T P_1 Z)^{-1} Z^T P_1 PW(W^T QW)^{-1} W^T Qe/N$

$= \text{plim}\ \{e^T QW/N\}(W^T QW/N)^{-1}\{W^T PZ/N\}(Z^T P_1 Z/N)^{-1}\{Z^T P_1 PW/N\}$

times $(W^T QW/N)^{-1}\{W^T Qe/N\}$

$= \text{plim}\ \{e^T QW/N\}\ \text{plim}\ (W^T QW/N)^{-1}$

times $\text{plim}\ \{W^T PZ/N\}\text{plim}\ (Z^T P_1 Z/N)^{-1}$

times $\text{plim}\ \{Z^T P_1 PW/N\}\ \text{plim}\ (W^T QW/N)^{-1}\ \text{plim}\ \{W^T Qe/N\}$

$= 0$ as $N \to \infty$.

9) $\text{plim}\ (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1} Z^T P(e + u)/N$

$= \text{plim}\ \{(e + u)^T PP_1 Z/N\}(Z^T P_1 Z/N)^{-1}\{Z^T P(e + u)/N\}$

$= \text{plim}\ \{(e + u)^T PP_1 Z/N\}\ \text{plim}\ (Z^T P_1 Z/N)^{-1}$

times $\text{plim}\ \{Z^T P(e + u)/N\}$

$= 0$ as $N \to \infty$.

10) $\text{plim}\ (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1} Z^T PW(W^T QW)^{-1} W^T Qe/N$

$= \text{plim}\ \{(e + u)^T PP_1 Z/N\}(Z^T P_1 Z/N)^{-1}$

times $\{Z^T PW/N\}(W^T QW/N)^{-1}\{W^T Qe/N\}$

$= \text{plim}\ \{(e + u)^T PP_1 Z/N\}\ \text{plim}\ (Z^T P_1 Z/N)^{-1}\ \text{plim}\ \{Z^T PW/N\}$

times $\text{plim}\ (W^T QW/N)^{-1}\ \text{plim}\ \{W^T Qe/N\}$

$= 0$ as $N \to \infty$.

11) $\text{plim}\ (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1} Z^T PZ(Z^T P_1 Z)^{-1} Z^T P_1 P(e + u)/N$

$= \text{plim}\ \{(e + u)^T PP_1 Z/N\}(Z^T P_1 Z/N)^{-1}\{Z^T PZ/N\}$

times $(Z^T P_1 Z/N)^{-1}\{Z^T P_1 P(e + u)/N\}$

$$= \text{plim } \{(e + u)^T PP_1 Z/N\} \text{ plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T PZ/N\}$$

$$\text{times plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T P_1 P(e + u)/N\}$$

$$= 0 \quad \text{as } N \to \infty.$$

12) $\text{plim } (e + u)^T PP_1 Z(Z^T P_1 Z)^{-1} Z^T PZ(Z^T P_1 Z)^{-1}$

times $Z^T P_1 PW(W^T QW)^{-1} W^T Qe/N$

$$= \text{plim } \{(e + u)^T PP_1 Z/N\}(Z^T P_1 Z/N)^{-1}\{Z^T PZ/N\}(Z^T P_1 Z/N)^{-1}$$

$$\text{times } \{Z^T P_1 PW/N\}(W^T QW/N)^{-1}\{W^T Qe/N\}$$

$$= \text{plim } \{(e + u)^T PP_1 Z/N\} \text{ plim } (Z^T P_1 Z/N)^{-1} \text{ plim}\{Z^T PZ/N\}$$

$$\text{times plim}(Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T P_1 PW/N\}$$

$$\text{times plim } (W^T QW/N)^{-1} \text{ plim } \{W^T Qe/N\}$$

$$= 0 \quad \text{as } N \to \infty.$$

13) $\text{plim } e^T QW(W^T QW)^{-1} W^T PP_1 Z(Z^T P_1 Z)^{-1} Z^T P(e + u)/N$

$$= \text{plim } \{e^T QW/N\}(W^T QW/N)^{-1}\{W^T PP_1 Z/N\}$$

$$\text{times } (Z^T P_1 Z/N)^{-1}\{Z^T P(e + u)/N\}$$

$$= \text{plim } \{e^T QW/N\} \text{ plim } (W^T QW/N)^{-1} \text{ plim } \{W^T PP_1 Z/N\}$$

$$\text{times plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T P(e + u)/N\}$$

$$= 0 \quad \text{as } N \to \infty.$$

14) $\text{plim } e^T QW(W^T QW)^{-1} W^T PP_1 Z(Z^T P_1 Z)^{-1} Z^T PW(W^T QW)^{-1} W^T Qe/n$

$$= \text{plim } \{e^T QW/N\} \text{ plim } (W^T QW/N)^{-1} \text{ plim } \{W^T PP_1 Z/N\}$$

$$\text{times plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T PW/N\}$$

$$\text{times plim } (W^T QW/N)^{-1} \text{ plim } \{W^T Qe/N\}$$

$$= 0 \quad \text{as } N \to \infty.$$

15) $\text{plim } e^T QW(W^T QW)^{-1} W^T PP_1 Z(Z^T P_1 Z)^{-1}$

times $Z^T PZ(Z^T P_1 Z)^{-1} Z^T P_1 P(e + u)/N$

$$= \text{plim } \{e^T QW/N\}(W^T QW/N)^{-1}\{W^T PP_1 Z/N\}(Z^T P_1 Z/N)^{-1}$$
$$\text{times } \{Z^T PZ/N\}(Z^T P_1 Z/N)^{-1}\{Z^T P_1 P(e + u)/N\}$$
$$= \text{plim } \{e^T QW/N\} \text{ plim } (W^T QW/N)^{-1} \text{ plim } \{W^T PP_1 Z/N\}$$
$$\text{times } \text{plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T PZ/N\}$$
$$\text{times } \text{plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T P_1 P(e + u)/N\}$$
$$= 0 \quad \text{as } N \to \infty.$$

16) $$\text{plim } \{ e^T QW(W^T QW)^{-1}W^T PP_1 Z(Z^T P_1 Z)^{-1}Z^T PZ(Z^T P_1 Z)^{-1}$$
$$\text{times } Z^T P_1 PW(W^T QW)^{-1}W^T Qe/N \}$$
$$= \text{plim } \{e^T QW/N\}(W^T QW/N)^{-1}\{W^T PP_1 Z/N\}$$
$$\text{times } \text{plim } (Z^T P_1 Z/N)^{-1} \{Z^T PZ/N\}(Z^T P_1 Z/N)^{-1}$$
$$\text{times } \text{plim } \{Z^T P_1 PW/N\}(W^T QW/N)^{-1}\{W^T Qe/N\}$$
$$= \text{plim } \{e^T QW/N\} \text{ plim } (W^T QW/n)^{-1} \text{ plim } \{W^T PP_1 Z/N\}$$
$$\text{times } \text{plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T PZ/N\}$$
$$\text{times } \text{plim } (Z^T P_1 Z/N)^{-1} \text{ plim } \{Z^T P_1 PW/N\}$$
$$\text{times } \text{plim } (W^T QW/N)^{-1} \text{ plim } \{W^T Qe/N\}$$
$$= 0 \quad \text{as } N \to \infty.$$

Q.E.D.

2.8 Conclusions

In this chapter, we have considered a linear regression model which contains unobserved individual effects. Given panel data, this model may be estimated in a variety of ways, depending on what is assumed about the correlation between the regressors and the effects. We have given a survey of the literature, tidying up a few loose ends, and we have introduced the analytical framework to be used in the rest of the thesis. In the next chapter, we will extend the analysis of this chapter to a model which contains unobservable time effects as well as individual effects.

CHAPTER 3

Individual and Time Effects

3.1 Introduction

In this chapter, we extend the linear regression model considered in the previous chapter to include unobservable time effects. We again assume that the data consists of T time-series observations on each of N individuals; we distinguish regressors which vary over time and individuals from those that are either time-invariant or individual-invariant; and now we assume the presence of unobservable time-invariant individual effects, unobservable individual-invariant time effects, and the usual statistical noise.

We write the model to be considered in this chapter as

$$(3.1.1) \quad y_{it} = X_{it}B + W_tC + Z_iD + u_i + v_t + e_{it}, \quad i = 1,\ldots,N;\ t = 1,\ldots,T.$$

where y_{it} is the dependent variable, X_{it} is a vector (of dimension 1 x g) of explanatory variables which vary both over time and over individuals, Z_i is a vector (of dimension 1 x k) of time-invariant explanatory variables, W_t is a vector (of dimension 1 x h) of individual-invariant explanatory variables, and B, D, and C are vectors of parameters to be estimated. The errors e_{it} are iid with mean zero and variance σ_e^2. Both the individual effects u_i and the time effects v_t are unobservable, and various assumptions about them will be made. However, in all cases the individual effects will be treated as time-invariant and the time effects will be treated as individual-invariant.

The plan of this chapter is as follows. In section 3.2 we review the geometry which is used in subsequent analyses. We then consider the estimation of the model under various assumptions. In section 3.3 we consider the fixed effects model, in which the individual effects are treated as fixed parameters to be estimated. The point of this model is to remove the potential bias caused by correlation of the regressors with the omitted individual-invariant and time-invariant variables. In section 3.4 we consider the random effects model, in which the individual and time effects are treated as random and uncorrelated with the regressors. Under these assumptions there is no problem of bias, and efficiency of estimation is our central concern. In section 3.5 we consider an extended version of the model of Hausman and Taylor (1981), in which the individual effects are treated as random but potentially correlated with the regressors. Since many currently available panel data sets are characterized by having many observations but for only a relatively few time periods, in section 3.6 we consider the previous two models for the case when N is large but T is fixed. Finally, in sections 3.7 and 3.8 we consider the problem of consistent estimation of the variances of the noise, the individual effects, and the time effects. Such estimates are necessary to implement the feasible weighted least squares estimators considered in section 3.4 and 3.5.

This chapter applies the Hausman and Taylor method of instrumental variables estimation to the panel data model extended to included both individual as well as time effects, and derives the subsequent estimator. In addition, it provides a survey of the existing literature on this extended model. The analysis of the regression models considered in this chapter is done using the approach introduced in chapter 2.

3.2 Geometry

A useful fact, and one to be used throughout the remainder of this chapter, is that the equation (3.1.1) can be written, equivalently, as the four orthogonal equations

$$(3.2.1)\quad (y_{it} - y_{.t} - y_{i.} + y_{..}) = (X_{it} - X_{.t} - X_{i.} + X_{..})B + (e_{it} - e_{.t} - e_{i.} + e_{..})$$

$$(3.2.2)\quad (y_{i.} - y_{..}) = (X_{i.} - X_{..})B + (Z_i - Z_.)D + (u_i - u_.) + (e_{i.} - e_{..})$$

$$(3.2.3)\quad (y_{.t} - y_{..}) = (X_{.t} - X_{..})B + (W_t - W_.)C + (v_t - v_.) + (e_{.t} - e_{..})$$

$$(3.2.4)\quad y_{..} = X_{..}B + W_.C + Z_.D + u_. + v_. + e_{..}$$

where $y_{i.} = (1/T)\sum_{t=1}^{T} y_{it}$, $y_{.t} = (1/N)\sum_{i=1}^{N} y_{it}$, and $y_{..}$

$= (1/NT)\sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$. Equation (3.2.2) expresses the data in terms of its individual averages over time with the grand mean subtracted, while equation (3.2.3) expresses the data in terms of its averages over individuals for each period of time with the grand mean subtracted. Equation (3.2.1) expresses the data in terms of its deviations around both the mean for each individual and for each time period with the grand mean added; equation (3.2.4) expresses the data in terms of its grand mean.

Writing equation (3.1.1) in matrix form we have

$$(3.2.5) \qquad y = XB + WC + ZD + u + v + e$$

where y, u, v, and e denote (NT x 1) dimensioned vectors; and X, W, and Z denote (NT x g), (NT x k), and (NT x h) dimensioned matrices, respectively. Again, following the convention of Hausman and Taylor (1981), the observations are ordered first by individuals and then by time, so that v and each column of W are (NT x 1) dimensioned vectors consisting of N blocks, with each block containing the same T entries.

To achieve the same decomposition as was accomplished above, we define the same four symmetric, idempotent, mutually orthogonal matrices used by Fuller and Battese (1974). These orthogonal projections are

$$(3.2.6) \qquad Q_1 = I_{NT} - Q_2 - Q_3 - Q_4$$

$$(3.2.7) \qquad Q_2 = (I_N \otimes j_T j_T^T/T) - (j_{NT} j_{NT}/NT)$$

$$(3.2.8) \qquad Q_3 = (j_N j_N^T/N \otimes I_T) - (j_{NT} j_{NT}^T/NT)$$

(3.2.9) $Q_4 = (j_N j_N^T/N \otimes j_T j_T^T/T) = (j_{NT} j_{NT}^T/NT)$

where $j_T = (1,\ldots 1)^T$ is $(T \times 1)$. The transformation Q_4 determines the grand mean for the NT observations repeated NT times. The transformation Q_2 determines the means for each of the individual groups, subtracts the grand mean, and repeats these N observations T times; the transformation Q_3 determines the means for each of the time periods, subtracts the grand mean, and repeats these T observations N times. The transformation Q_1 transforms each observation into the difference between itself and both its respective individual group mean and time mean, and then adds the grand mean. Explicitly, the (i,t) element of Q_1y, Q_2y, Q_3y, and Q_4y can be written as

(3.2.10) $(Q_1y)_{it} = y_{it} - y_{i.} - y_{.t} + y_{..}$

(3.2.11) $(Q_2y)_{it} = y_{i.} - y_{..}$

(3.2.12) $(Q_3y)_{it} = y_{.t} - y_{..}$

(3.2.13) $(Q_4y)_{it} = y_{..}$,

respectively.

Since W contains variables that are constant across all individual observations for a given time period, $Q_1W = 0$. Similiarly, $Q_2W = 0$. The elements of the columns of W are, on the other hand, expressed as deviations from their respective grand means by the transformation Q_3. Analogous results hold true for the time effects v; i.e. $Q_1v = 0$ and

$Q_2 v = 0$. Likewise, since Z contains variables that are constant across all time period observations for a given individual, $Q_1 Z = 0$ and $Q_3 Z = 0$. And similiarly, $Q_1 u = 0$ and $Q_3 u = 0$. Thus, the original equation (3.1.5) can now be written equivalently as the four orthogonal equations

(3.2.14) $Q_1 y = Q_1 XB + Q_1 e$

(3.2.15) $Q_2 y = Q_2 XB + Q_2 ZD + Q_2 (u + e)$

(3.2.16) $Q_3 y = Q_3 XB + Q_3 WC + Q_3 (v + e)$

(3.2.17) $Q_4 y = Q_4 XB + Q_4 WC + Q_4 ZD + Q_4 (u + v + e)$

3.3 Fixed Effects

In this section, we discuss the estimation of the linear regression equation (3.2.5) when both the individual-specific effects and the time-specific effects are treated as fixed constants. The standard approach is to use dummy variables for individuals and for time periods as regressors, and then to apply least squares. This yields the following estimator for B:

(3.3.1) $b_w = (X^T Q_1 X)^{-1} X^T Q_1 y$.

The estimator b_w is the familiar within-group estimator; it uses only the variation within each individual group and each time period. The estimator is unbiased, and it is consistent as either N or T (or both) approaches infinity. These are all well-known results; for example, see Judge et al. (1985, pp. 338).

A problem with this estimation procedure is that it is not possible to obtain estimates of either the coefficients of the time-invariant regressors (Z) or the coefficients of the individual-invariant regressors (W). The time invariant regressors are perfectly collinear with the individual dummy variables and the individual-invariant regressors are perfectly collinear with the time dummy variables; equivalently, they are removed by the transformation of the data by the matrix Q_1. If the original model contained no time-invariant regressors, the estimated coefficients of the individual dummy variables are

(3.3.2) $u_w = Q_2 y - Q_2 X b_w$,

and these estimates of the individual effects are consistent as T approaches infinity. If the original model contained time-invariant regressors, then u_w defined above is interpreted as an estimate of $(Q_2 ZD + Q_2 u)$ rather than of just u. Simliarly, if the original model contained no individual-invariant regressors, the estimated coefficients of the time period dummy variables are

(3.3.3) $v_w = Q_3 y - Q_3 X b_w$,

and these estimates of the time effects are consistent as N approaches infinity. If the original model contained individual-invariant regressors, then v_w defined above is interpreted as an estimate of $(Q_3 WC + Q_3 v)$ rather than of just v.

An equivalent derivation of the within estimator b_w is to define it as the least squares estimator in equation (3.2.14), ignoring equation (3.2.15), (3.2.16), and (3.2.17). Similiarly, the estimator u_w is least squares applied to (3.2.15), after setting B = b_w; ignoring the time-invariant variables Z. And, the estimator v_w is least squares applied to (3.2.16), after setting B = b_w, and ignoring the individual-invariant variables W.

Using only one part of equation (3.2.5), namely equation (3.2.1), when estimating B has the advantage of being computationally more convenient than estimating the whole of equation (3.2.5). This approach also makes explicit the statement that b_w ignores the between-group variation and the between time period variation; i.e. it ignores the cross-sectional variation in equation (3.2.15) and the time series variation in equation (3.2.16).

3.4 Random Effects, not Correlated with Regressors

In the previous section, we discussed the estimation of a linear regression model when the individual effects (the u_i) and the time effects (the v_t) are treated as fixed constants. In this section, we treat the individual and time effects similiarly to the way we treat the error term e_{it}; we assume both the u_i and the v_t to be random variables uncorrelated with the regressors. The N individuals are now interpreted as also being drawn from some larger population, and so too the effects u_i can be viewed as a random sample from some distribution. Similiarly, the T time periods are now interpreted as being drawn from some larger population, and so the effects v_t can be viewed as a random sample from some distribution.

We assume specifically that the u_i are iid with mean zero and variance σ_u^2, the v_t are iid with mean zero and variance σ_v^2, and the u_i and v_t are assumed to be uncorrelated with each other as well as with e_{it}. We also assume that X, Z, and W are uncorrelated with both u and v. The model is written as

$$(3.4.1)\quad y_{it} = X_{it}B + Z_iD + W_tC + u_i + v_t + e_{it}$$
$$= X_{it}B + Z_iD + W_tC + s_{it},\ i = 1,\ldots,N;\ t = 1,\ldots,T$$

The variance of y_{it}, conditional on X_{it}, Z_i, and W_t is

$$(3.4.2)\quad \text{var}(y_{it}) = \text{var}(s_{it}) = \sigma_u^2 + \sigma_v^2 + \sigma_e^2.$$

Therefore, this model is often referred to as the generalized error-components or generalized variance-components model.

The presence of the random effects u_i and v_t in the disturbance term results in correlation among the errors for a given individual as well as among different time series. This can be made explicit if we let s_i denote the (T x 1) dimensioned error vector $(s_{i1},\ldots,s_{iT})^T$. The covariance matrix of s_i is then the matrix

$$(3.4.3)\quad \text{Cov}(\ s_i\ ) = \sigma_u^2(j_T j_T^T) + \sigma_v^2 I_T + \sigma_e^2 I_T$$

where $j_T = (1,\ldots,1)^T$ is a (T x 1) vector of 1's. Furthermore, the covariance between s_i and s_j is given by the matrix

(3.4.4) $Exp(s_i s_j^T) = \sigma_v^2 I_T$.

3.4.1 Within Estimation

The Within-group estimator can be used regardless of whether the u_i's and v_t's are viewed as fixed constants or as random variables. The Within estimator of B can be viewed as least squares applied to equation (3.2.14), and neither the individual effects nor the time effects appear in this equation. So, whether the u_i and v_t are treated as nonstochastic or stochastic, the estimator b_w is still unbiased and consistent. However, as pointed out by Judge et al. (1985), the Within estimator is inefficient when the effects are random and uncorrelated with the regressors.

3.4.2 Generalized Least Squares Estimation

As was shown above, since the s_{it} in different time periods but for the same individual both contain u_i, the errors in the equation

(3.4.5) $y_{it} = X_{it}B + Z_iD + W_tD + u_i + v_t + e_{it}$

$= X_{it}B + Z_iD + W_tD + s_{it}$, $i = 1,\ldots,N$; $t = 1,\ldots,T$

are autocorrelated, and since the s_{it} in the same time period but for different individuals both contain v_t, the errors are intertemporally correlated. Efficient estimation requires that we use the generalized least squares method. Following Fuller and Battese (1974), we write

$$(3.4.6) \quad S = \mathrm{Cov}(s) = pQ_1 + qQ_2 + rQ_3 + kQ_4$$

where $p = \sigma_e^2$, $q = (\sigma_e^2 + T\sigma_u^2)$, $r = (\sigma_e^2 + N\sigma_v^2)$, and $k = (\sigma_e^2 + T\sigma_u^2 + N\sigma_v^2)$. Since the four matrices Q_1, Q_2, Q_3, and Q_4 are idempotent and orthogonal, it follows that

$$(3.4.7) \quad S^{-1} = (1/p)Q_1 + (1/q)Q_2 + (1/r)Q_3 + (1/k)Q_4$$

Now, if we rewrite equation (3.2.5) as

$$(3.4.7) \quad y = XB + ZC + WC + u + v + e$$
$$= RA + s$$

where $R = (X, Z, W)$ and $A = (B^T, C^T, D^T)^T$, and if we assume that σ_u^2, σ_v^2, and σ_e^2 are known, the generalized least squares estimator of A from equation (3.4.7) is simply

$$(3.4.8) \quad a_{GLS} = (R^T S^{-1} R)^{-1} R^T S^{-1} y.$$

Equivalently, the GLS estimator is ordinary least squares of $(S^{-1/2}y)$ on $(S^{-1/2}R)$. Fuller and Battese (1974, pp. 77) show that, up to a factor of proportionality,

$$(3.4.9) \quad S^{-1/2} = I_{NT} - (1-c_2)Q_2 - (1-c_3)Q_3 + (1+c_4)Q_4$$

where $c_2 = (p/q)^{1/2}$, $c_3 = (p/r)^{1/2}$, and $c_4 = (p/k)^{1/2}$, so that the GLS estimator can more conviently be calculated using the transformed variables

$$(3.4.10) \quad S^{-1/2}y = y - (1-c_2)Q_2 y - (1-c_3)Q_3 y + (1+c_4)Q_4 y$$

(and similiarly for R). For example,

$$(3.4.11)\quad (S^{-1/2}y)_{it} = y_{it} - (1-c_2)y_{i.} - (1-c_3)y_{.t} + (1+c_4)y_{..}$$

and this differs from the within transformation to the extent that the scalars c_2, c_3, and c_4 are nonzero. As pointed out by Hsiao (1986), the GLS estimator converges to the Within estimator when N - > oo, T -> oo, and the ratio of N over T trends to a non-zero constant. It can be shown that c_2 tends to zero as T gets large, that c_3 tends to zero as N gets large, and c_4 tends to zero as T gets large and the ratio of N over T is bounded from above.

The GLS estimator is consistent, as pointed out by Judge et al. (1985), when both N -> oo and T -> oo; it is not consistent as N -> oo for T fixed or as T -> oo for N fixed. The case when N -> oo for T fixed will be discussed in more detail in section 3.6.1.

3.4.3 Weighted Least Squares Estimation

As an alternative approach to the generalized least squares estimator of A, consider the equations which result from the decomposition of equation (3.4.7). These orthogonal equations can be written as

$$(3.4.12)\quad Q_1 y = Q_1 XB + Q_1 e$$

$$(3.4.13)\quad Q_2 y = Q_2 XB + Q_2 ZC + Q_2 (u + e)$$

$$(3.4.14)\quad Q_3 y = Q_3 XB + Q_3 WD + Q_3 (v + e)$$

$$(3.4.15)\quad Q_4 y = Q_4 XB + Q_4 ZC + Q_4 WD + Q_4 (u + v + e)$$

We note that the covariance matrices associated with the errors in the above equations may be written (respectively) as

(3.4.16) $\text{Cov}(\ Q_1 s\ ) = Q_1 S Q_1 = pQ_1$

(3.4.17) $\text{Cov}(\ Q_2 s\ ) = Q_2 S Q_2 = qQ_2$

(3.4.18) $\text{Cov}(\ Q_3 s\ ) = Q_3 S Q_3 = rQ_3$

(3.4.19) $\text{Cov}(\ Q_4 s\ ) = Q_4 S Q_4 = kQ_4$

Each of these four covariance matrices is of the form of a constant times an idempotent matrix. These four constants may be equated by multiplying equations (3.4.12), (3.4.13), (3.4.14), and (3.4.15), by the weights (1/p), (1/q), (1/r), and (1/s), respectively. Moreover, it follows from Lemma (2.1) that least squares applied to the system so weighted yields the best (minimum variance) unbiased estimator within the class containing all least squares estimator of the parameter vector A from any further transformation of these equations or, in fact, the original equation (3.4.7). We will refer to the least squares estimator of A from the system of orthogonal equations

(3.4.20) $(1/p)Q_1 y = (1/p)Q_1 XB + (1/p)Q_1$

(3.4.21) $(1/q)Q_2 y = (1/q)Q_2 XB + (1/q)Q_2 ZC + (1/q)Q_2 s$

(3.4.22) $(1/r)Q_3 y = (1/r)Q_3 XB + (1/r)Q_3 WD + (1/r)Q_3 s$

$$(3.4.23) \quad (1/k)Q_4 y = (1/k)Q_4 XB + (1/k)Q_4 ZC + (1/k)Q_4 WD + (1/k)Q_4 k$$

as the weighted least squares estimator of A. This estimator may be written as

$$(3.4.24) \quad a_{WLS} = (R^T Q_1 R/p + R^T Q_2 R/q + R^T Q_3 R/r + R^T Q_4 R/k)^{-1} \text{ times } (R^T Q_1/p + R^T Q_2/q + R^T Q_3/r + R^T Q_4/k)y$$

The decomposition of the original equation by the transformations, Q_1, Q_2, Q_3, and Q_4, has the effect of isolating the correlations found in the non-block diagonal covariance matrix of its error vector, S, to the particular orthogonal space. Since these transformations are orthogonal, and their sum is the identity matrix, equation (3.4.7) is said to have been reduced by the quadruple (Q_1, Q_2, Q_3, Q_4) into the four orthogonal equations (3.4.12), (3.4.13), (3.4.14), and (3.4.15). Since these four equations contain exactly the same information as the orthogonal equation, we would expect that the minimum variance unbiased estimator from the four equations would be equivalent to the generalized least squares estimator from the original equation. This result is stated in the following theorem.

Theorem (3.1): The weighted least squares estimator, a_{WLS}, is equal to the generalized least squares estimator, a_{GLS}.

<u>Proof</u>:

The generalized least squares estimator of A from equation (3.4.7) can be rewritten as

(3.4.25) a_{GLS}

$$= (R^T S^{-1} R)^{-1} R^T S^{-1} y$$

$$= (R^T[(1/p)Q_1 + (1/q)Q_2 + (1/r)Q_3 + (1/k)Q_4]R)^{-1}$$

$$\text{times} \quad R^T[(1/p)Q_1 + (1/q)Q_2 + (1/r)Q_3 + (1/k)Q_4]y$$

$$= (R^T Q_1 R/p + R^T Q_2 R/q + R^T Q_3 R/r + R^T Q_4 R/k)^{-1}$$

$$\text{times} \quad (R^T Q_1/p + R^T Q_2/q + R^T Q_3/r + R^T Q_4/k)y$$

Therefore, $a_{WLS} = a_{GLS}$. Q.E.D.

Now least squares applied to equation (3.4.13) is called the between-individual estimator; $a_I = (R^T Q_2 R)^{-1} R^T Q_2 y$. It utilizes the variation between individuals. Least squares applied to equation (3.4.14) is called the between-time period estimator; $a_T = (R^T Q_3 R)^{-1} R^T Q_3 y$. It utilizes variation the between time periods. Recall that the within estimator can be viewed as least squares applied to equation (3.4.12); it utilizes the residual variation. Maddala (1971) claims that the generalized least squares estimator can be viewed as an efficient combination of the above three estimators. The optimal weights for the three different sets of variation are the constants being used to normalize each of the equations; i.e. the reciprocal of the variances $p = \sigma_e^2$, $q = \sigma_e^2 + T\sigma_u^2$, and $r = \sigma_e^2 + N\sigma_v^2$ for the respective equations, (3.4.12), (3.4.13), (3.4.14). Since the weighted least squares estimator has been shown to be equivalent to the generalized least squares estimator, this would imply that equation (3.4.15) may be dropped and the weighted least squares estimator computed using the remaining three equations only. Indeed, equation (3.4.15) only determines the constant term and, therefore, dropping this equation and omitting the constant term, leaves the estimates for the other coefficients unchanged. We prove this in the next theorem.

Theorem (3.2): Weighted least squares applied to the set of equations (3.4.20), (3.4.21, (3.4.22), and (3.4.23) is equivalent to weighted least squares applied to the first three equations only.

Proof:

We rewrite equation (3.4.7) as

(3.4.26) $$y = RA + s = R_1 A_1 + R_2 A_2 + s$$

where $R_1 = (1,\ldots,1)^T$ is a (NT x 1) vector of 1's and A_1 is the constant term. From Schmidt (1983), the generalized least squares estimator of A_2 is

(3.4.27) $$a_2 = (R^T MR)R^T My$$

where $M = S^{-1} - S^{-1}A_1(A_1^T S^{-1} A_1)^{-1} A_1^T S^{-1}$. But for our S^{-1} and A_1, a straightforward calculation shows

(3.4.28) $$M = S^{-1} - (1/s)Q_4 = (1/p)Q_1 + (1/q)Q_2 + (1/r)Q_3$$ Q.E.D.

3.5 Random Effects, Correlated with Regressors

In some applications of the error-component model, there may be reasons to believe that either the individual-specific or the time-specific unobservable effects found in the error term may, in fact, be correlated with some of the included explanatory variables. If we take the view suggested earlier, that the random effects represent both omitted individual-specific and time-specific variables, this correlation would seem inevitable. When there is correlation between the random effects and the explanatory variables, the generalized least squares estimator is biased and inconsistent.

We consider the case in which the effects are correlated with some of the regressors. To consider this case, we first need to introduce some notation. Consider the equation

$$(3.5.1)\quad y_{it} = (X_{1it}, X_{2it}, X_{3it}, X_{4it})B + (Z_{1i}, Z_{2i})D + (W_{1t}, W_{2t})C + u_i + v_t + e_{it}$$

where X_{1it} represents the $(1 \times g_1)$ dimensioned vector of time and individual varying explanatory variables, Z_{1i} represents the $(1 \times k_1)$ dimesioned vector of time-invariant explanatory variables, and W_{1t}'s represent the $(1 \times h_1)$ dimensioned vector of individual-invariant explanatory variables, all of which are assumed to be uncorrelated with the three errors, u_i, v_t, and e_{it}. The $(1 \times g_2)$ dimensioned vector of time and individual-varying explanatory variables, X_{2it}, and the $(1 \times k_2)$ dimensioned vector of time-invariant explanatory variables, Z_{2i}, are both assumed to be correlated with u_i but uncorrelated with v_t and e_{it}. The $(1 \times g_3)$ dimensioned vector of time and individual-varying explanatory variables, X_{3it}, and the $(1 \times h_2)$ dimensioned vector of individual-invariant explanatory variables, W_{2i}, are both assumed to be correlated with v_t but uncorrelated with u_i and e_{it}. Finally, the $(1 \times g_4)$ dimensioned vector of time- and individual-varying explanatory variables, X_{4it}, is assumed to be correlated with u_i and v_t but uncorrelated with e_{it}. As before, the random noise component, e_{it}, the individual effects, u_i, and the time effects, v_t, are i.i.d. as well as independent of one another.

We note in passing that the variables X, which vary over both individuals and time, may be correlated or not with both the individual effects u_i and the time effects v_t. Thus there are four possible kinds of X's. However, the variables Z are time-invariant, and can not possibly be correlated with the time effects; there are only two kinds of Z's, correlated or not with the individual effects. Similiarly, the variables W are individual-invariant, and can not possibly be correlated with the individual effects; there are only two kinds of W's, correlated or not with the time effects.

The matrix form of equation (3.5.1) can be written as

$$(3.5.2) \quad y = (X_1, X_2, X_3, X_4)B + (Z_1, Z_2)D + (W_1, W_2)C + u + v + e$$

where y, u, v, and e are $(NT \times 1)$; X is $(NT \times g)$, $g = g_1 + g_2$

$+ g_3 + g_4$; Z is (NT x k), $k = k_1 + k_2$; and W is (NT x h), h $= h_1 + h_2$.

3.5.1 Weighted Least Squares Estimation

We decompose equation (3.5.2) into the three orthogonal equations

$$(3.5.3) \quad Q_1 y = Q_1 X_1 B_1 + Q_1 X_2 B_2 + Q_1 X_3 B_3 + Q_1 X_4 B_4 + Q_1 (e)$$

$$(3.5.4) \quad Q_2 y = Q_2 X_1 B_1 + Q_2 X_2 B_2 + Q_2 X_3 B_3 + Q_2 X_4 B_4 + Q_2 Z_1 D_1 + Q_2 Z_2 D_2 + Q_2 (e + u)$$

$$(3.5.5) \quad Q_3 y = Q_3 X_1 B_1 + Q_3 X_2 B_2 + Q_3 X_3 B_3 + Q_3 X_4 B_4 + Q_3 W_1 C_1 + Q_3 W_2 C_2 + Q_3 (e + v)$$

Now since $Q_1 v = 0$ and $Q_1 u = 0$, there is no correlation between errors and regressors in (3.5.3). Furthermore,

$$(3.5.6) \quad H_2 = [\ Q_2 X_1 ,\ Q_2 X_3 ,\ Q_2 Z_1\ ]$$

can readily be seen to be the largest available set of variables in equation (3.5.4) which have been assumed uncorrelated with the indivdual effects. Likewise,

$$(3.5.7) \quad H_3 = [\ Q_3 X_1 ,\ Q_3 X_2 ,\ Q_3 W_1\ ]$$

can readily be seen to be the largest available set of variables in equation (3.5.5) which have been assumed uncorrelated with the time effects. Projecting equation (3.5.4) onto the column space of H_2 and projecting equation (3.5.5) onto the column space of H_3, we have the set of orthogonal equations

$$(3.5.8) \qquad Q_1 y = Q_1 RA + Q_1(e)$$

$$(3.5.9) \qquad P_2 Q_2 y = P_2 Q_2 RA + P_2 Q_2(e + u)$$

$$(3.5.10) \qquad P_3 Q_3 y = P_3 Q_3 RA + P_3 Q_3(e + v)$$

where $P_2 = P[H_2]$ and $P_3 = P[H_3]$.

The covariance matrix associated with the errors in equations (3.5.8), (3.5.9), and (3.5.10) can be written as

$$(3.5.11) \qquad Cov(Q_1 e) = pQ_1$$

$$(3.5.12) \qquad Cov(P_2 Q_2(e + u)) = qP_2$$

and

$$(3.5.13) \qquad Cov(P_3 Q_3(e + v)) = rP_3 ,$$

respectively. We note that each of these three covariance matrices has the form of a constant times an idempotent matrix. Thus, Lemma (2.1) would imply that any further attempt at diagonalizing the covariance matrices in any of the equations would not improve the efficiency of the resulting estimator. Using the weights p, q, and r, the weighted least squares estimator of A from equations (3.5.8), (3.5.9), and (3.5.10) becomes

$$(3.5.14) \qquad a_{WLS} = \{(1/p)R^T Q_1 R + (1/q)R^T P_2 R + (1/r)R^T P_3 R\}^{-1} \text{ times } \{(1/p)R^T Q_1 + (1/q)R^T P_2 + (1/r)R^T P_3\}y$$

It is possible to derive the estimator (3.5.14) without decomposing the equation into orthogonal spaces, as follows. First, we multiply equation (3.5.2) by $S^{-1/2}$ to transform the error term so that it has a scalar covariance matrix. The transformed equation is simply

$$(3.5.15) \quad S^{-1/2}y = S^{-1/2}RA + S^{-1/2}s$$

Second, we note that the maximal set of available instruments for equation (3.5.2) may be written as

$$(3.5.16) \quad H^{*} = [\; Q_1,\; Q_2X_1,\; Q_2X_3,\; Q_2Z_1,\; Q_3X_1,\; Q_3X_2,\; Q_3W_1 \;].$$

We then follow the path of Hausman and Taylor (1981), by estimating (3.5.15) using IV with instrument set H^{*}. This yields

$$(3.5.17) \quad a_{IV} = \{ R^{T}V^{-1/2}P^{*}V^{-1/2}R \}^{-1}\{ R^{T}V^{-1/2}P^{*}V^{-1/2} \}y$$

where $P^{*} = P[\; H^{*} \;]$.

We can evaluate $P[\; H^{*} \;]$ using the following Lemma:

<u>Lemma (3.3)</u>: $P[\; H^{*} \;] = Q_1 + P[\; H_2 \;] + P[\; H_3 \;]$

$$\equiv Q_1 + P_2 + P_3$$

The efficient instrumental variables estimator of A from equation (3.5.2) can then be written as

$$(3.5.18) \quad a_{IV} = \{ R^{T}S^{-1/2}(Q_1 + P_2 + P_3)S^{-1/2}R \}^{-1}$$

$$\text{times } \{ R^{T}S^{-1/2}(Q_1 + P_2 + P_3)S^{-1/2} \}y$$

$$= \{ R^{T}(\; (1/p)Q_1 + (1/q)P_2 + (1/r)P_3)S^{-1/2}R \}^{-1}$$

$$\text{times } \{ R^{T}(\; (1/p)Q_1 + (1/q)P_2 + (1/r)P_3) \}y$$

But this is simply the weighted least squares instrumental variables estimator. We have therefore proved the following theorem.

<u>Theorem (3.4)</u>: The efficient instrumental variables estimator of A equals weighted least squares applied to equations (3.5.8), (3.5.9), and (3.5.10).

3.5.2 Counting Rules for Identification

Following Hausman and Taylor and corresponding to the familiar rank condition we have the theorem:

<u>Theorem (3.5)</u>: A necessary and sufficient condition that the vector of parameters A be identified in equation (3.4.7) is that the matrix $R^T P^* R$ be non-singular.

Corresponding to the order condition, we have the following theorem:

<u>Theorem (3.6)</u>: A necessary condition for the identification of A in equation (3.4.7) is that (i) $g_1 + g_3 \geq k_2$ and (ii) $g_1 + g_2 \geq h_2$.

<u>Proof</u>:

Since $P^* R = (Q_1 + P_2 + P_3)(\ X\ Z\ W\ ) = (\ Q_1 X\ 0\ 0\ ) + (\ P_2 X\ P_2 Z\ 0\ ) + (\ P_3 X\ 0\ P_3 W\ ) = (\ P^* X\ P_2 Z\ P_3 W\ )$, $\text{rank}(P^* R) = \text{rank}(P^* X) + \text{rank}(P_2 Z) + \text{rank}(P_3 W)$. It follows that a necessary condition for the matrix $R^T P^* R$ to be non-singular is that $\text{rank}(P^* R) = g$, $\text{rank}(P_2 R) = k$, and $\text{rank}(P_3 R) = h$. Now $\text{rank}(P_2 R) = \min \{\ \text{rank}(P_2),\ \text{rank}(R)\ \} = g_1 + g_2 + k_1$.

Similiarly, rank(P_3R) = $g_1 + g_3 + h_1$ and rank(P^*R) = g. Thus, a necessary condition for identification is that $g_1 + g_2 + k_1 \geq k$ and $g_1 + g_3 + h_1 \geq h$. Q.E.D.

Therefore, to insure that the parameters of the model are identified requires that the parameters in each of the three equations, (3.5.3), (3.4.4), and (3.4.5), separately be identified.

Theorem (3.7): Given the rank condition of Theorem (3.6), weighted least squares applied to equations (3.5.8), (3.5.9), and (3.5.10) is a consistent estimator for A.

Proof:

Weighted least squares applied to equations (3.5.8), (3.5.9), and (3.5.10) can be written as

$$a_{WLS}$$

$$= \{(1/p)R^TQ_1R + (1/q)R^TP_2R + (1/r)R^TP_3R\}^{-1}$$

$$\text{times } \{(1/p)R^TQ_1 + (1/q)R^TP_2 + (1/r)R^TP_3\}y$$

$$= A + \{ (1/p)R^TQ_1R + (1/q)R^TP_2R + (1/r)R^TP_3R \}^{-1}$$

$$\text{times } \{(1/p)R^TQ_1 + (1/q)R^TP_2 + (1/r)R^TP_3\}(u + v + e)$$

Since the estimator exists,

$\lim \{ (1/p)R^TQ_1R + (1/q)R^TP_2R + (1/r)R^TP_3R \}^{-1}$ is finite as both N -> oo and T -> oo. Next, consider

$$\{ (1/p)R^TQ_1 + (1/q)R^TP_2 + (1/r)R^TP_3 \}(u + v + e)/NT$$

$$= (1/p)R^TQ_1(u + v + e)/NT + (1/q)R^TP_2(u + v + e)/NT$$
$$+ (1/r)R^TP_3(u + v + e)/NT$$

$$= (1/p)R^T Q_1(u + v + e)/NT$$

$$+ (1/q)(R^T H_2/NT)(H_2^T H_2/NT)^{-1} H_2^T(u + v + e)/NT$$

$$+ (1/r)(R^T H_3/NT)(H_3^T H_3/NT)^{-1} H_3^T(u + v + e)/NT$$

where

$$R^T Q_1(u + v + e) = \begin{bmatrix} X^T Q_1 e \\ 0 \\ 0 \end{bmatrix},$$

$$H_2^T(u + v + e) = \begin{bmatrix} X_1^T Q_2(u + e) \\ X_3^T Q_2(u + e) \\ Z_1^T Q_2(u + e) \end{bmatrix}, \text{ and}$$

$$H_2^T(u + v + e) = \begin{bmatrix} X_1^T Q_3(v + e) \\ X_2^T Q_3(v + e) \\ W_1^T Q_3(v + e) \end{bmatrix}.$$

As we can easily show,

$$\text{plim } X^T Q_1 e/NT = 0 \quad \text{as} \quad N \to \infty \text{ or } T \to \infty,$$

$$\text{plim } X_1^T Q_2(u + e)/NT = 0 \quad \text{as} \quad N \to \infty,$$

$$\text{plim } X_3^T Q_2(u + e)/NT = 0 \quad \text{as} \quad N \to \infty,$$

$$\text{plim } Z_1^T Q_2(u + e)/NT = 0 \quad \text{as} \quad N \to \infty,$$

$$\text{plim } X_1^T Q_3(v + e)/NT = 0 \quad \text{as} \quad T \to \infty,$$

$$\text{plim } X_2^T Q_3(v + e)/NT = 0 \quad \text{as} \quad T \to \infty,$$

and

$$\text{plim } W_1^T Q_3(v + e)/NT = 0 \quad \text{as} \quad T \to \infty.$$

Therefore, $\text{plim } R^T Q_1(u + v + e)/NT = \text{plim } H_2^T(u + v + e)/NT$ $= \text{plim } H_3^T(u + v + e)/NT = 0$ as $N \to \infty$ and as $T \to \infty$.

Since the estimator exists, lim $(R^T H_2/NT)(H_2{}^T H_2/NT)^{-1}$ is finite as N -> oo and lim $(R^T H_3/NT)(H_3{}^T H_3/NT)^{-1}$ is finite as T -> oo. Thus,

$$\begin{aligned} & \text{plim } (1/p)R^T Q_1(u + v + e)/NT \\ &= \text{plim } (1/q)(R^T H_2/NT)(H_2{}^T H_2/NT)^{-1} H_2{}^T(u + v + e)/NT \\ &= \text{plim } (1/r)(R^T H_3/NT)(H_3{}^T H_3/NT)^{-1} H_3{}^T(u + v + e)/NT \\ &= 0 \end{aligned}$$

as both N -> oo and T -> oo.

It follows that, plim a_{WLS} = A as both N - > oo and T -> oo. Q.E.D.

The weighted least squares estimator is not consistent as N -> oo for T fixed or when T -> oo for N fixed. The case when N -> oo for T fixed will be discussed in more detail in section 3.6.2.

3.6 Random Effects when T is Fixed

In the previous two sections, we have derived GLS and IV estimators which are useful only when both N and T are large. In this section, we will be concerned with the case in which N is large and T is small. This is the situation most common in panel or longitudinal data.

(3.6.1) Random Effects not Correlated with the Regressors

For the present we will assume that the regressors (X, Z, and W) are all uncorrelated with the error components e, u, and v. Now unbiased estimation is still possible in the case of small (or fixed) T. The problem which does arise is the inability to obtain consistent estimates from applying either least squares or generalized least squares to the above equation. To see this, consider equation (3.4.7) multiplied by $S^{-1/2}$. We then have

$$(3.6.1) \quad S^{-1/2}y = S^{-1/2}XB + S^{-1/2}ZD + S^{-1/2}WC + S^{-1/2}(e + u + v)$$

where $S^{-1/2} = (Q_1/p^* + Q_2/q^* + Q_3/r^* + Q_4/k^*)$, $p^* = (\sigma^2_e)^{1/2}$; $q^* = (\sigma^2_e + T\sigma^2_u)^{1/2}$, $r^* = (\sigma^2_e + N\sigma^2_u)^{1/2}$, and $k^* = (\sigma^2_e + \sigma^2_u + \sigma^2_v)^{1/2}$.

Evaluating the probability limit of the cross product of the transformed regressor $S^{-1/2}X$ and the transformed error component $S^{-1/2}v$, we find that

$$(3.6.2) \quad \text{plim } X^T S^{-1} v/NT$$

$$= \text{plim } X^T (Q_1/p + Q_2/q + Q_3/r + Q_4/k)v/NT$$

$$= \text{plim } X^T (Q_3 v/r + Q_4 v/k)/NT$$

$$\text{since } Q_1 v = Q_2 v = 0$$

$$= \text{plim } (1/r)X^T (j_N j_N^T/N \otimes I_T)v/NT$$

$$+ \text{plim } (1/r)X^T Q_4 v/NT - \text{plim } (1/k)X^T Q_4 v/NT$$

This probability limit does not equal zero as N -> oo since

$$(3.6.3) \quad \text{plim } (1/r)X^T (j_N j_n^T/N \otimes I_T)v/NT$$

$$= \text{plim } (1/r)\sum_{t=1}^{T} (\sum_{i=1}^{N} X_{jt}/N)^T v_t/T = 0$$

only as T -> oo. Therefore, we have the problem of the regressors being correlated with one of the error components in the sense that their cross-moments have a nonzero probability limit for the case when only N -> oo. Thus, for the case of fixed T, the generalized least squares estimator of the coefficient vectors in equation (3.4.7) will not be consistent. Furthermore, exactly the same problem arises with ordinary least squares. For example, $X^T v/NT$ has a non-zero probability limit as N -> oo with T fixed. Only if both N and T -> oo will ordinary least squares be consistent.

A proposed solution to this problem is to apply weighted least squares to a subset of the equations in the decomposition of equation (3.4.7). Unfortunately, the coefficient vector C is no longer estimable. The weighted least squares estimator we derive is for the vector of coefficients $(B^T, D^T)^T$. Consider the decomposition of equation

$$(3.6.4) \quad y = XB + ZD + WC + u + v + e$$

into the four orthogonal equations

$$(3.6.5) \quad Q_1 y = Q_1 XB + Q_1 e$$

$$(3.6.6) \quad Q_2 y = Q_2 XB + ZD + Q_2 (u + e)$$

$$(3.6.7) \quad Q_3 y = Q_3 XB + WC + Q_3 (v + e)$$

$$(3.6.8) \quad Q_4 y = Q_4 XB + Q_4 ZD + Q_4 WC + Q_4 (u + v + e)$$

Using Theorem (3.2), we know equation (3.6.8) may be dropped without affecting the estimation of the remaining equations. In addition, equation (3.6.7) must be dropped for it is the source of the present problem. That is, it is equation

(3.6.7) from which comes the matrix of cross moments that has a nonzero probability limit unless T gets large. Thus, the estimator of $(B^T, D^T)^T$ will be derived by applying weighted least squares to the two remaining equations; namely equations (3.6.5) and (3.6.6).

Let R = (X, Z). Then the weighted least squares estimator of $(B^T, D^T)^T$ from equation (3.6.5) and (3.6.6) can be written as

$$(3.6.9) \quad \begin{bmatrix} b_{WLS} \\ d_{WLS} \end{bmatrix} = (R^T Q_1 R/p + R^T Q_2 R/q)^{-1}(R^T Q_1/p + R^T Q_2/q)y.$$

__Assumption (3.5)__: $\lim (R^T Q_1 R/p + R^T Q_2 R/q)$ as $N \to \infty$ is finite and nonsingular.

__Theorem (3.6)__: The weighted least squares estimator of $(B^T, D^T)^T$ from equations (3.6.5) and (3.6.6) is consistent as $N \to \infty$, with T fixed.

__Proof__:

The weighted least squares estimator can be written as

$$\begin{matrix} b_{WLS} \\ d_{WLS} \end{matrix} = (R^T Q_1 R/q + R^T Q_2 R/q)^{-1}(R^T Q_1/p + R^T Q_2/q)y$$

$$= \begin{bmatrix} B \\ D \end{bmatrix} + (R^T Q_1 R/p + R^T Q_2 R/q)^{-1}(R^T Q_1/p + R^T Q_2/q)(u + e).$$

$$= \begin{bmatrix} B \\ D \end{bmatrix}$$

$$+ \quad \{((1/p)R^T Q_1 R/NT + (1/q)R^T Q_2 R/NT)^{-1}$$

$$\text{times } ((1/p)R^T Q_1/NT + (1/q)R^T Q_2/NT)(u + e)\}.$$

Consider the second term. First, ($(1/p)R^T Q_1 R/NT$ + $(1/q)R^T Q_2 R/NT$)$^{-1}$ is by assumption finite as N gets large. Next, we can write ($R^T Q_1/p + R^T Q_2/q$)(u + e) = $(1/p)R^T Q_1 e$ + $(1/q)R^T Q_2(u + e)$, where

$$(3.6.10) \qquad R^T Q_1 e = \begin{bmatrix} X^T Q_1 e \\ 0 \end{bmatrix}$$

and

$$(3.6.11) \qquad R^T Q_2(u + e) = \begin{bmatrix} X^T Q_2(u + e) \\ Z^T Q_2(u + e) \end{bmatrix} .$$

As we can easily show,

(3.6.13) plim $X^T Q_1 e/NT$ = 0 as N -> oo or T -> oo

(3.6.14) plim $X^T Q_2(u + e)/NT$ = 0 as N -> oo

and

(3.6.14) plim $Z^T Q_1(u + e)/NT$ = 0 as N -> oo.

Thus, plim $R^T Q_1(u + e)/NT$ = plim $R^T Q_2(u + e)/NT$ = 0 as N -> oo. Hence, plim {$(1/p)R^T Q_1 e/NT + (1/q)R^T Q_2(u + e)/NT$} = 0 as N -> oo.

Therefore, plim $\begin{bmatrix} b_{WLS} \\ d_{WLS} \end{bmatrix} = \begin{bmatrix} B \\ D \end{bmatrix}$ as N -> oo. Q.E.D.

(3.6.2) Random Effects Correlated with the Regressors

We again consider the case when T is fixed but, in addition, we consider the case in which the effects are correlated with some of the regressors. To this end, we first re-introduce some notation. Consider the equation

$$(3.6.15) \quad y = XB + ZD + WC + u + v + e$$

where we again assume that $X = (X_1, X_2, X_3, X_4)$, $Z = (Z_1, Z_2)$, and $W = (W_1, W_2)$. That is, X_1, Z_1, and W_1 denote $(NT \times g_1)$, $(NT \times k_1)$, and $(NT \times h_1)$ dimensioned matrices, respectively, all assumed to be uncorrelated with e, u, and v; X_2 and Z_2 denote $(NT \times g_2)$ and $(NT \times k_2)$ dimensioned matrices, respectively, both assumed to be correlated with u but uncorrelated with e and v; X_3 and W_2 denote $(NT \times g_3)$ and $(NT \times h_2)$ dimensioned matrices, respectively, both assumed to be correlated with v but uncorrelated with e and u; and X_4 denotes a $(NT \times g_4)$ dimensioned matrix which is assumed to be correlated with both u and v but uncorrelated with e.

Now, not only is the weighted least squares estimator of A biased, but so is the weighted least squares estimator of $(B^T, D^T)^T$ derived in section 3.6.1. This bias is due to the presence of regressors which are assumed correlated with the equation's error term. One approach to consistent estimation of $(B^T, D^T)^T$ is to apply the instrumental variables, weighted least squares method to the equations (3.6.5) and (3.6.6). In this section we will derive such an estimator and show it to be consistent.

First we consider the equations

$$(3.6.16) \quad Q_1 y = Q_1(X_1, X_2, X_3, X_4)B + Q_1 e$$

$$(3.6.17) \quad Q_2 y = Q_2(X_1, X_2, X_3, X_4)B + Q_2(Z_1, Z_2)D + Q_2(u + e)$$

Since $Q_1 u = Q_1 v = 0$ and $Q_1 W = 0$, there is no problem of correlation between errors and regressors in (3.6.16). Furthermore, as we show in appendix A, the set

$$(3.6.18) \quad H_0 = [\ Q_2 X_1,\ Q_2 X_3,\ Q_2 Z_1\ ]$$

contains legitimate instruments for equation (3.6.17). It can readily be seen that H_0 is the largest available set of variables in equation (3.6.17) which have been assumed uncorrelated with the individual effects found in that equation. Projecting equation (3.6.17) onto the column space of H_0, we have the set of orthogonal equations

$$(3.6.19) \quad Q_1 y = Q_1(X_1, X_2, X_3, X_4)B + Q_1 e$$

$$(3.6.20) \quad P_2 y = P_2(X_1, X_2, X_3, X_4)B + P_2(Z_1, Z_2)D + P_2(u + e)$$

where $P_2 = P[H_0]$.

The covariance matrix associated with the errors in equations (3.6.19) and (3.6.20) can be written as

$$(3.6.21) \quad \mathrm{Cov}(\ Q_1 e\ ) = pQ_1$$

and

(3.6.22) $\text{Cov}(P_2(u + e)) = qP_2$,

respectively. We note that each of these two covariance matrices has the form of a constant times an idempotent matrix. Thus, Lemma (2.1) would imply that any further attempt at diagonalizing the covariance matrices in either equation (3.6.19) or (3.6.20) would not improve the efficiency of the resulting estimator. Using the weights p and q, the weighted least estimator of $(B^T, D^T)^T$ from equations (3.6.19) and (3.6.20) becomes

(3.6.23) $$\begin{bmatrix} b_{IV} \\ d_{IV} \end{bmatrix} = \{R^T Q_1 R/p + R^T P_2 R/q\}^{-1}\{R^T Q_1 y/p + R^T P_2 y/q\}$$

where $R = (X_1, X_2, X_3, X_4, Z_1, Z_2)$.

We first derive the necessary conditions for the existence of the above estimator, and then show it to be consistent for fixed T.

Corresponding to the order condition, we have the following theorem:

Theorem (3.7): A necessary condition for the weighted least squares estimator of $(B^T, D^T)^T$ from equations (3.6.19) and (3.6.20) to exist is that $g_1 + g_3 \geq k_2$.

Proof:

The existence of the IV estimator depends on the matrix

$$\begin{bmatrix} Q_1 R \\ P_2 R \end{bmatrix} = \begin{bmatrix} Q_1 X_1 & Q_1 X_2 & Q_1 X_3 & Q_1 X_4 & 0 & 0 \\ P_2 X_1 & P_2 X_2 & P_2 X_3 & P_2 X_4 & P_2 Z_1 & P_2 Z_2 \end{bmatrix}$$

being of full rank. But for this matrix to be of full rank it is necessary that (P_2Z_1, P_2Z_2) be of full rank. And since $(P_2Z_1, P_2Z_2) = (Z_1, P[Q_1X_1, Q_1X_3, Q_1Z_1]Z_2)$, it follows that a necessary condition for the existence of the estimator is that $\text{rank}(X_1, X_3) \geq \text{rank}(Z_2)$; or that $g_1 + g_3 \geq k_2$.

Theorem (3.8): Given the rank condition of theorem (3.7), weighted least squares applied to equations (3.6.19) and (3.6.20) is a consistent estimator for $(B^T, D^T)^T$ when T is fixed.

Proof:

Weighted least squares applied to equations (3.6.19) and (3.6.20) can be written as

$$\begin{bmatrix} b_{IV} \\ d_{IV} \end{bmatrix} = \{R^TQ_1R/p + R^TP_2R/q\}^{-1}\{R^TQ_1/p + R^TP_2/q\}y$$

$$= \begin{bmatrix} B \\ D \end{bmatrix} + \{R^TQ_1R/p + R^TP_2R/q\}^{-1}\{R^TQ_1/p + R^TP_2/q\}(u + e)$$

Since the estimator exists, $\lim \{R^TQ_1R/p + R^TP_2R/q\}^{-1}$ is finite as $N \to \infty$. Next, consider

$$\{R^TQ_1/p + R^TP_2/q\}(u + e)/NT$$
$$= (1/p)R^TQ_1(u + e)/NT + (1/q)R^TP_2(u + e)/NT$$
$$= (1/p)R^TQ_1e/NT + (1/q)R^TH_0(H_0^TH_0)^{-1}H_0^T(u + e)/NT$$
$$= (1/p)R^TQ_1e/NT + (1/q)(R^TH_0/NT)(H_0^TH_0/NT)^{-1}(H_0^T(u + e)/NT)$$

where

$$R^TQ_1e = \begin{bmatrix} X^TQ_1e \\ 0 \end{bmatrix}$$

and

$$H_0^T(u + e) = \begin{bmatrix} X_1^T Q_2(u + e) \\ X_3^T Q_2(u + e) \\ Z_1^T Q_2(u + e) \end{bmatrix}$$

As we can easily show,

$$\text{plim } X^T Q_1 e/NT = 0 \text{ as } N \to \infty \text{ or } T \to \infty,$$

$$\text{plim } X_1^T Q_2(u + e)/NT = 0 \text{ as } N \to \infty,$$

$$\text{plim } X_3^T Q_2(u + e)/NT = 0 \text{ as } N \to \infty,$$

and

$$\text{plim } Z_1^T Q_2(u + e)/NT = 0 \text{ as } N \to \infty.$$

Therefore,

$$\text{plim } (1/p)R^T Q_1(u + e)/NT = \text{plim } H_0^T(u + e)/NT = 0$$

as $N \to \infty$. Since the estimator exists,

$\lim (R^T H_0/NT)(H_0^T H_0/NT)^{-1}$ is finite as $N \to \infty$. Thus,

$$\text{plim } \{ (1/p)R^T Q_1(u + e)/NT + (1/q)(R^T H_0)(H_0^T H_0)^{-1}(H_0^T(u + e)/NT \} = 0 \text{ as } N \to \infty.$$

It follows that, $\text{plim} \begin{bmatrix} b_{IV} \\ d_{IV} \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}$ as $N \to \infty$.

Q.E.D.

(3.6.3) An Alternative Approach to Estimation

The above approach to estimating $(B^T, D^T)^T$ is an extension of the analytical method used throughout this chapter. Instead, we could follow a naive extension of the analytical method used in chapter 2. In the simple model of chapter 2 for the case when random indiviual effects are present consistency of the least squares estimator requires

$N \to \infty$. There the problem was that we had present regressors correlated with one of the error components in the sense that their cross-moments have a nonzero probabiltiy limit for the case when only $T \to \infty$. This is because the effect of the random component u_i can be averaged out only in the direction of that component. That is, probability limits of terms like $X^T Pu/NT = \sum_{i=1}^{N} (\sum_{t=1}^{T} X_{it}/T)u_i/N$ are to equal zero only as $N \to \infty$. A solution to this problem was to construct a transformation P which determined the means for each of the individual groups and repeats these N observations T times. The within transformation, $Q = I_{NT} - P$, then transforms each observation into the difference between itself and its respective indiviual group mean. Premultiplying equation (2.4.7) by the within transformation eliminated the individual random effects and so the need for $N \to \infty$. Thus, least squares applied to the transformed equation turns out to be consistent as $T \to \infty$.

Now however we are not interested in individual effects but rather in eliminating time effects and the need for $T \to \infty$ so we construct a projection similiar to P but in the other direction. To this end, we define

(3.6.24) $P^* = (j_N j_N^T/N \otimes I_T)$ and $Q^* = I_{NT} - P^*$

where $j^T = (1,\ldots,1)^T$ is (T x 1). The transformation P^* determines the means for each of the time periods and repeats each of these T observations N times. The transformation Q^* transforms each observation into the difference between itself and its respective time period mean. Explicitly, the (i,t) elements of P^*y and Q^*y can be written as

$$(3.6.25) \qquad (P^*y)_{it} = y_{.t} \quad \text{and} \quad (Q^*y)_{it} = y_{it} - y_{.t},$$

respectively. We note that in terms of our previous notation $P^* = Q_3 + Q_4$ and $Q^* = Q_1 + Q_2$.

Since W contains variables that are constant across all individual observations for a given time period, $Q^*W = 0$. The elements of the columns of W are, on the other hand, unaffected by the transformation P^*; that is, $P^*W = W$. Analogous results hold true for the time effects v; that is, $Q^*v = 0$ and $P^*v = v$. Thus, the original equation (3.6.4) can now be written equivalently as the two orthogonal equations

$$(3.6.26) \qquad P^*y = P^*XB + P^*ZD + P^*WC + P^*(u + v + e)$$

$$(3.6.27) \qquad Q^*y = Q^*XB + Q^*ZD \qquad\qquad + P^*(u + e)$$

Equation (3.6.27) represents the original model after being purged of the time effects. Of course the coefficients of the individual-invariant regressors cannot be estimated but OLS applied to equation (3.6.27) would yield a consistent estimator $(B^T, D^T)^T$.

<u>**Assumption (3.9)**</u>: $\lim R^TQ^*R$ as $N \to \infty$ is finite and nonsingular.

<u>**Theorem (3.10)**</u>: The least squares estimator of $(B^T, D^T)^T$ from equation (3.6.27) is consistent as $N \to \infty$.

Proof:

The least squares estimator can be written as

$$\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix} = (R^T Q^* R)^{-1} R^T Q^* y$$

$$= \begin{bmatrix} B \\ D \end{bmatrix} + (R^T Q^* R)^{-1} R^T Q^* (u + e).$$

Consider the second term. First, $(R^T Q^* R)^{-1}$ is by assumption finite as N gets large. Next, we can write

$(R^T Q^*)(u + e) = R^T Q_1 (u + e) + R^T Q_2 (u + e)$, where

$$(3.6.28) \qquad R^T Q_1 (u + e) = \begin{bmatrix} X^T Q_1 e \\ 0 \end{bmatrix}$$

and

$$(3.6.29) \qquad R^T Q_2 (u + e) = \begin{bmatrix} X^T Q_2 e \\ Z^T Q_2 (u + e) \end{bmatrix}.$$

As we can easily show,

$$(3.6.30) \quad \text{plim } X^T Q_1 e/NT = 0 \text{ as } N \to \infty \text{ or } T \to \infty,$$

$$(3.6.31) \quad \text{plim } X^T Q_2 (u + e)/NT = 0 \text{ as } N \to \infty,$$

and

$$(3.6.32) \quad \text{plim } Z^T Q_2 (u + e)/NT = 0 \text{ as } N \to \infty.$$

Thus, plim $R^T Q_1 (u + e)/NT$ = plim $R^T Q_2 (u + e)/NT = 0$ as $N \to \infty$. Hence, plim $\{ R^T Q^* (u + e)/NT \} = 0$ as $N \to \infty$.

Therefore, plim $\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix} = \begin{bmatrix} B \\ D \end{bmatrix}$ as $N \to \infty$. Q.E.D.

The OLS estimator from equation (3.6.27) can be viewed as unweighted version of the WLS estimator from equations (3.6.5) and (3.6.6) as we can see by comparing the following with equation (3.6.9).

$$\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix} = (R^T Q^* R)^{-1} R^T Q^* y$$

$$= (R^T Q_1 R + R^T Q_2 R)^{-1} (R^T Q_1 + R^T Q_2) y$$

Since WLS weights the two equations (3.6.5) and (3.6.6) optimally, we would expect weighted least squares to be efficient relative to least squares. This is shown in the following theorem.

__Theorem (3.11)__: The weighted least squares estimator of $(B^T , D^T)^T$ from equations (3.6.5) and (3.6.6) is asymptotically efficient (as $N \to \infty$) relative to the least squares estimator from equation (3.6.27). If p and q are known then weighted least squares is also efficient relative to least squares in finite samples.

__Proof__:

We prove the finite sample case; the other case is similiar. Let $Cov(u + e) \equiv \Omega_* = pQ_1 + qQ_2$ so $\Omega_*^{-1} = (1/p)Q_1 + (1/q)Q_2$. Then

$$\text{Cov}\begin{bmatrix} b_{WLS} \\ d_{WLS} \end{bmatrix}$$

$$= (R^T Q_1 R/p + R^T Q_2 R/q)^{-1} (R^T Q_1/p + R^T Q_2/q)(pQ_1 + qQ_2)$$

$$\text{times} \quad (Q_1 R/p + Q_2 R/q)(R^T Q_1 R/p + R^T Q_2 R/q)^{-1}$$

$$= (R^T Q_1 R/p + R^T Q_2 R/q)^{-1} (R^T Q_1 R/p + R^T Q_2 R/q)$$

$$\text{times} \quad (R^T Q_1 R/p + R^T Q_2 R/q)^{-1}$$

$$= (R^T Q_1 R/p + R^T Q_2 R/q)^{-1} = (R^T (Q_1/p + Q_2/q) R)^{-1}$$

$$= (R^T O_*^{-1} R)^{-1}$$

And

$$\text{Cov}\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix} = (R^T Q^* R)^{-1} R^T Q^* O_* Q^* R (R^T Q^* R)^{-1}$$

Now to show

$$\text{Cov}\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix} - \text{Cov}\begin{bmatrix} b_{WLS} \\ d_{WLS} \end{bmatrix}$$

is psd it is sufficient to show that

$$\text{Cov}\begin{bmatrix} b_{WLS} \\ d_{WLS} \end{bmatrix}^{-1} - \text{Cov}\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix}^{-1}$$

is psd. But the latter expression can be written as

$$(R^T O_*^{-1} R) - R^T Q^* R (R^T Q^* O_* Q^* R)^{-1} R^T Q^* R$$

$$= (R^T O_*^{-1} R) - R^T Q^* R (R^T Q^* O_* Q^* R)^{-1} R^T Q^* R$$

$$= R^T O_*^{-1/2} [I - O_*^{1/2} Q^* R (R^T Q^* O_* Q^* R)^{-1} R^T Q^* O_*^{1/2}] O_*^{-1/2} R$$

$$= R^T O_*^{-1/2} [I - D(D^T D)^{-1} D^T] O_*^{-1/2} R$$

where $D = O_*^{1/2}Q^*R$, which we see to be a quadratic form in an idempotent matrix; hence, our expression is psd. Q.E.D.

A similiar approach can be applied to the case when T is fixed but, in addition, the effects are correlated with <u>some</u> of the regressors. Using the notation of section 3.6.2 we consider the equation

$$(3.6.37) \quad y = (X_1, X_2, X_3, X_4)B + (Z_1, Z_2)D + (W_1, W_2)C + u + v + e$$

Here again, X_1, Z_1, and W_1 denote $(NT \times g_1)$, $(NT \times k_1)$, and $(NT \times h_1)$ dimensioned matrices, respectively, all assumed to be uncorrelated with e, u, and v; X_2 and Z_2 denote $(NT \times g_2)$ and $(NT \times k_2)$ dimensioned matrices, respectively, both assumed to be correlated with u but uncorrelated with e and v; X_3 and W_2 denote $(NT \times g_3)$ and $(NT \times h_2)$ dimensioned matrices, respectively, both assumed to be correlated with v but uncorrelated with e and u; and X_4 denotes a $(NT \times g_4)$ dimensioned matrix which is assumed to be correlated with both u and v but uncorrelated with e.

As shown in section 3.6.2, the weighted least squares estimator of $(B^T, D^T)^T$ derived in section 3.6.1 is biased due to the presence of regressors assumed correlated with the equation's error term. An alternative approach to consistent estimation of $(B^T, D^T)^T$ is to transform equation (3.6.39) by Q^* and then apply the instrumental variables method. In the remainder of this section we will derive two such estimators and discuss their consistency and relative efficiency.

First we consider the decomposition of equation (3.6.37) into two orthogonal equations

$$(3.6.38)\quad P^{*}y = P^{*}(X_1, X_2, X_3, X_4)B + P^{*}(Z_1, Z_2)D + P^{*}(W_1, W_2)C + P^{*}(u + v + e)$$

$$(3.6.39)\quad Q^{*}y = Q^{*}(X_1, X_2, X_3, X_4)B + Q^{*}(Z_1, Z_2)D + Q^{*}(u + e)$$

Since $Q^{*}v = 0$ and $Q^{*}W = 0$, time effects are eliminated from equation (3.6.39) but there still exists a problem of correlation between errors and the regressors X_2, X_4, and Z_2. The largest set of legitimate instruments for equation (3.6.39) would appear to be

$$(3.6.40)\quad H_{*} = [\ Q^{*}X_1,\ Q^{*}X_3,\ Q^{*}Z_1\ ]$$

Unfortunately, by comparing H_{*} to the list of instruments used for the instrumental variables estimator given in equation (3.6.23) it can be seen that H_{*} is not the largest available set of instruments available. Although not apparent in equation (3.6.39), both Q_1X_3 and Q_1X_4 are available instruments being excluded. Following White (1984, section IV.3), the efficiency of an instrumental variables estimator is not decreased by adding more instruments. Hence, an instrumental variables estimator using the instrument set H_{*} would not lead to a more efficient estimator than the instrumental variables estimator given in equation (3.6.23).

It is interesting to note that the existence of the instrumental variables estimator using H_* depends on the matrix $P_+R = [\ P[Q^*H_*]X,\ P[Q_2H_*]Z\ ]$ being of full rank; $P_+ = P[H_{**}]$. It follows that a necessary condition for the existence of the above estimator is that rank(X_1, X_3, Z_1) $\geq$ rank(X) and rank(X_1, X_3, Z_1) $\geq$ rank(Z); or that $k_1 \geq g_2 + g_4$ and $g_1 + g_3 \geq k_2$. Thus, not only is it necessary to have enough X_1's and X_3's to identify the coefficients of the Z_2's but now we must have enough Z_1's available to identify the coefficients of the X_2's and X_4's.

If instead of using the instrument set H_*, we use the instrument set

$$(3.6.41)\qquad H_{**} = [\ Q^*X_1,\ Q_1X_2,\ Q^*X_3,\ Q_1X_4,\ Q_2Z\ ]$$

we would be using the same list of instrument used in the instrumental variables estimator given in equation (3.6.23). Projecting equation (3.6.39) onto the column space of H_{**}, we have the equation

$$(3.6.42)\qquad P_+y = P_+(X_1,\ X_2,\ X_3,\ X_4)B + P_+(Z_1,\ Z_2)D + P_+(u + e).$$

It can be shown that $P_+ = P_1 + P_2$, where $P_1 = P[\ Q_1X,\ 0\ ]$, $P_2 = P[\ H_0\ ]$, and H_0 is the instrument set given in equation (3.6.18). The least squares estimator of $(B^T,\ D^T)^T$ from equation (3.6.42) can be written as

$$(3.6.43)\qquad \begin{bmatrix} b_{*IV} \\ d_{*IV} \end{bmatrix} = (R^TP_+R)^{-1}R^TP_+y$$

where $R = (X_1 , X_2 , X_3 , X_4 , Z_1 , Z_2)$.

We first derive the necessary conditions for the existence of the above estimator, and then show it to be consistent for fixed T.

Corresponding to the order condition, we have the following theorem:

Theorem (3.12): A necessary condition for the least squares estimator of $(B^T , D^T)^T$ from equation (3.6.42) to exist is that $g_1 + g_3 \geq k_2$.

Proof:

The existence of the least squares estimator from equation (3.6.42) depends on the matrix $P_+ R$ being of full rank. And since $P_+ R = P_1 R + P_2 R = (P_+ X, P_2 Z)$, it follows that a necessary condition for the existence of the estimator is that rank$(P_2 Z) \geq k$. But this requires that the rank$(P_2 Z\ X_1 , X_3 , Z_1) > k$; or that $g_1 + g_3 \geq k_2$. Q.E.D.

Theorem (3.13): Given the rank condition of theorem (3.12), least squares applied to equations (3.6.42) is a consistent estimator for $(B^T , D^T)^T$ when T is fixed.

Proof:

Least squares applied to equation (3.6.42) can be written as

$$\begin{bmatrix} b_{*IV} \\ d_{*IV} \end{bmatrix} = \{ R^T P_+ R \}^{-1} R^T P_+ y = \begin{bmatrix} B \\ D \end{bmatrix} + \{R^T P_+ R\}^{-1} R^T P_+ u + e)$$

Since the estimator exists, $\lim \{ R^T P_+ R \}^{-1}$ is finite as $N \to \infty$. Next, consider

$$\{ R^T P_+ \}(u + e)/NT$$

$$= R^T P_1 (u + e)/NT + R^T P_2 (u + e)/NT$$

$$= R^T Q_1 R (R^T Q_1 R)^{-1} R^T Q_1 (u + e) + R^T H_0 (H_0^T H_0)^{-1} H_0^T (u + e)$$

$$= R^T Q_1 (u + e)/NT + (R^T H_0/NT)(H_0^T H_0/NT)^{-1}(H_0^T (u + e)/NT)$$

where

$$R^T Q_1 (u + e) = \begin{bmatrix} X^T Q_1 e \\ 0 \end{bmatrix}$$

and

$$H_0^T (u + e) = \begin{bmatrix} X_1^T Q_2 (u + e) \\ X_3^T Q_2 (u + e) \\ Z_1^T Q_2 (u + e) \end{bmatrix}$$

As we can easily show,

$$\text{plim } X^T Q_1 e/NT = 0 \text{ as } N \to \infty \text{ or } T \to \infty,$$

$$\text{plim } X_1^T Q_2 (u + e)/NT = 0 \text{ as } N \to \infty,$$

$$\text{plim } X_3^T Q_2 (u + e)/NT = 0 \text{ as } N \to \infty,$$

and

$$\text{plim } Z_1^T Q_2 (u + e)/NT = 0 \text{ as } N \to \infty.$$

Therefore,

$\text{plim } R^T Q_1 (u + e)/NT = \text{plim } H_0^T (u + e)/NT = 0$

as $N \to \infty$. Since the estimator exists,

$\lim (R^T H_0/NT)(H_0^T H_0/NT)^{-1}$ is finite as $N \to \infty$.

Thus, plim $\{ R^T Q_1 (u + e)/NT + (R^T H_0)(H_0^T H_0)^{-1} H_0^T (u + e)/NT \}$ $= 0$ as $N \to \infty$.

It follows that, $\operatorname{plim} \begin{bmatrix} b_{*IV} \\ d_{*IV} \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}$ as $N \to \infty$.

Q.E.D.

The least squares estimator in equation (3.6.43) can be viewed as an unweighted version of the weighted least squares estimator from equations (3.6.19) and (3.6.20), as follows.

$$\begin{bmatrix} b_{*IV} \\ d_{*IV} \end{bmatrix} = (R^T P_+ R)^{-1} R^T P_+ y$$

$$= (R^T P_1 R + R^T P_2 R)^{-1} (R^T P_1 + R^T P_2) y$$

$$= (R^T Q_1 R (R^T Q_1 R)^{-1} R^T Q_1 R + R^T H_0 (H_0^T H_0)^{-1} H_0^T R)^{-1}$$

$$\text{times } (R^T Q_1 R (R^T Q_1 R)^{-1} R^T Q_1 + R^T H_0 (H_0^T H_0)^{-1} H_0^T) y$$

$$= (R^T Q_1 R + R^T P_2 R)^{-1} (R^T Q_1 + R^T P_2) y$$

Since the weighted least squares estimator weights the equations (3.6.19) and (3.6.20) optimally, we would expect that weighted least squares is efficient relative to ordinary least squares. This is shown in the following.

Theorem (3.14): The weighted least squares estimator of $(B^T, D^T)^T$ from equations (3.6.19) and (3.6.20) is asymptotically (as $N \to \infty$) efficient relative to the least squares estimator from equation (3.6.43). If p and q are known then weighted least squares is also efficient relative to least squares in finite samples.

Proof:

We prove the finite sample case; the other case is similiar. Again, let Cov(u + e) $\equiv \Omega_* = pQ_1 + qQ_2$ and $\Omega_*^{-1} = (1/p)Q_1 + (1/q)Q_2$. Then

$$\mathrm{Cov}\begin{bmatrix} b_{IV} \\ d_{IV} \end{bmatrix}$$

$$= (R^T Q_1 R/p + R^T P_2 R/q)^{-1} (R^T Q_1/p + R^T P_2/q)(pQ_1 + qP_2)$$

$$\text{times} \quad (Q_1 R/p + P_2 R/q)(R^T Q_1 R/p + R^T P_2 R/q)^{-1}$$

$$= (R^T Q_1 R/p + R^T P_2 R/q)^{-1} (R^T Q_1 R/p + R^T P_2 R/q)$$

$$\text{times} \quad (R^T Q_1 R/p + R^T P_2 R/q)^{-1}$$

$$= (R^T Q_1 R/p + R^T P_2 R/q)^{-1} = (R^T (Q_1/p + P_2/q) R)^{-1}$$

$$= (R^T \Omega_*^{-1} R)^{-1}$$

And

$$\mathrm{Cov}\begin{bmatrix} b_{*IV} \\ d_{*IV} \end{bmatrix} = (R^T P_+ R)^{-1} R^T P_+ \Omega_* P_+ R (R^T P_+ R)^{-1}$$

Now to show

$$\mathrm{Cov}\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix} - \mathrm{Cov}\begin{bmatrix} b_{WLS} \\ d_{WLS} \end{bmatrix}$$

is psd it is sufficient to show that

$$\mathrm{Cov}\begin{bmatrix} b_{WLS} \\ d_{WLS} \end{bmatrix}^{-1} - \mathrm{Cov}\begin{bmatrix} b_{OLS} \\ d_{OLS} \end{bmatrix}^{-1}$$

is p.s.d. But the latter expression can be written as

$$(R^T \Omega_*^{-1} R) - (R^T P_+ R)^{-1} R^T P_+ \Omega_* P_+ R (R^T P_+ R)^{-1}$$

$$= R^T \Omega_*^{-1/2} [I - \Omega_*^{1/2} P_+ R (R^T P_+ \Omega_* P_+ R)^{-1} R^T P_+ \Omega_*^{1/2}] \Omega_*^{-1/2} R$$

$$= R^T \Omega_*^{-1/2} [I - D(D^T D)^{-1} D^T] \Omega_*^{-1/2} R$$

where $D = \Omega_*^{-1/2} P_* R$, can be seen as a quadratic form in an idempotent matrix and hence, our expression is psd. Q.E.D.

3.7 Variance Estimation when the Random Effects are not Correlated with the Regressors

When discussing the generalized least squares estimator, we have implicitly assumed that the variance components, σ_e^2, σ_u^2, and σ_v^2, were known. In practice, this is not the case; the variance components are usually unknown and, therefore, must be estimated. When estimates of the variance components are used in place of the actual values, we have an example of feasible generalized least squares.

Under mild regularity conditions, Fuller and Battese (1973) have shown that the feasible generalized least squares estimator is consistent and has the same asymptotic distribution as the generalized least squares estimator with known variance components. This result holds true for either large N or large T. Swamy and Arora (1972) caution that, for small samples, the feasible generalized least squares estimator could have larger variances than either the least squares estimator if the variance components σ_u^2 and σ_v^2 are small, or the within estimator if σ_u^2 and σ_v^2 are very large.

Efficiency in the estimation of the variance components and its subsequent effect on the efficiency of the feasible generalized least squares has been discussed by Amemiya (1971).

In the following discussion, we rewrite equation

(3.4.12), (3.4.13), and (3.4.14) as

$$(3.7.1) \quad Q_1 y = R_1 A_1 + Q_1 s$$

where $R_1 = (Q_1 X)$, $A_1 = (B^T)^T$, and rank$(R_1) = g$;

$$(3.7.2) \quad Q_2 y = R_2 A_2 + Q_2 s$$

where $R_2 = (Q_2 X, Q_2 Z)$, $A_2 = (B^T, D^T)^T$, and rank(R_2) $= g + k$;

$$(3.7.3) \quad Q_3 y = R_3 A_3 + Q_3 s$$

where $R_3 = (Q_3 X, Q_3 W)$, $A_3 = (B^T, C^T)^T$, and rank(R_3) $= g + h$.

If feasible weighted least squares is to be implemented instead of the equivalent feasible least squares procedure, the weights p, q, and r are the parameters we need to estimate. One approach to estimating these weights is to estimate $p = \sigma_e^2$ using residuals from equation (3.7.1), $q = \sigma_e^2 + T\sigma_u^2$ using residuals from equation (3.7.2), and $r = \sigma_e^2 + N\sigma_v^2$ using residuals from equation (3.7.3). The groundwork for such an approach is laid by Maddala (1971), Nerlove (1971), and Swamy and Arora (1972). We now proceed to show that estimators so defined are both unbiased and consistent.

We define the sum of squared residuals from equation (3.7.1) as

$$(3.7.4) \quad SSE_1 = (Q_1 - R_1 a_1)^T (Q_1 - R_1 a_1)$$

where the residuals have been computed using the least squares estimates of the coefficents in equation (3.7.1), namely

(3.7.5) $a_1 = (R_1^T R_1)^{-1} R_1^T y.$

We also define the sum of squared residuals from equation (3.7.2) as

(3.7.6) $SSE_2 = (Q_2 - R_2 a_2)^T (Q_2 - R_2 a_2)$

where the least squares estimates of the coefficents in equation (3.7.2) are given as

(3.7.7) $a_2 = (R_2^T R_2)^{-1} R_2^T y.$

And we define the sum of squared residuals from equation (3.7.3) as

(3.7.8) $SSE_3 = (Q_3 - R_3 a_3)^T (Q_3 - R_3 a_3)$

where the least squares estimates of the coefficents in equation (3.7.3) are given as

(3.7.9) $a_3 = (R_3^T R_3)^{-1} R_3^T y.$

Theorem (3.9):

Let $s_1^2 = SSE_1/\{(N-1)(T-1) - g\}$,

$s_2^2 = SSE_2/\{(T-1) - g - k\}$,

and $s_3^2 = SSE_3/\{(N-1) - g - h\}$.

Then s_1^2, s_2^2, and s_3^2 are unbiased estimators of p, q, and r, respectively.

Proof:

Let P_1 represent the projection onto the column space of the regressors in equation (3.7.1); i.e. $P_1 = P[R_1] = R_1(R_1^T R_1)^{-1} R_1^T$. Then $Q_1 P_1 = P_1 Q_1 = P_1$, $P_1 R_1 = R_1$, $P_1^T = P_1$, and P_1 is orthogonal to Q_2, Q_3, and Q_4.

First we write the residual from equation (3.7.1) as

$$
\begin{aligned}
\text{Residual}_1 &= (Q_1 y - Q_1 P_1 y) = R_1 A_1 + Q_1 s - P_1 Q_1 y \\
&= R_1 A_1 + Q_1 s - P_1 R_1 A_1 - P_1 Q_1 s \\
&= R_1 A_1 - R_1 A_1 + Q_1 s - P_1 s \\
&= (Q_1 - P_1)s
\end{aligned}
$$

We then form the expression

$$
\begin{aligned}
SSE_1 &= (Q_1 y - Q_1 P_1 y)^T (Q_1 y - Q_1 P_1 y) \\
&= s^T (Q_1 - Q_1 P_1)^T (Q_1 - Q_1 P_1) s \\
&= s^T (Q_1 - P_1 Q_1 - Q_1 P_1 + P_1 Q_1 P_1) s \\
&= s^T (Q_1 - P_1 - P_1 + P_1) s \\
&= s^T (Q_1 - P_1) s
\end{aligned}
$$

Taking the expectation of SSE_1, we write

$$
\begin{aligned}
\text{Exp}\{ SSE_1 \} &= \text{Exp}\{ s^T (Q_1 - P_1) s \} \\
&= \text{Exp}\{ \text{trace}\{ s^T (Q_1 - P_1) s \} \} \\
&= E\{ \text{trace}\{ (Q_1 - P_1) s s^T \} \}
\end{aligned}
$$

since trace(AB) = trace(BA) if AB and BA are both defined and square.

$$
\begin{aligned}
&= \text{trace}\{ (Q_1 - P_1) \text{Exp}\{ s s^T \} \} \\
&= \text{trace}\{ (Q_1 - P_1)\{ pQ_1 + qQ_2 + rQ_3 + kQ_4 \} \\
&= (p)\text{trace}\{ (Q_1 - P_1) \}
\end{aligned}
$$

$$= (p)\text{rank}(Q_1 - P_1)$$

since trace(A) = rank(A) if A is idempotent

$$= (p)\{\text{rank}(Q_1) - \text{rank}(R_1)\}$$

Thus, $\text{Exp}\{ s_1^2 \} = p$.

Now, let P_2 represent the projection onto the column space of the regressors in equation (3.7.2); i.e. $P_2 = P[R_2] = R_2(R_2^T R_2)^{-1} R_2^T$. Then $Q_2 P_2 = P_2 Q_2 = P_2$, $P_2 R_2 = R_2$, $P_2^T = P_2$, and P_2 is orthogonal to Q_1, Q_3, and Q_4.

First we write the residual from equation (3.7.2) as

$$\begin{aligned} \text{Residual}_2 &= (Q_2 y - Q_2 P_2 y) = R_2 A_2 + Q_2 s - P_2 Q_2 y \\ &= R_2 A_2 + Q_2 s - P_2 R_2 A_2 - P_2 Q_2 s \\ &= R_2 A_2 - R_2 A_2 + Q_2 s - P_2 s \\ &= (Q_2 - P_2)s \end{aligned}$$

We then form the expression

$$\begin{aligned} SSE_2 &= (Q_2 y - Q_2 P_2 y)^T (Q_2 y - Q_2 P_2 y) \\ &= s^T (Q_2 - Q_2 P_2)^T (Q_2 - Q_2 P_2) s \\ &= s^T (Q_2 - P_2 Q_2 - Q_2 P_2 + P_2 Q_2 P_2) s \\ &= s^T (Q_2 - P_2 - P_2 + P_2) s \\ &= s^T (Q_2 - P_2) s \end{aligned}$$

Taking the expectation of SSE_2, we write

$$\begin{aligned} \text{Exp}\{ SSE_2 \} &= \text{Exp}\{ s^T (Q_2 - P_2) s \} \\ &= \text{Exp}\{ \text{trace}\{ s^T (Q_2 - P_2) s \} \} \\ &= E\{ \text{trace}\{ (Q_2 - P_2) s s^T \} \} \end{aligned}$$

since trace(AB) = trace(BA) if AB and BA are both defined and square.

$$= \text{trace}\{ (Q_2 - P_2)\text{Exp}\{ ss^T \} \}$$

$$= \text{trace}\{ (Q_2 - P_2)\{ pQ_1 + qQ_2 + rQ_3 + kQ_4 \}$$

$$= (q)\text{trace}\{ (Q_2 - P_2) \}$$

$$= (q)\text{rank}(Q_2 - P_2)$$

since trace(A) = rank(A) if A is idempotent

$$= (q)\{\text{rank}(Q_2) - \text{rank}(R_2)\}$$

Thus, $\text{Exp}\{ s_2^2 \} = q$.

Finally, let P_3 represent the projection onto the column space of the regressors in equation (3.7.3); i.e. $P_3 = P[R_3] = R_3(R_3^T R_3)^{-1}R_3^T$. Then $Q_3P_3 = P_3Q_3 = P_3$, $P_3R_3 = R_3$, $P_3^T = P_3$, and P_3 is orthogonal to Q_1, Q_2, and Q_4.

First we write the residual from equation (3.7.3) as

$$\begin{aligned}\text{Residual}_3 &= (Q_3y - Q_3P_3y) = R_3A_3 + Q_3s - P_3Q_3y \\ &= R_3A_3 + Q_3s - P_3R_3A_3 - P_3Q_3s \\ &= R_3A_3 - R_3A_3 + Q_3s - P_3s \\ &= (Q_3 - P_3)s\end{aligned}$$

We then form the expression

$$\begin{aligned}SSE_3 &= (Q_3y - Q_3P_3y)^T(Q_3y - Q_3P_3y) \\ &= s^T(Q_3 - Q_3P_3)^T(Q_3 - Q_3P_3)s \\ &= s^T(Q_3 - P_3Q_3 - Q_3P_3 + P_3Q_3P_3)s \\ &= s^T(Q_3 - P_3 - P_3 + P_3)s \\ &= s^T(Q_3 - P_3)s\end{aligned}$$

Taking the expectation of SSE_3, we write

$$\text{Exp}\{ SSE_3 \} = \text{Exp}\{ s^T(Q_3 - P_3)s \}$$

$= \text{Exp}\{ \text{trace}\{ s^T(Q_3 - P_3)s \} \}$

$= E\{ \text{trace}\{ (Q_3 - P_3)ss^T \} \}$

since trace(AB) = trace(BA) if AB and BA are both defined and square.

$= \text{trace}\{ (Q_3 - P_3)\text{Exp}\{ ss^T \} \}$

$= \text{trace}\{ (Q_3 - P_3)\{ pQ_1 + qQ_3 + rQ_3 + kQ_4 \}$

$= (r)\text{trace}\{ (Q_3 - P_3) \}$

$= (r)\text{rank}(Q_3 - P_3)$

since trace(A) = rank(A) if A is idempotent

$= (r)\{\text{rank}(Q_3) - \text{rank}(R_3)\}$

Thus, $\text{Exp}\{ s_3^2 \} = r$. Q.E.D.

__Theorem (3.10)__:

Let $s_1^2 = SSE_1/\{(N-1)(T-1) - g\}$,

$s_2^2 = SSE_2/\{(T-1) - g - k\}$,

and $s_3^2 = SSE_3/\{(N-1) - g - h\}$.

Then

a) s_1^2 is a consistent estimator of p as N or T -> oo,

b) s_2^2 is a consistent estimator of $q = \sigma_e^2 + T\sigma_u^2$ as N -> oo , and

c) s_3^2 is a consistent estimator of $r = \sigma_e^2 + N\sigma_v^2$ as T -> oo .

__Proof__:

a) $\text{plim } s_1^2 = \text{plim } SSE_1/\{\text{rank}(Q_1) - \text{rank}(R_1)\}$

$= \text{plim } SSE_1/(N-1)(T-1)$

$= \text{plim } s^T(Q_1 - P_1)s/(N-1)(T-1)$

$= \text{plim } s^TQ_1s/(N-1)(T-1) - \text{plim } s^TP_1s/(N-1)(T-1)$

The last term is zero since

$$s^T P_1 s/(N-1)(T-1)$$

$$= [s^T R_1/(N-1)(T-1)][R_1^T R_1/(N-1)(T-1)]^{-1} R_1^T s/(N-1)(T-1)$$

and $R_1^T s/(N-1)(T-1) \to 0$ as $(N-1)(T-1) \to \infty$ (as either $N \to \infty$ or $T \to \infty$)

The first term equals σ_e^2 because $s^T Q_1 s$ can be shown to be distributed as $\sigma_e^2 \chi^2_{(N-1)(T-1)}$ using standard results (e.g. Rao (1973, p 185)) on the distribution of idempotent quadractic forms in normals.

b) $\text{plim } s_2^2 = \text{plim } SSE_2/\{\text{rank}(Q_2) - \text{rank}(R_2)\}$

$$= \text{plim } SSE_2/(N-1)$$

$$= \text{plim } s^T(Q_2 - P_2)s/(N-1)$$

$$= \text{plim } s^T Q_2 s/(N-1) - \text{plim } s^T P_2 s/(N-1)$$

The last term is zero since $s^T P_2 s/(N-1) = [s^T R_2/(N-1)][R_2^T R_2/(N-1)]^{-1} R_2^T s/(N-1)$ and $R_2^T s/(N-1) \to 0$ as $N \to \infty$.

The first term equals $q = \sigma_e^2 + T\sigma_u^2$ because $s^T Q_2 s$ can be shown to be distributed as $q\chi^2_{(N-1)}$ using standard results (e.g. Rao (1973, p 185)) on the distribution of idempotent quadractic forms in normals.

c) $\text{plim } s_3^2 = \text{plim } SSE_3/\{\text{rank}(Q_3) - \text{rank}(R_3)\}$

$$= \text{plim } SSE_3/(T-1)$$

$$= \text{plim } s^T(Q_3 - P_3)s/(T-1)$$

$$= \text{plim } s^T Q_3 s/(T-1) - \text{plim } s^T P_3 s/(T-1)$$

The last term is zero since $s^T P_3 s/(T-1)$ $= [s^T R_3/(T-1)][R_3^T R_3/(T-1)]^{-1} R_3^T s/(T-1)$ and $R_3^T s/(T-1) \to 0$ as $T \to \infty$.

The first term equals $q = \sigma_e^2 + N\sigma_u^2$ because $s^T Q_3 s$ can be shown to be distributed as $q\chi^2_{(T-1)}$ using standard results (e.g. Rao (1973, p 185)) on the distribution of idempotent quadractic forms in normals. **Q.E.D.**

3.8 Variance Estimation when the Random Effects are Correlated with the Regressors

So far we have considered variance estimation for the feasible weighted least squares estimator only. We now turn our attention to the model of section 3.5, in which some of the regressors are correlated with the random effects. Once again we will need to estimate the weights p, q, and r, since they are needed to implement the weighted instrumental variables estimator. The estimate of p based on the within residuals, discussed in section 3.7, is still consistent in this model. However, the estimate of $q = \sigma_e^2 + T\sigma_u^2$ and $r = \sigma_e^2 + N\sigma_v^2$ which was discussed in section 3.7 is not consistent, since it was based on the residuals from least squares applied to (3.5.3) and (3.5.4), and these least squares estimator are inconsistent when regressors are correlated with either equation's error term.

We therefore turn our attention to the problem of finding consistent estimates of B, D, and C. Then, using these consistent estimates of $A_2 = (B^T, D^T)^T$ and $A_3 = (B^T, C^T)^T$, we derive consistent estimate of q and r. The background for this approach is the work of Hausman and Taylor (1981), who suggest the estimate of q which we discussed in section 2.7. However, they do not give a rigorous proof that it is consistent nor do they discuss the estimation of r.

The following assumptions will be made.

<u>Assumption (3.11)</u>: Let $H_2 = [\ Q_2X_1,\ Q_2X_3,\ Z_1\ ]$ and $H_3 = [\ Q_3X_1,\ Q_3X_2,\ W_1\ ]$. Then we assume that

(i) plim $X^TQ_1e/(N-1)(T-1) = 0$ as either $N \to \infty$ or $T \to \infty$.

(ii.a) plim $H_2{}^TQ_2(u + e)/N = 0$ as $N \to \infty$.

(ii.b) plim $H_3{}^TQ_3(v + e)/T = 0$ as $T \to \infty$.

(iii) plim $(X^TQ_1X)/(N-1)(T-1)$ is finite and nonsingular as either $N \to \infty$ or $T \to \infty$.

(iv.a) plim $(H_2{}^TZ_1)/N$ is finite as $N \to \infty$.

(iv.b) plim $(H_3{}^TW_1)/T$ is finite as $T \to \infty$.

(v.a) plim $(H_2{}^TX)/N$ is finite as $N \to \infty$.

(v.b) plim $(H_3{}^TX)/T$ is finite as $T \to \infty$.

Even after the introduction of X_2, X_3, X_4, Z_2, and W_2 - regressors assumed correlated with the effects - the within estimator is still a consistent estimator of B; no correlation exists between the disturbance and the regressors in equation (3.5.3). So the problem of finding a consistent estimator of A is reduced to finding a consistent estimator of D and C. The two regression equations introduced in the following Lemma will be used in deriving such estimators.

Lemma (3.12): Let $f_2^* = Q_2(y - Xb_w)$ and $f_3^* = Q_3(y - Xb_w)$. Then

$$(3.8.1) \quad f_2^* = ZD + (Q_2 - Q_2X(X^TQ_1X)^{-1}X^TQ_1)s$$

and

$$(3.8.2) \quad f_3^* = WC + (Q_3 - Q_3X(X^TQ_1X)^{-1}X^TQ_1)s$$

Proof:

f_2^*

$= Q_2(y - Wb_w) = Q_2y - Q_2Xb_w = Q_2y - Q_2X(X^TQ_1X)^{-1}X^TQ_1y$

$= Q_2(XB + ZD + WC + s) - Q_2X(X^TQ_1X)^{-1}X^TQ_1(XB + ZD + WC + s)$

$= Q_2(XB + ZD + WC + s) - Q_2X(X^TQ_1X)^{-1}X^T(Q_1XB + Q_1s)$

$= Q_2(XB + ZD + WC + s) - Q_2X(X^TQ_1X)^{-1}X^TQ_1XB$

$\quad + Q_2X(X^TQ_1X)^{-1}X^TQ_1s$

$= Q_2XB + Q_2ZD + Q_2s - Q_2XB + Q_2X(X^TQ_1X)^{-1}X^TQ_1s$

$= ZD + (Q_2 - Q_2X(X^TQ_1X)^{-1}X^TQ_1)s$

f_3^*

$= Q_3(y - Wb_w) = Q_3y - Q_3Xb_w = Q_3y - Q_3X(X^TQ_1X)^{-1}X^TQ_1y$

$= Q_3(XB + ZD + WC + s) - Q_3X(X^TQ_1X)^{-1}X^TQ_1(XB + ZD + WC + s)$

$= Q_3(XB + ZD + WC + s) - Q_3X(X^TQ_1X)^{-1}X^T(Q_1XB + Q_1s)$

$= Q_3(XB + ZD + WC + s) - Q_3X(X^TQ_1X)^{-1}X^TQ_1XB$

$\quad + Q_3X(X^TQ_1X)^{-1}X^TQ_1s$

$= Q_3XB + Q_3WC + Q_3s - Q_3XB + Q_3X(X^TQ_1X)^{-1}X^TQ_1s$

$= WC + (Q_3 - Q_3X(X^TQ_1X)^{-1}X^TQ_1)s$ Q.E.D.

Since part of Z is correlated with the error term, least squares applied to equation (3.8.1) does not yield a consistent estimator of D. Likewise, since part of W is correlated with the error term, least squares applied to equation (3.8.2) does not yield a consistent estimator of C. But, using $H_2 = (Q_2X_1, Q_2X_3, Z_1)$ as a set of instruments, the instrumental variable estimator of D from equation (3.8.1) is defined as

(3.8.3) $d_{IV} = (X^T P[H_2]X)^{-1}X^T P[H_2]f_2^*$.

Similiarly, using $H_3 = (Q_3X_1, Q_3X_2, W_1)$ as a set of instruments, the instrumental variable estimator of C from equation (3.8.2) is defined as

(3.8.4) $c_{IV} = (X^T P[H_3]X)^{-1}X^T P[H_3]f_3^*$.

It is interesting to note that using $f_3^{**} = (y - Xb_W)$ instead of $f_3^* = Q_3(y - Xb_W)$ would not increase the efficiency of the estimator, c_{IV}. Indeed, since $P[H_3]Q_3 = Q_3P[H_3] = P[H_3]$, $Z_1^TQ_1 = 0$, and the first order condition (i.e. the "normal equations") defining b_W imply that $(X_1Q_1, X_2^TQ_1)(y - Xb_W) = 0$,

(3.8.5) $H_3^TQ_3(y - Xb_W) = H_3^T(y - Xb_W)$;

thus the <u>same</u> estimator would result if we used f_3^{**} in place of f_3^*.

Given the estimators d_{IV} and c_{IV}, the next question is whether these estimators are, indeed, consistent estimates of D and C. But first, we consider the conditions necessary to assure that both d_{IV} and c_{IV} do exist.

3.8.1 Necessary Conditions for the Existence of d_{IV} and c_{IV}

A necessary condition for the existence of d_{IV} is that the rank of H_2 be at least is large as the rank of Z; that is, there must be at least as many instruments as regressors. This requires that $g_1 + g_3 + k_1 \geq k$, or $g_1 + g_3 \geq k_2$. Intuitively, Q_2X_1 and Q_2X_3 are serving as instruments for Z_2, and so there must be at least as many variables in X_1 and X_3 as in Z_2. Similiarly, a necessary condition for the existence of c_{IV} is that the rank of H_3 be at least is large as the rank of W; that is, there must be at least as many instruments as regressors. This requires that $g_1 + g_2 + h_1 \geq h$, or $g_1 + g_2 \geq h_2$. Here, Q_3X_1 and Q_3X_2 are serving as instruments for W_2, and so there must be at least as many variables in X_1 and X_2 as in W_2. The fact that f_1^* and f_2^* are calculated from the within-groups residuals suggests that if b_w is not fully efficient, then d_{IV} and c_{IV} may not be fully efficient either.

3.8.2 Consistency of d_{IV} and c_{IV}.

Lemma (3.13): Given assumption (3.11),

(1.a) $\text{plim } Z^T P_2 Q_2 (u + e)/N = 0$ as $N \to \infty$,

(1.b) $\text{plim } W^T P_3 Q_3 (v + e)/T = 0$ as $T \to \infty$,

(2.a) $\text{plim } Z^T P_2 Z/N$ is finite and non-singular as $N \to \infty$,

(2.b) $\text{plim } W^T P_3 W/T$ is finite and non-singular as $T \to \infty$,

(3.a) $\text{plim } Z^T P_2 X/N$ is finite as $N \to \infty$,

(3.b) $\text{plim } W^T P_3 X/T$ is finite as $T \to \infty$.

Lemma (3.13) can be easily proved by noting that P_2

$= P[H_2] = H_2(H_2^T H_2)^{-1} H_2^T$, where $H_2 = Q_2(X_1, X_3, Z_1)$, and that $P_3 = P[H_3] = H_3(H_3^T H_3)^{-1} H_3^T$, where $H_3 = Q_3(X_1, X_2, W_1)$.

Theorem (3.14): The instrumental variable estimator d_{IV} is a consistent estimator of D as N gets large and the instrumental variable estimator c_{IV} is a consistent estimator of C as T gets large.

Proof:

First, we rewrite d_{IV} as

$$\begin{aligned} d_{IV} &= (Z^T P_2 Z)^{-1} Z^T P_2 d_2^* \\ &= (Z^T P_2 Z)^{-1} Z^T P_2 (ZD + (Q_2 - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1)s) \\ &= (Z^T P_2 Z)^{-1} Z^T P_2 ZD \\ &\quad + (Z^T P_2 Z)^{-1} Z^T P_2 (Q_2 - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1)s \\ &= D + (Z^T P_2 Z)^{-1} Z^T P_2 Q_2 s - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 s \\ &= D + (Z^T P_2 Z/N)^{-1} \{Z^T P_2 Q_2 (e + u)/N\} \\ &\quad - (Z^T P_2 Z/N)^{-1} \{Z^T P_2 Q_2 X/N\} (X^T Q_1 X/N)^{-1} \{X^T Q_1 e/N\} \end{aligned}$$

By Assumption (3.11), plim $X^T Q_1 e/N = 0$ as $N \to \infty$ and plim $(X^T Q_1 X)/N$ is finite and nonsingular as $N \to \infty$. Using Lemma (3.13), it follows that plim $Z^T P_2 Q_2 (e + u)/N = 0$ as $N \to \infty$, plim $(Z^T P_2 Z)/N$ is finite and nonsingular as $N \to \infty$, and plim $(Z^T P_2 Q_2 X)/N$ is finite as $N \to \infty$. Thus,

$$\begin{aligned} &\text{plim } d_{IV} \\ &= D + \{\text{finite}\}\{0\} - \{\text{finite}\}\{\text{finite}\}\{\text{finite}\}\{0\} \\ &= D \end{aligned}$$

Next, we rewrite c_{IV} as

$$
\begin{aligned}
c_{IV} &= (W^T P_3 W)^{-1} W^T P_3 d_3{}^* \\
&= (W^T P_3 W)^{-1} W^T P_3 (WC + (Q_2 - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1)s) \\
&= (W^T P_3 W)^{-1} W^T P_3 WC \\
&\quad + (W^T P_3 W)^{-1} W^T P_3 (Q_3 - Q_3 X(X^T Q_1 X)^{-1} X^T Q_1)s \\
&= C + (W^T P_3 W)^{-1} W^T P_3 Q_3 s - Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 s \\
&= C + (W^T P_3 W/T)^{-1} \{W^T P_3 Q_3 s/T\} \\
&\quad - (W^T P_3 W/T)^{-1} \{W^T P_3 Q_3 X/T\}(X^T Q_3 X/T)^{-1} \{X^T Q_3 s/T\}
\end{aligned}
$$

By Assumption (3.11), plim $X^T Q_1 e/T = 0$ as $T \to \infty$ and plim $(X^T Q_1 X)/T$ is finite and nonsingular as $T \to \infty$. Using Lemma (3.13), it follows that plim $W^T P_3 Q_3 (e + v)/T = 0$ as $T \to \infty$, plim $(W^T P_3 W)/T$ is finite and nonsingular as $T \to \infty$, and plim $(W^T P_3 Q_3 X)/T$ is finite as $T \to \infty$. Thus,

$$
\begin{aligned}
&\text{plim } c_{IV} \\
&= C + \{\text{finite}\}\{0\} - \{\text{finite}\}\{\text{finite}\}\{\text{finite}\}\{0\} \\
&= C
\end{aligned}
$$

Q.E.D.

3.8.3 <u>Consistent Estimates of q and r</u>

Using as a consistent estimate of $A_2 = (B^T , D^T)^T$ the estimator b_W and d_{IV}, we will now form a vector of residuals. We will then show that the sum of the squared terms of this residual vector, divided by N, is a consistent estimator of $q = \sigma_e{}^2 + T\sigma_u{}^2$. Similiarly, using as a consistent estimate of $A_3 = (B^T , C^T)^T$ the estimator b_W and c_{IV}, we will form a vector of residuals and then show that the sum of the squared terms of this residual vector, divided by T, is a consistent estimator of $r = \sigma_e^2 + N\sigma_v^2$.

Lemma (3.15): Let

(3.8.6) $\text{Residual}_2 = Q_2 y - Q_2 X b_W - Q_2 Z d_{IV}$

and

(3.8.7) $\text{Residual}_3 = Q_3 y - Q_3 X b_W - Q_3 W c_{IV}$

Then

$$\begin{aligned}\text{Residual}_2 = \; & Q_2(e + u) - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e \\ & - Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u) \\ & + Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e\end{aligned}$$

and

$$\begin{aligned}\text{Residual}_3 = \; & Q_3(e + v) - Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e \\ & - Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v) \\ & + Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\end{aligned}$$

Proof:

First, we rewrite Residual_2 as

$$\begin{aligned}\text{Residual}_2 &= Q_2 y - Q_2 X b_W - Q_2 Z d_{IV} \\ &= Q_2 y - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 y - Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 d_2^* \\ &= Q_2 \{ XB + ZD + WC + s \} \\ & \quad - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 \{ XB + ZD + WC + s \} \\ & \quad - Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 \\ & \quad \quad \text{times}\{ ZD + (Q_2 - Q_2 X(X^T Q_1 X)^{-1} X^T Q_1)(e + u) \}\end{aligned}$$

$$= Q_2XB + ZD + Q_2(e + u)$$

$$- Q_2X(X^TQ_1X)^{-1}X^TQ_1XB - Q_2X(X^TQ_1X)^{-1}X^TQ_1e$$

$$- Q_2Z(Z^TP_2Z)^{-1}Z^TP_2ZD - Q_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2(e + u)$$

$$+ Q_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2X(X^TQ_1X)^{-1}X^TQ_1e$$

$$= Q_2XB + ZD + Q_2(e + u)$$

$$- Q_2XB - Q_2X(X^TQ_1X)^{-1}X^TQ_1(e + u)$$

$$- Q_2ZD - Q_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2(e + u)$$

$$+ Q_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2X(X^TQ_1X)^{-1}X^TQ_1e$$

$$= Q_2(e + u) - Q_2X(X^TQ_1X)^{-1}X^TQ_1e$$

$$- Q_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2(e + u)$$

$$+ Q_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2X(X^TQ_1X)^{-1}X^TQ_1e$$

Next, we rewrite $Residuals_3$ as

$$Residuals_3 = Q_3y - Q_3Xbw - Q_3Wc_{IV}$$

$$= Q_3y - Q_3X(X^TQ_1X)^{-1}X^TQ_1y - Q_3W(W^TP_3W)^{-1}W^TP_3d_3^*$$

$$= Q_3\{ XB + ZD + WC + s \}$$

$$- Q_3X(X^TQ_1X)^{-1}X^TQ_1\{ XB + ZD + WC + s \}$$

$$- Q_3W(W^TP_3W)^{-1}W^TP_3$$

$$\text{times } \{ WC + (Q_3 - Q_3X(X^TQ_1X)^{-1}X^TQ_1)(e + u) \}$$

$$= Q_3XB + WC + Q_3(e + u)$$

$$- Q_3X(X^TQ_1X)^{-1}X^TQ_1XB - Q_3X(X^TQ_1X)^{-1}X^TQ_1e$$

$$- Q_3W(W^TP_3W)^{-1}W^TP_3WC - Q_3W(W^TP_3W)^{-1}W^TP_3Q_3(e + u)$$

$$+ Q_3W(W^TP_3W)^{-1}W^TP_3Q_3X(X^TQ_1X)^{-1}X^TQ_1e$$

$$= Q_3XB + WC + Q_3(e + u)$$

$$- Q_3XB - Q_3X(X^TQ_1X)^{-1}X^TQ_1(e + u)$$

$$- Q_3WC - Q_3W(W^TP_3W)^{-1}W^TP_3Q_3(e + u)$$

$$+ Q_3W(W^TP_3W)^{-1}W^TP_3Q_3X(X^TQ_1X)^{-1}X^TQ_1e$$

$$= Q_3(e + u) - Q_3X(X^TQ_1X)^{-1}X^TQ_1e$$
$$- Q_3W(W^TP_3W)^{-1}W^TP_3Q_3(e + u)$$
$$+ Q_3W(W^TP_3W)^{-1}W^TP_3Q_3X(X^TQ_1X)^{-1}X^TQ_1e \qquad \text{Q.E.D.}$$

We now define consistent estimators for both q and r. Using the definitions found in Lemma (3.15), we define SSE_2^* as the sum of squared residual terms found in $Residual_2$ and SSE_2^* as the sum of squared residual terms found in $Residual_3$:

(3.8.8) $SSE_2^* = (Residual_2)^T(Residual_2)$

(3.8.9) $SSE_3^* = (Residual_3)^T(Residual_3)$

Our estimators for q and r are then SSE_2^*/N and SSE_3^*/T, respectively.

<u>Theorem (3.16)</u>:

$$\text{plim } SSE_2^*/N = \sigma_e^2 + T\sigma_u^2 \quad \text{as} \quad N \to \infty$$
$$\text{plim } SSE_3^*/T = \sigma_e^2 + N\sigma_v^2 \quad \text{as} \quad T \to \infty$$

<u>Proof</u>:

First, SSE_2^* can be written as

SSE_2^*

$$= (Residual_2)^T(Residual_2)$$
$$= (e + u)^TQ_2(e + u) - (e + u)^TQ_2X(X^TQ_1X)^{-1}X^TQ_1e$$
$$- (e + u)^TQ_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2(e + u)$$
$$+ (e + u)^TQ_2Z(Z^TP_2Z)^{-1}Z^TP_2Q_2X(X^TQ_1X)^{-1}X^TQ_1e$$
$$- e^TQ_1X(X^TQ_1X)^{-1}X^TQ_2(e + u)$$
$$+ e^TQ_1X(X^TQ_1X)^{-1}X^TQ_2X(X^TQ_1X)^{-1}X^TQ_1e$$

$+ e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u)$

$- e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e$

$- (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 (e + u)$

$+ (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e$

$+ (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u)$

$- (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e$

$+ e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 (e + u)$

$- e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e$

$- e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u)$

$+ e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 Z(Z^T P_2 Z)^{-1}$

times $Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e$

Now, from the above expression, taking the probability limit of $SSE_2{}^*$ as N gets large is equivalent to taking the probability limit of the sum of sixteen different terms. Evaluation of these sixteen terms shows that the first term has a probability limit equal to q and that the remaining fifteen terms each have a probability limit equal to zero with all limits being taken as $N \to \infty$. These probability limits are evaluated below.

1) $\text{plim } (e + u)^T Q_2 (e + u)/N = \text{plim } e^T Q_2 e/N + \text{plim } u^T Q_2 u/N$

Consider these term by term. First,

$$e^T Q_2 e/N = T\sum_{i=1}^{N} e_{i.}{}^2/N.$$

Each term $e_{i.}{}^2$ has a mean of $\sigma_e{}^2/T$, and the terms are independent. Therefore, $e^T Q_2 e/N \to T\sigma_e{}^2/T = \sigma_e{}^2$ as $N \to \infty$.

Second,

$$u^T Q_2 u/N \;=\; T\sum_{i=1}^{N} u_i^2/N \;\to\; T\sigma_u^2 \quad \text{as} \quad N \to \infty.$$

Third,

$$e^T Q_2 u/N \;=\; T\sum_{i=1}^{N} e_i . u_i/N \;\to\; 0 \quad \text{as} \quad N \to \infty$$

because e and u are uncorrelated. Therefore,

$$(e + u)^T Q_2 (e + u)/N \;\to\; \sigma_e^2 + T\sigma_u^2 \quad \text{as} \quad N \to \infty.$$

2.) $\text{plim } (e + u)^T Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$= \text{plim } \{(e + u)^T Q_2 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_1 e/N\}$

$= \text{plim } \{(e + u)^T Q_2 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_1 e/N\}$

$= 0$ as $N \to \infty$

3) $\text{plim } (e + u)^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u)/N$

$= \text{plim } \{(e + u)^T Q_2 Z/N\}(Z^T P_2 Z/N)^{-1}\{Z^T P_2 Q_2 (e + u)/N\}$

$= \text{plim } \{(e + u)^T Q_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

times $\text{plim } \{Z^T P_2 Q_2 (e + u)/N\}$

$= 0$ as $N \to \infty$

4) $\text{plim } (e + u)^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$= \text{plim } \{(e + u)^T Q_2 Z/N\}(Z^T P_2 Z/N)^{-1}\{Z^T P_2 Q_2 X/N\}$

time $(X^T Q_1 X/N)^{-1}\{X^T Q_1 e/N\}$

$= \text{plim } \{(e + u)^T Q_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

times $\text{plim } \{Z^T P_2 Q_2 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1}$

times $\text{plim } \{X^T Q_1 e/N\}$

$= 0$ as $N \to \infty$

5) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 (e + u)/N$

$$= \text{plim } \{e^T Q_1 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_2 (e + u)/N\}$$

$$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 (e + u)/N\}$$

$$= 0 \text{ as } N \to \infty$$

6) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$$= \text{plim } \{e^T Q_1 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_2 X/N\}$$

$$\text{times } (X^T Q_1 X/N)^{-1}\{X^T Q_1 e/N\}$$

$$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 X/N\}$$

$$\text{times plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_1 e/N\}$$

$$= 0 \text{ as } N \to \infty$$

7) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u)/N$

$$= \text{plim } \{e^T Q_1 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_2 Z/N\}(Z^T P_2 Z/N)^{-1}$$

$$\text{times } \{Z^T P_2 Q_2 (e + u)/N\}$$

$$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 Z/N\}$$

$$\text{times plim } (Z^T P_2 Z/N)^{-1} \text{ plim } \{Z^T P_2 Q_2 (e + u)/N\}$$

$$= 0 \text{ as } N \to \infty$$

8) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$$= \text{plim } \{e^T Q_1 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_2 Z/N\}(Z^T P_2 Z/N)^{-1}$$

$$\text{times } \{Z^T P_2 Q_2 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_1 e/N\}$$

$$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 Z/N\}$$

$$\text{times plim } (Z^T P_2 Z/N)^{-1} \text{ plim } \{Z^T P_2 Q_2 X/N\}$$

$$\text{times plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_1 e/N\}$$

$$= 0 \text{ as } N \to \infty$$

9) $\text{plim } (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 (e + u)/N$

$$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\}(Z^T P_2 Z/N)^{-1}\{Z^T Q_2 (e + u)/N\}$$

$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

$\text{times plim } \{Z^T Q_2 (e + u)/N\}$

$= 0 \text{ as } N \to \infty$

10) $\text{plim } (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\}(Z^T P_2 Z/N)^{-1}$

$\text{times } \{Z^T Q_2 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_1 e/N\}$

$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

$\text{times plim } \{Z^T Q_2 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_1 e/N\}$

$= 0 \text{ as } N \to \infty$

11) $\text{plim } (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u)/N$

$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\}(Z^T P_2 Z/N)^{-1}\{Z^T Q_2 Z/N\}$

$\text{times } (Z^T P_2 Z/N)^{-1}\{Z^T P_2 Q_2 (e + u)/N\}$

$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

$\text{times plim } \{Z^T Q_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

$\text{times plim } \{Z^T P_2 Q_2 (e + u)/N\}$

$= 0 \text{ as } N \to \infty$

12) $\text{plim } (e + u)^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 Z(Z^T P_2 Z)^{-1}$

$\text{times } Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\}(Z^T P_2 Z/N)^{-1}\{Z^T Q_2 Z/N\}(Z^T P_2 Z/N)^{-1}$

$\text{times } \{Z^T P_2 Q_2 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_1 e/N\}$

$= \text{plim } \{(e + u)^T Q_2 P_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

$\text{times plim } \{Z^T Q_2 Z/N\} \text{ plim } (Z^T P_2 Z/N)^{-1}$

$\text{times plim } \{Z^T P_2 Q_2 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1}$

$\text{times plim } \{X^T Q_1 e/N\}$

$= 0 \text{ as } N \to \infty$

13) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 (e + u)/N$

$= \text{plim } \{e^T Q_1 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_2 P_2 Z/N\}$

$\text{times } (Z^T P_2 Z/N)^{-1}\{Z^T Q_2 (e + u)/N\}$

$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 P_2 Z/N\}$

$\text{times plim } (Z^T P_2 Z/N)^{-1} \text{ plim } \{Z^T Q_2 (e + u)/N\}$

$= 0$ as $N \to \infty$

14) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 P_2 Z/N\}$

$\text{times plim } (Z^T P_2 Z/N)^{-1} \text{ plim } \{Z^T Q_2 X/N\}$

$\text{times plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_1 e/N\}$

$= 0$ as $N \to \infty$

15) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1}$

$\text{times } Z^T Q_2 Z(Z^T P_2 Z)^{-1} Z^T P_2 Q_2 (e + u)/N$

$= \text{plim } \{e^T Q_1 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_2 P_2 Z/N\}(Z^T P_2 Z/N)^{-1}$

$\text{times } \{Z^T Q_2 Z/N\}(Z^T P_2 Z/N)^{-1}\{Z^T P_2 Q_2 (e + u)/N\}$

$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 P_2 Z/N\}$

$\text{times plim } (Z^T P_2 Z/N)^{-1} \text{ plim } \{Z^T Q_2 Z/N\}$

$\text{times plim } (Z^T P_2 Z/N)^{-1} \text{ plim } \{Z^T P_2 Q_2 (e + u)/N\}$

$= 0$ as $N \to \infty$

16) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_2 P_2 Z(Z^T P_2 Z)^{-1} Z^T Q_2 Z(Z^T P_2 Z)^{-1}$

$\text{times } Z^T P_2 Q_2 X(X^T Q_1 X)^{-1} X^T Q_1 e/N$

$= \text{plim } \{e^T Q_1 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_2 P_2 Z/N\}$

$\text{times plim } (Z^T P_2 Z/N)^{-1} \{Z^T Q_2 Z/N\}(Z^T P_2 Z/N)^{-1}$

$\text{times plim } \{Z^T P_2 Q_2 X/N\}(X^T Q_1 X/N)^{-1}\{X^T Q_1 e/N\}$

$= \text{plim } \{e^T Q_1 X/N\} \text{ plim } (X^T Q_1 X/N)^{-1} \text{ plim } \{X^T Q_2 P_2 Z/N\}$

$\text{times plim } (Z^T P_2 Z/N)^{-1} \text{ plim } \{Z^T Q_2 Z/N\}$

times plim $(Z^T P_2 Z/N)^{-1}$

times plim $\{Z^T P_2 Q_2 X/N\}$ plim $(X^T Q_1 X/N)^{-1}$

times plim $\{X^T Q_1 e/N\}$

$= 0$ as $N \to \infty$

Next, $SSE_3{}^*$ can be written as

$$
\begin{aligned}
SSE_3{}^* &\\
&= (\text{Residuals}_3)^T(\text{Residuals}_3)\\
&= (e + v)^T Q_3 (e + v) - (e + v)^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\\
&\quad - (e + v)^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)\\
&\quad + (e + v)^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\\
&\quad - e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 (e + v)\\
&\quad + e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\\
&\quad + e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)\\
&\quad - e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\\
&\quad - (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 (e + v)\\
&\quad + (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\\
&\quad + (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)\\
&\quad - (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\\
&\quad + e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 (e + v)\\
&\quad - e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e\\
&\quad - e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)\\
&\quad + e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 W(W^T P_3 W)^{-1}\\
&\quad \text{times } W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e
\end{aligned}
$$

Now, from the above expression, taking the probability limit of $SSE_3{}^*$ as T gets large is equivalent to taking the probability limit of the sum of sixteen different terms.

Evaluation of these sixteen terms shows that the first term has a probability limit equal to r and that the remaining fifteen terms each have a probability limit equal to zero with all limits being taken as T -> oo. These probability limits are evaluated below.

1) $\text{plim } (e + v)^T Q_3 (e + v)/T \;=\; \text{plim } e^T Q_3 e/T + \text{plim } v^T Q_3 v/T$

Consider these term by term. First,

$$e^T Q_3 e/T \;=\; N\sum_{t=1}^{T} e_{.t}^2/T.$$

Each term $e_{.t}^2$ has a mean of σ_e^2/N, and the terms are independent. Therefore, $e^T Q_2 e/N \to N\sigma_e^2/N = \sigma_e^2$ as $T \to \infty$. Second,

$$v^T Q_3 v/T \;=\; N\sum_{t=1}^{T} v_i^2/T \;\to\; N\sigma_v^2 \quad \text{as} \quad T \to \infty.$$

Third,

$$e^T Q_3 v/T \;=\; N\sum_{t=1}^{T} e_i . v_i/T \;\to\; 0 \quad \text{as} \quad T \to \infty$$

because e and v are uncorrelated. Therefore,

$$(e + v)^T Q_3 (e + v)/T \;\to\; \sigma_e^2 + N\sigma_v^2 \quad \text{as} \quad T \to \infty.$$

2) $\text{plim } (e + v)^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$= \text{plim } \{(e + v)^T Q_3 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_1 e/T\}$

$= \text{plim } \{(e + v)^T Q_3 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_1 e/T\}$

$= 0 \quad \text{as} \quad T \to \infty$

3) $\text{plim } (e + v)^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)/T$

$= \text{plim } \{(e + v)^T Q_3 W/T\}(W^T P_3 W/T)^{-1}\{W^T P_3 Q_3 (e + v)/T\}$

$= \text{plim } \{(e + v)^T Q_3 W/T\} \text{ plim } (W^T P_3 W/T)^{-1}$

$\text{times plim } \{W^T P_3 Q_3 (e + v)/T\}$

$= 0 \text{ as } T \to \infty$

4) $\text{plim } (e + v)^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$= \text{plim } \{(e + v)^T Q_3 W/T\}(W^T P_3 W/T)^{-1}\{W^T P_3 Q_3 X/T\}$

$\text{times } (X^T Q_1 X/T)^{-1}\{X^T Q_1 e/T\}$

$= \text{plim } \{(e + v)^T Q_3 W/T\} \text{ plim } (W^T P_3 W/T)^{-1}$

$\text{times plim } \{W^T P_3 Q_3 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1}$

$\text{times plim } \{X^T Q_1 e/T\}$

$= 0 \text{ as } T \to \infty$

5) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 (e + v)/T$

$= \text{plim } \{e^T Q_1 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_3 (e + v)/T\}$

$= \text{plim } \{e^T Q_1 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_3 (e + v)/T\}$

$= 0 \text{ as } T \to \infty$

6) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$= \text{plim } \{e^T Q_1 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_3 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_1 e/T\}$

$= \text{plim } \{e^T Q_1 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_3 X/T\}$

$\text{times plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_1 e/T\}$

$= 0 \text{ as } T \to \infty$

7) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)/T$

$= \text{plim } \{e^T Q_1 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_3 W/T\}(W^T P_3 W/T)^{-1}$

$\text{times } \{W^T P_3 Q_3 (e + v)/T\}$

$= \text{plim } \{e^T Q_1 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_3 W/T\}$

$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T P_3 Q_3 (e + v)/T\}$

$= 0 \text{ as } T \to \infty$

8) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$$= \text{plim } \{e^T Q_1 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_3 W/T\}(W^T P_3 W/T)^{-1}$$

$$\text{times } \{W^T P_3 Q_3 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_1 e/T\}$$

$$= \text{plim } \{e^T Q_1 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_3 W/T\}$$

$$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T P_3 Q_3 X/T\}$$

$$\text{times plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_1 e/T\}$$

$$= 0 \quad \text{as} \quad T \to \infty$$

9) $\text{plim } (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 (e + v)/T$

$$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\}(W^T P_3 W/T)^{-1}\{W^T Q_3 (e + v)/T\}$$

$$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\} \text{ plim } (W^T P_3 W/T)^{-1}$$

$$\text{times plim } \{W^T Q_3 (e + v)/T\}$$

$$= 0 \quad \text{as} \quad T \to \infty$$

10) $\text{plim } (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\}(W^T P_3 W/T)^{-1}\{W^T Q_3 X/T\}$$

$$\text{times } (X^T Q_1 X/T)^{-1}\{X^T Q_1 e/T\}$$

$$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\} \text{ plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T Q_3 X/T\}$$

$$\text{times plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_1 e/T\}$$

$$= 0 \quad \text{as} \quad T \to \infty$$

11) $\text{plim } (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)/T$

$$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\}(W^T P_3 W/T)^{-1}\{W^T Q_3 W/T\}$$

$$\text{times } (W^T P_3 W/T)^{-1}\{W^T P_3 Q_3 (e + v)/T\}$$

$$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\} \text{ plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T Q_3 W/T\}$$

$$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T P_3 Q_3 (e + v)/T\}$$

$$= 0 \quad \text{as} \quad T \to \infty$$

12) $\text{plim } (e + v)^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 W(W^T P_3 W)^{-1}$

$\text{times } W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\}(W^T P_3 W/T)^{-1}\{W^T Q_3 W/T\}(W^T P_3 W/T)^{-1}$

$\text{times } \{W^T P_3 Q_3 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_1 e/T\}$

$= \text{plim } \{(e + v)^T Q_3 P_3 W/T\} \text{ plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T Q_3 W/T\}$

$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T P_3 Q_3 X/T\}$

$\text{times plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_1 e/T\}$

$= 0 \text{ as } T \to \infty$

13) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 (e + v)/T$

$= \text{plim } \{e^T Q_1 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_3 P_3 W/T\}$

$\text{times } (W^T P_3 W/T)^{-1}\{W^T Q_3 (e + v)/T\}$

$= \text{plim } \{e^T Q_1 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_3 P_3 W/T\}$

$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T Q_3 (e + v)/T\}$

$= 0 \text{ as } T \to \infty$

14) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$= \text{plim } \{e^T Q_1 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_3 P_3 W/T\}$

$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T Q_3 X/T\}$

$\text{times plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_1 e/T\}$

$= 0 \text{ as } T \to \infty$

15) $\text{plim } e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1}$

$\text{times } W^T Q_3 W(W^T P_3 W)^{-1} W^T P_3 Q_3 (e + v)/T$

$= \text{plim } \{e^T Q_1 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_3 P_3 W/T\}(W^T P_3 W/T)^{-1}$

$\text{times } \{W^T Q_3 W/T\}(W^T P_3 W/T)^{-1}\{W^T P_3 Q_3 (e + v)/T\}$

$= \text{plim } \{e^T Q_1 X/T\} \text{ plim } (X^T Q_1 X/T)^{-1} \text{ plim } \{X^T Q_3 P_3 W/T\}$

$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T Q_3 W/T\}$

$\text{times plim } (W^T P_3 W/T)^{-1} \text{ plim } \{W^T P_3 Q_3 (e + v)/T\}$

$= 0$ as $T \to \infty$

16) $\text{plim}\ e^T Q_1 X(X^T Q_1 X)^{-1} X^T Q_3 P_3 W(W^T P_3 W)^{-1} W^T Q_3 W(W^T P_3 W)^{-1}$

times $W^T P_3 Q_3 X(X^T Q_1 X)^{-1} X^T Q_1 e/T$

$= \text{plim}\ \{e^T Q_1 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_3 P_3 W/T\}$

times $\text{plim}\ (W^T P_3 W/T)^{-1}\ \{W^T Q_3 W/T\}(W^T P_3 W/T)^{-1}$

times $\text{plim}\ \{W^T P_3 Q_3 X/T\}(X^T Q_1 X/T)^{-1}\{X^T Q_1 e/T\}$

$= \text{plim}\ \{e^T Q_1 X/T\}\ \text{plim}\ (X^T Q_1 X/T)^{-1}\ \text{plim}\ \{X^T Q_3 P_3 W/T\}$

times $\text{plim}\ (W^T P_3 W/T)^{-1}\ \text{plim}\ \{W^T Q_3 W/T\}$

times $\text{plim}\ (W^T P_3 W/T)^{-1}\ \text{plim}\ \{W^T P_3 Q_3 X/T\}$

times $\text{plim}\ (X^T Q_1 X/T)^{-1}\ \text{plim}\ \{X^T Q_1 e/T\}$

$= 0$ as $T \to \infty$ Q.E.D.

3.9 Conclusions

In this chapter, we have considered a linear regression model which contains unobservable time effects as well as individual effects. Given panel data, this model may be estimated in a variety of ways, depending on what is assumed about the correlation between the regressors and the effects. We have given a survey of the literature; we introduced HT-like estimators for the coefficients of the linear regression when the effects are assumed to be random and correlated with some of the regressors, and we introduced estimators for the variances of the different error components. We also introduced estimators for the above model that are consistent as $N \to \infty$ for fixed T. These estimators may be useful because a common problem with panel data is that N is large but T is small. In the next chapter, we consider the linear simultaneous equations model with effects.

CHAPTER 4

Simultaneous Equations with Effects

4.1 Introduction

In this chapter, we consider a linear simultaneous equations model with individual effects. Within this context we investigate the problem of simultaneity, defined as the case in which some of the explanatory variables are correlated with the noise component of the error. We assume that for each of the M structural equations the data again consists of T time-series observations on each of N individuals; we distinguish regressors which vary over time and individuals from those which vary over individuals but are time-invariant; and we assume the presence of unobservable, time-invariant individual effects as well as the usual statistical noise. We will refer to a variable as endogenous if it is correlated with the noise and exogenous if it is uncorrelated with the noise.

We write the model to be considered in this chapter as a set of M simultaneous equations:

$$(4.1.1) \quad y_{itg} = Y_{itg}D_g + X_{itg}B_g + Z_{ig}C_g + u_{ig} + e_{itg}$$

$$i = 1,\ldots,N; \quad t = 1,\ldots,T; \quad g = 1,\ldots,M.$$

where there are M equations determining the M endogenous variables $y_{it1},\ldots,y_{itM}$; Y_{itg} is a vector (of dimension $1 \times H_g$) of endogenous explanatory variables; X_{itg} is a vector (of dimension $1 \times G_g$) of exogenous variables which vary both over time and individuals; Z_{ig} is a vector (of dimension $1 \times K_g$) of time-invariant exogenous variables; and both D_g, B_g, and C_g are vectors to be estimated. The individual effects u_{ig} are unobservable and will be treated as time-invariant.

Writing each of the M simultaneous equations in matrix form we have

(4.1.2) $$y_g = Y_g D_g + X_g B_g + Z_g C_g + u_g + e_g$$

where y_g, u_g, and e_g denote (NT x 1) dimensioned vectors; Y_g denotes the (NT x H_g) dimensioned matrix of endogenous variables; and X_g and Z_g denote (NT x G_g), and (NT x K_g) dimensioned matrices of exogenous variables, respectively. Again, following the convention of Hausman and Taylor, the observations are ordered first by individuals and then by time, so that u_g and each column of Z_g are (NT x 1) dimensioned vectors consisting of T blocks, with each block containing the same N entries.

Rewrite equation (4.1.2) as

(4.1.3) $$y_g = R_g A_g + s_g$$

where $R_g = [\ Y_g,\ X_g,\ Z_g\ ]$ and $A_g = (\ D_g^T,\ B_g^T,\ C_g^T\ )^T$. Now consider the set of all M equations

$$(4.1.4) \quad y_* = R_* A_* + s_*$$

where

$$y_* = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_M \end{bmatrix}, \qquad s_* = \begin{bmatrix} s_1 \\ \cdot \\ \cdot \\ \cdot \\ s_M \end{bmatrix}, \qquad A_* = \begin{bmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_M \end{bmatrix}, \text{ and}$$

$$R_* = \begin{bmatrix} R_1 & & 0 \\ & \ddots & \\ 0 & & R_M \end{bmatrix}.$$

We make the usual assumptions about the error terms. That is, we assume

$$(4.1.5) \qquad \begin{bmatrix} u_{i1} \\ \cdot \\ \cdot \\ \cdot \\ u_{iM} \end{bmatrix}$$

is iid $N(0, \Sigma_u)$, and

$$(4.1.6) \qquad \begin{bmatrix} e_{it1} \\ \cdot \\ \cdot \\ \cdot \\ e_{itM} \end{bmatrix}$$

is iid $N(0, \Sigma_e)$, where Σ_u and Σ_e are both $(M \times M)$ positive definite matrices. In addition, we assume the e's are uncorrelated with both the u's and with the (exogenous) X's and Z's.

For a single equation, say the first equation, the covariance structure is

$$(4.1.7) \quad S_{11} = Cov(\ u_1 + e_1\ ) = \Sigma_{e,11} I_{NT} + \Sigma_{u,11}(TP)$$
$$= \Sigma_{e,11} Q + (\Sigma_{e,11} + T\Sigma_{u,11})P = \sigma_1^2 Q + \sigma_2^2 P$$

where Q and P are the same two idempotent matrices given in chapter 2, $\sigma_1^2 = \Sigma_{e,11}$, and $\sigma_2^2 = (\Sigma_{e,11} + T\Sigma_{u,11})$; and so

$$(4.1.8) \quad S_{11}^{-1} = (1/\sigma_1^2)Q + (1/\sigma_2^2)P$$

and

$$(4.1.9) \quad S_{11}^{-1/2} = (1/\sigma_1)Q + (1/\sigma_2)P.$$

And for the system, the covariance structure is

$$(4.1.10) \quad S = Cov(\ u_* + e_*\ ) = (\ \Sigma_e \otimes I_{NT}\ ) + (\ \Sigma_u \otimes (TP)\ )$$
$$= (\ \Sigma_1 \otimes Q\ ) + (\ \Sigma_2 \otimes P\ )$$

where $\Sigma_1 = \Sigma_e$ and $\Sigma_2 = (\Sigma_e + T\Sigma_u)$.

Throughout this chapter we will consider a natural extension of the Hausman and Taylor model to a linear simultaneous equations model with random effects by allowing some of the explanatory variables to be correlated with the individual effects. The plan of this chapter is as follows. In section 4.2 we consider the estimation of the coefficients of a single linear equation from a simultaneous equations model. In section 4.3 we consider the estimation of the coefficients of a system of simultaneous equations. An interesting problem arises for the linear simultaneous equations model with random effects when some of the explanatory variables are correlated with the individual random effects; namely, the instruments need not be the same for every equation. This is the topic discussed in section 4.4. We summarize our results in section 4.5.

This chapter applies the Hausman and Taylor method of instrumental variables estimation to the simultaneous equations panel data model, derives the subsequent estimators, and discusses their relative efficiency. In addition, it provides a survey of the current literature on simultaneous equations with effects and translates those estimators into the notation of this thesis.

4.2 Single Equation Estimation

Let us now turn to the problem of estimating the coefficients of a single equation, say the first equation. That is, we wish to estimate the equation

$$(4.2.1)\quad y_1 = R_1 A_1 + (u_1 + e_1).$$

This is a generalization of the estimation problem considered in chapter 2 in the sense that, in addition to the "inside" instruments (i.e. instruments from within the equation itself), we now have available instruments from "outside" the equation. Now we need to introduce some notation, but first we must agree on the type of explanatory variables permitted. Amemiya and MaCurdy (1986) have considered a simultaneous equation model with random effects correlated with the endogenous variables, but in a somewhat non-standard way. The basic point of view in this thesis is that all variables correlated with the noise should also be correlated with the individual effects, but not conversely. That is, only exogenous variables can be uncorrelated with the individual effects. This point of view can be justified by consideration of a system in which every structural equation contains unobserved individual effects. By standard algebra such a system would imply a reduced form, in which each reduced form equation has an individual effect which is a linear combination of the individual effects in the structural equations. It therefore follows that every endogenous variable will be correlated with the individual effect in every equation, just as it is correlated with every structural error term. Thus, all endogenous variables must be correlated with the effects.

On the otherhand, if we follow a natural extention of the point of view in Hausman and Taylor, there are two kinds of exogenous variables possible; namely, those uncorrelated and those possibly correlated with the individual effects. That is, if we let X and Z represent the matrices of all time-varying and time-invariant exogenous variables, respectively, we can then write X and Z as

(4.2.2) $$X = [\ X_{(1)},\ X_{(2)}\ ]$$

(4.2.3) $$Z = [\ Z_{(1)},\ Z_{(2)}\ ]$$

where $X_{(1)}$ and $Z_{(1)}$ represents the doubly exogenous variables, meaning variables uncorrelated with the individual effects as well as the noise; and $X_{(2)}$ and $Z_{(2)}$ represents the singly exogenous variable, meaning variables uncorrelated with noise but possibly correlated with individual effects. It is important to note that $X_{(1)}$ is not the same as X_1. That is, X_1 is the matrix of time-varying exogenous variables that appear in the first equation, and since X_1 may consist of doubly as well as singly exogenous variables it may have elements in both $X_{(1)}$ and $X_{(2)}$. On the other hand, X contains both the doubly and singly exogenous variables from every equation, not just the first, so both $X_{(1)}$ and $X_{(2)}$ may contain elements not in X_1. A similiar relationship holds between Z_1, $Z_{(1)}$, $Z_{(2)}$, and Z.

It is an important observation that will be used later that each instrument set considered in this chapter is of the form

$$(4.2.4) \qquad H = [\ QX,\ PE\ ]$$

where the set E will vary. Given this form we can evaluate P[H] using the following Lemma:

Lemma (4.1): $P[H] = P[QX] + P[PE]$

Proof:

$$P[H] = (\ QX\ \ PE\ ) \begin{bmatrix} X^T Q \\ E^T Q \end{bmatrix} (\ QX\ \ PE\ ) \Bigg]^{-1} \begin{bmatrix} X^T Q \\ E^T Q \end{bmatrix}$$

$$= (\ QX\ \ PE\ ) \begin{bmatrix} X^T QX & 0 \\ 0 & E^T QE \end{bmatrix}^{-1} \begin{bmatrix} X^T Q \\ E^T P \end{bmatrix}$$

$$= QX(X^T QX)^{-1} X^T Q + PE(E^T PE)^{-1} E^T P$$

$$\equiv \quad P[QX] \quad + \quad P[PE] \qquad \text{Q.E.D.}$$

The obvious generalization of the analysis of Hausman and Taylor would be to choose $E = [\ X_{(1)},\ Z_{(1)}\ ]$ so that the instrument set is

(4.2.5) $H = [\ QX,\ X_{(1)},\ Z_{(1)}\ ]$.

But we could also consider $E = [\ X^*_{(1)},\ Z_{(1)}\ ]$, which is essentially the instrument set suggested by Amemiya and McCurdy. As explained by Breusch, Mizon, and Schmidt (1987), the matrix $X^*_{(1)}$ displays each variable <u>separately</u> for t = 1,2,...,T. That is, for any T x L panel data matrix S, the T x LT matrix S^* is defined by

$$\text{(4.2.6)} \quad S = \begin{bmatrix} s_{11} \\ \vdots \\ s_{1T} \\ \hline \vdots \\ \hline s_{N1} \\ \vdots \\ s_{NT} \end{bmatrix}, \quad S^* = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1T} \\ \vdots & \vdots & & \vdots \\ s_{11} & s_{12} & \cdots & s_{1T} \\ \hline \vdots & \vdots & & \vdots \\ \hline s_{N1} & s_{N2} & \cdots & s_{NT} \\ \vdots & \vdots & & \vdots \\ s_{N1} & s_{N2} & \cdots & s_{NT} \end{bmatrix}.$$

This leads to the instrument set

(4.2.7) $H_{AM} = [\ QX,\ X^*_{(1)},\ Z_{(1)}\ ]$

A third possibility is $E = [\ X^*_{(1)},\ Z_{(1)},\ QX^*_{(2)}\ ]$, which implies the instrument set

(4.2.8) $H_{BMS} = [\ QX,\ X^*_{(1)},\ Z_{(1)},\ QX^*_{(2)}\ ]$

suggested by BMS. For our purposes the list of instruments given in (4.2.5) will suffice, since the algebra in the other cases is the same.

(4.2.1) Two-Stage Least Squares

We derive the two-stage least squares estimator as follows. First, we multiply equation (4.2.1) by $S_{11}^{-1/2}$ to transform the error to a scalar covariance matrix. The transformed equation is simply

(4.2.9) $S_{11}^{-1/2}y_1 = S_{11}^{-1/2}R_1A_1 + S_{11}^{-1/2}(u_1 + e_1).$

We then follow the path of Hausman and Taylor, by estimating (4.2.9) using IV with instrument set H. This yields the following definition:

Definition (4.2): The two-stage least squares (2SLS) estimator of A_1 from equation (4.2.1) is the instrumental variables estimator of equation (4.2.9), using the instrument set H. Explicitly,

(4.2.10) $a_{1,2SLS}$

$$= [\ R_1^T S_{11}^{-1/2} P[H] S_{11}^{-1/2} R_1\ ]^{-1}$$

$$\text{times}\quad R_1^T S_{11}^{-1/2} P[H] S_{11}^{-1/2} y_1 .$$

It is an interesting detail that although we have transformed equation (4.2.1), we have used the untransformed instruments, H. Following White (1984, section IV.3), the optimal IV estimator is derived by transforming the equation to be estimated so that its error covariance is scalar (as we have done), and then using whatever instruments are optimal. Thus, in general, the question of whether H or $S_{11}^{-1/2}H$ is preferable depends on which instrument set better explains the endogenous variables contained in $S_{11}^{-1/2}R_1$. As Breusch, Mizon, and Schmidt point out, however, in the present context transforming the instruments by $S_1^{-1/2}$ makes no difference; either instrument set leads to the same estimator. This is implied by the following Lemma:

Lemma (4.3): Given H = [QX, PE] defined in (4.2.5) and $S_{11}^{-1/2}$ defined in (4.1.9),

(4.2.11) $P[H] = P[S_{11}^{-1/2}H]$.

Proof:

$$
\begin{aligned}
P[S_{11}^{-1/2}H] &= P[\ (1/\sigma_1)QX + (1/\sigma_2)PE\ ] \\
&= P[\ (1/\sigma_1)QX\ ] + P[\ (1/\sigma_2)PE\ ] \\
&= \{(1/\sigma_1)QX\}\{(1/\sigma_1^2)X^TQX\}^{-1}\{(1/\sigma_1)QX\}^T \\
&\quad + \{(1/\sigma_2)PE\}\{(1/\sigma_2^2)E^TPE\}^{-1}\{(1/\sigma_2)PE\}^T \\
&= QX\{X^TQX\}^{-1}X^TQ + PE\{E^TPE\}^{-1}E^TP \\
&= P[\ QX\ ] + P[\ PE\ ] = P[\ H\ ]
\end{aligned}
$$

Q.E.D.

(4.2.2) An Orthogonality Condition Derivation of the 2SLS Estimator

Following Hausman, Newey, and Taylor (1987), we consider an interpretation of the 2SLS estimator implied by the instrument-residual orthogonality condition written as plim $f_1/NT = 0$, where

$$(4.2.12) \qquad f_1 = H^T S_{11}^{-1/2}(y_1 - R_1 A_1).$$

Now the covariance structure of f_1 is

$$(4.2.13) \qquad \mathrm{Cov}(f_1) \equiv C_1 = H^T S_{11}^{-1/2} S_{11} S_{11}^{-1/2} H = H^T H$$

The instrumental variables estimator (also known as the "Generalized Method of Moments" estimator) then is the solution to the problem of minimizing with respect to A_1 the quadratic distance from zero of f_1:

$$(4.2.14) \qquad f_1^T C_1^{-1} f_1 = (y_1 - R_1 A_1)^T W_1 (y_1 - R_1 A_1)$$

where

$$(4.2.15) \qquad W_1 = S_{11}^{-1/2} H C_1^{-1} H^T S_{11}^{-1/2}$$
$$= S_{11}^{-1/2} H (H^T H)^{-1} H^T S_{11}^{-1/2}$$

is a quadratic form. This solution can be written as

$$(4.2.16) \qquad a_{1,IV} = [\, R_1^T W_1 R_1 \,]^{-1} R_1^T W_1 y_1$$
$$= [\, R_1^T S_{11}^{-1/2} H (H^T H)^{-1} H^T S_{11}^{-1/2} R_1 \,]^{-1}$$
$$\text{times} \quad R_1^T S_{11}^{-1/2} H (H^T H)^{-1} H^T S_{11}^{-1/2} y_1$$
$$= [\, R_1^T S_{11}^{-1/2} P[H] S_{11}^{-1/2} R_1 \,]^{-1}$$
$$\text{times} \quad R_1^T S_{11}^{-1/2} P[H] S_{11}^{-1/2} y_1 .$$

It can readily be seen that $a_{1,IV}$ is equal to the 2SLS estimator of A_1 given in (4.2.10).

It is an interesting result that $S_{11}^{-1/2}$ in the orthogonality condition given in (4.2.12) is superfluous. To see this, consider the simpler orthogonality condition plim $f_2/NT = 0$, where

$$(4.2.17) \quad f_2 = H^T(y_1 - R_1A_1).$$

Noting that

$$(4.2.18) \quad Cov(\ f_2\ ) \equiv C_2 = H^TS_{11}H,$$

the problem of minimizing with respect to A_1 the quadratic distance from zero of f_2,

$$(4.2.19) \quad f_2^TC_2^{-1}f_2 = (y_1 - R_1A_1)^TW_2(y_1 - R_1A_1)$$

where the quadratic form $W_2 = H(H^TS_{11}H)^{-1}H^T$, yields the solution

$$(4.2.20) \quad a_{2,IV} = [\ R_1^TW_2R_1\ ]^{-1}R_1^TW_2y_1.$$

Now we can write W_2 as

$$(4.2.21) \quad W_2$$

$$= H(H^TS_{11}H)^{-1}H^T$$

$$= (\ QX\ \ PE\ )\left\{ \begin{bmatrix} X^TQ \\ E^TQ \end{bmatrix} (\ \sigma_1^2Q + \sigma_2^2Q\ )(\ QX\ \ PE\ )\ \right\}^{-1} \begin{bmatrix} X^TQ \\ E^TQ \end{bmatrix}$$

$$= (\ QX\ \ PE\ ) \begin{bmatrix} \sigma_1^2X^TQX & 0 \\ 0 & \sigma_2^2E^TQE \end{bmatrix}^{-1} \begin{bmatrix} X^TQ \\ E^TP \end{bmatrix}$$

$$= (1/\sigma_1^2)QX(X^TQX)^{-1}X^TQ \quad + \quad (1/\sigma_2^2)PE(E^TPE)^{-1}E^TP$$

$$= (1/\sigma_1^2)P[QX] \quad + \quad (1/\sigma_2^2)P[PE].$$

On the other hand, W_1 given in (4.2.15) can be written as

$$(4.2.22) \quad W_1 = S_{11}^{-1/2}H(H^TH)^{-1}H^TS_{11}^{-1/2}$$

$$= [\ (1/\sigma_1^2)Q + (1/\sigma_2^2)P\ ]P[H]$$

$$\text{times } (1/\sigma_1^2)Q + (1/\sigma_2^2)P\ ]$$

$$= (1/\sigma_1^2)P[QX] \quad + \quad (1/\sigma_2^2)P[PE].$$

Therefore, $W_1 = W_2$ and the two estimators are the same. Substituting W_1 from (4.2.22) into the 2SLS estimator given in (4.2.16), we can rewrite the estimator as

$$(4.2.23) \quad a_{1,2SLS}$$

$$= [\ (1/\sigma_1^2(QR_1)^TP[QX](QR_1)$$

$$+ \ (1/\sigma_2^2(PR_1)^TP[PE](PR_1)\ ]^{-1}$$

$$\text{times } [\ (1/\sigma_1^2(QR_1)^TP[QX](Qy_1)$$

$$+ \ (1/\sigma_2^2(PR_1)^TP[PE](Py_1)\ ].$$

Now the same line of proof used above would show that you would get the same estimator based on the orthogonality conditions

$$(4.2.24) \quad f_3 = H^TS_{11}^{-1}(y_1 - R_1A_1).$$

This is in any case obvious because it corresponds to transforming both the equation and the instruments by $S_{11}^{-1/2}$, which we have shown above to be the same as transforming only the equation and not the instruments.

(4.2.3) Baltagi's Error-Component Two-Stage Least Squares Estimator

Baltagi (1981) considers a simultaneous equations model with effects which, in addition to individual effects, contains time effects as well. In contrast to the model considered in this chapter, Baltagi's does not distinguish between doubly and singly exogenous variables; implicitly he assumes that only doubly exogenous variables exist among the explanatory variables. In Baltagi's notation the Error-Component Two-Stage Least Squares (EC2SLS) estimator can be written as

(4.2.25)
$$a_{1,EC2SLS} = \{ \sum_{h=1}^{3} (Z_1^{(h)})^T P[X^{(h)}] Z_1^{(h)} / \sigma_{11}^{(h)2} \}^{-1}$$
$$\text{times} \quad \{ \sum_{h=1}^{3} (Z_1^{(h)})^T P[X^{(h)}] Z_1^{(h)} / \sigma_{11}^{(h)2} \}.$$

On the other hand, the 2SLS estimator given in (4.2.23) can again be written as

(4.2.26)
$$a_{1,2SLS} = [(1/\sigma_2^2 (QR_1)^T P[QX](QR_1) + (1/\sigma_1^2 (PR_1)^T P[PE](PR_1)]^{-1}$$
$$\text{times} \quad [(1/\sigma_2^2 (QR_1)^T P[QX](Qy_1) + (1/\sigma_1^2 (PR_1)^T P[PE](Py_1)].$$

This is "essentially" Baltagi's estimator translated into our notation. We use the word "essentially" in the preceding sentence because we do not include time effects in our model. Now, if we assume only individual effects, we have the translation as follows: Baltagi's Z_1 is our R_1, his δ_1 is our A_1, his X is our (X, Z), his $Z_1^{(1)}$ is our PR_1, his $X^{(1)}$ is our PX, his $\sigma_{11}^{(1)}$ is our σ_2^2, and both his $\sigma_{11}^{(2)}$ and $\sigma_{11}^{(3)}$ are our σ_1^2. Since time effects are not present the distinction between the two terms $X^{(2)}$ and $X^{(3)}$ is irrelevant so in Baltagi's notation $X^{(2)} + X^{(3)}$ is our QX. Similiarly, his $Z_1^{(2)} + Z_1^{(3)}$ is equal to our QR_1. Therefore, Baltagi's EC2SLS estimator can be written using our notation as

$$(4.2.27) \quad a_{1,EC2SLS} = [\ (1/\sigma_2^2(QR_1)^T P[QX](QR_1) + (1/\sigma_1^2(PR_1)^T P[P(X,\ Z)](PR_1)\ ]^{-1} \text{ times } [\ (1/\sigma_2^2(QR_1)^T P[QX](Qy_1) + (1/\sigma_1^2(PR_1)^T P[P(X,\ Z)](Py_1)\ ].$$

It is easily seen that this estimator is the same as the $a_{1,2SLS}$ when E = (X, Z); that is, when there are no singly exogenous variables.

4.3 System Estimation

In section 4.2 we discussed "single-equation" methods of estimation in the sense that the estimators there operated on each equation separately. This section will discuss "systems" methods of estimation, which estimate all equations jointly. The motivation for considering joint estimation is of course that the joint estimates are generally more (asymptotically) efficient than the single-equation procedures.

Again, let us consider the set of all M equations

(4.3.1) $$y_* = R_* A_* + s_*$$

where of course

$$y_* = \begin{bmatrix} y_1 \\ . \\ y_M \end{bmatrix}, \quad s_* = \begin{bmatrix} s_1 \\ . \\ s_M \end{bmatrix}, \quad A_* = \begin{bmatrix} A_1 \\ . \\ A_M \end{bmatrix}, \text{ and}$$

$$R_* = \begin{bmatrix} R_1 & & 0 \\ & . & \\ 0 & & R_M \end{bmatrix} .$$

Note that the covariance matrix of s_* is

(4.3.2) $$S = Cov(u_* + e_*) = (\Sigma_e \otimes I_{NT}) + (\Sigma_u \otimes (TP)) = (\Sigma_1 \otimes Q) + (\Sigma_2 \otimes P)$$

(4.3.3) $$S^{-1} = (\Sigma_1^{-1} \otimes Q) + (\Sigma_2^{-1} \otimes P)$$

(4.3.4) $$S^{-1/2} = (\Sigma_1^{-1/2} \otimes Q) + (\Sigma_2^{-1/2} \otimes P)$$

where $\Sigma_1 = \Sigma_e Q$ and $\Sigma_2 = (\Sigma_e + T\Sigma_u)$; Q and P are, again the two idempotent matrices used before.

Recall that X and Z represent the matrices of all time-varying and time-invariant exogenous variables, respectively, and that we can write X and Z as

(4.3.5) $$X = [X_{(1)}, X_{(2)}]$$

(4.3.6) $$Z = [Z_{(1)}, Z_{(2)}]$$

where $X_{(1)}$ and $Z_{(1)}$ represent the <u>doubly</u> <u>exogenous</u> variables, meaning uncorrelated with the individual effects as well as the noise; and $X_{(2)}$ and $Z_{(2)}$ represent the <u>singly exogenous</u> variable, meaning uncorrelated with noise but possibly correlated with individual effects. Note that the decomposition in both equations (4.3.5) and (4.3.6) are without reference to a particular equation. This is because we are assuming that we have the same instruments in every equation; that is, if a variable is <u>doubly exogenous</u> in one equation then it is <u>doubly exogenous</u> in every equation and likewise, if a variable is <u>singly exogenous</u> in one equation then it is <u>singly exogenous</u> in every equation. We will consider the more complicated case when the instruments may differ from equation to equation in section 4.4. Finally, recall that our instrument set is of the form $H = [\ QX,\ PE\ ]$, where $E = (\ X_{(1)},\ Z_{(1)}\ )$.

(4.3.1) <u>Three-Stage Least Squares</u>

We derive the three-stage least squares as follows. First, we multiply equation (4.3.1) by $S_*^{-1/2}$ to transform the error to a scalar covariance matrix. The transformed equation is simply

$$(4.3.7)\quad S_*^{-1/2}y_* = S_*^{-1/2}R_*A_* + S_*^{-1/2}s_*.$$

We then follow the path of Hausman and Taylor, by estimating (4.3.7) using IV with instrument set (I O H). This yields the following definition:

Definition (4.4): The three-stage least squares (3SLS) estimator of A_* from equation (4.3.1) is the instrumental variables estimator of equation (4.3.7), using the instrument set $(I \otimes H)$. Explicitly,

(4.3.8) a_{3SLS}

$$= [\ R_*^T (\ \Sigma_1^{-1} \otimes P[QX]\ ) + (\ \Sigma_2^{-1} \otimes P[PE]\ ) R_*\ ]^{-1}$$
$$\text{times } R_*^T (\ \Sigma_1^{-1} \otimes P[QX]\ ) + (\ \Sigma_2^{-1} \otimes P[PE]\ ) R_* y_* .$$

(4.3.2) __Instrumental Variables Estimation__

Following Hausman, Newey, and Taylor (1987), we consider an interpretation of the 3SLS estimator implied by the instrument-residual orthogonality condition written as $\text{plim } f_*/NT = 0$, where

$$(4.3.9) \quad f_* = \begin{bmatrix} H^T(y_1 - R_1 A_1) \\ \vdots \\ H^T(y_M - R_M A_M) \end{bmatrix}$$
$$= (I \otimes H^T)(y_* - R_* A_*).$$

The covariance structure of f_* is

$$(4.3.10) \quad \text{Cov}(f_*) \equiv C_* = (I \otimes H^T) S_* (I \otimes H)$$
$$= (I \otimes H^T)(\ \Sigma_1 \otimes Q\ ) + (\ \Sigma_2 \otimes P\ )(I \otimes H)$$
$$= (\ \Sigma_1 \otimes H^T Q H\ ) + (\ \Sigma_2 \otimes H^T P H\ ).$$

To assist in the simplification of the estimators considered below we need the following Lemma:

Lemma(4.5): Suppose T_1 and T_2 are positive definite, nonsingular matrices and $H = [\ QX,\ PE\ ]$. Then

$$(4.3.11) \quad \{ (T_1 \otimes H^T QH) + (T_2 \otimes H^T PH) \}^{-1} = T_1^{-1} \otimes (H^T QH)^{-1} + T_2^{-1} \otimes (H^T PH)^{-1}.$$

Proof:

Using Baltagi's lemma (Baltagi (1980), p. 1548), it is sufficient to show that

$$(H^T QH)(H^T PH)$$

$$= \begin{bmatrix} X^T Q \\ E^T P \end{bmatrix} Q(\ QX\ \ PE\ ) \begin{bmatrix} X^T Q \\ E^T P \end{bmatrix} P(\ QX\ \ PE\ )$$

$$= \begin{bmatrix} X^T QX & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & E^T PE \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0. \qquad \text{Q.E.D.}$$

As before the instrumental variables estimator (also known as the "Generalized Method of Moments" estimator) is then the solution to the problem of minimizing with respect to A_* the quadratic distance from zero of f_*,

$$(4.3.12) \quad f_*^T C_*^{-1} f_* = (y_* - R_* A_*)^T W_* (y_* - R_* A_*)$$

where

$$(4.3.13) \quad W_* = (I \otimes H) C_*^{-1} (I \otimes H^T)$$

is a quadratic form. By Lemma (4.5), C_*^{-1} can be written as

(4.3.14) $$C_*^{-1} = \Sigma_1^{-1} \otimes (H^T QH)^{-1} + \Sigma_2^{-1} \otimes (H^T PH)^{-1}$$

so we can rewrite W_* as

(4.3.15)
$$\begin{aligned} W_* &= (I \otimes H)(\Sigma_1^{-1} \otimes (H^T QH)^{-1} \\ &\qquad + \Sigma_2^{-1} \otimes (H^T PH)^{-1})(I \otimes H^T) \\ &= (\Sigma_1^{-1} \otimes H(H^T QH)^{-1} H^T) \\ &\qquad + (\Sigma_2^{-1} \otimes H(H^T PH)^{-1} H^T) \\ &= (\Sigma_1^{-1} \otimes P[QH]) + (\Sigma_2^{-1} \otimes P[PH]) \\ &= (\Sigma_1^{-1} \otimes P[QX]) + (\Sigma_2^{-1} \otimes P[PE]). \end{aligned}$$

The solution can be written as

(4.3.16) a_*
$$\begin{aligned} &= [R_*^T W_* R_*]^{-1} R_*^T W_* y_* \\ &= [R_*^T (\Sigma_1^{-1} \otimes P[QX]) + (\Sigma_2^{-1} \otimes P[PE]) R_*]^{-1} \\ &\quad \text{times} \quad R_*^T (\Sigma_1^{-1} \otimes P[QX]) + (\Sigma_2^{-1} \otimes P[PE]) y_* \end{aligned}$$

It can readily be seen that a_* is equal to the 3SLS estimator of A_* given in (4.3.8).

An alternative estimator can be derived from the instrument-residual orthogonality conditions given in (4.3.9) if, in place of the quadratic form W_* we use instead

(4.3.17)
$$\begin{aligned} W_2 &= (I \otimes H)[(\mathrm{diag}(\Sigma_1) \otimes H^T QH) \\ &\qquad + (\mathrm{diag}(\Sigma_2) \otimes H^T PH)]^{-1}(I \otimes H^T). \\ &= \mathrm{diag}(\Sigma_1)^{-1} \otimes H(H^T QH)^{-1} H^T \\ &\qquad + \mathrm{diag}(\Sigma_2)^{-1} \otimes H(H^T PH)^{-1} H^T \\ &= \mathrm{diag}(\Sigma_1)^{-1} \otimes P[QH] + \mathrm{diag}(\Sigma_2)^{-1} \otimes P[PH] \end{aligned}$$

where $\mathrm{diag}(\Sigma_1)$ and $\mathrm{diag}(\Sigma_2)$ are diagonal matrices whose diagonal entries are the diagonal entries of Σ_1 and Σ_2, respectively. Thus, we do not take account of the fact the covariance structure of f_* is $(\Sigma_1 \otimes H^T QH) + (\Sigma_2 \otimes H^T PH)$ rather than $(\mathrm{diag}(\Sigma_1) \otimes H^T QH) + (\mathrm{diag}(\Sigma_2) \otimes H^T PH)$. This yields the estimator

(4.3.18) a_{*2}

$$= [R_*^T W_2 R_*]^{-1} R_*^T W_2 y_*$$

$$= [R_*^T(\mathrm{diag}(\Sigma_1)^{-1} \otimes P[QH] + \mathrm{diag}(\Sigma_2)^{-1} \otimes P[PH])R_*]^{-1}$$

$$\text{times} \quad R_*^T(\mathrm{diag}(\Sigma_1)^{-1} \otimes P[QH] + \mathrm{diag}(\Sigma_2)^{-1} \otimes P[PH])y_*.$$

Since R_*, $\mathrm{diag}(\Sigma_1)^{-1} \otimes P[QH]$, and $\mathrm{diag}(\Sigma_2)^{-1} \otimes P[PH]$ are block diagonal and since $\Sigma_{1ii} P[QH] + \Sigma_{1i2} P[PH]$ $= S_{ii}^{-1} P[H] = S_{ii}^{-1/2} P[H] S_{ii}^{-1/2}$, we can rewrite (4.3.18) as

(4.3.19) a_{*2}

$$= \begin{bmatrix} [R_1^T S_{11}^{-1/2} P[H] S_{11}^{-1/2} R_1]^{-1} & & \\ & \ddots & \\ & & [R_M^T S_{MM}^{-1/2} P[H] S_{MM}^{-1/2} R_M]^{-1} \end{bmatrix}$$

$$\text{times} \begin{bmatrix} R_1^T S_{11}^{-1/2} P[H] S_{11}^{-1/2} y_1 \\ \vdots \\ R_M^T S_{MM}^{-1/2} P[H] S_{MM}^{-1/2} y_M \end{bmatrix}$$

which can be seen equal to be 2SLS applied to each equation separately.

Since W_2 is a suboptimal weighting matrix, this is one way of proving 3SLS efficient relative to 2SLS.

Still another estimator can be derived if we consider, instead of (4.3.9), the instrument-residual orthogonality conditions written as plim $f_3/NT = 0$, where

$$(4.3.20) \quad f_3 = (I \otimes H^T)S^{-1}(y_* - R_* A_*).$$

The covariance structure of f_3 is written as

$$(4.3.21) \quad C_3 \equiv \mathrm{Cov}(\ f_3\ ) = (I \otimes H^T)S^{-1}(I \otimes H).$$

It is an interesting result that S^{-1} in the orthogonality condition given above is superfluous. To see this, consider the problem of minimizing with respect to A_* the quadratic distance from zero of f_3,

$$(4.3.22) \quad f_3{}^T C_3{}^{-1} f_3 = (y_* - R_* A_*)^T W_3 (y_* - R_* A_*)$$

using the quadratic form

$$(4.3.23) \quad W_3 = S^{-1}(I \otimes H)[(I \otimes H^T)S^{-1}(I \otimes H)]^{-1}(I \otimes H^T)S^{-1}.$$

The solution to this problem yields the estimator

$$(4.3.24) \quad a_{*3} = [\ R_*{}^T W_3 R_*\ ]^{-1} R_*{}^T W_3 y_*.$$

Now we can write $C_3{}^{-1}$ as

$$(4.3.25) \quad C_3{}^{-1}$$

$$= \{\ (I \otimes H^T)S^{-1}(I \otimes H)\ \}^{-1}$$

$$= \{\ (I \otimes H^T)(\ \Sigma_1{}^{-1} \otimes Q + \Sigma_2{}^{-1} \otimes P\ )(I \otimes H)\ \}^{-1}$$

$$= \{\ \Sigma_1{}^{-1} \otimes H^T Q H + \Sigma_2{}^{-1} \otimes H^T P H\ \}^{-1}$$

$$= \Sigma_1 \otimes (H^T QH)^{-1} + \Sigma_2 \otimes (H^T PH)^{-1}$$

using Lemma (4.5). Then W_3 can be written as

(4.3.26) W_3

$$= S^{-1}(I \otimes H)C_3^{-1}(I \otimes H^T)S^{-1}$$

$$= S^{-1}(I \otimes H)\{ \Sigma_1 \otimes (H^T QH)^{-1} + \Sigma_2 \otimes (H^T PH)^{-1} \}(I \otimes H^T)S^{-1}$$

$$= \Sigma_1^{-1}\Sigma_1\Sigma_1^{-1} \otimes H(H^T QH)^{-1}H^T + \Sigma_2^{-1}\Sigma_2\Sigma_2^{-1} \otimes H(H^T PH)^{-1}H^T$$

$$= \Sigma_1^{-1} \otimes P[QH] + \Sigma_2^{-1} \otimes P[PH])$$

$$= \Sigma_1^{-1} \otimes P[QX] + \Sigma_2^{-1} \otimes P[PE]).$$

Therefore, comparing W_3 given above to W_* given in (4.3.15) it is clear that $W_* = W_2$, so the two problems are the same and the presence of S^{-1} in the orthogonality condition of (4.3.20) is irrelevant.

(4.3.3) <u>Special Cases</u>

Consider the model

(4.3.27) $y_* = R_* A_* + s_*$

We showed in section 4.3.2 that the 3SLS estimator can be interpreted as an IV estimator using the instrument set $(I \otimes H)$. As we have shown before, $H = [QX, PE]$ where $E = (X_1, Z_1)$ contains the doubly exogonous variables present in the model.

For our first special case, suppose that there are no doubly exogonous variables; i.e. all exogenous variables are correlated with the individual effects. Then the set E is empty and we should have fixed effects. Suppose also there are no time-invariant variables Z since estimation of their coefficients would now be impossible. Our estimator becomes

$$(4.3.28) \quad a_{3SLS} = [R_*^T (\Sigma_1^{-1} \otimes P[QX] R_*]^{-1} \text{ times } R_*^T(\Sigma_1^{-1} \otimes P[QX])y_*.$$

An alternative approach to fixed effects estimation would be to derive the 3SLS estimator by first premultiplying equation (4.3.27) by $(I \otimes Q)$, a system-wide within transformation, yielding

$$(4.3.29) \quad (I \otimes Q)y_* = (I \otimes Q)R_*A_* + (I \otimes Q)s_*$$

and then using the instrument set H = (QX). This fixed effects (within) estimator becomes

$$(4.3.30) \quad a_{FE} = [R_*^T(\Sigma_1^{-1} \otimes P[QX])R_*]^{-1}R_*^T(\Sigma_1^{-1} \otimes P[QX])R_*y_*.$$

This estimator can be seen as equal to our 3SLS estimator when there are no <u>doubly</u> <u>exogenous</u> variables.

Estimation of the panel-data simultaneous equations model with fixed effects have been considered by Cornwell and Schmidt (1987). There they show that in a simultaneous equation model in which the same exogenous variables in each equation have coefficients which vary over individuals, the MLE, the conditional MLE and the marginal MLE coincide. This is obviously a more general model than the one being considered here, but their model does simplifiy to a fixed-effects version of the simultaneous equations model with individual effects. In effect, they show that the MLE, CMLE and MMLE coincide in a simultaneous equation model with fixed effects. Their results imply that just as in the single equation case, the coefficients of the time and individual varying explanatory variables are determined by the "within" component of the likelihood and that the coefficients of the time invariant or the individual-invariant explanatory variables is determined by the appropriate "between" component of likelihood.

For our second special case, suppose there are no <u>singly exogenous</u> variables so all the exogenous variables are assumed uncorrelated with the individual effects. This is the Baltagi case; that is, the case when H = [QX PE] where E = (X, Z). In Baltagi's notation his error-component three-stage least squares (EC3SLS) estimator can be written as

(4.3.31) $_{EC3SLS}$

$$= \{ \sum_{h=1}^{3} (Z_1^{(h)})^T (\Sigma^{(h)} \otimes P[X^{(h)}]) Z_1^{(h)} \}^{-1}$$

$$\text{times} \ \{ \sum_{h=1}^{3} (Z_1^{(h)})^T (\Sigma^{(h)} \otimes P[X^{(h)}]) Z_1^{(h)} \}.$$

On the other hand, the 3SLS estimator given in (4.3.8) can again be written as

(4.3.32) $_{a3SLS}$

$$= [R_*^T (\Sigma_1^{-1} \otimes P[QX]) + (\Sigma_2^{-1} \otimes P[PE]) R_*]^{-1}$$

$$\text{times} \ R_*^T (\Sigma_1^{-1} \otimes P[QX]) + (\Sigma_2^{-1} \otimes P[PE]) R_* y_* .$$

This is essentially Baltagi's estimator translated into our notation. That is, if we assume only individual effects, note that his $\Sigma^{(1)}$ is our Σ_2^{-1} and his $\Sigma^{(2)} + \Sigma^{(3)}$ is our Σ_1^{-1}, and further, use the translation of section 4.2.3, then Baltagi's EC3SLS estimator in our notation is

(4.3.33) a_{EC3SLS}

$$= [\ R_*^T(\ \Sigma_1^{-1} \otimes P[QX]\ ) + (\ \Sigma_2^{-1} \otimes P[PE]\ )R_*\ ]^{-1}$$

$$\text{times } R_*^T(\ \Sigma_1^{-1} \otimes P[QX]\ ) + (\ \Sigma_2^{-1} \otimes P[PE]\ )R_* y_*.$$

which is the same as a_{3SLS}.

(4.4) <u>3SLS with Different Instruments</u>

We now allow different instruments to exist in different equations. To this end we need to introduce some notation. Let H_1 = [QX PE_1] be the instrument set for equation 1, H_2 = [QX PE_2] be the instrument set for equation 2, etc. Note that as before each instrument set is of the form H = [QX PE], but the E's differ across equations. This is because they contain variables that are <u>doubly</u> <u>exogenous</u> but only with respect to each particular equation. In this section a variable which is <u>doubly</u> <u>exogenous</u> for one equation may not be <u>doubly</u> <u>exogenous</u> in another.

Recall that in section 4.2 we derived the 2SLS estimator for the first equation by considering the instrument-residual orthogonality condition based on

(4.4.1) $f_1 = H_1^T(y_1 - R_1 A_1).$

Using

(4.4.2) $C_1 \equiv \mathrm{Cov}(\ f_1\ ) = H_1{}^T S_{11} H_1$,

the solution to the problem of minimizing the quadratic distance from zero to f_1,

(4.4.3) $f_1{}^T (H_1{}^T S_{11} H_1)^{-1} f_1$

$= (y_1 - R_1 A_1) H_1 (H_1{}^T S_{11} H_1)^{-1} H_1{}^T (y_1 - R_1 A_1$

$= (y_1 - R_1 A_1)((1/\sigma_1{}^2)\ P[QX] + (1/\sigma_2{}^2) P[PE_1])(y - R_1 A_1)$

yields the estimator

(4.4.4) $a_{1,2SLS}$

$= [R_1{}^T W_1 R_1]^{-1} R_1{}^T W_1 y_1$

$= [R_1{}^T ((1/\sigma_1{}^2)\ P[QX] + (1/\sigma_2{}^2\ P[PE]) R_1]^{-1}$

times $R_1{}^T ((1/\sigma_1{}^2)\ P[QX] + (1/\sigma_2{}^2)\ P[PE]) y_1$

where

(4.4.5) $W_1 = H_1 C_1{}^{-1} H_1{}^T = H_1 (H_1{}^T S_{11} H_1)^{-1} H_1{}^T$,

and the covariance structure for our 2SLS estimator is given by

(4.4.6) $\mathrm{Cov}(a_{1,2SLS})$

$= [R_1{}^T W_1 R_1]^{-1}$

$= [R_1{}^T ((1/\sigma_1{}^2)\ P[QX] + (1/\sigma_2{}^2) P[PE])\ R_1]^{-1}$.

Now we will derive the joint 2SLS estimator; a system estimator with 2SLS applied to each equation separately. Let

(4.4.7) $$H_* = \begin{bmatrix} H_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & H_M \end{bmatrix}$$

Then we write the instrument-residual orthogonality conditions as plim $f_{*1}/NT = 0$, where

(4.4.8) $$f_{*1} = H_1^T(y_* - R_* A_*).$$

Although the covariance structure is

(4.4.9) $$C_* = H_*^T S H_*$$

we used instead the sub-optimal weighting matrix

(4.4.10) $$W_{*1} = H_*(H_*^T \mathrm{blg}(S) H_*)^{-1} H_*^T$$

where

(4.4.11) $$\mathrm{blg}(S) = \begin{bmatrix} S_{11} & & & \\ & \cdot & 0 & \\ & & \cdot & \\ & 0 & & \cdot & \\ & & & & S_{MM} \end{bmatrix}.$$

We minimize the quadratic distance

(4.4.12) $$f_{*1}^T (H_*^T \mathrm{blg}(S) H_*)^{-1} f_{*1}$$

which yields the joint 2SLS estimator

(4.4.13) $$\begin{aligned} a_{*2SLS} &= [R_*^T W_{*1} R_*]^{-1} \; R_1^T W_{*1} y_* \\ &= [R_*^T H_* (H_*^T \mathrm{blg}(S) H_*)^{-1} H_1^T{}_{*1} R_*]^{-1} R_1^T W_{*1} y_* \end{aligned}$$

Because R_*, A_*, and blg(S) are block diagonal, we have

(4.4.14) a_{*2SLS}

$$= \begin{bmatrix} [R_1^T H_1 (H_1 S_{11} H_1)^{-1} H_1^T R_1]^{-1} & & \\ & \ddots & \\ & & [R_M^T H_M (H_M S_{MM} H_M)^{-1} H_M^T R_M]^{-1} \end{bmatrix}$$

$$\text{times} \begin{bmatrix} R_1^T H_1 (H_1^T S_{11} H_1)^{-1} H_1^T y_1 \\ \vdots \\ R_M^T H_M (H_M^T S_{MM} H_M)^{-1} H_M^T y_M \end{bmatrix}$$

which can be seen as 2SLS applied to each equation separately. And

(4.4.15) $Cov(a_{*2SLS})$

$$= [R_*^T W_{*1} R_*]^{-1} R_*^T W_{*1} S W_*^T R_* [R_*^T W_{*1} R_*]^{-1}$$

$$= [R_*^T H_* (H_*^T blg(S) H_*)^{-1} H_*^T R_*]^{-1} R_*^T W_* S$$

$$\text{times } W_*^T R_* [R_*^T H_* (H_*^T blg(S) H_*)^{-1} H_*^T R_*]^{-1}$$

Now consider again the instrument residual orthogonality conditions given in (4.4.8). Using the correct covariance structure, the problem of minimizing the quadratic distance from zero of f_{*1},

(4.4.16) $f_{*1}^T W_{*2} f_{*1}$

where

(4.4.17) $W_{*2} = H_* C_*^{-1} H_*^T$

$$= H_* (H_*^T S H_*)^{-1} H_*^T$$

yields the 3SLS estimator

$$(4.4.18)\quad a_{*3SL} = [R_* W_{*2} R_*]^{-1} R_* W_{*2} y_*$$
$$= [R_* H_* C_*^{-1} H_*^T R_*]^{-1} R_* H_* C_*^{-1} H_*^T y_*$$

with covariance matrix

$$(4.4.19)\quad Cov(a_{*3SLS}) = (R_*^T W_{*2} R_*)^{-1}$$
$$= (R_*^T H_* C_*^{-1} H_*^T R_*)^{-1}.$$

It is a standard result that this estimator is efficient relative to the 2SLS estimator given above. And when $H_* = (I \otimes H)$, it is easy to show that 3SLS given in (4.4.18) simplifies to 3SLS given in section 4.3.

Theorem (4.6): When $H_* = (I \otimes H)$, the 3SLS estimator given in (4.4.18) reduces to the 3SLS estimator given in (4.3.8).

Proof:

Note that when $H_* = (I \otimes H)$ where $H = [QX\ PE]$, the weighting matrix in (4.4.17), using Lemma (4.5), reduces to the matrix

$$(4.4.20)\quad W_{*2} = H_* [H_*^T S H_*] H_*^T$$
$$= (I \otimes H)[I \otimes H^T) S\ (I \otimes H)]^{-1} (I \otimes H^T)$$
$$= (I \otimes H)(\ \Sigma_1^{-1} \otimes (H^T H)^{-1}$$
$$+ \Sigma_2^{-1} \otimes (H^T H)^{-1}\ )(I \otimes H^T)$$
$$= \Sigma_1^{-1} \otimes P[QH] + \Sigma_2^{-1} \otimes P[PE]$$
$$= \Sigma_1^{-1} \otimes P[QX] + \Sigma_2^{-1} \otimes P[PE]$$

Since this reduced to the same weighting matrix used in

(4.3.8), the result follows. **Q.E.D.**

The question we now ask is whether our EC3SLS estimator for the more general model allowing instruments to vary across equations is efficient. We ask whether the instrument-residual orthogonality conditions given in (4.4.8) can be mixed using the cross equation covariances as weights.

Consider a positive definite matrix C (of dimension M x M) and the vector

$$(4.4.21) \quad f_{*3} = H_*^T (C \otimes I_T)(y_* - R_* A_*)$$

$$= \begin{bmatrix} \sum_{i=1}^{M} c_{1i} H_1^T (y_i - A_i R_i) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{M} c_{Mi} H_M^T (y_i - A_i R_i). \end{bmatrix}$$

Thus premultiplying the instruments H_* by a matrix of the form $(C \otimes I)$ would "mix" the equations (unless C was diagonal) and introduce terms like the cross-product $H_1^T(y_i - R_i A_i)$ whose probability limit we implicitly assumed was not zero for at least one j = 1,...., M. (If not then we have the special case when $H_1 = H_2 = \ldots = H_M$.) Thus, f_{*3} does not represent <u>true</u> instrument-residual orthogonal conditions so even consistency of any resulting estimator would be in doubt. Therefore, the orthogonality conditions f_{*3} would not lead to an improved estimator.

The question of whether we can improve the 3SLS estimator derived from f_{*2} must be addressed by searching among estimators derived from transformations of f_{*2} which do not create new and illegitimate cross-products. We pursue this line of reasoning in the remainder of this section.

Now there are two ways to order the instrument-residual orthogonality conditions given by f_{*2}. We can order first by residuals and then by instruments (which has been the method used so far) or we can order first by instruments and then by residuals. We will address the question of transforming the orthogonality conditions ordered in each of the two ways and consider the effect, if any, on the resulting GMM estimator.

First, we need to introduce some notation. Let

$$(4.4.22) \quad H = [\ h_1, \ . \ . \ . \ , h_L \ ]$$

be the set of <u>all</u> instruments; L denotes the total number of instruments. Then define

$$(4.4.23) \quad U_* = \begin{bmatrix} U_1 & & \\ & \ddots & \\ & & U_M \end{bmatrix}$$

to be a selection matrix where $HU_i = H_i$; each matrix U_i (of dimensions L_i x m_i) selects from H the instruments orthogonal to the residual s_i; and m_i equals the number of instruments orthogonal to residual s_i. We can now write

$$(4.4.24) \quad (I_M \otimes H)U_* = \begin{bmatrix} HU_1 & & \\ & \ddots & \\ & & HU_M \end{bmatrix} = \begin{bmatrix} H_1 & & \\ & \ddots & \\ & & H_M \end{bmatrix}$$

$$= H_* .$$

It would follow that the sum $m_1 + . . . + m_M$ is the total number of instrument residual orthogonality conditions found in (4.4.6) and that the matrix U_* is of dimension $(m_1 + . . . + m_M) \times ML$.

We can now write (4.4.6) as

(4.4.25) $f_{*4} = U_*^T vec(H^T s) = U_*^T (I_M \otimes H^T) vec(s)$

$$= \begin{bmatrix} H_1^T & & \\ & \ddots & \\ & & H_M^T \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_M \end{bmatrix}$$

and the covariance matrix of f_{*4} as

(4.4.26) $C_{*4} \equiv Cov(f_{*4}) = E\{ U_*^T vec(H^T s) vec(H^T s)^T U_* \}$

$$= E\{ U_*^T (I_M \otimes H^T) vec(s) vec(s)^T (I_M \otimes H) U_* \}$$

$$= U_*^T (I_M \otimes H^T)(\Sigma_1 \otimes Q + \Sigma_2 \otimes P)(I_M \otimes H) U_*$$

$$= U_*^T (\Sigma_1 \otimes H^T Q H + \Sigma_2 \otimes H^T P H) U_*$$

since $E\{ vec(s) vec(s)^T \} = \Sigma_1 \otimes Q + \Sigma_2 \otimes P$. The quadratic distance to zero of f_{*4} can be written as

(4.4.27) $f_{*4}^T C_{*4}^{-1} f_{*4}$

$$= vec(s)^T (I_M \otimes H) U_* C_{*4} U_*^T (I_M \otimes H^T) vec(s)$$

Now define the matrices T_i ($i = 1,.., M$) where T_i is a positive definite square matrix of order m_i. Then

$$(4.4.28)\quad T_* = \begin{bmatrix} T_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & T_M \end{bmatrix}$$

is a positive definite, square matrix of order equal to the total number of restrictions. We can then transform the orthogonality conditions in (4.4.25) by T_* and write them as

$$(4.4.29)\quad \underline{f}_{*4} = T_*^T U_*^T \text{vec}(H^T s) = T_*^T U_*^T (I_M \otimes H^T)\text{vec}(s)$$

$$= \begin{bmatrix} T_1^T U_1^T H^T & & & \\ & \cdot & & \\ & & \cdot & \\ & & & T_M^T U_M^T H^T \end{bmatrix} \begin{bmatrix} s_1 \\ \cdot \\ \cdot \\ \cdot \\ s_M \end{bmatrix}$$

$$= \begin{bmatrix} T_1^T H_1^T & & & \\ & \cdot & & \\ & & \cdot & \\ & & & T_M^T H_M^T \end{bmatrix} \begin{bmatrix} s_1 \\ \cdot \\ \cdot \\ \cdot \\ s_M \end{bmatrix} .$$

We should note that each block, $T_i^T H_i^T S_i$, is a mix of the cross-products beween the instruments in H_i and the residual s_i. Since every instrument in H_i is orthogonal to residual s_i, $T_i^T H_i^T S_i$ represents a mixing of only legitimate instrument-residual orthogonality conditions. In this mixing cross-products which have nonzero probability limit are not introduced.

Now consider

$$(4.4.30)\quad \underline{C}_{*4}$$

$$\equiv \text{Cov}(\underline{f}_{*4})$$

$= E\{ T_*^T U_*^T vec(H^T s) vec(H^T s)^T U_* T_* \}$

$= E\{ T_*^T U_*^T (I_M \otimes H^T) vec(s) vec(s)^T (I_M \otimes H) U_* T_* \}$

$= T_*^T C_{*4} T_*$

We then write the quadratic distance to zero from $\underline{f}_{*4}$ as

$$(4.4.31) \quad (\underline{f}_{*4})^T (\underline{C}_{*4})^{-1} \underline{f}_{*4}$$

$$= vec(H^T s)^T U_* T_* [T_*^T C_{*4} T_*]^{-1} T_*^T U_*^T vec(H^T s)$$

$$= vec(H^T s)^T U_* T_* T_*^{-1} C_{*4}^{-1} (T_*^T)^{-1} T_*^T U_*^T vec(H^T s)$$

$$= vec(H^T s)^T U_* C_{*4}^{-1} U_*^T vec(H^T s)$$

$$= vec(s)^T (I_M \otimes H) U_* C_{*4} U_*^T (I_M \otimes H^T) vec(s)$$

which is the same as in (4.4.27). Thus, the GMM derived from either f_{*4} or $\underline{f}_{*4}$ would be the same. Therefore, mixing the instrument-residual orthogonality conditons having a common residual will have no effect on the resulting estimating.

We next consider mixing the orthogonality conditions in (4.4.6) within subgroups with a common instrument but first we need to introduce some additional notation. Let

$$(4.4.32) \quad s_+ = (s_1, \ . \ . \ . \ , s_M)$$

be the matrix (of dimension T x M) containing the residuals. Then define

$$(4.4.33) \quad s_{(i)} = s_+ V_i, \quad i = 1, \ . \ . \ . \ , L,$$

as the matrices (of dimension T x l_i) containing only those residuals assumed orthogonal to the instrument h_i. $s_{(i)}$ is the matrix which selects the l_i residuals from the list in (4.4.32) where l_i denotes the number of residuals orthogonal to instrument h_i. Note that there are as many $s_{(i)}$'s as there are instruments.

The matrix containing all the selection matrices can be written as

$$(4.4.34) \quad V_* = \begin{bmatrix} V_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \quad V_L \end{bmatrix}.$$

We can now rearrange the instrument-residual orthogonality conditions found in (4.4.25) first by instrument and then by residuals. The orthogonal conditions reordered in such a manner can be written as

$$(4.4.35) \quad f_{*5} = V_*^T vec(s^T H)$$
$$= V_*^T (I_L \otimes s^T) vec(H).$$

It should be pointed out that this rearrangement has in no way effected the orthogonal conditions; the same instrument-residual orthogonal conditions contained in (4.4.25) are still found in (4.4.35) but now in a different order.

The covariance structure of f_{*5} is written

$$(4.4.36) \quad C_{*5} \equiv Cov(f_{*5}) = E\{ V_*^T vec(s^T H) vec(s^T H)^T V_* \}$$
$$= E\{ V_*^T (H^T \otimes I_M) vec(s^T) vec(s^T)^T (H \otimes I_M) V_* \}$$
$$= V_*^T (H^T \otimes I_M)(Q \otimes \Sigma_1 + P \otimes \Sigma_2)(H \otimes I_M) V_*$$
$$= V_*^T (H^T QH \otimes \Sigma_1 + H^T PH \otimes \Sigma_2) V_*$$

since $E\{ vec(s) vec(s)^T \} = Q \otimes \Sigma_1 + P \otimes \Sigma_2$. So the quadratic distance from f_{*5} to zero can be written as

$$(4.4.37)\quad (f_{*5})^T(C_{*5})^{-1}f_{*5}$$

$$= \mathrm{vec}(s^T H)^T V_* C_{*5}^{-1} V_*^T \mathrm{vec}(s^T H).$$

Now define the matrices T_i ($i = 1, . . ., L$) where T_i is positive definite, square matrix of order l_i and so

$$(4.4.38)\quad T_* = \begin{bmatrix} T_1 & & & \\ & \ddots & & \\ & & & T_L \end{bmatrix}$$

is a positive definite, square matrix of order equal to the total number of restrictions. We can then transform the orthogonality conditions in (4.4.35) by T_* and then write them as

$$(4.4.39)\quad \underline{f}_{*5} = V_*^T \mathrm{vec}(s^T H)$$

$$= \begin{bmatrix} T_1^T V_1^T s^T h_1 \\ \vdots \\ T_L^T V_L^T s^T h_L \end{bmatrix} .$$

We note that each block, $T_i^T V_i^T s^T h_i$, is a mixing of the instrument-residual orthogonality conditions but for only a single instrument h_i. That is, we are mixing the cross-products of $s_{(i)}$ and h_i which are all assumed to have a probability limit equal to zero. Thus, we have not introduced illegitimate instrument-residual cross-products whose probability limit may be nonzero.

Now consider

(4.4.40) $\underline{C}_{*5}$

$$\equiv Cov(\underline{f}_{*5})$$

$$= E\{ T_*^T V_*^T vec(s^T H) vec(s^T H)^T V_* T_* \}$$

$$= E\{ T_*^T V_*^T (H^T \otimes I_M) vec(s^T) vec(s^T)^T (H \otimes I_M) V_* T_* \}$$

$$= T_*^T C_{*5} T_*$$

We then write the quadratic distance to zero from $\underline{f}_{*5}$ as

(4.4.41) $(\underline{f}_{*5})^T (\underline{C}_{*5})^{-1} \underline{f}_{*5}$

$$= vec(s^T H)^T V_* T_* [T_*^T C_{*5} T_*]^{-1} T_*^T V_*^T vec(s^T H)$$

$$= vec(s^T H)^T V_* T_* T_*^{-1} C_{*5}^{-1} (T_*^T)^{-1} T_*^T V_*^T vec(s^T H)$$

$$= vec(s^T H)^T V_* C_{*5}^{-1} V_*^T vec(s^T H).$$

By comparing the above quadratic distance to that given in (4.4.37), we find that as long as the T_i's are nonsingular so T_*^{-1} exists, transforming the orthogonality conditions in (4.4.37) by T_* will have no effect on the resulting GMM estimator.

In summary, when transforming the instrument-residual orthogonality conditions when the instruments are different for each residual, we must restrict ourselves to transformations which do not create new and illegitimate cross-products. Unfortunately, when we do so it turns out that there is no gain for doing such transformation - we just get back the 3SLS estimator.

4.5 Conclusions

This chapter applies the Hausman and Taylor method of instrumental variables estimation to the simultaneous equations panel data model, and derives the subsequent estimator. Throughout, we attempt to improve our instrumental-variables estimator by transforming the error so to change its error covariance or by transforming the instruments so to improve their explanatory abilitiy of the endogenous variables. We consider a natural extension of the Hausman-Taylor model to a linear simultaneous equation model with random effects by allowing the effects to be potentially correlated with some of the regressors. We then consider the affect on our instrumental-variables estimator when the instrument sets are not the same for each equation in the system.

CHAPTER 5

Conclusion

In this thesis, I have considered the specification and estimation of linear models in the presence of panel data. The previous literature on this topic can be organized according to the following four distinctions: first, the nature of the model, such as single equation versus simulataneous equation model: second, whether there are assumed to be individual and time effects, or just one or the other; third, whether the effects are assumed to be fixed or random, and, if they are random, whether they are assumed to be correlated with some or all of the explanatory variables; and fourth, whether asymptotic properties of the estimators depend on a large number of individuals (large N), a large number of time periods (large T), or both. Existing papers cover some but not all of the possible combinations of these assumptions, and the basic purpose of this thesis is to fill in some of the more obvious gaps in the literature by considering plausible and important combinations of assumptions not previously considered. However, another purpose of the thesis is to advance a particular mathematical framework for the analysis and to demonstrate its usefulness.

There are three substantive contributions of the thesis. The first is to extend the analysis of Hausman and Taylor (1981) to a model containing individual and time effects correlated with some or all of the regressors, under the assumption of large N and small T. I consider random individual and time effects, and allow the regressors to be correlated or not with either or both types of effects. The analysis is similiar to that of Hausman and Taylor, but it is algebraically more complicated because there are more different types of exogeneity assumptions to consider. It should also be noted that all previous treatments of models with both individual and time effects assume large N and large T. I consider this case in detail, but I also consider separately the case of large N and small T (as assumed by Hausman and Taylor).

The second contribution of the thesis is to extend the analysis of Hausman and Taylor to a single equation in a simultaneous equations system; that is, to a regression model in which some of the regressors are correlated with the random noise component of the error. This case has previously been analyzed by Amemiya and MaCurdy (1987), but in an unsatisfactory way. I follow Hausman and Taylor and Amemiya and MaCurdy in considering random individual effects (no time effects) which may be correlated with some or all of the exogenous regressors, and in assuming large N and small T. I propose 2SLS estimators based on instrument sets proposed by Hausman and Taylor, Amemiya and MaCurdy, and

Breusch, Mizon, and Schmidt (1987).

The third contribution of the thesis is to propose full-information (3SLS) estimators for a simultaneous equations system with random individual effects correlated with some or all of the exogenous variables. These estimators are shown to reduce to the usual fixed-effects treatment if all exogenous variables are correlated with the effects, and to reduce to an estimator previously proposed by Baltagi (1981) if none of the exogenous variables are correlated with the effects. I also consider the case in which some exogenous variables may be correlated with the effects in some equations but not in others, so that the available instrument set varies from equation to equation.

The line of research followed in this dissertation can be extended in a straightforward fashion by considering additional new combinations of the assumptions underlying previous work. One obvious and interesting task would be to analyze a simultaneous equations model when there are both individual and time effects that may be correlated with the exogenous variables. A second possible topic of future research is to consider single equation models in which the random noise component of the error has a non-scalar covariance matrix. Finally, although this direction of research is less clearly defined, I hope to extend the analyses of this dissertation to nonlinear models.

BIBLIOGRAPHY

BIBLIOGRAPHY

Arnold, S. F. The Theory of Linear Models and Multivariate Analysis. New York: Wiley, 1981

Amemiya, T., and T. E. MaCurdy. "Instrumental-Variable Estimation of an Error-Components Model." Econometrica 45 (1986): 869-881.

Baltagi, B. H. "On Seemingly Unrelated Regressions with Error Components." Econometrica 48 (1981): 1547-1551.

Baltagi, B. H. "Simultaneous Equations with Error Components." Journal of Econometrics 17 (1981): 189-200.

Breusch, T. S., G. E. Mizon and P. Schmidt. "Some Results on Panel Data." unpublished manuscript.

Cornwell, C. and P. Schmidt. "Models for which the MLE and the Conditional MLE Coincide." unpublished manuscript.

Fuller, W. and G. Battese. "Estimation of Linear Models with Crossed-Error Structure." Journal of Econometrics 2 (1974): 67-78.

Halmos, P. R. Finite-Dimensional Vector Spaces. Princeton: Van Nostrand, 1958.

Hausman, J. A., W. K. Newey and W.E. Taylor. "Efficient Estimation and Identification of Simultaneous Equation Models with Covariance Restrictions." Econometrica 55 (1987): 849-874.

Hausman, J. A. and W. E. Taylor. "Panel Data and Unobsevable Individual Effects." Econometrica 49 (1981): 1377-1398.

Judge, George G., et al. The Theory and Practice of Econometrics. New York: Wiley, 1980.

Maddala, G. S. "The Use of Variance Components Models in Pooling Cross Section and Time Series Data." Econometrica 39 (1971): 341-358.

Mundlak, Y. "On the Pooling of Time Series and Cross Section Data." Econometrica 46 (1978): 69-85.

Mundlak, Y. "Models with Variable Coefficients-Integration and Extension." Annales de l'Insee 30/31 (1978): 483-510.

Nerlove, M. "A Note on Error Components Model." Econometrica 39 (1971): 383-396.

Rao, C. R. Linear Statistical Inference and Its Applications. New York: Wiley, 1973.

Schmidt, P. Econometrics. New York: Marcel Dekker, 1976.

Schmidt, P. "A Note on a Fixed Effects Model with Arbitrary Interpersonal Covariance." Journal of Econometrics 22 (1983): 391-393.

Swamy, P. and S. Arora. "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models." Econometrica 40 (1972): 261-275.

White, H. Asymtotic Theory for Econometricians. New York: Academic Press, 1984.