ESTIMATION AND TESTING OF DYNAMIC ECONOMETRIC MODELS FROM PANEL DATA

by

Manuel Arellano Gonzalez

1985

Thesis submitted for the degree of

Doctor of Philosophy

at the

London School of Economics

University of London

This thesis is dedicated to my parents.

ABSTRACT

This research develops methods of estimation and test statistics of dynamic single equation models from panel data when the errors are serially correlated. It is assumed that the number of time periods is fixed while the number of cross-section observations is large. This makes it possible to consider prediction equations of the initial observations based on the exogenous variables corresponding to all periods available in the sample, as well as to leave unrestricted the covariances of the prediction errors with the remaining errors in the model.

The concentrated likelihood function is derived both for cases where the prediction error is left unrestricted and where it is assumed to have the marginal distribution of the stationary process. The performance of maximum likelihood methods is investigated, either for correct models or under several misspecifications, by resorting to Monte Carlo methods using antithetic variates.

Dynamic models from panel data can be seen as a specialisation of a triangular system with covariance restrictions. In this context, the asymptotic distribution of the estimators that maximise the gaussian likelihood function is derived when normality holds and also when the errors are non-normal. In particular, it is shown that in the latter case the estimator that takes into account the covariance restrictions is not generally more efficient than the estimator that leaves the covariance matrix unrestricted.

The possibility of obtaining consistent estimates of the unrestricted intertemporal covariance matrix is used to develop test statistics of covariance restrictions arising from various random effects specifications. A Wald test and a minimum chi-square test, which are robust to the non-normality of the errors, and appropriate asymptotic probability limits for the quasi-likelihood ratio test are proposed. Monte Carlo experiments are conducted to study the performance of these test criteria. In order to illustrate these procedures, QML estimates of dynamic earnings functions from the Michigan Panel are obtained.

Joint minimum distance estimators of slope and covariance parameters are defined that are generally efficient relative to QML estimators when normality is not imposed and the covariance matrix is restricted. Finally, it is shown that there exist separate minimum distance estimators of the covariance parameters and generalised least squares estimators of the slope parameters that are efficient. A simulation is also carried out to examine the performance of these methods.

ACKNOWLEDGEMENTS

I wish to thank my supervisor Professor J.D. Sargan for his constant help, advice and encouragement. His many comments and suggestions have been a constant source of inspiration for which I am particularly grateful. But above all I must express my gratitude to Professor Sargan for all I have learnt from him throughout the whole period of my research which has profoundly influenced my understanding of econometrics.

Special thanks are due to Olympia Bover for her help, encouragement and patience during the three years in which I worked on this thesis. I have benefited from innumerable discussions with Olympia and her careful reading of the manuscript has also led to numerous improvements for which I am most grateful.

I am indebted to various other people for assistance in the preparation of this thesis. I thank A. Bhargava for the use of his maximum likelihood computer program at the early stages of my work; G. Chowdhury for making her data base available; and F. Srba for allowing me to go through the program NAIVE and for his readiness to help. I am especially grateful to J. Magnus for his help and useful comments. I would also like to thank S. Peters and S. Marwaha for their help in my computing work.

Financial support from the London School of Economics in the form of a Suntory-Toyota Scholarship, the Caja de Ahorros de Barcelona and the ESRC DEMEIC Econometric Project at L.S.E. is gratefully acknowledged.

Finally, I am very happy to thank Christine Wills for typing the bulk of the manuscript and Marianna Tappas for typing Chapter 4 in such an efficient way.

CONTENTS

ABSTRACT 3

ACKNOWLEDGEMENTS 5

CHAPTER 1 DYNAMIC ECONOMETRIC MODELS FROM PANEL DATA

1.1 Introduction 9
1.2 The Model of Interest 10
1.3 The Problem of the Initial Observations 12
1.4 Estimating Models with Unrestricted Prediction Equation for y_{i0} 19
1.5 Serial Correlation and Unrestricted Intertemporal Covariance 24
1.6 Correlation between the Explanatory Variables and the Individual Effects 28
Notes 36

CHAPTER 2 QML ESTIMATION OF DYNAMIC MODELS WITH SERIALLY CORRELATED ERRORS

2.1 Introduction 38
2.2 Three Alternative Models with ARMA Errors 39
2.3 Quasi-Maximum Likelihood Estimation 43
2.4 Estimation with Arbitrary Intertemporal Covariance 52
2.5 Experimental Evidence 55
2.6 A Model with Arbitrary Heteroscedasticity Over Time 61
Notes 70
Appendix 2.A The Determinant and the Inverse of the Covariance Matrix for the Stationary ARMA(1,1) Process 71

CHAPTER 3 THE ASYMPTOTIC PROPERTIES OF QML ESTIMATORS IN A TRIANGULAR MODEL WITH COVARIANCE RESTRICTIONS

3.1 Introduction 74
3.2 The Model and the Limiting Distribution of the QML Estimator 75
3.3 The Asymptotic Variance Matrix of the QML Estimators when the Errors are Normal 83

3.4 The Asymptotic Variance Matrix of the QML Estimators when the Errors are Possibly Non-normal 88
3.5 A Simple Two Equation Model 94
Notes 101
Appendix 3.A The First and Second Derivatives of the Quasi-log-likelihood Function with respect to θ and τ, and the Matrices Φ_u, Φ_r, Θ_u, Θ_r and Φ_u^{-1} 102
Appendix 3.B Derivation of $W_{r,11}$ and $W_{r,22}$ 110

CHAPTER 4 WALD AND QUASI-LIKELIHOOD RATIO TESTS OF RANDOM EFFECTS SPECIFICATIONS IN DYNAMIC MODELS
4.1 Introduction 114
4.2 The AVM of Ω-Unrestricted QML Estimators for Panel Data 115
4.3 Wald and Quasi-Likelihood Ratio Tests 119
4.4 Simulation Results 123
4.5 Estimation of Earnings Functions for the US 125
Notes 139

CHAPTER 5 MINIMUM DISTANCE AND GLS ESTIMATION OF TRIANGULAR MODELS WITH COVARIANCE RESTRICTIONS
5.1 Introduction 141
5.2 Joint Minimum Distance Estimation of Slope and Covariance Parameters 143
5.3 The Asymptotic Variance Matrix of the Optimal Joint MDE 149
5.4 Efficient MDE of Covariance Parameters and Minimum Chi-Square Specification Tests of Covariance Constraints 152
5.5 Generalised Least Squares Estimation of Regression Coefficients 155
5.6 Subsystem Estimation 161
Notes 166
Appendix 5.A The Limiting Distribution of the Minimum Distance Estimator 169

CHAPTER 6 EFFICIENT MD AND GLS ESTIMATORS APPLIED TO PANEL DATA

6.1 Introduction 171

6.2 The Estimators 172

6.3 Numerical Results 178

CONCLUSION 187

REFERENCES 189

CHAPTER 1

DYNAMIC ECONOMETRIC MODELS FROM PANEL DATA

1.1 Introduction

The dynamic error components model has been a major subject of attention for econometricians ever since economists began to make use of panel data to estimate economic relationships. A reason why this interest has so scarcely materialised in applied work is that, despite recent relevant contributions, a complete answer to the problem of estimating and testing dynamic models from panel data still does not exist. It is the purpose of this research to present a further contribution to this end.

The fact that typically a panel involves a large number of individuals, but only over a short number of time periods, makes it necessary to rely only upon the increase in the number of individual units in developing the asymptotic properties of the statistical methods under consideration. Treating the number of time periods as fixed creates different problems to those encountered in time series analysis, particularly a careful specification of initial conditions is required, but it is also the basis of new and fruitful ways of approaching dynamic modelling.

In what follows we introduce the dynamic error components model, and we will survey the relevant literature as we discuss the implications of different assumptions concerning the specification of the basic equation.

1.2 The Model of Interest

The typical dynamic error components model assumes that the endogenous variable y_{it} satisfies

(1.2.1) $$y_{it} = \alpha\, y_{i(t-1)} + \beta'\, x_{it} + \gamma'\, z_i + u_{it} ,$$

(1.2.2) $$u_{it} = \eta_i + v_{it} ,$$

where x_{it} is a n x 1 vector of time-varying observable variables, z_i is a m x 1 vector of time-invariant observable variables (if required we may have $z_{1i} = 1$ for all i). α is an unknown scalar coefficient, and β and γ are n x 1 and m x 1 vectors of unknown coefficients, respectively. It is assumed that α, β and γ remain constant over all time periods and individual units. Of course, there may be lagged values of x_{it} and additional lags of y_{it}, but this simple formulation does not miss any essential feature of the problem and thus most of the discussion will be conducted on the basis of this model. Equally, although we assume that there is no a priori information on $\delta' = (\alpha\ \beta'\ \gamma')$ no new essential complications would arise if δ is subject to restrictions. η_i and v_{it} are unobservable random variables identically and independently distributed across individuals. It is also assumed that η_i and v_{it} have zero mean and are uncorrelated to each other. Thus, $E(\eta_i) = E(v_{it}) = 0$, $E(\eta_i\, v_{jt}) = 0$, $E(\eta_i^2) = \sigma_\eta^2$ and $E(v_{it}^2) = \sigma_v^2$ for all i, j and t. η_i is meant to capture individual specific shocks and other unobservable factors that influence y_{it} and remain the same over time. Equally, v_{it} would capture omitted time-varying effects of various kinds that we assume can be well represented by a random error with the same properties for different individuals, though the effects embodied in v_{it} can induce serial correlation. We could also assume a time specific component in u_{it}, but as we shall consider inference for a fixed number of time observations, it will not be a problem to condition on the time specific effects that are in the sample if desired (by treating them as a further set of coefficients to be estimated), and therefore we omit them for simplicity in this general discussion.

There remains the question of what properties to attribute to the observable variables x_{it} and z_i. The simplest possibility is to assume that x_{it} and z_i are stochastic variables independent of u_{it}. In this case we would be conducting inference conditional on the values of x_{it} and z_i that are in the sample, and so there is no difference if we regard these sample values as being fixed. Moreover, this provides a more natural framework since we shall encounter many cases where some exogenous variables cannot be considered as random. Indeed, this is the assumption that we shall make throughout the remaining chapters of this work. However, if we think of η_i as a latent variable representing relevant but unobserved characteristics, it would be reasonable to assume that some or all the observed explanatory variables are correlated with η_i. This situation has been extensively studied for static models in the literature (cf. Mundlak (1978) and Hausman and Taylor (1981), among others). In fact, some authors would point to the ability of controlling unobserved individual heterogeneity as one of the main purposes in using panel data. This point will be discussed further in Section 1.6.

We assume that our sample consists of N individual units $(i=1,...,N)$ observed through $(T+1)$ consecutive time periods $(t=0,1,...,T)$. Nevertheless, there is no reason to believe that the process of the dependent variable started at the same time at which we started sampling, and even in this case it would be unreasonable to assume that the individual effects η_i did not play a role in determining y_{i0}.

If $|\alpha|<1$ and the process of y_{it} started in the distant past, our model implies the following equation for y_{i0}

$$(1.2.3) \qquad y_{i0} = \sum_{k=0}^{\infty} \alpha^k \beta' x_{i(-k)} + \gamma^{*\prime} z_i + \eta_i^* + v_{i0}^*$$

with $\gamma^* = (1-\alpha)^{-1} \gamma$, $\eta_i^* = (1-\alpha)^{-1} \eta_i$ and $v_{i0}^* = \sum_{k=0}^{\infty} \alpha^k v_{i(-k)}$.
It is the presence of time-varying exogenous variables in our original equation what complicates matters, as it makes y_{i0} to depend on the entire past history of such variables. In this sense, even if we know the distribution of u_{it}, further assumptions about the initial observations are required to be able to define maximum likelihood estimators of $(\alpha\ \beta\ \gamma)$. In the next Section we shall discuss different solutions that have been proposed in the literature to circumvent this problem and their implications for panel data.

1.3 The Problem of the Initial Observations

If we complete the model (1.2.1) by assuming that the values of y_{i0} are fixed for all i, x_{it} and z_i are nonstochastic and there is no serial correlation among the v_{it}, we have the model of Balestra and

Nerlove (1966). Further assuming that the u_{it} are normally distributed, Balestra and Nerlove defined the maximum-likelihood estimator and also a generalised least squares estimator (the 'two round' estimator). This model and related estimation methods have been the object of detailed analysis in a series of Monte Carlo studies by Nerlove and Maddala (cf. Nerlove (1967), Nerlove (1971) and Maddala (1971)). The difficulty here is that these methods of estimation will only be consistent (as $N \to \infty$) if the y_{io} are truly fixed or stochastic but independent of η_i; otherwise they will fail to control for the lack of orthogonality between y_{io} and u_{i1} in the equation for y_{i1}. However, we have seen that in the 'model of interest' there is no connection between the starting time of the process and the sampling starting time, and in any event it would be unrealistic to assume independence between y_{io} and η_i.

The assumption of fixed initial observations is a common one in time series models and its implications are rather different when T tends to infinity. Although in our context T is fixed, it is worth stressing that the source of the difficulty is the correlation between y_{io} and u_{i1}; in other words, if $\eta_i = 0$ for all i and the v_{it} are white noise errors, we could safely take y_{io} as fixed and still being able to estimate consistently the model of interest as $N \to \infty$ for constant T.

These problems have been pointed out by Pudney (1979) and Anderson and Hsiao (1981 and 1982). Furthermore, Anderson and Hsiao (1982) discuss various likelihood functions that arise from a variety of assumptions about the initial conditions of the process. They show how the properties of estimators vary from one sampling plan to another and also depend on the way in which the sample becomes large ($N \to \infty$ or $T \to \infty$). However, it would be useful to discuss the relative merits of different assumptions about the initial observations as approximations to what we call the model of interest.

The solution adopted in this work is to complete the system in Section 1.2 with an unrestricted prediction equation of the form

$$(1.3.1) \qquad y_{i0} = \mu_0' x_{i0} + \ldots + \mu_T' x_{iT} + \xi' z_i + u_{i0}^+ \qquad (i=1,\ldots,N).$$

This alternative has been advocated by Bhargava and Sargan (1983) and Chamberlain (1984), although rationalising equation (1.3.1) in different frameworks. On the one hand, Chamberlain assumes $(y_i' \; x_i' \; z_i')$ where $y_i' = (y_{i0},\ldots,y_{iT})$, $x_i' = (x_{i0}',\ldots,x_{iT}')$, to be independent and identically distributed according to some common multivariate distribution with finite moments up to the fourth order. Under this sole assumption there is no reason a priori for the regression function $E(y_{i0}|x_i, z_i)$ to be linear, but a minimum mean-square error linear predictor can always be specified

$$E^*(y_{i0}|x_i, z_i) = \mu_0' x_{i0} + \ldots + \mu_T' x_{iT} + \xi' z_i = y_{i0}^*, \text{ say.}$$

Furthermore, if $E(y_{i0}|x_i, z_i) \neq E^*(y_{i0}|x_i, z_i)$, u_{i0}^+ will be heteroscedastic, since it will contain $\{E(y_{i0}|x_i, z_i) - E^*(y_{i0}|x_i, z_i)\}$. On the other hand, Bhargava and Sargan assume u_{it} to be normally distributed, and x_{it} and z_i to be independent of u_{it}. Then if we let $\bar{y}_{i0}$ to be systematic part of (1.2.3)

$$\bar{y}_{i0} = \sum_{k=0}^{\infty} \alpha^k \beta' x_{i(-k)} + \gamma^{*\prime} z_i ,$$

they assume y^*_{i0} to be the optimal predictor of $\bar{y}_{i0}$ conditional upon x_i and z_i, where $\varepsilon_i = \bar{y}_{i0} - y^*_{i0}$ is also normally and identically distributed for all i with variance σ^2_ε. These assumptions allow Bhargava and Sargan not only to ensure the homoscedasticity of u^+_{i0} but also to characterise the form of its variance and the covariances with $u_{i1},\ldots,u_{iT}$; that is, since

(1.3.2) $$u^+_{i0} = \varepsilon_i + \eta^*_i + v^*_{i0} \quad ,$$

if $|\alpha|<1$ and v_{it} is stationary we then have

(1.3.3) $$E(u^{+2}_{i0}) = \sigma^2_\varepsilon + \frac{\sigma^2_\eta}{(1-\alpha)^2} + E(v^{*2}_{i0}) \quad ,$$

(1.3.4) $$E(u^+_{i0}\ u_{it}) = \frac{\sigma^2_\eta}{1-\alpha} + E(v^*_{i0}\ v_{it}) \qquad (t=1,\ldots,T).$$

Furthermore, if v_{it} is white noise (the case considered by Bhargava and Sargan), $E(v^{*2}_{i0}) = \sigma^2_v/(1-\alpha^2)$ and $E(v^*_{i0}\ v_{it}) = 0$ for $t=1,\ldots,T$.

In any event, notice that our model, once completed with the prediction equation for y_{i0} above, can be written as

(1.3.5) $$B(\alpha)\ y_i + C_x(\mu,\beta)\ x_i + C_z(\xi,\gamma)\ z_i = u_i \quad (i=1,\ldots,N)$$

where $B(\alpha)$ is the (T+1) square matrix

$$B(\alpha) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\alpha & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \end{bmatrix},$$

$C_x(\mu,\beta)$ is the (T+1) x (T+1)n matrix

$$C_x(\mu,\beta) = -\begin{bmatrix} \mu_0' & \mu_1' & \cdots & \mu_T' \\ 0 & \beta' & & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \beta' \end{bmatrix},$$

$C_z(\xi,\gamma)$ is the (T+1) x m matrix

$$C_z(\xi,\gamma) = -\begin{bmatrix} \xi' \\ \gamma' \\ \vdots \\ \gamma' \end{bmatrix},$$

and $u_i' = (u_{i0}^{+}\ u_{i1}\ \ldots\ u_{iT})$. The variance matrix of u_i is given by

$$\text{(1.3.6)} \quad E(u_i u_i' | x_i, z_i) = \left[\begin{array}{c:c} E(u_{i0}^{+2}|x_i,z_i) & E(u_{i0}^{+}u_{i1}|x_i,z_i) \ldots E(u_{i0}^{+}u_{iT}|x_i,z_i) \\ \vdots & \\ & \sigma_\eta^2\, \iota\iota' + V \end{array} \right]$$

where V is the variance matrix of $(v_{i1} \ldots v_{iT})$ and ι is a T x 1 vector of ones.

Thus our model can be expressed as a simultaneous triangular system of T+1 equations with linear restrictions linking the coefficients of the last T equations. Indeed, since the parameters of interest (α, β, γ) are confined to this subset of equations (the 'structural block') and the prediction equation for y_{i0} is unrestricted, explicit estimation of the coefficients of the latter will not usually be required. We postpone the discussion about the implications of the covariance matrix structure to the next Section.

Nerlove (1971) and Pudney (1979) have proposed to use lagged values of the exogenous variables as instruments for the lagged dependent variable in order to ensure consistent estimation of the model of interest. It has also been pointed out that, following this approach, the best choice of instruments is not obvious, since the number of time observations available depends upon the number of instruments chosen. It is therefore of some interest to highlight the instrumental variables implications of the unrestricted prediction equation for y_{i0} discussed above.

The simplest consistent estimator of a set of equations in a simultaneous system is well known to be the IV estimator that ignores the fact that the covariance matrix is not a scalar matrix (what in the absence of cross-equation restrictions is simply equivalent to 2SLS estimation of separate equations). This is Sargan's Crude Instrumental Variable estimator (CIV), and we can define the CIV estimator of $\delta' = (\alpha\ \beta'\ \gamma')$ for the block of the last T equations in the system (1.3.5) by choosing $\hat{\delta}_{CIV}$ to

$$(1.3.7) \qquad \min_{\delta}\ [\mathrm{vec}(U)]'\ (Z^*(Z^{*\prime}\ Z^*)^{-1}\ Z^{*\prime} \otimes I_T)\ \mathrm{vec}(U)$$

where $U' = (u_1 \; u_2 \; \ldots \; u_N)$, $z_i^{*\prime} = (x_i', z_i')$ and $Z^{*\prime} = [z_1^* \; \ldots \; z_N^*]$.[1] i.e. Z^* is an N-rowed matrix of observations on the exogenous variables with $(T+1)n+m$ columns.

In order to relate $\hat{\delta}_{CIV}$ to other estimators suggested in the literature, let us introduce the following NT-vectors

$$y' = [y_{11} \; \ldots \; y_{1T} \; \ldots \; y_{N1} \; \ldots \; y_{NT}] \; ,$$

$$y_{-1}' = [y_{10} \; \ldots \; y_{1(T-1)} \; \ldots \; y_{N0} \; \ldots \; y_{N(T-1)}] \; ,$$

$$u' = [u_{11} \; \ldots \; u_{1T} \; \ldots \; u_{N1} \; \ldots \; u_{NT}] \; ;$$

we also define the NT x n matrix of time-varying exogenous variables

$$X' = [x_{11} \; \ldots \; x_{1T} \; \ldots \; x_{N1} \; \ldots \; x_{NT}]$$

and the NT x m matrix of time-invariant exogenous variables

$$Z' = [z_1 \; \ldots \; z_1 \; \ldots \; z_N \; \ldots \; z_N] \; .$$

Furthermore, introducing the NT x (n+m+1) matrix $X^+ = [y_{-1} \vdots X \vdots Z]$, we can write our structural block of T equations in the usual regression form

(1.3.8) $\quad y = X^+ \delta + u \; .$

Now, a simple explicit expression of $\hat{\delta}_{CIV}$ can be obtained by noting that vec(U) = u. Then straightforward minimisation of

$$(y - X^{+} \delta)'(Z^*(Z^{*\prime}Z^*)^{-1} Z^{*\prime} \otimes I_T)(y - X^{+} \delta)$$

gives

$$(1.3.9) \qquad \hat{\delta}_{CIV} = [X^{+\prime}(Z^*(Z^{*\prime}Z^*)^{-1} Z^{*\prime} \otimes I_T) X^{+}]^{-1} X^{+\prime}(Z^*(Z^{*\prime}Z^*)^{-1} Z^{*\prime} \otimes I_T)y.$$

Clearly, the matrix of instruments in regression form is the NT x T[(T+1)n+m] matrix $(Z^* \otimes I_T)$. This is an optimal set of instruments in the absence of any prior knowledge about the way in which the process of y_{it} started off when T is fixed. Obviously, this requires Z* to be of full column rank for identification and therefore if a subset of variables in the vector x_{it} do not vary across individuals, they must not be used as instruments in the definition of Z*, although they will be included in X^{+} (cf. Bhargava and Sargan (1982) for a discussion on macro-variables).

1.4 Estimating Models with Unrestricted Prediction Equation for y_{i0}

In this Section we survey the methods that have so far been proposed for estimating models with unrestricted prediction equations for the initial observations.

Under Chamberlain's assumptions $E(u_{i0}^{+2}|x_i,z_i)$ and $E(u_{i0}^{+} u_{it}|x_i,z_i)$ (t=1,...,T) will be arbitrary functions of x_i and z_i and therefore, in general, heteroscedastic. Writing model (1.3.5) in reduced form we have

$$(1.4.1) \qquad y_i = (P_x \vdots P_z) \begin{pmatrix} x_i \\ z_i \end{pmatrix} + \nu_i$$

where $P = (P_x \vdots P_z) = - B^{-1}(\alpha)(C_x \vdots C_z)$ and $\nu_i = B^{-1}(\alpha)\, u_i$. The constraints in B, C_x and C_z imply a set of non-linear restrictions on P. Let $\hat{P}$ be the unrestricted least squares estimator of P

$$\hat{P} = \sum_{i=1}^{N} y_i \, z_i^{*\prime} \left(\sum_{i=1}^{N} z_i^* \, z_i^{*\prime} \right)^{-1} .$$

Allowing $E(\nu_i \, \nu_i' | x_i, z_i)$ to be an arbitrary function of x_i and z_i, Chamberlain (1982 and 1984) follows White (1980)'s approach to show that

$$\sqrt{N} \, vec(\hat{P}-P) \xrightarrow{D} N(0, W)$$

where

$$W = E[\nu_i \, \nu_i' \otimes M^{-1} \, (z_i^* \, z_i^{*\prime}) M^{-1}]$$

and

$$M = E(z_i^* z_i^{*\prime}) \ .$$

A consistent estimator of W can be obtained from the corresponding sample moments

$$\hat{W} = \frac{1}{N} \sum_{i=1}^{N} [\hat{\nu}_i \hat{\nu}_i' \otimes \hat{M}^{-1} (z_i^* z_i^{*\prime}) \hat{M}^{-1}]$$

with

$$\hat{M} = \frac{1}{N} \sum_{i=1}^{N} z_i^* z_i^{*\prime} .$$

Then Chamberlain proposes to use a minimum distance procedure to impose the nonlinear restrictions on P. That is, we minimise the following criterion function with respect to the free parameters in P[2]

$$[\mathrm{vec}(\hat{P}-P)]' \hat{W}^{-1} \mathrm{vec}(\hat{P}-P) .$$

Alternatively, Chamberlain suggests to use the structural form and apply a 'generalised three-stage least squares' estimator, which achieves the same limiting distribution as the minimum distance estimator sketched above; the advantage of the latter is that as the restrictions in the structural form are linear there is no need to use numerical methods of optimisation.

Under Bhargava and Sargan's assumptions, the variance matrix of the error, Ω^* say, is fully specified and remains the same across individual units. Relying upon the assumption of normality they specify the log-likelihood function for the complete system

$$L = k - \tfrac{1}{2} N \log \det(\Omega^*) - \tfrac{1}{2} \sum_{i=1}^{N} u_i' \Omega^{*-1} u_i .$$

Bhargava and Sargan discuss two types of estimators. First, assuming Ω^* to be an arbitrary $(T+1)\times(T+1)$ symmetric matrix, they consider the likelihood function concentrated with respect to Ω^* and the coefficients in the prediction equation ξ and μ as an straightforward application of the LIML method from the classical simultaneous equations theory. Second, concentrating only ξ and μ out of the likelihood function they enforce on Ω^* the error components restrictions given by (1.3.3), (1.3.4) and (1.3.6). This likelihood function has to be maximised with the restriction that $|\alpha|<1$.

As it stands, the comparison of the two approaches suggests a trade-off between robustness and efficiency. If the errors are truly normally distributed we may expect maximum likelihood estimators to make an optimal use of the constraints in the covariance matrix, thus leading to efficient estimators of all parameters in the model. However, since in many practical situations there are no particular reasons to assume normality (and frequently sample measures of skewness and kurtosis will contradict this assumption) it is of interest to investigate the properties of the estimators obtained by maximising the gaussian likelihood function when the assumption of normality is false.

The early work from the Cowles Commission (cf. Koopmans, Rubin and Leipnik (1950)) demonstrated that maximum likelihood estimators of the simultaneous equations model with unrestricted covariance matrix maintained the same asymptotic distribution even when the errors are non-normal, and called them quasi-maximum likelihood (QML) estimators. However, as the discussion in Chapter 3 will make it clear, this is not the case in the presence of a priori knowledge about the covariance matrix. In this situation, the QML estimator does not make optimal use of the prior information and its asymptotic distribution depends on higher order moments of the errors. Moreover, it will be shown that it can be the case that the covariance restricted-QML estimator of the slope coefficients is less efficient than the QML estimator that leaves the covariance matrix unrestricted.

Chamberlain's procedure is very robust in the sense that by allowing the reduced form errors variance matrix to be an arbitrary function of x_i and z_i, we can make consistent inferences about $(\alpha, \beta', \gamma')$ in a wide variety of situations. However, while there are no particular reasons to believe that u^+_{i0} is homoscedastic, the variance components structure for $u_{i1},\ldots,u_{iT}$ is one of the basic features of the model that we are interested in testing. Moreover, if a structure of this kind (possibly including an autoregressive-moving average scheme for the transitory component v_{it}) is not rejected, the implied constraints can be exploited in order to obtain more efficient estimates of $(\alpha, \beta', \gamma')$.

In view of this considerations, we shall make the simplifying assumption that $E(u^{+2}_{i0})$ and $E(u^+_{i0}\ u_{it})$, $t=1,\ldots,T$, do not depend upon z^*_i (possibly accompanied by a White (1980a)'s heteroscedasticity test), while assuming that $(u^+_{i0}\ u_{i1},\ldots,u_{iT})$ are independently and identically distributed according to some multivariate distribution, not necessarily normal, with finite moments up to the fourth order. Furthermore, we can replace Bhargava and Sargan's stationarity assumptions about the covariances between u^+_{i0} and $u_{i1},\ldots,u_{iT}$, ω_{01}, $\omega_{02},\ldots,\omega_{0T}$ say, by assuming that they are a further set of T arbitrary coefficients. Note that the variance of u^+_{i0}, ω_{00}, is already effectively unrestricted in Bhargava and Sargan's formulation given the presence of σ^2_ε. This approach has been introduced in Arellano (1983, 1984) and has various advantages.[3] First, it allows us to consider nonstationary schemes for the v_{it} taking advantage of the availability of a large number of observations per-period (see Section 1.5 below), and it also makes unnecessary to restrict α to lie inside the unit circle. Second, under this formulation, the error covariance matrix is constrained but independent of the slope parameters, what will lead to enormously simplified efficient methods of estimation for both regression and covariance parameters (see Chapters 5 and 6). In particular, it is worth noticing that in the QML context, it is possible to concentrate $\omega_{00}, \omega_{01}, \ldots, \omega_{0T}$ out of the likelihood function, thus leading to simpler manipulation and lesser computational burden.

1.5 Serial Correlation and Unrestricted Intertemporal Covariance

The effect of random shocks acting through the time-varying errors, while deteriorating over time, may persist longer than one period . The v_{it} may also include unobservable variables which are serially correlated. Since both situations are likely to occur in practice, it is unrealistic to make the assumption that the v_{it} are white noise errors. This fact has been acknowledged for a long time, and many researchers have allowed for serial correlation - mostly first order autoregressive schemes - in the estimation of static equations from panel data (cf. Lillard and Willis (1978), Bhargava, Franzini and Narendranathan (1982), Chowdhury and Nickell (1982), and MaCurdy (1982), among others).

However, given the fact that T is fixed and N is large we are able to estimate arbitrary intertemporal covariance matrices, thus avoiding to place restrictions in the form of the serial correlation of u_{it}.[4] Therefore, if the objective is simply 'to allow for' serial correlation, a robust solution is to obtain Ω^* unrestricted estimates of α, β and γ. The problem of specifying serial correlation in panel data then becomes a problem of modelling Ω^* (for example, researchers can be interested in testing the existence and the magnitude of a permanent component in the error term). A consequence of this is that it is possible to consider a broader family of models than in time series models. Various kinds of non-stationarity can be introduced, like autoregressive schemes with the roots on or outside the unit circle, arbitrary forms of time heteroscedasticity, or ARMA schemes with changing coefficients (cf. MaCurdy (1982)). Nevertheless, it is convenient to preserve the interpretability of the error structures under consideration and in this regard the models that display a stationary correlation pattern are the more interesting. Incidentally, notice that it is possible to allow for arbitrary heteroscedasticity over time and at the same time to specify a stationary serial correlation pattern for v_{it}; this can be achieved by setting

$$v_{it} = \sigma_t v^*_{it}$$

where v^*_{it} follows some stationary ARMA process with i.i.d.(0,1) white noise errors. Now we have $Cov(v_{it}, v_{is}) = \sigma_t \sigma_s cov(v^*_{it}, v^*_{is})$ and thus $Corr(v_{it}, v_{is}) = Corr(v^*_{it}, v^*_{is})$ for any i. This is not the case if we consider instead an ARMA process where the variance of the white noise error is varying over time. However, more general non-stationary models are still possible, and see MaCurdy (1982) and Tiao and Ali (1971) for suggestions about the treatment of initial conditions when stationary correlation is not assumed.

MaCurdy (1982) has also proposed a method of selecting autoregressive-moving average schemes for the v_{it} in models that are not dynamic. His suggestion is to use least-squares residuals of the equation in first differences (thus avoiding the complications originated by the time invariant error component) to construct sample correlograms and sample partial correlation functions, which can be used as a basis for choosing an appropriate specification for the ARMA process generating the transitory components. Then since differencing simply introduces an unit root in the moving average polynomial, its effects can be undone in the sense that one can reconstruct the ARMA process associated with levels. MaCurdy also suggests an ingenious method to estimate simple and partial autocorrelations using a constrained seemingly unrelated equations procedure. Once a particular specification is chosen, he proposes to estimate the restricted covariance matrix by using conditional QML methods (see also MaCurdy (1981) for a discussion of its properties).

While this approach could be generalised to dynamic models (e.g. by basing the calculation of correlograms on three-stage least-squares residuals) and it can be of interest in indicating models to consider, the possibility of obtaining consistent estimates of Ω^* unrestricted suggests to base a formal specification search on a sequence of tests of particular structures against Ω^* unrestricted in increasing order of complexity. This is the approach advocated by Bhargava and Sargan (1983), which rely upon likelihood ratio statistics to test the white noise error components model against Ω^* unrestricted. Unfortunately, unlike the case of regression parameters constraints, likelihood ratio tests of covariance restrictions are asymptotically distributed as a chi-squared on the null hypothesis only under the assumption of normality of the error term (a point noticed by MaCurdy (1981) and that will be discussed in Chapter 4). The asymptotic distribution of the likelihood ratio test under the null hypothesis can still be calculated when the errors are non-normal, but it seems convenient to construct tests that are robust to the non-normality of the errors. Among these, we shall develop Wald tests (Chapter 4) and minimum chi-square tests (Chapters 5 and 6). The advantage of the former is that it only requires the estimation of the unrestricted model. However, as it is well known there are two different ways of expressing exact prior information. If ω is the $(T+1)(T+2)/2$ vector of different elements of Ω^*, a set of r constraints can be expressed as a set of equations of the form

$$f_j(\omega) = 0 \qquad (j=1,\ldots,r).$$

Alternatively, we may assume that the elements of ω are related functionally to a second set of $(T+1)(T+2)/2-r$ parameters τ

$$\omega = \omega(\tau).$$

Setting up Wald tests requires explicit expressions of the constraint equations f_j $(j=1,\ldots,r)$ which can be difficult to obtain in some cases. On the contrary, minimum chi-square statistics are expressed in terms of the constrained parameters τ, what, for our purposes, will usually be a straightforward way of handling the problem.

1.6 Correlation between the Explanatory Variables and the Individual Effects

A leading objective in the estimation of models from panel data has been to obtain estimates of the regression coefficients free of bias due to the omission of relevant individual-specific effects. In static models, this has traditionally been achieved by subtracting time means to individual observations, thus removing all time-invariant terms in the equation. Clearly, in this way the coefficients on the time-invariant variables γ cannot be estimated. In fact, if all the x_{it} and z_i are correlated with the individual effects, the γ's are not identified. Different alternatives arise if we are willing to assume that some of the included explanatory variables are uncorrelated with the individual effects. This is the case studied by Hausman and Taylor (1981) for static models and Bhargava and Sargan (1983) for dynamic models.

In the latter context, we still assume that x_{it} and z_i are independent of v_{it} but now we introduce the partitions $x'_{it} = (x'_{1it} \vdots x'_{2it})$ of dimension $(1 \times n_1, 1 \times n_2)$ and $z'_i = (z'_{1i} \vdots z'_{2i})$ of dimension $(1 \times m_1, 1 \times m_2)$ such that x_{1it} and x_{2it} are, respectively, vectors of variables uncorrelated and correlated with the η_i, and similarly for z_{1i} and z_{2i}. The suggestion of Hausman and Taylor is to use the individual means of the x_{1it} variables as instruments for the

z_{2i}, and thus a necessary condition for the identification of β and γ (in their model α is equal to zero) is that $n_1 \geq m_2$. If the rank condition does not fail and $n_1 > m_2$, the Hausman and Taylor's estimator of β differs from and is more efficient than the within-groups estimator, while if $n_1 = m_2$ the two estimators are identical. Incidentally, note that Hausman and Taylor's reduced form equation for z_{2i} can be replaced by

$$(1.6.1) \qquad z_{2i} = \phi_0' x_{1io} + \ldots + \phi_T' x_{1iT} + \psi' z_{1i} + \xi_i \ .$$

Indeed, they assume $\phi_0 = \ldots = \phi_T$. These restrictions can be appropriate if T is large and x_{1it} is stationary, but in general they are not justified in the present context (see Chamberlain (1980 and 1982) for a discussion of this point). Using a general reduced form for z_{2i}, each variable in x_{1it} provides a set of T+1 instruments for the z_{2i} and the order condition for identification becomes $(T+1)n_1 \geq m_2$.

Bhargava and Sargan adopt a similar approach for dynamic models, but they further assume that the deviations from time means of the x_{2it} are uncorrelated with η_i, what enables them to have a set of Tn_2 extra instruments at the expense of only n_2 new variables - the time means of the x_{2it} - that are correlated with η_i. Let us consider this model in some more detail. In general, we can write

$$(1.6.2) \qquad x_{2it} = \kappa_t \eta_i + \zeta_{it} \qquad (t=0,1,\ldots,T)$$

where ζ_{it} is independent of η_i. But if we assume $\kappa_0 = \ldots = \kappa_T$, (1.6.3) $\tilde{x}_{2it} = x_{2it} - \bar{x}_{2i} = \zeta_{it} - \bar{\zeta}_i$ $(t=1,\ldots,T)$

where $\bar{x}_{2i} = \frac{1}{T+1} \sum_{t=0}^{T} x_{2it}$ and $\bar{\zeta}_i = \frac{1}{T+1} \sum_{t=0}^{T} \zeta_{it}$, so that the $\tilde{x}_{2it}$ are independent of η_i. The vector of instrumental variables is now

$$z_i^{+'} = (x'_{1io}, \ldots, x'_{1iT}, \tilde{x}'_{2i1}, \ldots, \tilde{x}'_{2iT}, z'_{1i})$$

and the complete model can be expressed as

(1.6.4) $B(\alpha)\, y_i + C_x\, x_i + C_z\, z_i = u_i$

(1.6.5) $z_{2i} = F\, z_i^{+} + \xi_{zi}$,

(1.6.6) $\bar{x}_{2i} = G\, z_i^{+} + \xi_{xi}$,

(1.6.7) $x_{2it} = \tilde{x}_{2it} + \bar{x}_{2i}$ $(t=1,\ldots,T)$.

Substituting the last set of T identities into the first block of equations, we obtain a system of $(T+1) + m_2 + n_2$ equations whose endogenous variables are given by $y_i^{+'} = (y'_i, z'_{2i}, \bar{x}'_{2i})$. The identification of this model, as shown by Bhargava and Sargan, requires $T \geq 4$, $n_1 > 0$, $\text{plim}(\Sigma_{i=1}^{N}(z_i^{+} z_i^{+'})/N)$ to be positive definite and the matrix F_x to be of full rank, where $F = (F_x \vdots F_z)$ and F_z correspond to the columns of coefficients on z_{1i}. A crude instrumental variables estimator of $\delta' = (\alpha\ \beta'\ \gamma')$, equivalent to (1.3.9) minimises

(1.6.8) $[\text{vec}(U)]'\, (Z^{+}(Z^{+'} Z^{+})^{-1} Z^{+'} \otimes I_T)\, \text{vec}(U)$

where $Z^{+'} = (z_1^+, \ldots, z_N^+)$, and it is given by

$$(1.6.9) \quad \hat{\delta}_{CIV} = [X^{+'}(Z^+(Z^{+'}Z^+)^{-1}Z^{+'} \otimes I_T)X^+]^{-1} X^{+'}(Z^+(Z^{+'}Z^+)^{-1}Z^{+'} \otimes I_T)y.$$

Using CIV residuals, $\hat{u}_{it}$ say, we can compute an unrestricted estimate of the covariance matrix of $(u_{i1}, \ldots, u_{iT})$

$$(1.6.10) \quad \hat{\Omega} = \left\{ \frac{1}{N} \sum_{i=1}^{N} \hat{u}_{it} \hat{u}_{is} \right\} \qquad (t,s=1,\ldots,T)$$

which in turn can be used to construct a three stage least-squares estimator of δ

$$(1.6.11) \quad \hat{\delta}_{3SLS} = [X^{+'}(Z^+(Z^{+'}Z^+)^{-1}Z^{+'} \otimes \hat{\Omega}^{-1})X^+]^{-1} X^{+'}(Z^+(Z^{+'}Z^+)^{-1}Z^{+'} \otimes \hat{\Omega}^{-1})y.$$

$\hat{\delta}_{3SLS}$ is asymptotically equivalent to the Ω-unrestricted LIML estimator suggested by Bhargava and Sargan. They also apply to this case the constrained LIML procedures that enforce the error components restrictions on the covariance matrix.

None of these methods can be applied when the individual effects are suspected to be correlated with all the observed explanatory variables. Regrettably, the within-group estimates for dynamic models are inconsistent as N tends to infinity if T is kept fixed. Analytical expressions for these inconsistencies have been given by Nickell (1981). The problem is that transformations like deviations from time means or first differences fail to remove the correlation between the lagged endogenous variables and the disturbance term. However, they do remove the permanent component of the errors and so the source of correlation with the remaining explanatory variables. Thus, if proper account is taken of the correlation between lagged y's and errors, consistent estimation of the coefficients corresponding to time-varying variables is still possible. But this is precisely the problem that the methods introduced in Sections 1.3 and 1.4 are intended to solve, and they can be easily extended to cover such cases.

For example, transforming to first differences our original equation we have

(1.6.12) $y_{it} - y_{i(t-1)} = \alpha(y_{i(t-1)} - y_{i(t-2)}) + \beta'(x_{it} - x_{i(t-1)}) + u^*_{it}$,

(1.6.13) $u^*_{it} = v_{it} - v_{i(t-1)}$ $\qquad$ (t=2,...,T).

Now, the model has to be completed with prediction equations for y_{io} and y_{i1} (see Chamberlain (1984))

(1.6.14) $y_{io} = \mu'_{oo}\, x_{io} + \ldots + \mu'_{oT}\, x_{iT} + u^+_{io}$,

(1.6.15) $y_{i1} = \mu'_{1o}\, x_{io} + \ldots + \mu'_{1T}\, x_{iT} + u^+_{i1}$.

If v_{it} follows an ARMA(1,1) scheme with coefficients ϕ and λ , u^*_{it} will follow an ARMA(1,2) scheme of the form

(1.6.16) $u^*_{it} = \phi\, u^*_{i(t-1)} + \varepsilon_{it} - (1+\lambda)\varepsilon_{i(t-1)} + \lambda\, \varepsilon_{i(t-2)}$.

In particular, if v_{it} is white noise with variance σ^2, the covariance matrix of $(u^*_{i2},\ldots,u^*_{iT})$ will be a $(T-1) \times (T-1)$ matrix of the form

$$\sigma^2 \begin{bmatrix} 2 & -1 & 0 & \ldots & 0 \\ -1 & 2 & -1 & \ldots & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & -1 & 2 \end{bmatrix} = \sigma^2\, \Omega_0, \text{ say.}$$

Since T is fixed, none of the time series problems that appear in the estimation of models with moving average errors when the root lies on the unit circle are relevant here. If the v_{it} are known to be white noise errors, Ω_0 can be used to construct a 3SLS estimator of α and β by minimising

$$(1.6.17) \quad S(\alpha,\beta) = [\text{vec}(U^*)]' \; (X(X'X)^{-1} X' \otimes \Omega_0^{-1}) \; \text{vec}(U^*)$$

where $X' = (x_1,\ldots,x_N)$ and $U^{*\prime} = (u^*_1,\ldots,u^*_N)$ with $u^{*\prime}_i = (u^*_{i2},\ldots,u^*_{iT})$. Moreover, noting that $\text{vec}(U^*) = u^*$ where u^* is a $N(T-1) \times 1$ vector of errors in first differences

$$u^{*\prime} = (u^*_{12},\ldots,u^*_{1T},\ldots,u^*_{N2},\ldots,u^*_{NT})$$

and that

$$u^* = (I_N \otimes D)u$$

where u is defined in Section 1.3 and D is the $(T-1) \times T$ matrix

$$D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix},$$

we can write

$$S(\alpha,\beta) = u'(I_N \otimes D')(X(X'X)^{-1} X' \otimes \Omega_O^{-1})(I_N \otimes D)u$$

$$= u'(X(X'X)^{-1} X' \otimes D' \Omega_O^{-1} D)u.$$

However, since $\Omega_O = DD'$ and D is an orthogonal complement of ι, i.e. $D\iota = 0$ (cf. Sargan and Bhargava (1983)) it turns out that

$$D' \Omega_O^{-1} D = I_T - \frac{\iota\iota'}{T} = Q, \text{ say.}$$

Finally, using $Q = QQ$ we have

$$S(\alpha,\beta) = u'(I_N \otimes Q)(X(X'X)^{-1} X' \otimes I_T)(I_N \otimes Q)u$$

$$= u^{+'}(X(X'X)^{-1} X' \otimes I_T)u^+$$

where u^+ is the NT x 1 vector of errors in deviations from time means. Therefore, CIV in the model in deviations from times means is numerically the same as the 3SLS estimator that uses Ω_O in the model in first differences.

The estimator that replaces Ω_0 in (1.6.17) by an unrestricted estimate of the covariance matrix will be asymptotically equivalent to the minimiser of (1.6.17). However, it has the advantage that its asymptotic distribution remains unchanged when the v_{it} are serially correlated.

Note that the previous model in first differences is a particular case of the model

(1.6.18) $$y_{it} = \alpha_1 y_{i(t-1)} + \alpha_2 y_{i(t-2)} + \beta_0' x_{it} + \beta_1' x_{i(t-1)} + u_{it}$$

with the linear constraints $\alpha_1 + \alpha_2 = 1$ and $\beta_0 + \beta_1 = \underset{\sim}{0}$. Interestingly, a dynamic model in which the x_{it} are correlated with the individual effects can be seen as a special case of a more general dynamic model in which the x_{it} are completely exogenous variables and no individual effects are present.

The purpose of the previous discussion has been to emphasise the relevance of the basic dynamic model with exogenous variables and serially correlated errors to cover a variety of situations of interest.

NOTES

1 Given a matrix $A = \begin{pmatrix} a_1' \\ a_2' \\ \vdots \\ a_n' \end{pmatrix}$ with n rows, we define the vec operator as

$$\text{vec}(A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} .$$

If $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are matrices of arbitrary order we define the Kronecker product as $A \otimes B = \{a_{ij} B\}$. Two properties of the vec operation which will be useful are

$$\text{vec}(ABC) = (A \otimes C') \text{ vec}(B)$$

if the matrix product ABC exists, and

$$[\text{vec}(A)]' \text{ vec}(B) = \text{tr}(A'B) ,$$

where A and B are matrices of the same order.

2 If $\hat{W}$ is replaced by $\left(\frac{1}{N} \sum_{i=1}^{N} \hat{\nu}_i \hat{\nu}_i' \right) \otimes \hat{M}^{-1}$ we obtain Malinvaud (1970)'s minimum distance estimator.

3 Bhargava and Sargan also discussed a similar specification, but with $\omega_{01} = \omega_{02} = \ldots = \omega_{0T}$. These equality constraints will not be appropriate if the v_{it} are serially correlated.

4 The implications of this procedure were studied by Kiefer (1980) in the context of a 'fixed effects' treatment of a static model.

CHAPTER 2

QML ESTIMATION OF DYNAMIC MODELS WITH SERIALLY CORRELATED ERRORS

2.1 Introduction

This Chapter is concerned with the formulation and estimation by quasi-maximum likelihood (QML) methods of dynamic random effects models with first-order autoregressive-first-order moving average time-varying errors.

Quasi-maximum likelihood estimators are of interest because they provide a broad framework for the estimation of models under general constraints. For this reason they have been commonly used when prior information on covariance matrices is available. They are specially attractive in our context, i.e. that of a system of (T+1) equations, T of which are linked by linear constraints and the remaining one - the prediction equation for y_{i0} - is in reduced form, since the nuisance coefficients in the latter equation can be easily concentrated out of the likelihood function (what in fact is an application of the LIML technique for a subset of equations in a simultaneous system), thus leading to a criterion function that only depends on the parameters of interest and where we can still introduce constraints in the covariance coefficients. However, since normality is not assumed, we cannot rely upon maximum likelihood asymptotic theory in discussing the properties of these estimators.

This discussion will be the purpose of Chapter 3. Section 2.2 in this Chapter introduces the models arising from three different assumptions about the covariances between the errors in the equation for y_{i0} and the remaining disturbances in the model. Section 2.3 derives the concentrated likelihood functions for these three alternative models. Section 2.4 considers QML estimation with arbitrary intertemporal covariance and other alternative asymptotically equivalent methods. In Section 2.5, the performance of QML methods is investigated, either for correct models or under several misspecifications - though always using normal variates - by resorting to experimental evidence. Finally, Section 2.6 discusses an extended model with arbitrary heteroscedasticity over time.

2.2 Three Alternative Models with ARMA Errors

We assume the following model

(2.2.1) $$y_{it} = \alpha\, y_{i(t-1)} + \beta' x_{it} + \gamma' z_i + u_{it}$$

(2.2.2) $$u_{it} = \eta_i + v_{it}$$

(2.2.3) $$v_{it} = \phi\, v_{i(t-1)} + \varepsilon_{it} + \lambda \varepsilon_{i(t-1)} \qquad (i=1,\dots,N;t=1,\dots,T)$$

with $\eta_i \sim iid(0,\sigma^2_\eta)$, $\varepsilon_{it} \sim iid(0,\sigma^2)$ and $E(\eta_i \varepsilon_{jt})=0$ for all i,j,t.

x_{it} and z_i are observed constants: z_i is a m-vector of time-invariant exogenous variables and x_{it} is a n-vector of time-varying exogenous variables. β and γ are n x 1 and m x 1 vectors of unknown parameters, respectively, and α is an unknown scalar parameter. We also observe t=0, so that (T+1) time series observations are available on N cross-sectional units. It is also assumed that $|\phi| < 1$ and $|\lambda| < 1$ so that the error process is stationary and invertible.

It is useful to re-write (2.2.1) as an incomplete system of T simultaneous equations. Introducing the T x (T+1) matrix

$$B^{+} = \begin{pmatrix} -\alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \end{pmatrix}$$

and the vectors $y_i' = (y_{i0}\ y_{i1} \ldots y_{iT})$, $u_i' = (u_{i1} \ldots u_{iT})$, we then have

$$(2.2.4) \qquad A(\delta)\, d_i = u_i \qquad (i=1,\ldots,N)$$

where $d_i' = (y_i'\ x_i'\ z_i') = (y_i'\ z_i^{*\prime})$, $\delta' = (\alpha\ \beta'\ \gamma')$ and

$$A(\delta) = (B^{+} \vdots -I^{*} \otimes \beta' \vdots -\iota\gamma') = (B^{+} \vdots C).$$

ι is a T-vector of ones and $I^{*} = (0 \vdots I_T)$ (i.e. a T-unit matrix augmented by a column of zeroes).

We rule out the possibility that $x_{kit} = x_{kis}$ for all i and for some k and $t \neq s$. Indeed, we shall make the assumption that $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (z_i^{*} z_i^{*\prime})$ is a positive definite matrix.

In view of our assumptions, the covariance matrix of u_i is given by

$$(2.2.5) \qquad E(u_i u_i') = \Omega = \sigma^2 V + \sigma_\eta^2 \iota\iota'$$

where V is a Toeplitz matrix proportional to the serial covariance matrix of the ARMA (1,1) process, whose t,s th element is

$$v_{t-s} = \frac{1 + \lambda^2 + 2\phi\lambda}{1-\phi^2} \quad \text{if } t-s=0$$

$$= \phi^{|t-s|-1} \frac{(1+\phi\lambda)(\phi+\lambda)}{1-\phi^2} \quad \text{otherwise}$$

In our simultaneous equations analogue, Ω becomes the variance matrix of the errors of T structural equations, that is, serial correlation turns into correlation between disturbances from different equations, and so we end up with a simultaneous equations system with linear cross-equation restrictions and a restricted variance matrix.

We complete the model with the assumption that the initial observations are determined by a reduced form equation of the type

$$(2.2.6) \qquad y_{io} = \mu_o' x_{io} + \ldots + \mu_T' x_{iT} + \xi' z_i + u_{io} = \mu' z_i^* + u_{io}$$

where μ is a n(T+1)+m vector of unrestricted coefficients, and u_{io} is a random error with zero mean and arbitrary variance ω_{oo}. We will develop this model under three different assumptions about the covariances between u_{io} and $u_{i1}, \ldots, u_{iT}$, $E(u_{io} u_{it}) = \omega_{ot}$ $(t=1,\ldots,T)$:

(i) $\omega_{O1} = \ldots = \omega_{OT} = 0$. Thus y_{iO} can be regarded as an exogenous variable in the simultaneous system and so (2.2.4) becomes a complete model. This is equivalent to the assumption that the y_{iO} $(i=1,\ldots,N)$ are fixed and known constants.

(ii) Further assuming $|\alpha| < 1$ we may restrict $\omega_{O1},\ldots,\omega_{OT}$ on the lines suggested by Bhargava and Sargan (1983) for models with white noise errors. In this case we take u_{iO} to be

$$(2.2.7) \qquad u_{iO} = \zeta_i + \frac{\eta_i}{1-\alpha} + \sum_{k=0}^{\infty} \alpha^k \, v_{i(-k)}$$

where ζ_i is a prediction error defined as

$$(2.2.8) \qquad \zeta_i = \sum_{k=0}^{\infty} \alpha^k (\beta' x_{i(-k)} + \gamma' z_i) - \mu' z_i^*$$

which is assumed to have constant variance σ_ζ^2 for all i. So we have

$$(2.2.9) \qquad \omega_{Ot} = \frac{\sigma_\eta^2}{1-\alpha} + \phi^{t-1} \, \delta_2 \, \sigma^2 \qquad (t=1,\ldots,T)$$

with

$$(2.2.10) \qquad \delta_2 = \frac{(1+\phi\lambda)(\phi+\lambda)}{(1-\alpha\phi)(1-\phi^2)} \, ,$$

in particular if v_{it} follows purely a moving average scheme, $\phi=0$ and $\delta_2 = \lambda$, and then, except ω_{O1}, all covariances are equal to $\sigma_\eta^2/(1-\alpha)$. ω_{OO} is still unrestricted but it can be expressed in terms of σ_ζ^2 as follows

$$(2.2.11) \quad \omega_{00} = \sigma_\zeta^2 + \frac{\sigma_\eta^2}{(1-\alpha)^2} + \frac{\delta_1}{1-\alpha^2}\sigma^2$$

with

$$\delta_1 = \frac{(1+\lambda^2)(1+\alpha\phi) + 2\lambda(\alpha+\phi)}{(1-\alpha\phi)(1-\phi^2)} .$$

(iii) $\omega_{01},\ldots,\omega_{0T}$ will be simply unrestricted parameters. The advantages of this assumption were already discussed in Section 1.4.

We shall refer to these three cases as models a, b and c respectively. Let Ω^* be the variance matrix of the complete system comprising (2.2.4) and (2.2.6) ie $\Omega^* = E(u_i^* u_i^{*\prime})$ where $u_i^{*\prime} = (u_{i0} \vdots u_i^\prime)$. Models a, b and c will differ in the assumptions about the coefficients of the top row of Ω^*. In any event, we are assuming u_i^* to be i.i.d. according to some multivariate distribution with zero mean vector and covariance matrix Ω^*, and we further assume the third and fourth order moments to be finite and unrestricted.

2.3 Quasi-Maximum Likelihood Estimation

The log-likelihood function for the complete system of (T+1) equations, apart from a constant term, can be written as

$$(2.3.1) \quad L = -\frac{N}{2}\log\det(\Omega^*) - \tfrac{1}{2}\,\mathrm{tr}(\Omega^{*-1} U^{*\prime}U^*) + N\log|\det(B)|$$

where $U^{*\prime} = (u_1^*,\ldots,u_N^*)$. Alternatively we can partition $U^{*\prime}$ as

$$U^{*\prime} = \begin{pmatrix} u_0^\prime \\ U^\prime \end{pmatrix}$$

with $U' = (u_1,\ldots,u_N)$ and $u_0' = (u_{10},\ldots,u_{N0})$. B is the (T+1) x (T+1) matrix given by

$$B = \begin{bmatrix} 1 & 0 \ldots 0 \\ \vdots & B^+ \end{bmatrix} .$$

(2.3.1) is the likelihood function for a general simultaneous equations system. However, since B is lower triangular and all its diagonal elements are equal to 1, log det(B) = 0, and therefore the jacobian term will not occur here.

In the likelihood function, $U^* = (u_0 \vdots U)$ is a short-hand for

$$u_0 = y_0 - Z^*\mu$$

and

$$U' = A(\delta)D'$$

with $Z^{*\prime} = (z_1^*,\ldots,z_N^*)$ and $D' = (d_1,\ldots,d_N)$.
Since we are only interested in the estimation of the parameters corresponding to the structural block of the last T equations, we will concentrate the likelihood function with respect to μ.

It is convenient to introduce a general notation for the partition of Ω^{*-1}, namely

$$\Omega^{*-1} = \begin{bmatrix} \omega^{00} & \omega^{01} \\ \omega^{10} & \Omega^{11} \end{bmatrix}$$

Now by making use of the formulae for the determinant and the inverse of a partitioned matrix, after some manipulation, we can re-write (2.3.1) as[1]

$$(2.3.2) \qquad L = -\frac{N}{2}\log\det(\Omega) - \tfrac{1}{2}\operatorname{tr}(\Omega^{-1}U'U) + \frac{N}{2}\log\omega^{00}$$

$$- \frac{1}{2\omega^{00}}\,\omega^{01}(U'U)\omega^{10} - \frac{\omega^{00}}{2}(u_0'u_0) - \omega^{01}(U'u_0)$$

From the first order conditions for μ (note that μ only appears in the last two terms on the left hand side of (2.3.2)), its maximum likelihood estimator turns out to be

$$(2.3.3) \qquad \hat{\mu} = (Z^{*\prime}Z^{*})^{-1}Z^{*\prime}\left(y_0 + \frac{U\omega^{10}}{\omega^{00}}\right)$$

which is used to concentrate L, i.e. $L^{*}(\delta,\Omega^{*}) = L(\delta,\hat{\mu},\Omega^{*})$. Substituting and rearranging we have

$$(2.3.4) \qquad L^{*} = -\frac{N}{2}\log\det(\Omega) - \tfrac{1}{2}\operatorname{tr}(\Omega^{-1}U'U) + \frac{N}{2}\log\omega^{00}$$

$$- \frac{\omega^{00}}{2}(y_0'\,Q\,y_0) - \frac{1}{2\omega^{00}}\,\omega^{01}(U'QU)\omega^{10} - \omega^{01}(U'Qy_0)$$

where Q stands for the idempotent matrix

$$Q = I_N - Z^{*}(Z^{*\prime}Z^{*})^{-1}Z^{*\prime}$$

and so

$$U'Q = B^{+}\,Y'Q.$$

In what follows we specialise the likelihood function (2.3.4) to the models a, b and c introduced in the previous section.

Model a

In model a, $\underset{\sim}{\omega}_{01} = (\omega_{01},\ldots,\omega_{0T}) = (0,\ldots,0)$ so that $\omega^{00} = 1/\omega_{00}$, $\omega^{01} = 0$ and $\Omega^{11} = \Omega^{-1}$. Enforcing these restrictions in (2.3.4) we obtain

$$(2.3.5) \qquad L_a^+ = -\frac{N}{2}\log\det(\Omega) - \tfrac{1}{2}\,\mathrm{tr}(\Omega^{-1}U'U) - \frac{N}{2}\log\omega_{00} - \frac{1}{2\omega_{00}}(y_0'\,Q\,y_0)$$

Since the last two terms are irrelevant in so far as the maximisation with respect to δ and the constraint parameters in Ω is concerned, we may just consider maximising

$$(2.3.6) \qquad L_a = -\frac{N}{2}\log\det(\Omega) - \tfrac{1}{2}\,\mathrm{tr}(\Omega^{-1}U'U)$$

This is the kind of likelihood function that was considered by Balestra and Nerlove (1966). Note that $\mathrm{tr}(\Omega^{-1}U'U) = [\mathrm{vec}(U)]'(I_N \otimes \Omega^{-1})\,\mathrm{vec}(U)$, and using the stacked regression notation introduced in Chapter 1 we have

$$(2.3.7) \qquad \mathrm{tr}(\Omega^{-1}U'U) = (y - X^+\delta)'(I_N \otimes \Omega^{-1})(y - X^+\delta)$$

Here, we follow Bhargava and Sargan (1983) in parameterising Ω as

$$(2.3.8) \qquad \Omega = \sigma^2(V + \rho^2\,\iota\iota')$$

where $\rho = \sigma_\eta/\sigma$. This has the advantage that any finite value of ρ leads to a positive definite matrix Ω. The determinant and the inverse of Ω are thus given by

$$\text{(2.3.9)} \qquad \det(\Omega) = \sigma^{2T}(1 + \rho^2\, \iota' V^{-1}\iota)\, \det(V)$$

and

$$\text{(2.3.10)} \qquad \Omega^{-1} = \frac{1}{\sigma^2}\left[V^{-1} - \frac{\rho^2}{(1+\rho^2\iota' V^{-1}\iota)} V^{-1}\iota\iota' V^{-1}\right] = \frac{1}{\sigma^2}\,\bar{\Omega}^{-1}\ , \text{ say.}$$

The exact form of the determinant and the inverse of V have been obtained by Tiao and Ali (1971), who show that

$$\text{(2.3.11)} \qquad \det(V) = 1 + \frac{(\phi+\lambda)^2(1-\lambda^{2T})}{(1-\phi^2)(1-\lambda^2)}\ .$$

The exact inverse is highly nonlinear and the derivation of a computationally convenient expression is given in Appendix 2.A.

L_a becomes then

$$\text{(2.3.12)} \qquad L_a = -\frac{NT}{2}\log\sigma^2 - \frac{N}{2}\log\det(V) - \frac{N}{2}\log(1+\rho^2\,\iota' V^{-1}\iota)$$

$$- \frac{1}{2\sigma^2}\operatorname{tr}(V^{-1}U'U) + \frac{1}{2\sigma^2}\,\frac{\rho^2}{(1+\rho^2\iota' V^{-1}\iota)}\,\iota' V^{-1}U'U\, V^{-1}\iota$$

and concentrating the likelihood with respect to σ^2

$$\text{(2.3.13)} \qquad L_a^* = -\frac{N}{2}\log\det(V) - \frac{N}{2}\log\,(1 + \rho^2\,\iota' V^{-1}\iota) - \frac{NT}{2}\log s^2$$

where the maximum likelihood estimator of σ^2 is

$$(2.3.14) \quad s^2 = \frac{1}{NT}\,[tr(V^{-1}U'U) - \frac{\rho^2 \iota' V^{-1} U'UV^{-1} \iota}{1+\rho^2 \iota' V^{-1} \iota}]$$

$$= \frac{1}{NT}\,(y-X^{+}\delta)'(I_N \otimes (V+\rho^2 \iota\iota')^{-1})(y-X^{+}\delta).$$

L_a^* is a function of δ, ρ, ϕ and λ that can be maximised by using some numerical optimisation procedure, with the restrictions that $|\phi| < 1$ and $|\lambda| < 1$.

Model b

In model b

$$(2.3.15) \quad \underset{\sim}{\omega}_{01} = \sigma^2(\frac{\rho^2}{1-\alpha}\,\iota' + \delta_2\, q') = \sigma^2\, \underset{\sim}{\bar{\omega}}_{01}\ , \text{ say.}$$

where $q' = (1\ \phi\ \phi^2,\ldots,\phi^{T-1})$.
Using $\omega^{01} = -\,\omega^{00}(\omega_{01}\,\Omega^{-1})$ and further noting that $\omega_{01}\,\Omega^{-1} = \bar{\omega}_{01}\,\bar{\Omega}^{-1}$, we can write (2.3.4) as

$$(2.3.16)\ L_b = L_a + \frac{N}{2}\log\omega^{00} - \frac{\omega^{00}}{2}\,(y_0' - \bar{\omega}_{01}\,\bar{\Omega}^{-1}U')Q(y_0 - U\bar{\Omega}^{-1}\,\bar{\omega}_{10}).$$

So there is no difficulty in concentrating L_a with respect to σ^2 as above, and the remaining two terms with respect to ω^{00},

$$(2.3.17) \quad L_b^* = L_a^* - \frac{N}{2}\log[(y_0' - \bar{\omega}_{01}\,\bar{\Omega}^{-1}U')Q(y_0 - U\,\bar{\Omega}^{-1}\,\bar{\omega}_{10})]$$

where after some algebraic reductions we obtain

$$(2.3.18)\quad \bar{\omega}_{01}\,\bar{\Omega}^{-1} = (\delta_3\,\iota' + \delta_2\,q')V^{-1}$$

with

$$(2.3.19)\quad \delta_3 = \frac{\rho^2}{(1+\rho^2\iota'V^{-1}\iota)}\left(\frac{1}{1-\alpha} - \delta_2\,q'V^{-1}\iota\right)$$

In particular, if the process is purely autoregressive, $q'V^{-1}\iota = 1-\phi^2$ and $\iota'V^{-1}\iota = 2\phi(1-\phi) + T(1-\phi)^2$.

(2.3.17) is a convenient expression that enforces the constraints (2.3.15) and can be maximised as a function of δ, ρ, ϕ and λ with the restrictions that $|\phi| < 1$, $|\lambda| < 1$ and $|\alpha| < 1$. Alternatively we could parameterise ω_{00} as in (2.2.11) and then explicitly estimate the ratio σ_ζ/σ rather than concentrating ω^{00} out of the likelihood function (see Arellano (1983)).

Model c

In this case we enforce the random effects constraints on Ω but $\omega_{01},\ldots,\omega_{0T}$ are left unrestricted. Thus we only have constraints in the structural block of T equations and its variance matrix. Hence it is natural to concentrate (2.3.4) further with respect to ω^{00} and ω^{01}. To do so it is convenient to parameterise L* in terms of $f_{01} = -\,\omega^{01}/\omega^{00}$, what leads to

$$(2.3.20)\quad L^* = L_a + \frac{N}{2}\log\omega^{00} - \frac{\omega^{00}}{2}[(y_0'Qy_0) + f_{01}(U'QU)f_{10} - 2\,f_{01}(U'Qy_0)]$$

Differentiating L* with respect to f_{10}, its maximum likelihood estimator turns out to be

$$(2.3.21) \quad \hat{f}_{10} = (U'QU)^{-1} U'Qy_0$$

Substituting $\hat{f}_{10}$ into L* we have

$$(2.3.22) \quad L^{**} = L_a + \frac{N}{2} \log \omega^{00} - \frac{\omega^{00}}{2} [y_0'Qy_0 - y_0'QU(U'QU)^{-1}U'Qy_0] \; .$$

Now it only remains to concentrate L** with respect to ω^{00}, but clearly

$$(2.3.23) \quad \frac{1}{\hat{\omega}^{00}} = \frac{1}{N}[y_0'Qy_0 - y_0'QU(U'QU)^{-1}U'Qy_0]$$

$$= \frac{1}{N}(y_0'-\hat{f}_{01}U')Q(y_0-U\hat{f}_{10})$$

so that

$$(2.3.24) \quad L_c = L_a - \frac{N}{2} \log \frac{1}{N}[y_0'Qy_0 - y_0'QU(U'QU)^{-1}U'Qy_0]$$

Note that (2.3.24) directly compares to (2.3.17). However, using the formula for the determinant of a partitioned inverse, a computationally simpler expression can be found. We can write

$$(2.3.25) \quad y_0'Qy_0 - y_0'QU(U'QU)^{-1}U'Qy_0 = \frac{\det\begin{pmatrix} y_0'Qy_0 & y_0'QU \\ U'Qy_0 & U'QU \end{pmatrix}}{\det(U'QU)}$$

and this equals

$$\frac{\det\begin{bmatrix} Y_0'QY_0 & Y_0'QYB^{+\prime} \\ B^{+}Y'QY_0 & B^{+}Y'QYB^{+\prime} \end{bmatrix}}{\det(B^{+}Y'QYB^{+\prime})} = \frac{(\det B)\ \det(Y'QY)(\det B')}{\det(B^{+}Y'QYB^{+\prime})}$$

but since det(B) = 1 we end up with

$$(2.3.26)\quad L_c = L_a + \frac{N}{2}\log\det(B^{+}\ W\ B^{+\prime}) - \frac{N}{2}\log\det(W)$$

where W is the unrestricted estimate of the reduced form covariance matrix

$$W = \frac{1}{N}(Y'QY).$$

Hence, the concentrated likelihood with respect to σ^2 is given by

$$(2.3.27)\quad L_c^* = L_a^* + \frac{N}{2}\log\det(B^{+}WB^{+\prime}) - \frac{N}{2}\log\det(W).$$

If w_{ts} is the (t+1), (s+1) th element of W, the elements of $B^{+}\ W\ B^{+\prime}$ are of the form

$$w_{ts} - \alpha(w_{(t-1)s} + w_{t(s-1)}) + \alpha^2\ w_{(t-1)(s-1)} \qquad (t,s=1,\ldots,T)$$

Notice that L_b^* and L_c^* are of the same form as L_a^*, but an additional term is introduced in each case in order to correct for the correlation between u_{i0} and $(u_{i1},\ldots,u_{iT})$. In all cases the crude instrumental variables estimator introduced in Chapter 1 can be used to provide consistent initial values for δ.

2.4 Estimation with Arbitrary Intertemporal Covariance

We can define the QML estimator of δ that treats Ω^* as an arbitrary symmetric positive definite matrix. Since (2.3.26) has already been concentrated with respect to μ, ω_{00}, $\omega_{01},\ldots,\omega_{0T}$, the relevant likelihood here can be obtained simply by concentrating (2.3.26) further with respect to Ω, where L_a is now as given in (2.3.6). The maximum likelihood estimation of Ω is

$$(2.4.1) \qquad \hat{\Omega} = \frac{1}{N} U'U = \frac{1}{N} A(\delta)\left(\sum_{i=1}^{N} d_i d_i'\right)A'(\delta)$$

so that we obtain a likelihood function which only depends on δ (cf. Bhargava and Sargan (1983)) given by

$$(2.4.2) \qquad L_d(\delta) = -\frac{N}{2}\log\det\left(\frac{U'U}{N}\right) + \frac{N}{2}\log\det(B^{+}WB^{+\prime}) - \frac{N}{2}\log\det(w)$$

Since the covariance matrix is unrestricted, this is an application of the limited information (quasi) maximum likelihood (LIML) method to a subset of equations, and its asymptotic properties are well known in the literature.[2] In particular, it is asymptotically equivalent to the three stage least squares (3SLS) estimator applied to that subset of equations (e.g. see Sargan (1964)). The advantage of the latter is that, since the restrictions in $A(\delta)$ are linear, it is the solution of a set of linear equations and therefore it can be calculated without requiring iterative optimization techniques. However, we are still interested in the QML estimator as it will allow us to discuss (quasi) likelihood ratio tests of covariance restrictions at a later stage and, more generally, its relation to the QML procedures discussed in Section 2.3.

In any event, the 3SLS estimator of δ minimises

$$\text{(2.4.3)} \qquad [\mathrm{vec}(U)]'(Z^*(Z^{*\prime}Z^*)^{-1} Z^{*\prime} \otimes \tilde{\Omega}^{-1}) \ \mathrm{vec}(U)$$

where $\tilde{\Omega}$ is a consistent estimator of Ω, e.g. $\tilde{\Omega} = (1/N)A(\hat{\delta}_{CIV})(\sum_{i=1}^{N} d_i d_i')A'(\hat{\delta}_{CIV})$ and $\hat{\delta}_{CIV}$ is the crude instrumental variables estimator of δ introduced in Section 1.3. Again, using $\mathrm{vec}(U) = y - X^{+}\delta$, the explicit solution of (2.4.3) is given by

$$\text{(2.4.4)} \qquad \hat{\delta}_{3SLS} = [X^{+\prime}(Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}\otimes\tilde{\Omega}^{-1})X^{+}]^{-1} \ X^{+\prime}(Z^*(Z^{*\prime}Z^*)Z^{*\prime} \otimes \tilde{\Omega}^{-1})y$$

However, (2.4.4) is not an useful expression from a computational point of view. The reason is that since N is large and T is small, it is convenient to compute the second order moments data matrix $(1/N)\sum_{i=1}^{N}(d_i d_i')$ just once and construct from it the relevant statistics, thus avoiding the storage of arrays of dimension N, or having to perform successive summations of N products. First, notice that

$$X^{+} = \sum_{t=1}^{T} (X_t^{+} \otimes d_t)$$

where X_t^{+} is the N x (1+n+m) matrix $X_t^{+} = (y_{t-1} \ X_t \ Z)$ and d_t is a T-vector with one in the th position and zero elsewhere. Equally

$$y = \sum_{t=1}^{T} (y_t \otimes d_t)$$

so that

$$x^{+\prime}(Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} \otimes \tilde{\Omega}^{-1})\, x^{+} =$$

$$[\sum_{t=1}^{T}(x_t^{+\prime} \otimes d_t^{\prime})](Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} \otimes \tilde{\Omega}^{-1})[\sum_{s=1}^{T}(x_s^{+} \otimes d_s)]$$

Letting $\tilde{\Omega}^{-1} = \{\tilde{\omega}^{ts}\}$, and since $d_t^{\prime}\, \tilde{\Omega}^{-1}\, d_s = \tilde{\omega}^{ts}$ we have

$$(2.4.5) \qquad \tilde{\delta}_{3SLS} = (\sum_{t=1}^{T}\sum_{s=1}^{T} \tilde{\omega}^{ts}\, \hat{x}_t^{+\prime}\, \hat{x}_s^{+})^{-1} \sum_{t=1}^{T}\sum_{s=1}^{T} \tilde{\omega}^{ts}\, \hat{x}_t^{+\prime}\, y_s$$

where

$$(2.4.6) \qquad \hat{x}_t^{+} = [\hat{y}_{t-1}\ x_t\ z]$$

and

$$\hat{y}_t = Z^*(Z^{*\prime}Z^*)^{-1}(Z^{*\prime}y_t) \qquad (t=0,1,\ldots,T)$$

Moreover using the fact that $d_t^{\prime}\, d_s = 0$ for $t \neq s$ and $d_t^{\prime}\, d_t = 1$, the corresponding expression for the crude instrumental variables estimator is

$$(2.4.7) \qquad \tilde{\delta}_{CIV} = (\sum_{t=1}^{T} \hat{x}_t^{+\prime}\, \hat{x}_t^{+})^{-1} \sum_{t=1}^{T} \hat{x}_t^{+\prime}\, y_t \quad .$$

Finally, we make some remarks on identification issues. The basic identification condition is that $\lim_{N\to\infty}(1/N)(Z^{*\prime}Z^*)$ should be positive definite; if further $T \geq 1$ and at least some element of the vector μ_o is non-zero, the model with unrestricted covariance matrix is identified. Alternatively, if we are not willing to state conditions in terms of

μ_0, the requirements are that at least some element of β is non-zero and $T \geq 2$. Therefore, a necessary condition for identification is that $n \geq 1$. Turning to restricted models, if T=1, Ω^* has three different elements so that a random effects specification with white noise time-varying errors of type b is just identified. Identification of ARMA(1,1) models requires that $T \geq 3$.

2.5 Experimental Evidence

Given the existence in the literature of a certain amount of conflicting Monte-Carlo results on the performance of various maximum-likelihood methods for dynamic random effects models, (cf. Nerlove (1971), Maddala (1971), Bhargava and Sargan (1983)) it was decided to carry out some simulation experiments in order to investigate the practical performance of the methods introduced in this Chapter. We are particularly interested in the consequences of incorrectly specifying y_0 as exogenous when the errors are serially correlated. We also wish to obtain some insight into how the methods that do not constrain Ω^* compare to those in which the covariance restrictions are enforced, and how models b and c compare in turn. Finally, it is important to inspect the ability of these procedures in distinguishing among different serial correlation schemes and between dynamics (lagged endogenous variables) and serial correlation.

Five different sets of samples were generated from models of the form

$$y_{it} = 1 + \alpha\, y_{i(t-1)} + .15\, z_i + .35\, x_{it} + u_{it}$$

$$u_{it} = \eta_i + v_{it}$$

$$v_{it} = \phi\, v_{i(t-1)} + \varepsilon_{it} + \lambda\, \varepsilon_{i(t-1)}$$

$$(i = 1,\ldots,100\ ;\ t = 1,\ldots,20)$$

where $\eta_i \sim NID(0, .16)$, $\varepsilon_{it} \sim NID(0, .25)$ (i.e. $\rho^2 = .64$), and $y_{i0} = v_{i0} = 0$.

The exogenous variables were generated in a similar way as in previous studies

$$x_{it} = .1\, t + .5\, x_{i(t-1)} + p_{it}$$

$$z_i = .1\, x_{i4} + r_i$$

where $p_{it} \sim NID(0,1)$ and $r_i \sim NID(0,1)$. The first ten cross-sections were discarded so that y_0 is an endogenous variable in the system and the same process for v_{it} has been holding in the past.
We are thus left with T = 9 and N = 100.

The five sets of data correspond to the following values of α, ϕ and λ:

Data 1 : $\alpha = .5$, $\phi = .35$, $\lambda = .5$ (ARMA errors)

Data 2 : $\alpha = .5$, $\phi = .35$, $\lambda = 0$ (autoregressive errors)

Data 3 : $\alpha = .5$, $\phi = 0$, $\lambda = .35$ (moving average errors)

Data 4 : $\alpha = 0$, $\phi = .35$, $\lambda = 0$ (a static case with AR(1) errors)

Data 5 : $\alpha = .5$, $\phi = 0$, $\lambda = 0$ (a dynamic case without serial correlation)

TABLE 1

Description of Models to be Simulated

<table>
<tr><th colspan="2" rowspan="2"></th><th rowspan="2">Ω unrestricted</th><th colspan="4">$\Omega = \sigma^2 V + \sigma_\eta^2 \iota\iota'$</th></tr>
<tr><th>ARMA (1,1)</th><th>AR(1)</th><th>MA(1)</th><th>White Noise</th></tr>
<tr><td colspan="2">y_0 exog. ($\omega_{01}=0$)</td><td>un1</td><td>a1</td><td>a2</td><td>a3</td><td>a4</td></tr>
<tr><td rowspan="2">y_0 endog</td><td>ω_{01} unrestricted</td><td rowspan="2">un2</td><td>c1</td><td>c2</td><td>c3</td><td>-</td></tr>
<tr><td>ω_{01} restricted</td><td>b1</td><td>b2</td><td>b3</td><td>b4</td></tr>
</table>

Our aim was to obtain Monte Carlo estimates of the biases for the parameters of the thirteen models given in Table 1 for each of the five sets of data. However, given the size of the problem (several of the likelihood functions to be maximised are highly nonlinear), the possibility of finding more efficient Monte Carlo estimates than the sample-mean method was investigated.

If the bias is denoted by $\theta = E(\hat{\delta} - \delta)$, its standard Monte-Carlo estimate based on H replications is given by $\hat{\theta} = (1/H) \sum_{j=1}^{H} (\hat{\delta}_j - \delta)$, where $\hat{\delta}_j$ is the estimate of δ obtained in the jth replication, and $\hat{\theta}$ is unbiased for θ. $\hat{\delta}_j$ depends on a particular set of (0,1) normal variates obtained from some pseudo-random numbers generator, i.e. $\hat{\delta}_j = \hat{\delta}(u_j)$. In the antithetic variate technique a second unbiased estimator θ^* is sought, having negative correlation with $\hat{\theta}$. Then $\bar{\theta} = \frac{1}{2}(\hat{\theta} + \theta^*)$ will also be an unbiased estimator of θ with variance $Var(\bar{\theta}) = \frac{1}{4} Var(\hat{\theta}) + \frac{1}{4} Var(\theta^*) + \frac{1}{2} Cov(\hat{\theta},\theta^*)$. If θ^* is a sample mean that has been constructed from a further set of random replications, then $Cov(\hat{\theta},\theta^*) = 0$. However, since u_j are standard normal variates so are $-u_j$ and, clearly, an estimator θ^* of the form

$$\theta^* = \frac{1}{H} \sum_{j=1}^{H} (\hat{\delta}(-u_j) - \delta)$$

will also be an unbiased estimator of θ. Now since u_j and $-u_j$ are perfectly negatively correlated, it can be the case that a negative covariance is induced between $\hat{\theta}$ and θ^*, so that $\bar{\theta}$ would have a smaller variance than the sample mean estimator based on an equal number of replications (cf. Hammersley and Handscomb (1964) and Hendry (1984)).

In previous studies it has been noticed the difficulty of finding antithetic transformations which reduce the variance of Monte Carlo estimators for dynamic models (cf. Hendry and Harrison (1974)). However, the simultaneous equations analogue provides a different perspective in the case of models from panel data and in this context it seemed worth to re-use the random numbers in pairs of opposite sign.

The situation can be illustrated by mean of a simple example. If we take T = 1, n = 1 and m = 0, our general model becomes

$$y_{io} = \mu_o x_{io} + \mu_1 x_{i1} + u_{io}$$

$$y_{i1} = \alpha y_{io} + \beta x_{i1} + u_{i1} \qquad (i=1,\ldots,N)$$

This model is exactly identified so that the QML estimator of $\delta' = (\alpha\ \beta)$ that leaves Ω^* unrestricted is identical to the 2SLS estimator, and it equals

$$\hat{\delta} = (X'W)^{-1} X'y_1$$

with $X = (x_o\ x_1)$ and $W = (y_o\ x_1)$. After some manipulation we have

$$\hat{\alpha} - \alpha = \frac{\xi' u_1}{\xi' u_o + c}$$

where ξ is the vector of least squares residuals from regressing x_o and x_1, i.e. $\xi = x_o - b_{o1} x_1$ with $b_{o1} = x_o' x_1 / x_1' x_1$, which remains constant over the replications, and so is c, that is given by $c = \xi' X \mu$. Now, notice that a trial of $\hat{\alpha}-\alpha$ based on $(-u_o, -u_1)$ yields $\xi' u_1/(\xi' u_o - c)$, and although $(\xi' u_o/\xi' u_1)$ is invariant to this transformation, a negative covariance is still generated between these antithetic pairs.

Thus the results reported in Table 2 were obtained from 20 replications corresponding to 10 antithetic pairs $(u_{it}, -u_{it})$, i.e. every trial was performed twice, and the resulting estimates were averaged. In all cases, the non-derivative Gill-Murray-Pitfield algorithm EO4JBF implemented in the Nag Library, Mark IX was used to optimise the log-likelihood functions.

First considering the consequences of misspecifying the initial conditions, i.e. wrongly assuming that $\omega_{01} = \ldots = \omega_{0T} = 0$, our results for the set of samples with white noise time-varying errors (data 5) and models un1 and a4 fairly generally agree with those reported by Bhargava and Sargan (1983). Indeed, these biases are rather small with the exceptions perhaps of the intercept; for example, the bias of α is .0396 for model un1 and .0263 for model a4. However, as one would expect, the consequences of treating y_0 as exogenous are rather more serious when the v_{it} are also serially correlated. To take an extreme case, for the ARMA(1,1) samples (data 1) and model a4 the biases of α and γ_0 are ten times larger than those obtained with the same model for data 5, but even if the ARMA(1,1) structure is properly specified and model a1 is used, these biases still are between 5 and 6 times bigger.

The cases where the endogeneity of y_0 is properly specified (both for models b and c) and no misspecifications are present in v_{it}, perform extremely well and the biases are almost negligible. Turning to the comparison between model un2 and models b and c, the Monte Carlo finite sample standard deviation of the estimates (which is just $\sqrt{20}$ times the standard errors of bias) are slightly lower for models b and c in the case of α and γ_0, and roughly the same for γ_1 and β; on the other hand, it does not appear to be any noticeable difference, both in terms of bias and standard deviation, between models b and c. These results suggest that in the QML framework un2 is a highly convenient method of estimation at the early stages of model building and that if we are interested in the structure of Ω^*, models c, that leave $\omega_{01}, \ldots, \omega_{0T}$ unrestricted, can achieve similar results to models b at a lower computational cost.

Data 4 (with $\alpha=0$ and $\phi=.35$) were generated to check the ability of our simulated model to distinguish systematic dynamics from serial correlation, and at least in this case, the results turned out to be extremely satisfactory. No doubt, this ability will depend on the characteristics of the process generating the time-varying exogenous variables.

Finally, we remark that models b_1 and c_1 (those allowing for ARMA(1,1) errors) are able to identify the correct serial correlation scheme in every case and therefore they are useful in order to choose between purely autoregressive and purely moving average schemes.

2.6 A Model with Arbitrary Heteroscedasticity over Time

If the presence of heteroscedasticity over time in the random effects model is suspected, equation (2.2.3) can be extended on the lines suggested in Section 1.5 by assuming:

(2.6.1) $$v_{it} = \sigma_t \, v^*_{it},$$

(2.6.2) $$v^*_{it} = \phi v^*_{i(t-1)} + \varepsilon^*_{it} + \lambda \varepsilon^*_{i(t-1)} \qquad (i=1,\ldots,N;\ t=1,\ldots,T)$$

TABLE 2

Biases in the Estimates[a]

		y_0 exogenous				
		unl	a1	a2	a3	a4
γ_0	D_1[b]	-.4892 (.0507)[e]	-.4800 (.0446)	-.5020 (.0361)	-.6047 (.0371)	-.8004 (.0271)
	D_2	-.2767 (.0471)	-.2710 (.0408)	-.2693 (.0402)	-.3163 (.0373)	-.4809 (.0313)
	D_3	-.2325 (.0458)	-.2186 (.0390)	-.2008 (.0385)	-.2033 (.0375)	-.4232 (.0304)
	D_4	-.1406 (.0325)	-.1382 (.0301)	-.1311 (.0305)	-.1746 (.0288)	-.3278 (.0214)
	D_5	-.1180 (.0382)	-.1131 (.0322)	-.1083 (.0313)	-.1096 (.0315)	-.0804 (.0296)
γ_1	D_1	-.0425 (.0091)	-.0405 (.0096)	-.0429 (.0090)	-.0528 (.0087)	-.0842 (.0069)
	D_2	-.0262 (.0094)	-.0254 (.0099)	-.0252 (.0098)	-.0303 (.0094)	-.0494 (.0082)
	D_3	-.0223 (.0094)	-.0206 (.0099)	-.0186 (.0100)	-.0190 (.0099)	-.0434 (.0084)
	D_4	-.0127 (.0095)	-.0125 (.0103)	-.0119 (.0103)	-.0165 (.0099)	-.0346 (.0086)
	D_5	-.0119 (.0093)	-.0112 (.0101)	-.0109 (.0100)	-.0111 (.0100)	-.0080 (.0102)
β	D_1	-.0032 (.0043)	-.0023 (.0042)	-.0061 (.0045)	-.0106 (.0048)	-.0730 (.0071)
	D_2	-.0143 (.0042)	-.0134 (.0041)	-.0132 (.0041)	-.0181 (.0044)	-.0400 (.0051)
	D_3	-.0133 (.0044)	-.0114 (.0043)	-.0095 (.0044)	-.0102 (.0044)	-.0351 (.0053)
	D_4	-.0097 (.0045)	-.0092 (.0043)	-.0089 (.0042)	-.0142 (.0044)	-.0398 (.0049)
	D_5	-.0104 (.0040)	-.0098 (.0041)	-.0090 (.0041)	-.0092 (.0041)	-.0064 (.0042)

TABLE 2 continued

		y_0 exogenous				
		unl	a1	a2	a3	a4
α	D_1	.1319 (.0110)	.1283 (.0095)	.1370 (.0075)	.1676 (.0079)	.2677 (.0068)
	D_2	.0846 (.0106)	.0820 (.0088)	.0814 (.0087)	.0976 (.0080)	.1580 (.0072)
	D_3	.0722 (.0102)	.0667 (.0082)	.0605 (.0080)	.0617 (.0077)	.1390 (.0069)
	D_4	.0861 (.0123)	.0840 (.0111)	.0800 (.0111)	.1100 (.0105)	.2263 (.0064)
	D_5	.0396 (.0081)	.0376 (.0066)	.0358 (.0063)	.0362 (.0064)	.0263 (.0057)
ρ	D_1		-.3397 (.0335)	-.5633 (.0207)	-.3509 (.0294)	-.5044 (.0134)
	D_2		-.1858 (.0295)	-.1874 (.0296)	-.1958 (.0286)	-.3091 (.0218)
	D_3		-.1600 (.0302)	-.1872 (.0283)	-.1558 (.0296)	-.3056 (.0213)
	D_4		-.0761 (.0253)	-.0520 (.0245)	-.0574 (.0242)	-.1954 (.0159)
	D_5		-.0596 (.0239)	-.0702 (.0260)	-.0707 (.0261)	-.0532 (.0246)
ϕ	D_1		-.1123 (.0105)	.2491 (.0056)		
	D_2		-.1054 (.0203)	-.0725 (.0068)		
	D_3		-.0809 (.0191)	d		
	D_4		.0159 (.0205)	-.0859 (.0114)		
	D_5		c	-.0317 (.0060)		
λ	D_1		.0079 (.0072)		.1460 (.0052)	
	D_2		.0324 (.0163)		d	
	D_3		.0275 (.0159)		-.0423 (.0048)	
	D_4		-.1021 (.0221)		d	
	D_5		c		-.0332 (.0065)	

TABLE 2 continued

		y_0 endogenous							
			ω_{00} and ω_{01} unrestr.			Ω^* fully restricted			
		un2	c1	c2	c3	b1	b2	b3	b4
γ_0	D_1	.0223 (.0695)	.0324 (.0627)	-.1246 (.0425)	-.3374 (.0395)	.0308 (.0619)	-.1160 (.0408)	-.3657 (.0377)	-.7170 (.0279)
	D_2	.0196 (.0527)	.0244 (.0454)	.0235 (.0455)	-.0988 (.0374)	.0164 (.0458)	.0152 (.0458)	-.1308 (.0370)	-.3802 (.0296)
	D_3	.0193 (.0510)	.0255 (.0434)	.0491 (.0418)	.0075 (.0378)	.0178 (.0433)	.0535 (.0416)	.0016 (.0379)	-.3221 (.0287)
	D_4	.0150 (.0378)	.0180 (.0356)	.0148 (.0332)	-.0464 (.0331)	.0149 (.0352)	.0125 (.0323)	-.0579 (.0330)	-.3098 (.0217)
	D_5	.0068 (.0390)	.0089 (.0321)	.0078 (.0309)	.0068 (.0308)	.0045 (.0330)	-.0014 (.0310)	-.0026 (.0310)	.0012 (.0284)
γ_1	D_1	.0011 (.0114)	.0028 (.0123)	-.0101 (.0109)	-.0287 (.0099)	.0026 (.0123)	-.0096 (.0109)	-.0313 (.0096)	-.0751 (.0073)
	D_2	.0023 (.0109)	.0025 (.0115)	.0025 (.0115)	-.0090 (.0106)	.0017 (.0115)	.0017 (.0115)	-.0122 (.0103)	-.0389 (.0087)
	D_3	.0023 (.0108)	.0028 (.0113)	.0049 (.0114)	.0012 (.0110)	.0019 (.0113)	.0052 (.0114)	.0005 (.0110)	-.0329 (.0089)
	D_4	.0019 (.0104)	.0019 (.0112)	.0016 (.0111)	-.0040 (.0108)	.0017 (.0112)	.0013 (.0111)	-.0050 (.0108)	-.0327 (.0087)
	D_5	.0012 (.0102)	.0012 (.0107)	.0012 (.0108)	.0012 (.0108)	.0005 (.0108)	.0001 (.0107)	.0000 (.0107)	.0004 (.0106)
β	D_1	-.0004 (.0049)	-.0004 (.0047)	-.0000 (.0046)	-.0037 (.0048)	-.0004 (.0047)	-.0002 (.0047)	-.0042 (.0049)	-.0649 (.0069)
	D_2	.0005 (.0042)	.0008 (.0041)	.0008 (.0041)	-.0050 (.0043)	.0004 (.0041)	.0004 (.0041)	-.0067 (.0044)	-.0310 (.0050)
	D_3	.0004 (.0044)	.0009 (.0042)	.0018 (.0044)	.0003 (.0044)	.0005 (.0042)	.0019 (.0044)	-.0000 (.0044)	-.0262 (.0052)
	D_4	.0002 (.0046)	.0005 (.0043)	.0004 (.0044)	-.0033 (.0044)	.0003 (.0043)	.0004 (.0044)	-.0041 (.0044)	-.0376 (.0049)
	D_5	.0002 (.0039)	.0005 (.0041)	.0006 (.0041)	.0005 (.0041)	-.0000 (.0040)	-.0003 (.0041)	-.0004 (.0041)	.0000 (.0043)

TABLE 2 continued

		y_0 endogenous							
			ω_{00} and ω_{01} unrestr.			Ω* fully restricted			
		un2	c1	c2	c3	b1	b2	b3	b4
α	D_1	-.0050	-.0079	.0329	.0919	-.0074	.0309	.0997	.2395
		(.0144)	(.0129)	(.0083)	(.0073)	(.0126)	(.0076)	(.0069)	(.0059)
	D_2	-.0051	-.0069	-.0066	.0301	-.0044	-.0041	.0399	.1244
		(.0109)	(.0090)	(.0091)	(.0069)	(.0091)	(.0092)	(.0069)	(.0057)
	D_3	-.0050	-.0073	-.0142	-.0020	-.0049	-.0154	-.0002	.1054
		(.0104)	(.0085)	(.0081)	(.0068)	(.0084)	(.0079)	(.0069)	(.0054)
	D_4	-.0079	-.0098	-.0081	.0287	-.0080	-.0069	.0357	.2139
		(.0140)	(.0128)	(.0115)	(.0115)	(.0126)	(.0110)	(.0117)	(.0064)
	D_5	-.0015	-.0026	-.0023	-.0020	-.0010	.0008	.0011	-.0002
		(.0077)	(.0059)	(.0056)	(.0056)	(.0062)	(.0056)	(.0056)	(.0050)
ρ	D_1		.0503	-.4331	-.1040	.0727	-.3095	-.1565	-.4953
			(.0540)	(.0334)	(.0357)	(.0498)	(.0260)	(.0296)	(.0132)
	D_2		.0567	.0612	.0129	.0424	.0455	-.0377	-.2478
			(.0365)	(.0373)	(.0321)	(.0342)	(.0349)	(.0289)	(.0211)
	D_3		.0563	.0175	.0453	.0457	.0285	.0337	-.2391
			(.0363)	(.0332)	(.0328)	(.0338)	(.0314)	(.0298)	(.0202)
	D_4		.0496	.0473	.0693	.0387	.0399	.0491	-.1859
			(.0311)	(.0283)	(.0299)	(.0291)	(.0267)	(.0270)	(.0162)
	D_5		-.0284	.0425	.0420	-.0144	.0241	.0235	.0253
			(.0478)	(.0275)	(.0274)	(.0430)	(.0272)	(.0271)	(.0253)
ϕ	D_1		.0204	.3209		.0089	.2935		
			(.0155)	(.0075)		(.0161)	(.0070)		
	D_2		.0138	.0039		.0134	.0036		
			(.0176)	(.0100)		(.0174)	(.0102)		
	D_3		.0345	d		.0218	d		
			(.0210)			(.0233)			
	D_4		.0120	.0051		.0216	.0109		
			(.0145)	(.0128)		(.0160)	(.0133)		
	D_5		c	-.0040		c	-.0037		
				(.0065)			(.0066)		
λ	D_1		-.0103		.1724	.0061		.1722	
			(.0069)		(.0052)	(.0062)		(.0050)	
	D_2		-.0093		d	-.0097		d	
			(.0117)			(.0108			
	D_3		-.0306		-.0057	-.0189		-.0046	
			(.0155)		(.0054)	(.0172)		(.0050)	
	D_4		-.0041		d	-.0085		d	
			(.0174)			(.0190)			
	D_5		c		-.0048	c		-.0046	
					(.0064)			(.0065)	

NOTES TO TABLE 2

[a] N = 100, T = 9, 20 replications (10 pairs of antithetic variates).

[b] D_i = Data i.

[c] The ARMA(1,1) process degenerates into a white noise for any $\phi = -\lambda$. Therefore, if the process generating v_{it} is white noise (as in D_5) ϕ and λ are not identified for models a1, b1 and c1. For our 20 replications the results turned out to be the following

	Model a1	Model b1	Model c1
Converged to $\phi=\lambda=0$	11	13	16
Converged to $\phi=1$, $\lambda=-1$	3	3	3
Converged to $\phi=-1$, $\lambda=1$	1	0	1
Converged to other antithetic pairs	5	4	0

[d] When the true v_{it}'s are autoregressive (moving average) and the estimated model only allows for a moving average (autoregressive) scheme, the MA (AR) coefficient picks up the effect of the serial correlation, so that it cannot be regarded as an estimate of its (zero) true value.

[e] Standard errors of bias are in parentheses.

with $\varepsilon_{it} \sim$ iid (O,1). This reduces to the former case if $\sigma_t = \sigma$ for all t and we noted that with this formulation the serial correlation pattern remains stationary. The covariance matrix of u_i now becomes

$$(2.6.3) \qquad \Omega = S\ V\ S + \sigma_\eta^2\ \iota\iota'$$

where $S = \text{diag}\{\sigma_t\}$. Nevertheless in setting up the likelihood function it is convenient to parameterise Ω as

$$(2.6.4) \qquad \Omega = \sigma_\eta^2 (R^{-1}\ VR^{-1} + \iota\iota') = \sigma_\eta^2\ \bar{\Omega}$$

where $R = \text{diag}\{\rho_t\}$ and $\rho_t = (\sigma_\eta/\sigma_t)$.

As a consequence of the nonstationarity of the variance, the specification of models of type b is now more complicated. If, as in (2.2.7), we take

$$u_{iO} = \zeta_i + \frac{\eta_i}{1-\alpha} + \sum_{k=O}^{\infty} \alpha^k\ v_{i(-k)}$$

The covariances $E(u_{iO}\ u_{it})$ are given by

$$(2.6.5) \qquad \omega_{Ot} = \frac{\sigma_\eta^2}{1-\alpha} + \sigma_t\ \phi^{t-1}\ \frac{(1+\phi\lambda)(\phi+\lambda)}{(1-\phi^2)} \sum_{k=O}^{\infty} \sigma_{(-k)}\ (\alpha\ \phi)^k \qquad (t=1,\ldots,T)$$

Thus the terms ω_{Ot} would depend on the infinite series $\sigma_O, \sigma_{(-1)}, \ldots$ and cannot, in general, be estimated.

This is not the case when v^*_{it} follows a pure moving average process since then we have $\omega_{01} = \sigma^2_\eta/(1-\alpha) + \sigma_0\sigma_1\lambda$ and $\omega_{0t} = \sigma^2_\eta/(1-\alpha)$ for $t=2,\ldots,T$. In the general case, we can still assume that the infinite summation $\sum_{k=0}^{\infty} \sigma_{(-k)} \, (\alpha \, \phi)^k$ converges, and then treat it as a further parameter to be estimated. However, it seems reasonable to concentrate our attention to models of type c in defining likelihood methods for heteroscedastic models. Thus it is a matter of imposing the restrictions derived from (2.6.4) in

$$(2.6.6) \qquad L = -\tfrac{1}{2} N \log\det \Omega - \tfrac{1}{2} \operatorname{tr}(\Omega^{-1} U'U) + \frac{N}{2} \log\det(B^{+} W B^{+\prime}) - \frac{N}{2} \log\det(W)$$

(2.6.6) can be concentrated with respect to σ^2_η yielding

$$(2.6.7) \qquad L_e = -\tfrac{1}{2} N \log\det \bar{\Omega} - \tfrac{1}{2} NT \log s^2_\eta + \tfrac{1}{2} N \log\det(B^{+} W B^{+\prime}) - \frac{N}{2} \log\det(W)$$

where the QML estimator of σ^2_η is

$$(2.6.8) \qquad s^2_\eta = \frac{1}{NT} \operatorname{tr}(\bar{\Omega}^{-1} U'U).$$

The determinant and the inverse of $\bar{\Omega}$ are given by

$$(2.6.9) \qquad \det(\bar{\Omega}) = \{ \prod_{t=1}^{T} \rho_t^{-2} \} \det(V) \; [1 + \iota'(R \, V^{-1} R)\iota]$$

(2.6.10) $$\bar{\bar{\Omega}}^{-1} = R\, V^{-1}\, R - \frac{(RV^{-1}R)\,\iota\iota'(RV^{-1}R)}{1 + \iota'(RV^{-1}R)\iota}$$

L_e can be maximised numerically, as a function of δ, ϕ, λ, $\rho_1, \ldots, \rho_T$, with the restrictions that $|\phi| < 1$ and $|\lambda| < 1$.

NOTES

1 We make use of the following formulae:

$$\det(\Omega^*) = (\omega_{00} - \omega_{01}\ \Omega^{-1}\ \omega_{10})\ \det(\Omega) = \frac{1}{\omega^{00}} \det\ (\Omega)$$

and

$$\Omega^{-1} = \Omega^{11} - \frac{1}{\omega^{00}}\ \omega^{10}\ \omega^{01}$$

the latter being solved for Ω^{11}.

2 Remark that since an alternative 'structural' equation for y_{i0} is not available, LIML and FIML methods are equivalent here. The use of the term 'limited information' in this context simply reflects the technical fact that the likelihood function has been concentrated with respect to the coefficients in the equation for y_{i0}.

APPENDIX 2.A

The Determinant and the Inverse of the Covariance Matrix for the Stationary ARMA(1,1) Process

The process for v_{it} is given by

$$(2.A.1) \qquad v_{it} - \phi\, v_{i(t-1)} = \varepsilon_{it} + \lambda\varepsilon_{i(t-1)} \qquad (t=1,\ldots,T)$$

with

$$|\phi| < 1, \lambda \neq -\phi \text{ and } \varepsilon_{it} \sim \text{iid } (0, \sigma^2)$$

Now following Tiao and Ali (1971) we re-write (2.A.1) as:

$$(2.A.2) \qquad B(\phi)\, v_i = B(-\lambda)\, \varepsilon_i + \xi_i d_1$$

where $B(\theta)$ is an already familiar T x T matrix function such that

$$\begin{aligned} b_{ij}(\theta) &= 1 \quad \text{for } i=j, \\ &= -\theta \quad \text{for } i=j+1, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Also, $v_i' = (v_{i1},\ldots,v_{iT})$, $\varepsilon_i' = (\varepsilon_{i1},\ldots,\varepsilon_{iT})$, d_1 is the T-vector $d_1' = (1\ 0,\ldots,0)$, and $\xi_i = \phi\, v_{i0} + \lambda\, \varepsilon_{i0}$.

Furthermore, note that

$$B^{-1}(\theta) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \theta & 1 & \cdots & 0 \\ \theta^2 & \theta & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \theta^{T-1} & \theta^{T-2} & & 1 \end{pmatrix}$$

and that $\det B(\theta) = \det B^{-1}(\theta) = 1$.

Since we have assumed stationarity, ξ_i is distributed independently of ε_i with $E(\xi_i) = 0$ and

(2.A.3) $$E(\xi_i^2) = \sigma^2 \frac{(\phi+\lambda)^2}{1-\phi^2} = \sigma^2 \kappa \text{ , say.}$$

Now the covariance of (2.A.2) is

$$B(\phi)\ E(v_i v_i')\ B'(\phi) = \sigma^2 (B(-\lambda)\ B'(-\lambda) + \kappa\ d_1 d_1')$$

Hence

$$V = \frac{1}{\sigma^2} E(v_i v_i') = B^{-1}(\phi)\ [B(-\lambda)\ B'(-\lambda) + \kappa\ d_1 d_1']\ B'^{-1}(\phi)$$

or

(2.A.4) $$V = B^{-1}(\phi)\ B(-\lambda)\ (I + \kappa\ qq')\ B'(-\lambda)\ B'^{-1}(\phi)$$

where q holds the first column of $B^{-1}(-\lambda)$:

$$q = B^{-1}(-\lambda)d_1 = \begin{pmatrix} 1 \\ -\lambda \\ \vdots \\ (-\lambda)^{T-1} \end{pmatrix}$$

Therefore

$$\text{(2.A.5)} \qquad V^{-1} = B'^{-1}(-\lambda)\ B'(\phi)\ (I - p\ qq')\ B^{-1}(\lambda)\ B(\phi)$$

with

$$p = \frac{\kappa}{1 + \kappa\ q'q}$$

and also

$$\det(V) = \det(I + \kappa\ qq') = 1 + \kappa\ q'q$$

but since $q'q = 1 + \lambda^2 + \lambda^4 + \ldots + \lambda^{2(T-1)} = \dfrac{1-\lambda^{2T}}{1-\lambda^2}$

we then have

$$\text{(2.A.6)} \qquad \det(V) = 1 + \frac{(\phi+\lambda)^2}{(1-\phi^2)} \left(\frac{1-\lambda^{2T}}{1-\lambda^2} \right)$$

(2.A.5) and (2.A.6) are computationally convenient expressions for the inverse and the determinant of V, which can be used in the evaluation of the relevant likelihood functions. Explicit expressions of the elements of V^{-1} can be found in Tiao and Ali (1971).

CHAPTER 3

THE ASYMPTOTIC PROPERTIES OF QML ESTIMATORS IN A TRIANGULAR MODEL WITH COVARIANCE RESTRICTIONS

3.1 Introduction

We have seen that the structure of dynamic models from panel data is that of a simultaneous equations system of (T+1) equations where the matrix of coefficients of the endogenous variables has a triangular structure and the error covariance matrix is constrained. Generalising the problem, this Chapter examines the asymptotic properties of quasi-maximum likelihood estimators of triangular systems with general restrictions in both the slope and the covariance coefficients. However, we do not treat the case where there are restrictions relating the slope coefficients to those in the covariance matrix, so that the results presented below will primarily apply to random effects models with unrestricted covariances between initial observations errors and the remaining errors. In this context, it is natural to assume all predetermined variables to be exogenous and therefore standard central limit theorems for independent observations can be applied.

Normality is not imposed but we assume that the error vector is generated by a distribution where the third order moments vanish and the fourth order moments are finite. Indeed we are mainly concerned with the role of non-normal kurtosis when estimating models with covariance restrictions by quasi-maximum likelihood methods. This role is more relevant, both for slope and covariance parameters estimates, when covariance restrictions are enforced than when the covariance matrix is left unrestricted. Distributions with long tails are common in practice due to the presence of extreme observations in the sample, and they lead to large fourth moments relative to the variances and covariances.

The structure of this Chapter is as follows. Section 3.2 states the assumptions concerning the model and the quasi-log-likelihood function, and derives limiting matrices of second partial deviatives and products of first partials of the log-likelihood function when the covariance matrix is not restricted and also when restrictions are present. These results are extensively used below in deriving useful expressions for the asymptotic variance matrix of the QML estimates when normality holds (Section 3.3) and when the errors are possibly non-normal (Section 3.4). It turns out that in the latter case imposing the covariance restrictions may lead to an efficiency loss relative to the estimators that leave the covariance matrix unrestricted. We give an explicit condition on the fourth order moments to characterise this situation. Finally, Section 3.5 discusses a simple two equation model in order to illustrate our general results.

3.2 The Model and the Limiting Distribution of the QML Estimator

y_i is a n x 1 vector of dependent variables, and z_i is a k x 1 vector of nonstochastic exogenous variables. We assume that the y_i are explained by

$$(3.2.1) \qquad B(\bar{\theta})\, y_i + C(\bar{\theta}) z_i = A(\bar{\theta}) x_i = \bar{u}_i \qquad (i=1,\ldots,N)$$

where the elements of the $n \times (n+k)$ coefficient matrix $A(\theta) = (B(\theta) \vdots C(\theta))$ are continuous functions of a $p \times 1$ vector of parameters θ and $x_i' = (y_i' \; z_i')$. The $\bar{u}_i$ are $n \times 1$ vectors of independent and identically distributed random errors with finite moments up to the fourth order, such that

$$E(\bar{u}_i) = 0 \; ,$$

$$(3.2.2) \qquad E(\bar{u}_i \bar{u}_i') = \Omega(\bar{\tau}) \; .$$

The covariance matrix $\Omega(\bar{\tau})$ is assumed to be non-singular and its elements are continuous functions of a $q \times 1$ vector of parameters τ.

Before proceeding further, some conventions and notation must be introduced. For any nxn symmetric matrix $A = \{a_{ij}\}$, let $\nu(A)$ be the $\frac{1}{2} n(n+1)$ vector of distinct elements of A

$$[\nu(A)]' = (a_{11}, a_{21}, a_{22}, a_{31}, a_{32}, a_{33}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nn}) .$$

$\nu(A)$ and vec(A) can be connected defining a 0-1 matrix D, say, of order $n^2 \times \frac{1}{2}n(n+1)$ that maps $\nu(A)$ into vec(A), i.e. $D\, \nu(A) = \mathrm{vec}(A)$ and $D = \partial\, \mathrm{vec}(A)/\partial[\nu(A)]'$. Furthermore, since $(D'D)$ is not singular, we also have $\nu(A) = (D'D)^{-1} D' \mathrm{vec}(A) = D^{+} \mathrm{vec}(A)$.[1]

Now let Δ_3 be a $n \times \frac{1}{2} n(n+1)$ matrix of third order moments μ_{hjk} $\Delta_3 = E\{\bar{u}_i [\nu(\bar{u}_i \bar{u}_i')]'\}$, and let Δ_4 be a $\frac{1}{2} n(n+1) \times \frac{1}{2} n(n+1)$ matrix of fourth order moments $\mu_{hjk\ell}$ defined as $\Delta_4 = E\{\nu(\bar{u}_i \bar{u}_i')[\nu(\bar{u}_i \bar{u}_i')]'\}$. Under normality $\mu_{hjk} = 0$ $(h,j,k = 1,\ldots,n)$ and

(3.2.3) $\mu_{hjk\ell} = \bar{\omega}_{hj}\,\bar{\omega}_{k\ell} + \bar{\omega}_{hk}\,\bar{\omega}_{j\ell} + \bar{\omega}_{h\ell}\,\bar{\omega}_{jk}$ $(h,j,k,\ell = 1,\ldots,n)$

where $\bar{\omega}_{jk}$ is the (j,k) th element of $\Omega(\bar{\tau})$. Here we assume that $\Delta_3 = 0$, as in the normal case but Δ_4 is left unrestricted so that its elements do not necessarily satisfy (3.2.3). It is worth pointing out that in the present context assuming that the third order moments are zero is not too restrictive. In particular, as the proofs below will make it clear, the results in this Chapter are unaltered if we assume Δ_3 to be unrestricted and the exogenous variables to be in mean deviation form, i.e. $\lim_{N\to\infty}(1/N)\sum_{i=1}^{N} z_i = 0$.

We assume that $B(\bar{\theta})$ is a lower triangular and nonsingular nxn matrix and has the usual standarising restrictions (i.e. the diagonal elements are equal to -1), and we also assume that the kxk matrix

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} z_i z_i' = M$$

exists and is non-singular.

As a simplified notation we shall use $B(\bar{\theta}) = \bar{B}$, $C(\bar{\theta}) = \bar{C}$, $A(\bar{\theta}) = \bar{A}$ and $\Omega(\bar{\tau}) = \bar{\Omega}$ when referring to these matrices evaluated at the true values of θ and τ. The model is identified if

$$-\bar{B}^{-1}\,\bar{C} = -\,B^{-1}(\theta^*)\;C(\theta^*)$$

and

$$\bar{B}^{-1}\;\bar{\Omega}\;\bar{B}'^{-1} = B^{-1}(\theta^*)\;\Omega(\tau^*)\;B'^{-1}(\theta^*)$$

for some θ^* and τ^* in the parameter space, implies that $\theta^* = \bar{\theta}$ and $\tau^* = \bar{\tau}$. However, since we are also interested in considering the quasi-maximum likelihood estimator of $\bar{\theta}$ without imposing the restrictions in the covariance matrix, we assume that the model is identified by mean of the prior restrictions implicit in the matrix $A(\theta)$ alone.

Let L_u be the gaussian log-likelihood function that leaves Ω unrestricted, $L_u = \sum_{i=1}^{N} L_{u,i}$ with

$$(3.2.4) \qquad L_{u,i}(\theta,\omega) = k_o - \tfrac{1}{2} \log \det \Omega - \tfrac{1}{2} x_i' A'(\theta) \; \Omega^{-1} \; A(\theta) \; x_i$$

and let L_r be of the same form as L_u but in this case some set of constraints in Ω are enforced so that

$$(3.2.5) \qquad L_r = L_r(\theta,\tau) = L_u(\theta,\omega(\tau))$$

where $\omega(\tau) = \nu[\Omega(\tau)]$, and accordingly we set $\bar{\omega} = \nu(\bar{\Omega})$.

The first order conditions for the estimators of $\bar{\theta}$ and $\bar{\tau}$ that maximise L_r are given by

$$(3.2.6) \qquad R'(\theta)(\Omega^{-1}(\tau) \otimes X'X) \text{ vec } A(\theta) = 0$$

$$(3.2.7) \qquad G'(\tau) \; D'(\Omega^{-1}(\tau) \otimes \Omega^{-1}(\tau)) \text{ vec}[A(\theta) \; X'X \; A'(\theta) - \Omega(\tau)] = 0$$

where $R(\theta) = \dfrac{\partial \text{ vec } A(\theta)}{\partial \theta'}$, $G(\tau) = \dfrac{\partial \; \omega(\tau)}{\partial \tau'}$ and $X'X = \sum_{i=1}^{N} x_i x_i'$

(see Appendix 3.A). Let $\hat{\theta}_r$ and $\hat{\tau}$ be the QML estimators. These solve

(3.2.6) and (3.2.7) so that the two sets of equations are simultaneously reconciled. Moreover, let $\hat{\theta}_u$ be the QML estimate of $\bar{\theta}$ that leaves Ω unrestricted, and let $\hat{\omega} = \nu(\hat{\Omega})$ be the corresponding unrestricted estimate of $\bar{\omega}$.

If the restrictions in $\bar{A}$ are linear, so that $\mathrm{vec}(\bar{A}) = R\bar{\theta} - r$, and R and r are, respectively, a matrix and a vector of known constants, we may re-write (3.2.6) to have

$$\text{(3.2.8)} \qquad R'(\Omega^{-1}(\hat{\tau}) \otimes X'X)\, R\, \hat{\theta}_r = R'(\Omega^{-1}(\hat{\tau}) \otimes X'X) r$$

Therefore, the QML estimator of $\bar{\theta}$, $\hat{\theta}_r$, is the G.L.S. estimator that uses $\Omega(\hat{\tau})$ as the estimator of $\bar{\Omega}$. The basic discussion of GLS estimators in triangular systems without covariance restrictions is due to Lahiri and Schmidt (1978). In Chapter 5 we will consider estimators of the form of (3.2.8) but computed in two stages as a typical GLS estimator.

<u>The matrices Φ_u, Θ_u, Φ_r and Θ_r</u>

Now we define the matrices

$$\text{(3.2.9)} \qquad \Phi_u = \operatorname*{plim}_{N\to\infty} \left[-\frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 L_{u,i}}{\partial \psi_u\, \partial \psi_u'} \Big|\, \bar{\psi}_u \right] = \begin{pmatrix} \Phi_{u,11} & \Phi_{u,12} \\ \Phi_{u,21} & \Phi_{u,22} \end{pmatrix}$$

$$\text{(3.2.10)} \qquad \Theta_u = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{\partial L_{u,i}}{\partial \psi_u}\Big|\bar{\psi}_u \cdot \frac{\partial L_{u,i}}{\partial \psi_u'}\Big|\bar{\psi}_u \right] = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \Theta_{ui}$$

$$= \begin{bmatrix} \Theta_{u,11} & \Theta_{u,12} \\ \Theta_{u,21} & \Theta_{u,22} \end{bmatrix}$$

where $\bar{\psi}_u' = (\bar{\theta}' \; \bar{\omega}')$, a $p + \frac{1}{2}n(n+1)$ vector, and the partitions in Φ_u and Θ_u correspond to that of $\bar{\psi}_u$. Similarly, we may define Φ_r and Θ_r in the same way, but now $L_{u,i}$ is replaced by $L_{r,i}$ and $\bar{\psi}_u$ by the (p+q) vector $\bar{\psi}_r' = (\bar{\theta}' \; \bar{\tau}')$. Model (3.2.1) can also be written in the form

$$(3.2.11) \quad x_i = P^* z_i + \tilde{B} \bar{u}_i \qquad (i=1,\ldots,N)$$

where P^* and $\tilde{B}$ are (n+k) x k and (n+k) x n matrices, respectively, given by

$$P^* = \begin{pmatrix} -\bar{B}^{-1} \bar{C} \\ I_k \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} \bar{B}^{-1} \\ 0 \end{pmatrix}$$

Furthermore, for simplicity let $R = R(\bar{\theta})$ and $G = G(\bar{\tau})$ when referring to the matrices of partial derivatives of the coefficients evaluated at the true values. Then we have the following results

$$(3.2.12) \quad \Phi_{u,11} = R'(\bar{\Omega}^{-1} \otimes P^* M P^{*\prime})R + R'(\bar{\Omega}^{-1} \otimes \tilde{B} \bar{\Omega} \tilde{B}')R,$$

$$(3.2.13) \quad \Phi_{u,12} = - R'(\bar{\Omega}^{-1} \otimes \tilde{B})D,$$

$$(3.2.14) \quad \Phi_{u,22} = \tfrac{1}{2} D'(\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1})D.$$

We also have

$$(3.2.15) \quad \Phi_{r,11} = \Phi_{u,11} ,$$

$$(3.2.16) \quad \Phi_{r,12} = \Phi_{u,12} G ,$$

(3.2.17) $\Phi_{r,22} = G' \Phi_{u,22} G.$

With respect to the limiting matrices of products of first partial derivatives we have

(3.2.18) $\Theta_{u,11} = \Phi_{u,11} + \Phi_{u,12}[(\Delta_4 - \bar{\omega}\bar{\omega}') - \Phi^{-1}_{u,22}]\ \Phi_{u,21}$,

(3.2.19) $\Theta_{u,12} = \Phi_{u,12}\ (\Delta_4 - \bar{\omega}\bar{\omega}')\ \Phi_{u,22}$,

(3.2.20) $\Theta_{u,22} = \Phi_{u,22}\ (\Delta_4 - \bar{\omega}\bar{\omega}')\ \Phi_{u,22}$.

Note that for a triangular model $\Phi_{u,12}\ \bar{\omega} \equiv 0$,[2] so that (3.2.18) and (3.2.19) can be simplified further. Nevertheless, a well known result in matrix algebra is

(3.2.21) $\Phi^{-1}_{u,22} = 2\ D^{+}(\bar{\Omega} \otimes \bar{\Omega})D^{+\prime}$

(cf. Richard (1975) and Magnus and Neudecker (1980)). It can also be checked that if the $\mu_{hjk\ell}$ are as in (3.2.3) then

$$\Delta_4 = \bar{\omega}\bar{\omega}' + 2D^{+}(\bar{\Omega} \otimes \bar{\Omega})D^{+\prime}$$

so that under normality $\Delta_4 - \bar{\omega}\bar{\omega}' = \Phi^{-1}_{u,22}$, and the formulae above clearly show that in this case $\Theta_u = \Phi_u$. Moreover, since

(3.2.22) $\Theta_{r,11} = \Theta_{u,11}$

(3.2.23) $\Theta_{r,12} = \Theta_{u,12}\, G$

(3.2.24) $\Theta_{r,22} = G'\, \Theta_{u,22}\, G$

also $\Theta_r = \Phi_r$ under normality.

The proofs of (3.2.12) to (3.2.20) and (3.2.22) to (3.2.24) are given in Appendix 3.A.

Having assumed that our model is identified, by using the arguments in Sargan (1975), it follows that the quasi-maximum likelihood estimators of $\bar{\theta}$ and $\bar{\tau}$ obtained by maximising $L_r(\theta,\tau)$ are consistent even when the $\bar{u}_i$ are not normally distributed.

Now since $(\partial L_r/\partial\psi_r)|\hat{\psi}_r = 0$ by the definition of $\hat{\psi}_r$, an exact first order Taylor series expansion of $\partial L_r/\partial\psi_r$ about $\bar{\psi}_r$ yields

(3.2.25) $$-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 L_{r,i}}{\partial\psi_r\,\partial\psi_r'}\Big|\psi_r^*\ \sqrt{N}(\hat{\psi}_r-\bar{\psi}_r) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial L_{r,i}}{\partial\psi_r}\Big|\bar{\psi}_r$$

where ψ_r^* lies between $\hat{\psi}_r$ and $\bar{\psi}_r$.

Since $\partial L_{r,i}/\partial\psi_r|\bar{\psi}_r$ are independently distributed random vectors with zero mean and covariance matrices $\Theta_{r,i}$, given our assumptions, standard (Liapunov) central limit theorem (e.g. see Rao (1973), p.147) ensures that

(3.2.26) $$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial L_{r,i}}{\partial\psi_r}\Big|\bar{\psi}_r \xrightarrow{d} N(0,\ \Theta_r).$$

Finally, if we note that since $\hat{\psi}_r$ is consistent for $\bar{\psi}_r$ so is ψ^*_r and thus $\mathrm{plim}_{N\to\infty}\{-\ (1/N)\ \sum_{i=1}^{N} (\partial^2 L_{r,i}/\partial\psi_r \partial\psi_r')|_{\psi^*_r}\} = \Phi_r$, using the Cramer linear transformation theorem we have

$$(3.2.27)\quad \sqrt{N}(\hat{\psi}_r - \bar{\psi}_r) \overset{d}{\to} N(0,\ W_r)$$

with

$$(3.2.28)\quad W_r = \Phi_r^{-1}\ \Theta_r\ \Phi_r^{-1}.$$

Clearly, the same is true for $\hat{\psi}_u = (\hat{\theta}_u\ \hat{\omega})'$ as it can be regarded as a specialisation of the previous result to $\omega(\bar{\tau}) = \bar{\tau}$. So

$$(3.2.29)\quad \sqrt{N}(\hat{\psi}_u - \bar{\psi}_u) \overset{d}{\to} N(0,\ W_u)$$

with

$$(3.2.30)\quad W_u = \Phi_u^{-1}\ \Theta_u\ \Phi_u^{-1}.$$

3.3 The Asymptotic Variance Matrix of the QML Estimators when the Errors are Normal

Under normality, in view of the equivalence between Φ_r and Θ_r (they are, respectively, the hessian and outer product forms for the information matrix), the asymptotic variance matrix (AVM) of $\hat{\psi}_r$ reduces to $W_r = \Phi_r^{-1}$. In what follows we report some results concerning the partitions of Φ_u^{-1} and Φ_r^{-1}. Letting

$$\Phi_u^{-1} = \begin{bmatrix} \Phi_u^{11} & \Phi_u^{12} \\ \Phi_u^{21} & \Phi_u^{22} \end{bmatrix} , \qquad \Phi_r^{-1} = \begin{bmatrix} \Phi_r^{11} & \Phi_r^{12} \\ \Phi_r^{21} & \Phi_r^{22} \end{bmatrix}$$

we then have

$$(3.3.1) \qquad \Phi_u^{11} = [R'(\bar{\Omega}^{-1} \otimes P^*MP^{*\prime})R]^{-1} ,$$

$$(3.3.2) \qquad \Phi_u^{12} = - \Phi_u^{11}(\Phi_{u,12}\Phi_{u,22}^{-1}) ,$$

$$(3.3.3) \qquad \Phi_u^{22} = \Phi_{u,22}^{-1} + (\Phi_{u,22}^{-1} \Phi_{u,21}) \Phi_u^{11}(\Phi_{u,12} \Phi_{u,22}^{-1})$$

where

$$(3.3.4) \qquad \Phi_{u,22}^{-1} \Phi_{u,21} = - 2 D^+(I \otimes \bar{\Omega} \tilde{B}')R$$

$$= - 2 D^+(I \otimes \bar{\Omega} \bar{B}^{-1\prime})(\frac{\partial \text{ vec } B(\theta)}{\partial \theta'}\Big|_{\bar{\theta}})$$

and $\Phi_{u,22}^{-1}$ is given in (3.2.21). (3.3.2) and (3.3.3) are direct application of partitioned inverse formulae. The proofs of (3.3.1) and (3.3.4) are given in Appendix 3.A.

Turning to the restricted case, we state the results concerning the inverse of Φ_r in the form of a general lemma on partitioned inverses as it will prove useful in a different context below.

Lemma

Let A_u be a nonsingular matrix whose inverse is given by

$$A_u^{-1} = \begin{pmatrix} A_{u,11} & A_{u,12} \\ A_{u,21} & A_{u,22} \end{pmatrix}^{-1} = \begin{pmatrix} A_u^{11} & A_u^{12} \\ A_u^{21} & A_u^{22} \end{pmatrix}$$

where $A_{u,11}$ and $A_{u,22}$ are, respectively, p x p and s x s nonsingular symmetric submatrices, and let A_r be defined by

$$(3.3.5) \qquad A_r = \begin{pmatrix} A_{u,11} & A_{u,12}H \\ H'A_{u,21} & H'A_{u,22}H \end{pmatrix}$$

where H is a s x q matrix of rank q (so that $q \leq s$). Then, letting

$$A_r^{-1} = \begin{pmatrix} A_r^{11} & A_r^{12} \\ A_r^{21} & A_r^{22} \end{pmatrix} ,$$

we have

$$(3.3.6) \qquad A_r^{11} = A_u^{11} - A_{u,11}^{-1} A_{u,12}\{A_u^{22} - H[H'(A_u^{22})^{-1}H]^{-1}H'\}A_{u,21} A_{u,11}^{-1} ,$$

$$(3.3.7) \qquad A_r^{22} = [H'(A_u^{22})^{-1}H]^{-1} .$$

Proof:

From the formulae for partitioned inverses we know

$$A_u^{11} = A_{u,11}^{-1} + A_{u,11}^{-1} A_{u,12} A_u^{22} A_{u,21} A_{u,11}^{-1} ,$$

$$A_r^{11} = A_{r,11}^{-1} + A_{r,11}^{-1} A_{r,12} A_r^{22} A_{r,21} A_{r,11}^{-1}$$

but since $A_{r,11} = A_{u,11}$ and $A_{r,12} = A_{u,12}H$ we have

$$A_r^{11} = A_{u,11}^{-1} + A_{u,11}^{-1} A_{u,12}(H A_r^{22} H') A_{u,21} A_{u,11}^{-1}.$$

Now subtracting both expressions

$$A_r^{11} - A_u^{11} = A_{u,11}^{-1} A_{u,12}[H A_r^{22} H' - A_u^{22}] A_{u,21} A_{u,11}^{-1} ,$$

thus to complete the proof of (3.3.6) it only remains to prove (3.3.7). (3.3.7) is equivalent to $(A_r^{22})^{-1} = H'(A_u^{22})^{-1}H$, but using again formulae for partitioned inverses

$$H'(A_u^{22})^{-1} H = H'(A_{u,22} - A_{u,21} A_{u,11}^{-1} A_{u,12})H$$

$$= (H' A_{u,22} H) - (H' A_{u,21}) A_{u,11}^{-1} (A_{u,12}H)$$

$$= A_{r,22} - A_{r,21} A_{r,11}^{-1} A_{r,12} = (A_r^{22})^{-1}$$

Q.E.D.

Note that

$$(3.3.8) \qquad A_u^{22} - H[H'(A_u^{22})^{-1}H]^{-1} H' = (I-P_A)\, A_u^{22}(I-P_A')$$

where $P_A = H[H'(A_u^{22})^{-1}H]^{-1} H'(A_u^{22})^{-1}$. So that if A_u^{22} is positive definite (3.3.8) will be positive semi-definite.

Also note that an alternative expression for $(A_r^{11})^{-1}$ is given by

$$(3.3.6a) \qquad (A_r^{11})^{-1} = (A_u^{11})^{-1} + A_{u,12}[A_{u,22}^{-1} - H(H'\, A_{u,22}H)^{-1}H']A_{u,21}.$$

The proof of (3.3.6a) paralels that of (3.3.6), but now we use

$$(A_r^{11})^{-1} = A_{r,11} - A_{r,12}\, A_{r,22}^{-1}\, A_{r,21} = A_{u,11} - A_{u,12}H(H'A_{u,22}H)^{-1}H'A_{u,21}$$

and

$$(A_u^{11})^{-1} = A_{u,11} - A_{u,12}\, A_{u,22}^{-1}\, A_{u,21} ;$$

subtracting both expressions the result follows.

In view of this Lemma, for restricted models we have

$$(3.3.9) \qquad \Phi_r^{11} = \Phi_u^{11} - \Phi_{u,11}^{-1}\, \Phi_{u,12}\, \{\Phi_u^{22} - G[G'(\Phi_u^{22})^{-1}G]^{-1}G'\}\, \Phi_{u,21}\, \Phi_{u,11}^{-1}$$

$$(3.3.10) \qquad \Phi_r^{22} = [G'(\Phi_u^{22})^{-1}G]^{-1}$$

When normality holds, Φ_u^{11} and Φ_r^{11} are the asymptotic variance matrices of the QML estimators of $\bar{\theta}$ based on L_u and L_r, respectively, $\hat{\theta}_u$ and $\hat{\theta}_r$. Since $\{\Phi_u^{22} - G[G'(\Phi_u^{22})^{-1}G]^{-1}G'\}$ is positive semi-definite, (3.3.9) implies that $(\Phi_u^{11} - \Phi_r^{11})$ is positive semi-definite. So that under normality $\hat{\theta}_r$ is efficient relative to $\hat{\theta}_u$.[3] This result was first shown by Rothenberg and Leenders (1964) who analised the cases where $\bar{\Omega}$ is either a diagonal matrix or is completely known, and it was further investigated in Rothenberg (1973). Recently, a similar result for simultaneous (non-triangular) models has been given an instrumental variables interpretation, in the sense that when some a priori information on $\bar{\Omega}$ is available, the ML estimator is able to form better instruments for the endogenous variable (cf. Hausman, Newey and Taylor (1983)). Nevertheless, for a triangular model the QML estimator solves a set of "generalised" least squares equations and so, in this case, the efficiency gain of $\hat{\theta}_r$ with respect to $\hat{\theta}_u$ can be regarded as a direct consequence of the efficiency gain of $\hat{\Omega}(\tau)$ with respect to $\hat{\Omega}$ in estimating $\bar{\Omega}$. In particular, if $\bar{\Omega}$ is completely known, then $\Phi_r^{22} \equiv 0$ so that (3.3.9) becomes

$$\Phi_r^{11} = \Phi_u^{11} - \Phi_{u,11}^{-1}\,\Phi_{u,12}(\Phi_u^{22})\Phi_{u,21}\,\Phi_{u,11}^{-1} = \Phi_u^{11} - (\Phi_u^{11} - \Phi_{u,11}^{-1}) = \Phi_{u,11}^{-1} .$$

3.4 The Asymptotic Variance Matrix of the QML Estimators when the Errors are Possibly Non-normal

When the errors are not normal the AVM's of $\hat{\psi}_u$ and $\hat{\psi}_r$ are given by the general formulae $W_u = \Phi_u^{-1}\,\Theta_u\,\Phi_u^{-1}$ and $W_r = \Phi_r^{-1}\,\Theta_r\,\Phi_r^{-1}$, respectively. To be more specific, this will be the case when the fourth order moments do not satisfy (3.2.3) as this is the only non-normal feature of the actual distribution that is relevant in the present context.

Starting with Ω-unrestricted estimators, the partition of W_u is given by

$$(3.4.1) \qquad W_{u,11} = \Phi_u^{11}$$

$$(3.4.2) \qquad W_{u,12} = \Phi_u^{12}$$

$$(3.4.3) \qquad W_{u,22} = \Phi_u^{22} + [(\Delta_4 - \bar{\omega}\bar{\omega}') - \Phi_{u,22}^{-1}] = \Phi_u^{22} + (\Delta_4 - \Delta_4^n)$$

where Δ_4^n is the value of Δ_4 under normality, i.e. $\Delta_4^n = \bar{\omega}\bar{\omega}' + \Phi_{u,22}^{-1}$.

Proofs:

To prove (3.4.1) write

$$W_{u,11} = (\Phi_u^{11} \vdots \Phi_u^{12}) \begin{pmatrix} \Theta_{u,11} & \Theta_{u,12} \\ \Theta_{u,21} & \Theta_{u,22} \end{pmatrix} \begin{pmatrix} \Phi_u^{11} \\ \Phi_u^{21} \end{pmatrix}$$

$$= (\Phi_u^{11} \vdots \Phi_u^{12}) \left[\begin{pmatrix} (\Phi_u^{11})^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \Phi_{u,12} \\ \Phi_{u,22} \end{pmatrix} (\Delta_4 - \bar{\omega}\bar{\omega}')(\Phi_{u,21} \vdots \Phi_{u,22}) \right] \begin{pmatrix} \Phi_u^{11} \\ \Phi_u^{21} \end{pmatrix}$$

and simply noting that

$$(3.4.4) \qquad (\Phi_u^{11} \vdots \Phi_u^{12}) \begin{pmatrix} \Phi_{u,12} \\ \Phi_{u,22} \end{pmatrix} = 0 ,$$

we have

$$W_{u,11} = \Phi_u^{11} (\Phi_u^{11})^{-1} \Phi_u^{11} = \Phi_u^{11}.$$

Equally, to prove (3.4.2) we write

$$W_{u,12} = (\Phi_u^{11} \vdots \Phi_u^{12}) \left[\begin{pmatrix} (\Phi_u^{11})^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \Phi_{u,12} \\ \Phi_{u,22} \end{pmatrix} (\Delta_4 - \bar{\omega}\bar{\omega}')(\Phi_{u,21} \vdots \Phi_{u,22}) \right] \begin{pmatrix} \Phi_u^{12} \\ \Phi_u^{22} \end{pmatrix} ,$$

but using again (3.4.4), we have

$$W_{u,12} = \Phi_u^{11} (\Phi_u^{11})^{-1} \Phi_u^{12} = \Phi_u^{12} .$$

Finally, $W_{u,22}$ is given by

$$W_{u,22} = (\Phi_u^{21} \vdots \Phi_u^{22}) \left[\begin{pmatrix} (\Phi_u^{11})^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \Phi_{u,12} \\ \Phi_{u,22} \end{pmatrix} (\Delta_4 - \bar{\omega}\bar{\omega}')(\Phi_{u,21} \vdots \Phi_{u,22}) \right] \begin{pmatrix} \Phi_u^{12} \\ \Phi_u^{22} \end{pmatrix} .$$

Using that

$$(\Phi_u^{21} \vdots \Phi_u^{22}) \begin{pmatrix} \Phi_{u,12} \\ \Phi_{u,22} \end{pmatrix} = I \quad ,$$

we have

$$W_{u,22} = \Phi_u^{21} (\Phi_u^{11})^{-1} \Phi_u^{12} + (\Delta_4 - \bar{\omega}\bar{\omega}')$$

but in view that $\Phi_{u,22}^{-1} = \Phi_u^{22} - \Phi_u^{21} (\Phi_u^{11})^{-1} \Phi_u^{12}$, $W_{u,22}$ equals

$$W_{u,22} = \Phi_u^{22} - \Phi_{u,22}^{-1} + (\Delta_4 - \bar{\omega}\bar{\omega}')$$

what proves (3.4.3).

(3.4.1) shows that the A.V.M. of $\hat{\theta}_u$ (i.e. the QML estimator of $\bar{\theta}$ based on Ω unrestricted) equals Φ_u^{11} independently of non-normality. This is not the case, however, for the A.V.M. of $\hat{\omega}$ (i.e. the unrestricted QML estimator of $\bar{\omega}$): for a leptokurtic distribution (i.e. $\Delta_4 - \Delta_4^n > 0$, where the inequality sign is taken in the usual matrix sense) the A.V.M. of $\hat{\omega}$ will be larger as compared with the A.V.M. of $\hat{\omega}$ under normality, whereas for a platykurtic distribution the conclusion will be the opposite. Nevertheless, the first case is more likely to occur in practice than the latter, as the presence of outliers among the sample observations tends to increase the value of fourth order moments above their gaussian levels.

Turning to QML estimators based on Ω restricted, we have the following results:

(3.4.5) $$W_{r,11} = \Phi_u^{11} - \Phi_{u,11}^{-1} \Phi_{u,12}(I-P_\Phi)[\Phi_u^{22}-(\Delta_4-\Delta_4^n)](I-P_\Phi') \Phi_{u,21} \Phi_{u,11}^{-1}$$

where $P_\Phi = G[G'(\Phi_u^{22})^{-1}G]^{-1} G'(\Phi_u^{22})^{-1}$, and

(3.4.6) $$W_{r,22} = \Phi_r^{22} + \Phi_r^{22} G'(\Phi_u^{22})^{-1} (\Delta_4-\Delta_4^n)(\Phi_u^{22})^{-1} G \Phi_r^{22},$$

or equivalently

(3.4.6a) $$W_{r,22} = \Phi_r^{22} G'(\Phi_u^{22})^{-1} W_{u,22} (\Phi_u^{22})^{-1} G \Phi_r^{22}.$$

The proofs of (3.4.5), (3.4.6) and (3.4.6a) are given in Appendix 3.B.

The expression for $W_{r,11}$ in (3.4.5) has important implications. In the first place, notice that the A.V.M. of $\hat{\theta}_r$, contrary to what happens with $\hat{\theta}_u$, does depend on non-normality. But more relevant, now it is not generally guaranteed that the QML estimator of $\bar{\theta}$ that takes into account the constraints in Ω is efficient relative to the QML estimator of $\bar{\theta}$ that leaves Ω unrestricted. Actually, if the fourth order moments of the distribution of the errors are large enough, it can be the case that imposing the constraints in Ω implies an efficiency loss with respect to $\hat{\theta}_u$. We state this result formally in the following proposition.

Proposition

$\hat{\theta}_r$ is efficient relative to $\hat{\theta}_u$, i.e. $W_{r,11} < \Phi_u^{11}$, if and only if $H[\Phi_u^{22} - (\Delta_4 - \Delta_4^n)]H' > 0$, where $H = \Phi_{u,11}^{-1}\ \Phi_{u,12}\ (I-P_{\Phi})$. Therefore, a sufficient condition is

$$\Phi_u^{22} = \underset{N\to\infty}{\text{plim}} \left[-\frac{1}{N}\left(\frac{\partial^2 L_u}{\partial\omega\partial\omega'} - \frac{\partial^2 L_u}{\partial\omega\partial\theta'} \left(\frac{\partial^2 L_u}{\partial\theta\partial\theta'} \right)^{-1} \frac{\partial^2 L_u}{\partial\theta\partial\omega'} \right)\Bigg|\bar{\psi}_u \right]^{-1} > \Delta_4 - \Delta_4^n$$

or equivalently, $W_{u,22} < 2\ \Phi_u^{22}$.

These are verifiable conditions which may help to choose between estimators that take into account the restrictions in Ω and estimators that do not, when using QML methods.

Moreover, remark that this Proposition holds true independently on how far goes our knowledge about $\bar{\Omega}$. In particular, when $\bar{\Omega}$ is completely known, P_{Φ} vanishes, so that (3.4.5) becomes

$$(3.4.7) \qquad W_{r,11} = \Phi_u^{11} - \Phi_{u,11}^{-1}\ \Phi_{u,12}[\Phi_u^{22} - (\Delta_4 - \Delta_4^n)]\Phi_{u,21}\ \Phi_{u,11}^{-1} = \Phi_{u,11}^{-1}\ \Theta_{u,11}\ \Phi_{u,11}^{-1}$$

This simply shows the fact that an estimator of $\bar{\theta}$ that minimises det(U'U) is more robust to non-normality than one that minimises $\text{tr}(\bar{\Omega}^{-1}U'U)$.

The Proposition also questions, for example, well-known results on the relative efficiency of quasi-FIML with respect to 3SLS when the covariance matrix is diagonal. Since 3SLS is asymptotically equivalent to the QML estimator that leaves Ω unrestricted, only when our condition on the fourth order moments holds this will be true.

We have not commented on results (3.3.10) and (3.4.6a). However, they have important consequences on efficient estimation which will become apparent in Chapter 5 when we discuss minimum distance estimation of covariance parameters.

3.5 A Simple Two Equation Model

It seemed appropriate to illustrate the previous general discussion with a two equation model with a diagonal covariance matrix, since many of the key results take a very simple form in this context. The model is

$$(3.5.1)\qquad y_{1i} = \bar{\mu}_1 z_{1i} + \bar{\mu}_2 z_{2i} + \bar{u}_{1i}$$

$$(3.5.2)\qquad y_{2i} = \bar{\alpha}\, y_{1i} + \bar{\beta}\, z_{2i} + \bar{u}_{2i} \qquad\qquad (i=1,\dots,N)$$

We assume $\begin{pmatrix} \bar{u}_{1i} \\ \bar{u}_{2i} \end{pmatrix} \sim$ i.i.d. $(0, \bar{\Omega})$, where $\bar{\Omega} = \begin{pmatrix} \bar{\omega}_{11} & \bar{\omega}_{12} \\ \bar{\omega}_{21} & \bar{\omega}_{22} \end{pmatrix}$

and for the true value $\bar{\omega}_{12} = 0$. We further assume that the joint distribution of the errors is symmetric and that the fourth order moments exist. $z_i' = (z_{1i}\ z_{2i})$ are nonstochastic and $\lim_{N\to\infty} (1/N) \sum_{i=1}^{N} z_i z_i' = M = \{m_{ij}\}$ is finite and non-singular. $\mu' = (\mu_1\ \mu_2)$ and $\delta' = (\alpha\ \beta)$ are 2 x 1 vectors of coefficients, and $x_i' = (y_{1i}\ z_{2i})$ is the 2 x 1 vector of left hand side variables in the second equation. Thus, in our general notation $\theta' = (\delta'\ \mu')$ and $\tau' = (\omega_{11}\ \omega_{22})$.

This model can be regarded as a dynamic specification from panel data where two cross-sections are available and there is only one (time-varying) exogenous variable observed over the two time periods; the constraint $\bar{\omega}_{12} = 0$ effectively implies the exogeneity of the initial observations. Alternatively, we may simply think of this model as specifying the linear regression equation (3.5.2) where one of the two regressors is stochastic and given by (3.5.1), though uncorrelated with $\bar{u}_{2i}$ in virtue of the covariance restriction.

In the absence of the covariance constraint, the model is triangular and just-identified. If the covariance restriction holds, the second equation is overidentified and the model becomes recursive.

The QML estimator of $\bar{\delta}$ that leaves Ω unrestricted is given by the simple instrumental variables estimator

(3.5.3) $$\hat{\delta}_u = (Z'X)^{-1} Z'y_2$$

which also is the 2SLS and the 3SLS estimator, and where we have set $X = (y_1 \; z_2)$, $Z = (z_1 \; z_2)$ and y_1, y_2, z_1 and z_2 are N x 1 vectors of observations.

The QML estimator that takes into account the covariance restriction is just the ordinary least squares estimator

(3.5.4) $$\hat{\delta}_r = (X'X)^{-1} X'y_2 \; .$$

Moreover, the restricted and unrestricted QML estimators of $\bar{\mu}$ coincide and correspond to the O.L.S. estimator

$$(3.5.5)\qquad \hat{\mu}_r = \hat{\mu}_u = (Z'Z)^{-1} Z'y_1$$

so that the only nontrivial comparison is between $\hat{\delta}_u$ and $\hat{\delta}_r$.

In the notation of the previous sections, P* takes the form

$$P^* = \begin{pmatrix} \bar{\mu}_1 & \bar{\mu}_2 \\ \bar{\alpha}\bar{\mu}_1 & \bar{\alpha}\bar{\mu}_2+\bar{\beta} \\ 1 & 0 \\ 0 & 1 \end{pmatrix} ,$$

R and G are 8 x 4 and 3 x 2 0-1 matrices, respectively, given by

$$R = \frac{\partial\ \mathrm{vec}(A)}{\partial \theta'} = - \left(\begin{array}{c:c} 0 & \begin{matrix} 0 \\ I_2 \end{matrix} \\ \hdashline \begin{matrix} d_1 d_1' \\ d_2 d_2' \end{matrix} & 0 \end{array}\right) ,$$

$$G = \frac{\partial\ \omega}{\partial \tau'} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} ,$$

where $d_1' = (1, 0)$ and $d_2' = (0, 1)$. Furthermore, Δ_4 is the 3 x 3 matrix

$$(3.5.6)\qquad \Delta_4 = E\left[\begin{pmatrix} \bar{u}_{1i}^2 \\ \bar{u}_{1i}\cdot\bar{u}_{2i} \\ \bar{u}_{2i}^2 \end{pmatrix} (\bar{u}_{1i}^2, \bar{u}_{1i}\bar{u}_{2i}, \bar{u}_{2i}^2)\right] = \begin{pmatrix} \mu_{4,0} & \mu_{3,1} & \mu_{2,2} \\ \mu_{3,1} & \mu_{2,2} & \mu_{1,3} \\ \mu_{2,2} & \mu_{1,3} & \mu_{0,4} \end{pmatrix} ,$$

which under normality, given that $\bar{\omega}_{12} = 0$, equals to

$$(3.5.7) \qquad \Delta_4^n = \begin{bmatrix} 3\bar{\omega}_{11}^2 & 0 & \bar{\omega}_{11}\bar{\omega}_{22} \\ 0 & \bar{\omega}_{11}\bar{\omega}_{22} & 0 \\ \bar{\omega}_{11}\bar{\omega}_{22} & 0 & 3\bar{\omega}_{22}^2 \end{bmatrix} .$$

Noting that since $\bar{\Omega}$ is diagonal, $\bar{\Omega}^{-1} = \mathrm{diag}(1/\bar{\omega}_{ii})$, after some manipulation we have

$$(3.5.8) \qquad W_{u,11} = \Phi_u^{11} = \begin{bmatrix} \bar{\omega}_{22}(\Pi\ M\ \Pi')^{-1} & 0 \\ 0 & \bar{\omega}_{11}\ M^{-1} \end{bmatrix}$$

where $\Pi = \begin{bmatrix} \bar{\mu}_1 & \bar{\mu}_2 \\ 0 & 1 \end{bmatrix}$. This gives the AVM of $\hat{\theta}_u$. Equivalently, if we re-write (3.5.3) as $\sqrt{N}(\hat{\delta}_u - \bar{\delta}) = (Z'X/N)^{-1}\ Z'\bar{u}_2/\sqrt{N}$, the same result for $\hat{\delta}_u$ follows by noting that $\mathrm{plim}(Z'X/N) = M\Pi'$ and that, irrespective of non-normality, $Z'\bar{u}_2/\sqrt{N} \overset{d}{\to} N(0,\ \bar{\omega}_{22}\ M)$. Moreover, since

$$D^+ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} ,$$

straightforward algebra reveals that

$$\Phi_{u,22}^{-1} = \begin{bmatrix} 2\bar{\omega}_{11}^2 & 0 & 0 \\ 0 & \bar{\omega}_{11}\bar{\omega}_{22} & 0 \\ 0 & 0 & 2\bar{\omega}_{22}^2 \end{bmatrix} \quad \text{and} \quad \Phi_{u,21} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1/\bar{\omega}_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} ,$$

from which we have

$$(3.5.9) \qquad \Phi_u^{22} = \begin{pmatrix} 2\bar{\omega}_{11}^2 & 0 & 0 \\ 0 & [\bar{\omega}_{11}\bar{\omega}_{22}+\bar{\omega}_{11}^2 \text{ Avar}(\hat{\alpha}_u)] & 0 \\ 0 & 0 & 2\bar{\omega}_{22}^2 \end{pmatrix}$$

where $\text{Avar}(\hat{\alpha}_u)$ is the top diagonal element of $\bar{\omega}_{22}(\Pi \ M \ \Pi')^{-1}$, given by $\text{Avar}(\hat{\alpha}_u) = \bar{\omega}_{22}(m^{11}/\bar{\mu}_1^2)$ with $M^{-1} = \{m^{ij}\}$. Φ_u^{22} is the A.V.M. of $\hat{\omega}$ unrestricted under the normality assumption.

Turning to the calculation of the AVM of $\hat{\theta}_r$, $W_{r,11}$, we further require the matrices $\Phi_{u,11}^{-1}$ and P_Φ. $\Phi_{u,11}^{-1}$ is given by

$$\Phi_{u,11}^{-1} = \begin{pmatrix} \bar{\omega}_{22}F^{-1} & 0 \\ 0 & \bar{\omega}_{11}M^{-1} \end{pmatrix}$$

where

$$(3.5.10) \qquad F = \text{plim } (\frac{X'X}{N}) = \Pi \ M \ \Pi' + \bar{\omega}_{11}(d_1 d_1') \ ,$$

and given G and that Φ_u^{22} is a diagonal matrix, the projector P_Φ is simply

$$P_\Phi = G[G'(\Phi_u^{22})^{-1}G]^{-1} \ G'(\Phi_u^{22})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Hence

$$\Phi^{-1}_{u,11}\,\Phi_{u,12}(I-P_{\Phi}) = \begin{bmatrix} 0 & f^{11} & 0 \\ 0 & f^{21} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Calculation of $W_{r,11}$ also requires $\Phi^{22}_u - (\Delta_4 - \overset{n}{\Delta}_4)$. But in view of (3.5.9), (3.5.6) and (3.5.7) this is given by

$$(3.5.11)\quad \Phi^{22}_u - (\Delta_4 - \overset{n}{\Delta}_4) = \begin{bmatrix} 5\bar{\omega}^2_{11} - \mu_{4,0} & -\mu_{3,1} & \bar{\omega}_{11}\bar{\omega}_{22} - \mu_{2,2} \\ -\mu_{3,1} & \bar{\omega}_{11}\bar{\omega}_{22}(2+\bar{\omega}_{11}\frac{m^{11}}{\bar{\mu}^2_1}) - \mu_{2,2} & -\mu_{1,3} \\ \bar{\omega}_{11}\bar{\omega}_{22} - \mu_{2,2} & -\mu_{1,3} & 5\bar{\omega}^2_{22} - \mu_{0,4} \end{bmatrix}.$$

Finally, using (3.4.5) we obtain

$$(3.5.12)\quad W_{r,11} = \begin{bmatrix} \bar{\omega}_{22}(\Pi\ M\ \Pi')^{-1} & 0 \\ 0 & \bar{\omega}_{11}M^{-1} \end{bmatrix} - \phi \begin{bmatrix} F^{-1}\,d_1 d_1'\,F^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

where ϕ is the (2,2)nd element of $\Phi^{22}_u - (\Delta_4 - \overset{n}{\Delta}_4)$, namely

$$(3.5.13)\quad \phi = \bar{\omega}_{11}\bar{\omega}_{22}(2 + \bar{\omega}_{11}\frac{m^{11}}{\bar{\mu}^2_1}) - \mu_{2,2}\quad .$$

Equivalently, we can write (3.5.4) as $\sqrt{N}(\hat{\delta}_r - \bar{\delta}) = (X'X/N)^{-1}\,X'\bar{u}_2/\sqrt{N}$. Next we observe that $X'\bar{u}_2 = \sum_{i=1}^{N} x_i\,\bar{u}_{2i}$ and also that the $x_i\bar{u}_{2i}$ are independently distributed observations with zero mean (crucially depending on the restriction $\bar{\omega}_{12} = 0$) and variance given by

$$E(\bar{u}_{2i}^2 \; x_i x_i') = \bar{\omega}_{22} \; \Pi(z_i z_i')\Pi' + \mu_{2,2}(d_1 d_1')$$

Thus, $X'\bar{u}_2/\sqrt{N} \overset{d}{\rightarrow} N(0, \bar{\omega}_{22} \; \Gamma)$, with

$$(3.5.14) \qquad \Gamma = \Pi \; M \; \Pi' + (\mu_{2,2}/\bar{\omega}_{22})(d_1 d_1').$$

Therefore, $\sqrt{N}(\hat{\delta}_r - \bar{\delta}) \overset{d}{\rightarrow} N(0, \bar{\omega}_{22} \; F^{-1} \; \Gamma \; F^{-1})$. It can be shown that $\bar{\omega}_{22} \; F^{-1} \; \Gamma \; F^{-1} = \bar{\omega}_{22}(\Pi \; M \; \Pi')^{-1} - \phi \; F^{-1} \; d_1 d_1' \; F^{-1}$ as in (3.5.12). Remark that under normality or, more generally, if $\bar{u}_{1i}$ and $\bar{u}_{2i}$ are independent, $\mu_{2,2} = \bar{\omega}_{11}\bar{\omega}_{22}$ so that $\Gamma = F$ and also $\phi = \bar{\omega}_{11}\bar{\omega}_{22}(1 + \bar{\omega}_{11} m^{11}/\bar{\mu}_1^2)$ which is clearly non-negative.

However, in general

$$(3.5.15) \qquad \text{AVAR}(\hat{\delta}_r) = \text{AVAR}(\hat{\delta}_u) - \phi \; F^{-1} \; d_1 d_1' \; F^{-1}$$

and therefore

$$\text{AVAR}(\hat{\delta}_r) < \text{AVAR}(\hat{\delta}_u)$$

if any only if $\phi > 0$ or

$$(3.5.16) \qquad \frac{E(\bar{u}_{1i}^2 \bar{u}_{2i}^2)}{\bar{\omega}_{11}\bar{\omega}_{22}} < 2 + \bar{\omega}_{11} \frac{m^{11}}{\bar{\mu}_1^2}$$

given the fact that $F^{-1} \; d_1 d_1' \; F^{-1}$ is positive semi-definite. (3.5.16) particularises to this example the necessary and sufficient condition stated in the Proposition of Section 3.4.

NOTES

1 D is the duplication matrix whose properties are extensively studied in Magnus and Neudecker (1980), and D^+ is the Moore-Penrose inverse of D.

2 In view of (3.2.13) we have $\Phi_{u,12}\,\bar{\omega} = -R'(\bar{\Omega}^{-1} \otimes \tilde{B}) D\,\bar{\omega} = -R'(\bar{\Omega}^{-1} \otimes \tilde{B})\,\mathrm{vec}(\bar{\Omega}) = -R'\,\mathrm{vec}(\tilde{B}')$. A typical element of this vector is given by

$$-[\mathrm{vec}(A_j)]'\,\mathrm{vec}(\tilde{B}') = -\,\mathrm{tr}(A_j'\,\tilde{B}') = -\,\mathrm{tr}(\tilde{B}\,A_j)$$

where $A_j = \frac{\partial A(\theta)}{\partial \theta_j}\Big|_{\bar{\theta}} = (B_j \vdots C_j)$, $(j=1,\ldots,p)$. But $\tilde{B}\,A_j = \bar{B}^{-1}\,B_j$, and since B_j is strictly lower triangular and $\bar{B}^{-1}$ is lower triangular, the diagonal elements of $\bar{B}^{-1}\,B_j$ are all zero so that $\mathrm{tr}(\bar{B}^{-1}\,B_j) = 0$ $(j=1,\ldots,p)$.

3 Incidentally, notice that in a multivariate regression model (i.e. where $\bar{B}$ is a unit matrix) $\hat{\theta}_u$ and $\hat{\theta}_r$ are equivalent, as in this case $\Phi_{u,12} = 0$ so that $\Phi_r^{11} = \Phi_u^{11}$.

APPENDIX 3.A

The First and Second Derivatives of the Quasi-log-likelihood Function with respect to θ and τ, and the Matrices Φ_u, Φ_r, Θ_u, Θ_r and Φ_u^{-1}

Letting

$$L_{r,i} = k_0 - \tfrac{1}{2} \log \det \Omega(\tau) - \tfrac{1}{2} x_i' A'(\theta)\ \Omega^{-1}(\tau)\ A(\theta) x_i$$

the first and second derivatives of $L_{r,i}$ are

$$\text{(3.A.1)} \quad \frac{\partial L_{r,i}}{\partial \theta} = - R'(\theta)(\Omega^{-1}(\tau) \otimes x_i x_i') \operatorname{vec} A(\theta) = - R'(\theta)(\Omega^{-1}(\tau) \otimes I)(u_i \otimes x_i)$$

$$\text{(3.A.2)} \quad \frac{\partial L_{r,i}}{\partial \tau} = \tfrac{1}{2} G'(\tau)\ D'(\Omega^{-1}(\tau) \otimes \Omega^{-1}(\tau)) \operatorname{vec}[(u_i u_i') - \Omega(\tau)]$$

$$\text{(3.A.3)} \quad \frac{\partial^2 L_{r,i}}{\partial \theta \partial \theta'} = - R'(\theta)\ (\Omega^{-1}(\tau) \otimes x_i x_i')\ R(\theta) - [I \otimes \operatorname{vec}(\Omega^{-1}(\tau)\ u_i x_i')]' \frac{\partial}{\partial \theta'} \operatorname{vec} R'(\theta)$$

$$\text{(3.A.4)} \quad \frac{\partial^2 L_{r,i}}{\partial \theta \partial \tau'} = R'(\theta)\ (I \otimes x_i u_i')\ (\Omega^{-1}(\tau) \otimes \Omega^{-1}(\tau))\ D\ G(\tau)$$

$$\text{(3.A.5)} \quad \frac{\partial^2 L_{r,i}}{\partial \tau \partial \tau'} = - \tfrac{1}{2} G'(\tau)\ D'(\Omega^{-1}(\tau) \otimes \Omega^{-1}(\tau))\ D\ G(\tau) + \tfrac{1}{2}[I \otimes \operatorname{vec}(u_i u_i' - \Omega(\tau))]' \frac{\partial}{\partial \tau'} \operatorname{vec}[G'(\tau) D'(\Omega^{-1}(\tau) \otimes \Omega^{-1}(\tau))]$$

where

$$R(\theta) = \frac{\partial \text{ vec } A(\theta)}{\partial \theta'}, \quad G(\tau) = \frac{\partial \omega(\tau)}{\partial \tau'}$$

and $u_i = A(\theta)\, x_i$. Also note that $\text{vec}(u_i x_i') = u_i \otimes x_i$. To obtain the derivatives of $L_{u,i}$ with respect to θ and ω, we simply replace τ by ω in (3.A.1) to (3.A.5) and set $G(\tau)$ equal to a unit matrix.

<u>The matrices Φ_r and Φ_u</u>

From (3.A.3) we have

$$\Phi_{r,11} = \text{plim}_{N\to\infty} \left(- \frac{1}{N} \sum_{1}^{N} \frac{\partial^2 L_{r,i}}{\partial\theta\partial\theta'} \Big|\bar{\psi}_r \right) = R'(\bar{\Omega}^{-1} \otimes \text{plim}_{N\to\infty}(\frac{1}{N} \sum x_i x_i'))R$$

To see that the second term in (3.A.3) vanishes it is simpler to consider the second derivative with respect to the jth element of θ. This term is thus given by

$$[I \otimes \text{vec}(\Omega^{-1}(\tau)\, u_i x_i')]' \text{ vec } R_j'(\theta) = R_j'(\theta) \text{ vec}(\Omega^{-1}(\tau)\, u_i x_i')$$

where $R_j(\theta) = \frac{\partial R(\theta)}{\partial \theta_j}$. Then in view of

(3.A.6) $\quad \text{plim } \frac{1}{N}(\sum \bar{u}_i x_i') = \bar{\Omega}\, \tilde{B}'$

and the triangularity of $\bar{B}$ we obtain

$$\underset{N\to\infty}{\text{plim}}\ R_j'(\bar{\theta})\ \text{vec}(\bar{\Omega}^{-1}\ \frac{1}{N}\sum \bar{u}_i x_i') = R_j'(\bar{\theta})\ \text{vec}(\tilde{B}') = 0$$

since the same argument used for R' in note 2 applies to $R_j'(\bar{\theta})$ $(j=1,\ldots,p)$ in this case.

In what follows we shall make extensive use of the compact notation $x_i = P^* z_i + \tilde{B}\ \bar{u}_i$. Hence, $x_i x_i' = P^*(z_i z_i')P^{*\prime} + \tilde{B}(\bar{u}_i \bar{u}_i')\tilde{B}' + P^*(z_i \bar{u}_i')\tilde{B}' + \tilde{B}(\bar{u}_i z_i')P^{*\prime}$ and thus we have

$$\text{(3.A.7)}\qquad \text{plim}\ \frac{1}{N}(\sum x_i x_i') = P^*\ M\ P^{*\prime} + \tilde{B}\ \bar{\Omega}\ \tilde{B}'$$

from which (3.2.12) and (3.2.15) follow.

To prove (3.2.13) and (3.2.16) we make use again of (3.A.6)

$$\Phi_{r,12} = \underset{N\to\infty}{\text{plim}} \left(-\frac{1}{N}\sum_{1}^{N} \frac{\partial^2 L_{r,i}}{\partial\theta\,\partial\tau'}\bigg|_{\bar{\psi}_r} \right) = -\ R'(I \otimes \tilde{B}\ \bar{\Omega})(\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1})\ D\ G$$

$$= -\ R'(\bar{\Omega}^{-1} \otimes \tilde{B})\ D\ G = \Phi_{u,12}\ G$$

(3.2.14) and (3.2.17) come straightforwardly from (3.A.5) noticing that the second term vanishes in the limit since $\underset{N\to\infty}{\text{plim}}[(1/N)(\sum \bar{u}_i \bar{u}_i') - \bar{\Omega}] = 0$.

The matrices Θ_r and Θ_u

Turning to outer product matrices, from (3.A.1) we have

$$\Theta_{ri,11} = E\left(\frac{\partial L_{r,i}}{\partial \theta}\Big|_{\bar{\psi}_r} \frac{\partial L_{r,i}}{\partial \theta'}\Big|_{\bar{\psi}_r} \right) = R'(\bar{\Omega}^{-1} \otimes I)E(\bar{u}_i\bar{u}_i' \otimes x_i x_i')(\bar{\Omega}^{-1} \otimes I)R$$

furthermore

$$E(\bar{u}_i\bar{u}_i' \otimes x_i x_i') = (I \otimes P^*)\ E(\bar{u}_i\bar{u}_i' \otimes z_i z_i')(I \otimes P^{*\prime}) +$$

$$(I \otimes \tilde{B})E(\bar{u}_i\bar{u}_i' \otimes \bar{u}_i\bar{u}_i')(I \otimes \tilde{B}') + (I \otimes P^*)E(\bar{u}_i\bar{u}_i' \otimes z_i\bar{u}_i')(I \otimes \tilde{B}')$$

$$+ (I \otimes \tilde{B})\ E(\bar{u}_i\bar{u}_i' \otimes \bar{u}_i z_i')(I \otimes P^{*\prime})$$

and since we assumed the third order moments to be zero, this reduces to

$$\text{(3.A.8)} \qquad E(\bar{u}_i\bar{u}_i' \otimes x_i x_i') = (\bar{\Omega} \otimes P^*\ z_i z_i'\ P^{*\prime}) + (I \otimes \tilde{B})\ D\ \Delta_4\ D'(I \otimes \tilde{B}').$$

Nevertheless, remark that if we have non-zero third order moments we still obtain the same expression for $\lim_{N\to\infty}\{(1/N)\ \sum_{i=1}^{N} E(\bar{u}_i\bar{u}_i' \otimes x_i x_i')\}$ if $\lim_{N\to\infty}\{(1/N)\ \sum_{i=1}^{N} z_i\} = 0$. Therefore

$$\text{(3.A.9)} \qquad \Theta_{r,11} = \lim_{N\to\infty} \frac{1}{N}\sum_{1}^{N} \Theta_{ri,11}$$

$$= R'(\bar{\Omega}^{-1} \otimes P^*\ M\ P^{*\prime})R + R'(\bar{\Omega}^{-1} \otimes \tilde{B})D\ \Delta_4\ D'(\bar{\Omega}^{-1} \otimes \tilde{B}')R$$

what in view of (3.2.12) and (3.2.13) can be written as

$$\Theta_{r,11} = \Phi_{u,11} - R'(\bar{\Omega}^{-1} \otimes \tilde{B}\, \bar{\Omega}\, \tilde{B}')R + \Phi_{u,12}\, \Delta_4\, \Phi_{u,21} .$$

Now to prove (3.2.18) it suffices to show that

$$\text{(3.A.10)} \quad R'(\bar{\Omega}^{-1} \otimes \tilde{B}\, \bar{\Omega}\, \tilde{B}')R = \Phi_{u,12}\, \Phi^{-1}_{u,22}\, \Phi_{u,21} .$$

From (3.2.13) and (3.2.21) we have

$$\Phi_{u,12}\, \Phi^{-1}_{u,22}\, \Phi_{u,21} = 2\, R'(\bar{\Omega}^{-1} \otimes \tilde{B})DD^{+}(\bar{\Omega} \otimes \bar{\Omega})D^{+'}\, D'(\bar{\Omega}^{-1} \otimes \tilde{B}')R$$

Now using properties from Magnus and Neudecker (1980)

$$2\, DD^{+}(\bar{\Omega} \otimes \bar{\Omega})D^{+'}D' = 2(\bar{\Omega} \otimes \bar{\Omega})D^{+'}D' = (\bar{\Omega} \otimes \bar{\Omega})(I+K)$$

where K is the commutation matrix. Thus we have

$$\text{(3.A.11)} \quad \Phi_{u,12}\, \Phi^{-1}_{u,22}\, \Phi_{u,21} = R'(\bar{\Omega}^{-1} \otimes \tilde{B}\, \bar{\Omega}\, \tilde{B}')R + R'(I \otimes \tilde{B}\, \bar{\Omega})K(\bar{\Omega}^{-1} \otimes \tilde{B}')R.$$

The second term simplifies to $R'(\tilde{B}' \otimes \tilde{B})(\frac{\partial\, vec(A')}{\partial \theta'})$ which equals zero for $\bar{B}$ lower triangular. This establishes (3.A.10) and indeed (3.2.18) and (3.2.22). Incidentally, note that

$$\frac{\partial^2 \log\det B(\theta)}{\partial\theta\partial\theta'} = -\, R'(\tilde{B}' \otimes \tilde{B})(\frac{\partial\, vec(A')}{\partial\theta'})$$

which is non-zero if B is not triangular. This is a third extra term that appears in the formula for $\Phi_{u,11}$ when the model is not triangular, in order to take account of the Jacobian term in the quasi-log-likelihood function. But as (3.A.11) makes clear the result in (3.2.18) is still valid in non-triangular cases.

Considering the off-diagonal terms in Θ_r and Θ_u, we have

$$\Theta_{ri,12} = E\left\{ \frac{\partial L_{r,i}}{\partial \theta}\Big|_{\bar{\psi}_r} \frac{\partial L_{r,i}}{\partial \tau'}\Big|_{\bar{\psi}_r} \right\}$$

$$= - \tfrac{1}{2} R'(\bar{\Omega}^{-1} \otimes I)\ E(\bar{u}_i \bar{u}_i' \otimes x_i \bar{u}_i')(\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1})\ DG$$

$$+ \tfrac{1}{2} R'(\bar{\Omega}^{-1} \otimes I)\ \mathrm{vec}[E(\bar{u}_i x_i')][\mathrm{vec}(\bar{\Omega}^{-1})]'\ DG$$

and given that $E(\bar{u}_i x_i') = \bar{\Omega}\ \tilde{B}'$ and $E(\bar{u}_i \bar{u}_i' \otimes x_i \bar{u}_i') = (I \otimes \tilde{B}) D\ \Delta_4\ D'$ we obtain

$$\Theta_{r,12} = \lim_{N\to\infty} \frac{1}{N} \sum_1^N \Theta_{ri,12} = - \tfrac{1}{2} R'(\bar{\Omega}^{-1} \otimes \tilde{B}) D\ \Delta_4\ D'(\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}) DG$$

$$+ \tfrac{1}{2} R'\ \mathrm{vec}(\tilde{B}')[\mathrm{vec}(\bar{\Omega}^{-1})]'\ DG\ .$$

The second term vanishes since $R'\ \mathrm{vec}(\tilde{B}') = 0$ (see note 2), and using (3.2.13) and (3.2.14)

$$\Theta_{r,12} = \Phi_{u,12}\ \Delta_4\ \Phi_{u,22}\ G = \Theta_{u,12}\ G$$

what proves (3.2.19) and (3.2.23).

Finally, noting that

$$E\{vec(\bar{u}_i\bar{u}_i' - \bar{\Omega})[vec(\bar{u}_i\bar{u}_i' - \bar{\Omega})]'\} = E(\bar{u}_i\bar{u}_i' \otimes \bar{u}_i\bar{u}_i') - vec(\bar{\Omega})[vec(\bar{\Omega})]'$$

$$= D(\Delta_4 - \bar{\omega}\,\bar{\omega}')D',$$

we have

$$\Theta_{ri,22} = E\left\{ \frac{\partial L_{r,i}}{\partial \tau}\Big|\bar{\psi}_r \; \frac{\partial L_{r,i}}{\partial \tau'}\Big|\bar{\psi}_r \right\}$$

$$= \frac{1}{4} G'D'(\bar{\bar{\Omega}}^{-1} \otimes \bar{\bar{\Omega}}^{-1})D(\Delta_4 - \bar{\bar{\omega}}\bar{\bar{\omega}}')D'(\bar{\bar{\Omega}}^{-1} \otimes \bar{\bar{\Omega}}^{-1})DG \;,$$

and in view of (3.2.14), then (3.2.20) and (3.2.24) follow.

<u>The matrix Φ_u^{-1}</u>

In order to prove (3.3.1) we use the partitioned inverse result

$$\Phi_u^{11} = (\Phi_{u,11} - \Phi_{u,12}\,\Phi_{u,22}^{-1}\,\Phi_{u,21})^{-1}$$

and this yields

$$\Phi_u^{11} = [R'(\bar{\bar{\Omega}}^{-1} \otimes P^*MP^{*\prime})R + R'(\bar{\bar{\Omega}}^{-1} \otimes \tilde{B}\,\bar{\Omega}\,\tilde{B}')R - \Phi_{u,12}\,\Phi_{u,22}^{-1}\,\Phi_{u,21}]^{-1}.$$

But in view of (3.A.10) the last two terms cancel and the result is proved (our previous remark also applies here: this result is still valid if the model is not triangular). (3.3.4) is proved in a similar way to (3.A.10). In this case we have

$$\Phi^{-1}_{u,22}\ \Phi_{u,21} = -\ 2\ D^{+}(\bar{\Omega} \otimes \bar{\Omega})\ D^{+'}\ D'(\bar{\Omega}^{-1} \otimes \tilde{B}')R$$

and given that $(\bar{\Omega} \otimes \bar{\Omega})D^{+'}D' = DD^{+}(\bar{\Omega} \otimes \bar{\Omega})$ (cf. Magnus and Neudecker (198o))

$$\Phi^{-1}_{u,22}\ \Phi_{u,21} = -\ 2\ D^{+}DD^{+}(I \otimes \bar{\Omega}\ \tilde{B}')R$$

Finally, noting that $D^{+}D = (D'D)^{-1}\ D'D = I$, (3.3.4) follows.

Note that the jth column of $\Phi^{-1}_{u,22}\ \Phi_{u,21}$ is then given by

$$-\ 2\ D^{+}(I \otimes \bar{\Omega}\ \tilde{B}')\ \text{vec}(\bar{A}_j) = -\ 2\ D^{+}\ \text{vec}(\bar{A}_j\ \tilde{B}\ \bar{\Omega}) \qquad (j=1,\ldots,p)$$

with

$$\bar{A}_j = \frac{\partial A(\theta)}{\partial \theta_j}\Big|\bar{\theta} = \left(\frac{\partial B(\theta)}{\partial \theta_j}\Big|\bar{\theta}\ \vdots\ \frac{\partial C(\theta)}{\partial \theta_j}\Big|\bar{\theta} \right) = (\bar{B}_j\ \vdots\ \bar{C}_j)$$

but since $A_j\ \tilde{B}\ \bar{\Omega} = \bar{B}_j\ \bar{B}^{-1}\ \bar{\Omega}$, the jth column can also be written as

$$-\ 2\ D^{+}(I \otimes \bar{\Omega}\ \bar{B}^{-1'})\ \text{vec}(\bar{B}_j)$$

and thus also

$$\Phi^{-1}_{u,22}\ \Phi_{u,21} = -\ 2\ D^{+}(I \otimes \bar{\Omega}\ \bar{B}^{-1'})\left(\frac{\partial\ \text{vec}\ B(\theta)}{\partial \theta'}\Big|\bar{\theta}\right).$$

APPENDIX 3.B

Derivation of $W_{r,11}$ and $W_{r,22}$

Starting with $W_{r,11}$, we have

$$W_{r,11} = (\Phi_r^{11} \vdots \Phi_r^{12})\ \Theta_r \begin{pmatrix} \Phi_r^{11} \\ \Phi_r^{21} \end{pmatrix}$$

$$= (\Phi_r^{11} \vdots \Phi_r^{12}) \left[\begin{pmatrix} (\Phi_u^{11})^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \Phi_{u,12} \\ G'\Phi_{u,22} \end{pmatrix} (\Delta_4 - \bar{\omega}\bar{\omega}')(\Phi_{u,21} \vdots \Phi_{u,22}G) \right] \begin{pmatrix} \Phi_r^{11} \\ \Phi_r^{21} \end{pmatrix}$$

$$= \Phi_r^{11}\ (\Phi_u^{11})^{-1}\ \Phi_r^{11} + H(\Delta_4 - \bar{\omega}\ \bar{\omega}')H'$$

where

$$(3.B.1) \qquad H = \Phi_r^{11}\ \Phi_{u,12} + \Phi_r^{12}\ G'\ \Phi_{u,22}\ .$$

However, if we use

$$(3.B.2) \qquad \Phi_r^{12} = -\ \Phi_r^{11}\ \Phi_{r,12}\ \Phi_{r,22}^{-1} = -\ \Phi_r^{11}\ \Phi_{u,12}\ G(G'\ \Phi_{u,22}\ G)^{-1}\ ,$$

H can be written

$$(3.B.3) \qquad H = \Phi_r^{11}\ \Phi_{u,12}\ (I-H^*)$$

with

$$H^* = G(G'\ \Phi_{u,22}\ G)^{-1}\ G'\ \Phi_{u,22}\ .$$

Alternatively, using

$$(3.B.4)\qquad \Phi_r^{12} = -\,\Phi_{r,11}^{-1}\,\Phi_{r,12}\,\Phi_r^{22} = -\,\Phi_{u,11}^{-1}\,\Phi_{u,12}\,G(G'(\Phi_u^{22})^{-1}\,G)^{-1}\,,$$

and

$$\Phi_r^{11} = \Phi_{r,11}^{-1} + \Phi_{r,11}^{-1}\,\Phi_{r,12}\,\Phi_r^{22}\,\Phi_{r,21}\,\Phi_{r,11}^{-1}$$

$$= \Phi_{u,11}^{-1} + \Phi_{u,11}^{-1}\,\Phi_{u,12}\,G(G'(\Phi_u^{22})^{-1}\,G)^{-1}\,G'\,\Phi_{u,21}\,\Phi_{u,11}^{-1}\,,$$

and

$$(3.B.4a)\qquad \Phi_{u,22} = (\Phi_u^{22})^{-1} + \Phi_{u,21}\,\Phi_{u,11}^{-1}\,\Phi_{u,12}\,,$$

simple substitution in (3.B.1) reveals that

$$(3.B.5)\qquad H = \Phi_{u,11}^{-1}\,\Phi_{u,12}(I-P_\Phi)$$

with

$$P_\Phi = G(G'(\Phi_u^{22})^{-1}\,G)^{-1}\,G'(\Phi_u^{22})^{-1}\quad .$$

Now using the identity $\Delta_4 - \bar{\omega}\,\bar{\omega}' = \Phi_{u,22}^{-1} + (\Delta_4 - \Delta_4^n)$ and (3.B.3) we have

$$W_{r,11} = \Phi_r^{11}[(\Phi_u^{11})^{-1} + \Phi_{u,12}(I-H^*)\Phi_{u,22}^{-1}(I-H^{*\prime})\Phi_{u,21}]\Phi_r^{11} + H(\Delta_4-\Delta_4^n)H'$$

but since $(I-H^*)\Phi_{u,22}^{-1}(I-H^{*\prime}) = \Phi_{u,22}^{-1} - G(G'\,\Phi_{u,22}\,G)^{-1}\,G'$, in view of (3.3.6a) from the Lemma, we have

$$(3.B.6) \qquad W_{r,11} = \Phi_r^{11} + H(\Delta_4 - \overset{n}{\Delta}_4)H'$$

Thus, $(W_{r,11}-\Phi_r^{11})$ is positive semi-definite if $(\Delta_4-\overset{n}{\Delta}_4)$ is positive semi-definite. This establishes the comparison between $W_{r,11}$ and Φ_r^{11}. Furthermore, using (3.B.5), the expression for Φ_r^{11} in (3.3.9) can be written

$$\Phi_r^{11} = \Phi_u^{11} - H\,\Phi_u^{22}\,H'$$

and substitution in (3.B.6) yields

$$W_{r,11} = \Phi_u^{11} - H[\Phi_u^{22} - (\Delta_4 - \overset{n}{\Delta}_4)]H'$$

what proves (3.4.5).

Turning to $W_{r,22}$, we have

$$W_{r,22} = (\Phi_r^{21} \,\vdots\, \Phi_r^{22})\left[\begin{pmatrix} (\Phi_u^{11})^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \Phi_{u,12} \\ G'\Phi_{u,22} \end{pmatrix}(\Delta_4-\bar{\omega}\bar{\omega}')(\Phi_{u,21} \,\vdots\, \Phi_{u,22}G)\right]\begin{pmatrix} \Phi_r^{12} \\ \Phi_r^{22} \end{pmatrix}.$$

Using (3.B.4) and (3.B.4a) we obtain

$$(\Phi_r^{21} \,\vdots\, \Phi_r^{22})\begin{pmatrix} \Phi_{u,12} \\ G'\Phi_{u,22} \end{pmatrix} = \Phi_r^{22}\,G'(\Phi_u^{22})^{-1}$$

Thus,

$$W_{r,22} = \Phi_r^{21}(\Phi_u^{11})^{-1}\ \Phi_r^{12} + \Phi_r^{22}\ G'(\Phi_u^{22})^{-1}\ (\Delta_4 - \bar{\omega}\bar{\omega}')(\Phi_u^{22})^{-1}\ G\ \Phi_r^{22}$$

Again, using the identity $\Delta_4 - \bar{\omega}\bar{\omega}' = \Phi_{u,22}^{-1} + (\Delta_4 - \Delta_4^n)$ and noting that from (3.B.4) we obtain

$$\Phi_r^{12} = -\ (\Phi_{u,11}^{-1}\ \Phi_{u,12})G\ \Phi_r^{22} = \Phi_u^{12}(\Phi_u^{22})^{-1}\ G\ \Phi_r^{22},$$

we have

$$W_{r,22} = \Phi_r^{22}\ G'(\Phi_u^{22})^{-1}[\Phi_u^{21}(\Phi_u^{11})^{-1}\ \Phi_u^{12} + \Phi_{u,22}^{-1}](\Phi_u^{22})^{-1}\ G\ \Phi_r^{22}$$

$$+\ \Phi_r^{22}\ G'(\Phi_u^{22})^{-1}(\Delta_4 - \Delta_4^n)(\Phi_u^{22})^{-1}\ G\ \Phi_r^{22}$$

but since $\Phi_u^{22} = \Phi_u^{21}(\Phi_u^{11})^{-1}\ \Phi_u^{12} + \Phi_{u,22}^{-1}$ and $\Phi_r^{22} = [G'(\Phi_u^{22})^{-1}G]^{-1}$,

this simplifies to

$$W_{r,22} = \Phi_r^{22} + \Phi_r^{22}\ G'(\Phi_u^{22})^{-1}\ (\Delta_4 - \Delta_4^n)(\Phi_u^{22})^{-1}\ G\ \Phi_r^{22}$$

what proves (3.4.6). Alternatively, writing

$$W_{r,22} = \Phi_r^{22}\ G'(\Phi_u^{22})^{-1}[\Phi_u^{22} + (\Delta_4 - \Delta_4^n)](\Phi_u^{22})^{-1}\ G\ \Phi_r^{22},$$

In view of (3.4.3), (3.4.6a) is also proven.

CHAPTER 4

WALD AND QUASI-LIKELIHOOD RATIO TESTS OF RANDOM EFFECTS SPECIFICATIONS IN DYNAMIC MODELS

4.1 Introduction

This Chapter examines, in a quasi-maximum likelihood framework, the problem of testing covariance restrictions arising from various random effects specifications. We concentrate on dynamic models with unrestricted initial observations errors of the type c considered in Chapter 2. The availability of both Ω-restricted and Ω-unrestricted QML estimates suggests the use of straightforward quasi-likelihood ratio statistics. On the other hand, in cases where explicit expressions for the constraint equations are available, Wald tests can also be used as they only require the estimation of the Ω-unconstrained model. Nevertheless, as first pointed out by Box (1953), if the fourth order moments deviate from their gaussian values, the asymptotic size of tests on variances that are based on the assumption of normality will be incorrect. In our context, this has been made clear by the results of Chapter 3, which show the dependence of the asymptotic distribution of QML estimates of variance matrices on the value of the actual fourth order moments of the errors.

Therefore, we start in Section 4.2 by specialising our previous results to compute the limiting distribution of $\hat{\Omega}$ unrestricted for panel data models. Section 4.3 discusses a Wald test which is robust to the non-normality of the errors, and shows that appropriate asymptotic probability limits can still be calculated for the quasi-LR and the 'normal-Wald' tests when the errors are non-normal. In Section 4.4 a limited simulation is carried out in order to investigate how far the diagnostics from the proposed tests statistics are likely to be affected by the kurtosis measure of the errors being large compared to that of the normal distribution. Finally, in Section 4.5 we use the Michigan Panel to estimate an empirical earnings function for the US with serially correlated transitory errors. The final specification we choose is not rejected against the unrestricted model on the basis of formal tests statistics. In this case, controlling for non-normality of the errors is crucial as evidence is found that the distribution of earnings errors has long tails.

4.2 The AVM of Ω-Unrestricted QML Estimators for Panel Data

We begin by examining the variances of Ω-unrestricted estimators of the model developed in Chapter 2. Thus

(4.2.1) $$y_{i0} = \mu' z_i^* + u_{i0}$$

(4.2.2) $$y_{it} = \alpha y_{i(t-1)} + \beta' x_{it} + \gamma' z_i + u_{it} \qquad (t=1,\dots,T)$$

with $\delta' = (\alpha\, \beta'\, \gamma')$, $\theta' = (\mu'\delta')$, $u_i' = (u_{i1},\dots,u_{iT})$, $u_i^{*\prime} = (u_{i0} u_i')$, $E(u_i u_i') = \Omega$, $E(u_i^* u_i^{*\prime}) = \Omega^*$, $E(u_{i0}^2) = \omega_{00}$ and $E(u_{i0} u_i') = \omega_{01}$. Let $\omega^* = \nu(\Omega^*)$ be the $\frac{1}{2}(T+1)(T+2)$ vector of variances and covariances containing the lower triangle of Ω^*, and let $\hat{\theta}_u' = (\hat{\mu}_u' \; \hat{\delta}_u')$ and $\hat{\omega}^*$ be the Ω-unrestricted QML estimators of θ and ω^* (i.e. $\hat{\delta}_u$ maximises (2.4.2)). Following the discussion in Chapter 3, the AVM of $\hat{\theta}_u$ is given by (3.3.1) irrespective of non-normality. In this case the coefficient derivative matrix R is given by

$$R = \left[\begin{array}{c|c} O & O \\ I & \\ \hline O & \dfrac{\partial \text{ vec } A(\delta)}{\partial \delta'} \end{array}\right]$$

with $A(\delta)$ as in (2.2.4) and I is an unit matrix of order $n(T+1)+m$. $M = \text{plim}(Z^{*\prime}Z^{*}/N)$ and the variance matrix of the complete model is given by Ω^{*}. $(\partial \text{ vec } A(\delta)/\partial\delta') = S$, say, is a 0-1 matrix that maps the coefficients α, β, γ into the matrix $A(\delta)$. Using the formulae for partitioned inverses, we have after some manipulation

$$(4.2.3) \qquad \text{AVM}(\hat{\delta}_u) = [S'(\Omega^{-1} \otimes P^{*}MP^{*\prime})S]^{-1}$$

Furthermore since

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \begin{pmatrix} (Y'Z^{*})(Z^{*\prime}Z^{*})^{-1}(Z^{*\prime}Y) & Y'Z^{*} \\ Z^{*\prime}Y & Z^{*\prime}Z^{*} \end{pmatrix} = \underset{N\to\infty}{\text{plim}} \frac{1}{N} R_D = P^{*}MP^{*\prime}$$

a consistent estimte of AVM $(\hat{\delta}_u)$ is given by

$$(4.2.4) \qquad \widehat{\text{AVM}}(\hat{\delta}_u) = N[S'(\tilde{\Omega}^{-1} \otimes R_D)S]^{-1}$$

where $\tilde{\Omega}$ is a consistent estimate of Ω. An alternative expression for $\widehat{\text{AVM}}(\hat{\delta}_u)$ that uses the regression notation introduced in Section 1.3 can be shown to be

$$(4.2.4a) \qquad \widehat{\text{AVM}}(\hat{\delta}_u) = N[X^{+\prime}(Z^{*}(Z^{*\prime}Z^{*})^{-1}Z^{*\prime} \otimes \tilde{\Omega}^{-1})X^{+}]^{-1}$$

Now we proceed to evaluate the AVM of $\hat{\omega}^{*}$. In order to use the results in Chapter 3 we can either assume that the third order moments are zero or, alternatively, we may think of the variables in equation (4.2.2) as being expressed in deviations from cross-section means. Nevertheless, this assumes that the original model has time specific intercepts (i.e., if γ_{0t} is the intercept for period t, this formulation does not enforce the constraints $\gamma_{01} = \ldots = \gamma_{0T}$ which rules out the possibility of estimating the effect of particular individual-invariant variables).[1] Using (3.3.3) and (3.4.3), the AVM of $\hat{\omega}^*$ is given by

$$(4.2.5) \qquad W_{\omega\omega} = \text{AVM}(\hat{\omega}^*) = H \cdot \text{AVM}(\hat{\theta}_u)H' + \Delta_4 - \omega^*\omega^{*\prime}$$

where, in view of (3.3.4), H is given by

$$H = -2D^+(I \otimes \Omega^* B'^{-1}) \frac{\partial \text{ vec } B(\alpha)}{\partial \theta'}$$

However since B only depends on α, all columns of H are zero except the corresponding to the partial derivatives with respect to α. Let us introduce the following $(T+1)\times(T+1)$ matrix

$$B_\alpha = -\frac{\partial B(\alpha)}{\partial \alpha} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

then the non-zero column of H is given by

$$2D^{+}(I \otimes \Omega^{*}B'^{-1})\mathrm{vec}(B_{\alpha}) = 2D^{+}\mathrm{vec}(B_{\alpha}B^{-1}\Omega^{*})$$

$$= \nu(B_{\alpha}B^{-1}\Omega^{*} + \Omega^{*}B'^{-1}B'_{\alpha}) = q, \quad \text{say.}$$

Thus

$$(4.2.6) \qquad W_{\omega\omega} = \mathrm{Avar}(\hat{\alpha}_u)qq' + \Delta_4 - \omega^{*}\omega^{*'}$$

where $\mathrm{Avar}(\hat{\alpha}_u)$ is the top diagonal element of $\mathrm{AVM}(\hat{\delta}_u)$ in (4.2.3). Now letting $B_{\alpha}B^{-1}\Omega^{*} + \Omega^{*}B'^{-1}B'_{\alpha} = \{a_{ts}\}$ $(t,s = 0,\ldots,T)$ we then have

$$a_{00} = 0,$$

$$(4.2.7) \qquad a_{0t} = a_{t0} = \sum_{k=1}^{t} \alpha^{(k-1)}\omega_{(t-k)0} \qquad (t=1,\ldots,T)$$

$$a_{ts} = a_{st} = \sum_{k=1}^{t} \alpha^{(k-1)}\omega_{(t-k)s} + \sum_{\ell=1}^{s} \alpha^{(\ell-1)}\omega_{(s-\ell)t} \qquad (t,s=1,\ldots,T)$$

Hence the elements of $W_{\omega\omega}$ take the form

$$(4.2.8) \qquad \mathrm{Asy.\ cov}(\hat{\omega}_{ts},\hat{\omega}_{t's'}) = \mathrm{Avar}(\hat{\alpha}_u)a_{ts}a_{t's'} + \mu_{tst's'} - \omega_{ts}\omega_{t's'}$$

where the a_{ts} are given in (4.2.7) and $\mu_{tst's'} = E(u_{it}u_{is}u_{it'}u_{is'})$ $(t,s,t',s' = 0,\ldots,T)$.

An estimate of $\mathrm{Asy.\ cov}(\hat{\omega}_{ts},\hat{\omega}_{t's'})$ is obtained by replacing true values by their sample counterparts in (4.2.8). In particular, sample fourth order moments are given by

$$\hat{\mu}_{tst's'} = \frac{1}{N}\sum_{i=1}^{N} \hat{u}_{it}\hat{u}_{is}\hat{u}_{it'}\hat{u}_{is'}$$

where

$$\hat{u}_{i0} = y_{i0} - \hat{\mu}' z_i^*$$

and

$$\hat{u}_{it} = y_{it} - \hat{\alpha} y_{i(t-1)} - \hat{\beta}' x_{it} - \hat{\gamma}' z_i \qquad (t=1,\ldots,T)$$

If the u_{it} are normally distributed, (3.2.3) holds and (4.2.8) reduces to

$$\text{Asy. cov}(\hat{\omega}_{ts}, \hat{\omega}_{t's'}) = \text{Avar}(\hat{\alpha}_u) a_{ts} a_{t's'} + \omega_{tt'}\omega_{ss'} + \omega_{ts'}\omega_{st'}$$

and accordingly the AVM of $\hat{\omega}^*$ is given by

(4.2.9) $$\Xi_{\omega\omega} = \text{Avar}(\hat{\alpha}_u) qq' + 2D^+ (\Omega^* \otimes \Omega^*) D^{+\prime}$$

$\Xi_{\omega\omega}$ equals $W_{\omega\omega}$ if condition (3.2.3) is satisfied.

4.3 Wald and Quasi-Likelihood Ratio Tests

Suppose we wish to test a set of r restrictions in Ω^*, namely

(4.3.1) $$H_0: \; f(\omega^*) = 0$$

where f is an $r \times 1$ continuous vector function of ω^*. Alternatively we can parameterise the constraints so that $\omega^* = \omega^*(\tau)$, where τ is a $(T+1)(T+2)/2-r$ vector of constraint parameters. Let $\hat{\theta}_r$ and $\omega^*(\hat{\tau})$ be the restricted QML estimates of θ and ω^*, respectively. Denote the $r \times \frac{1}{2}(T+1)(T+2)$ first derivative matrix at $\hat{\omega}^*$ as

$$F = \left.\frac{\partial f(\omega^*)}{\partial \omega^{*\prime}}\right|_{\hat{\omega}^*}$$

Let $\hat{W}_{\omega\omega}$ and $\hat{\Xi}_{\omega\omega}$ be consistent estimates of $W_{\omega\omega}$ and $\Xi_{\omega\omega}$, respectively. Then we introduce the three following statistics

$$\text{(4.3.2)} \quad WT = Nf(\hat{\omega}^*)'(F\hat{W}_{\omega\omega}F')^{-1}f(\hat{\omega}^*),$$

$$\text{(4.3.3)} \quad NWT = Nf(\hat{\omega}^*)'(F\hat{\Xi}_{\omega\omega}F')^{-1}f(\hat{\omega}^*),$$

$$\text{(4.3.4)} \quad QLR = 2[L(\hat{\theta}_u,\hat{\omega}^*) - L(\hat{\theta}_r,\omega^*(\hat{\tau}))].$$

WT is a robust Wald criterion of the type discussed by White (1982) and on the null hypothesis it is distributed asymptotically as a χ^2 with r degrees of freedom. NWT, henceforth 'Normal-Wald', is an appropriate Wald criterion on the assumption of normality of the error term. Specifically, in addition that all constraints are satisfied, NWT also requires that condition (3.2.3) is true in order to be distributed asymptotically as a χ^2. Furthermore, since the quasi-likelihood ratio statistic has the same asymptotic distribution as NWT under the null hypothesis, similar remarks apply to the statistic QLR.[2]

However, if the matrices $\hat{W}_{\omega\omega}$ and F are available we still can compute the asymptotic distribution of the QLR and NWT statistics under the null hypothesis. To show this, note that if the restrictions are satisfied we may define a standardised r×1 random vector c such that

$$\text{(4.3.5)} \quad c = \sqrt{N}(F\hat{W}_{\omega\omega}F')^{-\frac{1}{2}}f(\hat{\omega}^*) \xrightarrow{d} N(0,I),$$

moreover, let us define the r×r matrix

$$(4.3.6) \qquad \Psi = (F\hat{W}_{\omega\omega}F')^{\frac{1}{2}}(F\hat{\Xi}_{\omega\omega}F')^{-1}(F\hat{W}_{\omega\omega}F')^{\frac{1}{2}}$$

so that

$$(4.3.7) \qquad NWT = c'\Psi c$$

The canonical form of Ψ is given by $\Psi = L\Lambda L'$, where Λ is a diagonal matrix containing the latent roots of Ψ and L is an orthogonal matrix. Because of $L'L = I$, the elements of the transformed vector $c^+ = L'c$ still are standard normal random variables in the limit. Therefore

$$(4.3.8) \qquad NWT = c^{+\prime}\Lambda c^+$$

Hence the asymptotic distribution of NWT is a linear combination of independent χ^2 variables with one degree of freedom and so is the asymptotic distribution of the quasi-likelihood ratio test, which can be evaluated numerically from the central Imhof computing procedure (see Imhof (1961) and Koerts and Abrahamse (1969)). The weights are given by the latent roots of $(F\hat{W}_{\omega\omega}F')(F\hat{\Xi}_{\omega\omega}F')^{-1}$ which is a similar matrix to Ψ (cf. Foutz and Srivastava (1977) and MaCurdy (1981)).

Summing up, likelihood ratio tests of covariance restrictions crucially depend on the assumption of normality of the error term for being asymptotically distributed as a χ^2 under the null hypothesis, unlike the case of regression parameter restrictions. For practical purposes, this means that in order to obtain a Wald or an LR test of covariance restrictions with appropriate asymptotic probability limits, we must compute the matrix $(F\hat{W}_{\omega\omega}F')$ and therefore explicit expressions of the constraint equations and their derivatives (or approximating first differences) are required. Furthermore, evaluating $\hat{W}_{\omega\omega}$ requires estimates of the matrix Δ_4 of fourth order moments as noted above.

The Constraint Equations for Moving Average Random Effects Covariance Matrices

In what follows we examine the form of the constraint equations implied by homoscedastic and heteroscedastic MA(1) random effects schemes of the type introduced in Chapter 2. Thus

$$(4.3.9) \qquad u_{it} = \eta_i + \varepsilon_{it} + \lambda(\sigma_t/\sigma_{t-1})\varepsilon_{i(t-1)} \qquad (t=1,\ldots,T)$$

where $\varepsilon_{it} \sim idd(0,\sigma_t^2)$, $\eta_i \sim iid(0,\sigma_\eta^2)$ and $\omega_{0t} = E(u_{i0}u_{it})$ $(t=1,\ldots,T)$ are unrestricted parameters.

The MA(1) homoscedastic structure (i.e. the case where $\sigma_t = \sigma$ for all t) imposes $\frac{1}{2}T(T+1) - 3$ linear restrictions in Ω^*. Namely

$$(4.3.10) \qquad \omega_{(k+1)(k+1)} - \omega_{kk} = 0, \qquad (k=1,\ldots,T-1)$$

$$(4.3.11) \qquad \omega_{(k+1)(k+2)} - \omega_{k(k+1)} = 0, \qquad (k=1,\ldots,T-2)$$

$$(4.3.12) \qquad \omega_{i(i+k)} - \omega_{j(j+s)} = 0 \qquad \text{for } i \neq j \text{ and } k,s > 1.$$

On the other hand, the MA(1) heteroscedastic structure imposes $\frac{1}{2}T(T-1) - 2$ implicit constraints on Ω^*. One possible way to write them out is given by

$$(4.3.13)\quad (\omega_{k(k+1)} - \omega_{1T})^2 (\omega_{(k+2)(k+2)} - \omega_{1T})$$
$$- (\omega_{(k+1)(k+2)} - \omega_{1T})^2 (\omega_{kk} - \omega_{1T}) = 0 \qquad (k=1,\ldots,T-2)$$

$$(4.3.14)\quad \omega_{i(i+k)} - \omega_{j(j+s)} = 0 \qquad \text{for } i \neq j \text{ and } k,s > 1.$$

In both cases derivatives are straightforward and so the F matrices can be evaluated analytically.

While generalising these constraint equations to higher order moving average schemes is straightforward, it is unclear how to set up corresponding equations for autoregressive and mixed schemes. Minimum chi-squared statistics which are based on constraint parameters will provide an alternative framework where this problem can be overcome (see Chapter 5).

4.4 Simulation Results

A limited simulation was carried out to study the performance of the proposed testing procedures. The main purpose of the experiments was to investigate how far the diagnostics from the various tests are likely to be affected by the kurtosis measure of the errors being large compared to that of the normal distribution for samples of the size encountered in practice. Thus we performed two experiments. In each case 50 samples were generated on the following model

$$y_{it} = 1. + .5y_{i(t-1)} + .35x_{it} + .15z_i + u_{it},$$

$$(4.4.1)\quad u_{it} = \eta_i + v_{it}, \qquad (i=1,\ldots,500;\ t=1,\ldots,20)$$

$$v_{it} = \varepsilon_{it} + .5\varepsilon_{i(t-1)}$$

where $\eta_i \sim \text{iid}(0,.16)$, $\varepsilon_{it} \sim \text{iid}(0,.25)$, $y_{i0} = v_{i0} = 0$. The exogenous variables were generated in the same manner as in the experiments reported in Section 2.5. The first ten cross-sections were discarded so that $T = 9$ and $N = 500$. Now let us denote the kurtosis measure by $\gamma_2 = \{E(u_{it}^4)/[E(u_{it}^2)]^2\}$. The only difference between the two experiments is that in the first one $\gamma_2 \simeq 12$ while in the second the kurtosis measure attains its normal value $\gamma_2 = 3$. The same set of pseudo-random numbers were used in both cases.

Before proceeding further, it is worth explaining how non-normal variates were generated. Let us consider a random variable X whose distribution function is contaminated normal

$$(4.4.2) \qquad F_k(x) = (1-\rho)\Phi(x) + \rho\Phi(x/k)$$

such that $\rho = 1/(k^2-1)$ (see Ali (1974)) and where $\Phi(x)$ is the standard normal cdf

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp(-\tfrac{1}{2}u^2)\,du.$$

Then X is symmetrically distributed with zero mean and variance equal to two for all k. However, since the kurtosis measure of X is $\gamma_2 = (3/4)(k^2+2)$, this enables us to increase γ_2 as much as we like while keeping Var(X) constant (of course, one cannot go too far without bearing a too large sample variance of γ_2 for any reasonable sample size).

In model (4.4.1), η_i has cdf $F_k(x\sqrt{2}/\sigma_\eta)$ and ε_{it} has cdf $F_k(x\sqrt{2}/\sigma_\varepsilon)$. In Experiment 1 we set $k^2 = 31.1$ which can be shown to lead to a kurtosis measure for u_{it} of about twelve. In Experiment 2, k^2 is simply set to two.

For each replication both restricted and unrestricted QML estimates were obtained and, from those, QLR, NWT and WT tests of the restrictions (4.3.10) to (4.3.12) were calculated. Finally, the Imhof routine was used to compute proper asymptotic limits for the QLR and the NWT statistics.[3] Tables 1 and 2 summarise the results. WT appears to be slightly upward biased (i.e. too rejecting) especially in Experiment 2, while NWT shows a smaller bias in the opposite direction. QLR lies in between and so it seems to be the best option, at least in our example. However, the outstanding feature of these results is the confirmation of the fact that in practice tests of covariance matrix restrictions are useless without controlling for departures of the errors from normality: when the distribution of the errors is long-tailed, the mean and the variance of the QLR test under the null hypothesis are far greater than their χ^2 counterparts (in our case, the mean is 42 and the variance 84).

4.5 Estimation of Earnings Functions for the US

The purpose of this Section is to estimate empirical earnings functions for panel data that take into account the dynamic features of the sample under consideration. We wish to choose a specification that is not rejected against a reasonably general maintained hypothesis on the basis of formal test statistics. A standard earnings function is best interpreted as a reduced form relation made up of a mixture of several supply and demand factors, among which personal characteristics, like years of education and work experience, play the central role. This

TABLE 1

Simulation Results for the Model with Long-Tailed Errors ($\gamma_2 \simeq 12$)

	QLR		NWT		WT
	Number of Rejections out of 50 Cases				
Size	I[a]	C[b]	I	C	C
0.10	5	47	3	47	9
0.05	2	45	1	46	5
0.01	0	42	0	39	0
Mean	97.842		90.812		45.921
Variance	1216.680		757.630		79.709

[a] According to calculated Imhof limits.

[b] According to a chi-square with 42 degrees of freedom.

TABLE 2

Simulation Results for the Model with Normal Errors ($\gamma_2 = 3$)

	QLR		NWT		WT
	Number of Rejections Out of 50 Cases				
Size	I[a]	C[b]	I	C	C
0.10	5	5	1	2	15
0.05	0	0	0	0	10
0.01	0	0	0	0	1
Mean	42.020		41.791		47.744
Variance	69.455		66.706		108.107

[a] According to calculated Imhof limits.

[b] According to a chi-square with 42 degrees of freedom.

means that some caution is required in drawing conclusions from the estimated coefficients unless we are prepared to think in terms of the mincerian schooling model.

Our sample corresponds to male heads of households observed through ten consecutive years (1967-1976) of the Michigan Panel of Income Dynamics, with the following characteristics: (i) were not included in the SEO sample since it was non-random, (ii) have remained the same over the sample period (i.e., no split-off family units where the head has changed), (iii) were not unemployed, retired or full-time students, (iv) reported positive annual hours and earnings throughout the sample period. In view of this, we are left with a sample of 742 individuals.

We decided to transform the data into deviations from cross-section means (i.e., $x^*_{it} = x_{it} - (1/N)\sum_{i=1}^{N} x_{it}$); this is equivalent to introducing a set of time dummies that capture the combined effect of all potential macroeconomic explanatory variables (like productivity changes).[4] Our dependent variable is the logarithm of the real hourly earnings (in 1967 dollars) and, apart from lagged earnings, we consider some of the explanatory variables that are included in the most conventional earnings functions: years of education, linear and quadratic age effects (note that $A_{it} - \bar{A}_t = A_{i0} - \bar{A}_0$), a dummy variable for non-whites and another binary variable for the professional and managerial occupational groups. Given the absence of other measures of work experience in the sample, this variable is usually defined as the time period that has elapsed since an individual left school (e.g., Exp = Age - Schooling - 6), but as this creates an artifical association between experience and schooling, nothing is lost by using instead the age variable itself. On the other hand, having an experience variable defined in such a way rules out the possibility of distinguishing a dynamic response of earnings to changes in experience from a static one.

Table 3 gives the results of estimates for the model in which Ω^* is treated as an unrestricted matrix. Crude instrumental variable estimates have been used as initial consistent values in order to start off the iterative procedure for the QML estimator. Asymptotic standard errors have been calculated using structural form estimates of P^* in (4.2.3). The actual estimates we have found are very similar to the results in other studies that assume exogenous regressors (e.g., Lillard and Willis (1978)); using years of education, age, age squared and race as explanatory variables, we obtain that an additional year of education leads to a 7.3 per cent higher earnings. However, the introduction of the occupational dummy (model 2 in Table 3) has the effect of reducing the mean schooling coefficient to 5.1 per cent while leaving all other coefficients unchanged.

Turning now to consider constrained models, in Tables 4 and 5 we present the results for the homoscedastic models, and in Table 6 for the heteroscedastic ones. In these cases, asymptotic standard errors have been calculated by using the differencing estimates of the second derivatives of the log-likelihood function, and thus they may be affected by the non-normality of the errors. The estimated coefficients in the homoscedastic cases show little variation in relation to those obtained prior to enforcing the restrictions. The results for the ARMA(1,1) specifications in Table 4 clearly point to models with moving average errors, which are presented in Table 5.

When we let the variances of the transitory errors vary over time, there is an increase in the moving average coefficient while the coefficient of the lagged dependent variable decreases and the derived mean effects of the independent variables remain fairly unchanged. This suggests some degree of indetermination in the contribution of the systematic component, relative to the moving average one, to the overall time dependencies in the sample (the estimated asymptotic correlation between the two coefficients is -.85). On the other hand, the estimated ratios $\hat{\rho}_t^2$, t=1,2,...,9, show some variation but there is no evidence of any systematic pattern. In fact, it should be noticed that, since the only time-varying explanatory variable we are including in the present application is a trending one, the model with lagged endogenous variable and moving average errors is not distinguishable from a static model with ARMA errors. So, all we can say about the motion of this model is that a stochastic individual level is determined by the observed characteristics and the unobservable effect $\eta_i^* = \eta_i/(1-\alpha)$ once for all before the start of the sample period; and then, as random shocks come out, the log-earnings of a particular individual evolve around its random mean following a serially correlated pattern controlled by α and λ. In particular, we obtain that unobserved permanent differences among individuals account for 61 per cent of total error variation.[5]

Now we proceed to test the covariance restrictions implied by our random effects specifications using the methods developed in Section 4.3. Table 7 gives the results.[6] Clearly, if we rely on the assumption of normality of the errors, and so we compare the QLR and the NWT criteria against χ^2 limits, both the homoscedastic and the heteroscedastic sets of constraints are rejected at any reasonable level of significance.

But the situation is the opposite when we look at the Wald test and when QLR or NWT are compared against limits calculated from their appropriate asymptotic distribution under the null hypothesis. Clearly, the constraints for the heteroscedastic model are not rejected and, while for the homoscedastic moving average scheme WT is somewhat higher than its expected value of 42, all three tests accept the restrictions at the 90 per cent level.

The values of the standardised fourth order cumulants of the errors for the unrestricted model (version 1) are given in Table 8. These values are rather high and this suggests that the distribution of the errors is long-tailed. In fact, the shape of the distribution as perceived by plotting the histogram of the errors is fairly normal. This points to the fact that QLR and NWT are not χ^2 variates mainly as a consequence of long tails in the density of the errors.

TABLE 3

QML Estimates with Unrestricted Covariance Matrix

Dependent Variable: Log Hourly Earnings[a]

	Model 1		Model 2	
	Estimates	Derived Mean Effects[e]	Estimates	Derived Mean Effects
Years of Education	.0112 (.0037)[b]	.0730	.0099 (.0034)	.0510
Age	.0046 (.0030)	.0301	.0058 (.0034)	.0297
Age Squared	-.000046 (.000031)	-.0003	-.000059 (.000035)	-.0003
Race[c]	.0213 (.0131)	.1391	.0269	.1384
Occupation[d]	-	-	.0469 (.0161)	.2413
Lagged Dependent Variable	.8469 (.0445)		.8057 (.0553)	

[a] Data in mean deviation form (N= 742, T=9, period 1967-1976).

[b] Standard errors in parentheses.

[c] Dummy variable: 1 if individual is white.

[d] Dummy variable: 1 if individual belongs to professional or managerial groups in 1967.

[e] Calculated as $\hat{\gamma}_k^* = \hat{\gamma}_k / (1-\hat{\alpha})$.

TABLE 4

QML Estimates of ARMA(1,1) Homoscedastic Models

Dependent Variable: Log Hourly Earnings

	Model 1		Model 2	
	Estimates	Derived Mean Effects	Estimates	Derived Mean Effects
Years of Education	.0118 (.0042)	.0730	.0090 (.0030)	.0500
Age	.0054 (.0031)	.0334	.0055 (.0030)	.0309
Age Squared	-.000054 (.000032)	-.000332	-.000057 (.000032)	-.000317
Race	.0219 (.0141)	.1352	.0237 (.0145)	.1322
Occupation	-	-	.0431 (.0156)	.2400
Lagged Dependent Variable	.8376 (.0508)		.8205 (.0492)	
ϕ	-.0072 (.0508)		-.0166 (.0553)	
λ	-.4189 (.0763)		-.3997 (.0792)	
ρ^2	.0532 (.0371)		.0624 (.0375)	
σ^2	.0681		.0676	
σ_η^2 [a]	.0036	$\sigma_{\eta*}^2$ = .1373[b]	.0042	$\sigma_{\eta*}^2$ = .1310

[a] $\sigma_\eta^2 = \rho^2\sigma^2$.

[b] $\sigma_{\eta*}^2 = \sigma_\eta^2/(1-\alpha)^2$.

TABLE 5

QML Estimates of Moving Average Homoscedastic Models

Dependent Variable: Log Hourly Earnings

	Model 1		Model 2	
	Estimates	Derived Mean Effects	Estimates	Derived Mean Effects
Years of Education	.0116 (.0036)	.0728	.0086 (.0028)	.0498
Age	.0053 (.0030)	.0331	.0052 (.0028)	.0304
Age Squared	-.000053 (.000031)	-.000333	-.000054 (.000029)	-.000314
Race	.0214 (.0137)	.1345	.0225 (.0137)	.1309
Occupation	-	-	.0413 (.0135)	.2398
Lagged Dependent Variable	.8410 (.0429)		.8279 (.0440)	
λ	-.4287 (.0303		-.4219 (.0295)	
ρ^2	.0511 (.0318)		.0579 (.0342)	
σ^2	.0681		.0676	
σ^2_η	.0035	$\sigma^2_{\eta*}$ = .1378	.0039	$\sigma^2_{\eta*}$ = .1322

TABLE 6

QML Estimates of Serially Correlated Heteroscedastic Models

Dependent Variable: Log Hourly Earnings (Model 1)

	ARMA(1,1) Errors		MA(1) Errors	
	Estimates	Mean Effects	Estimates	Mean Effects
Years of Education	.0198 (.0041)	.0754	.0176 (.0036)	.0748
Age	.0096 (.0032)	.0366	.0085 (.0030)	.0360
Age Squared	-.000091 (.000033)	-.000346	-.000081 (.000032)	-.000345
Race	.0400 (.0181)	.1525	.0353 (.0172)	.1500
Lagged Dependent Variable	.7374 (.0492)		.7650 (.0423)	
ϕ	-.0653 (.0577)		-	
λ	-.2969 (.0855)		-.3847 (.0320)	
ρ_1^2	.1632 (.0654)		.1331 (.0520)	
ρ_2^2	.1837 (.0774)		.1483 (.0606)	
ρ_3^2	.1399 (.0546)		.1144 (.0453)	
ρ_4^2	.1562 (.0639)		.1278 (.0512)	
ρ_5^2	.1375 (.0570)		.1123 (.0445)	
ρ_6^2	.1831 (.0776)		.1481 (.0612)	

.../continued

TABLE 6 continued

	ARMA(1,1) Errors		MA(1) Errors	
	Estimates	Mean Effects	Estimates	Mean Effects
ρ_7^2	.1367 (.0549)		.1122 (.0439)	
ρ_8^2	.1195 (.0475)		.0978 (.0382)	
ρ_9^2	.1320 (.0528)		.1074 (.0428)	
$\bar{\sigma}^2$ [a]	.0645		.0646	
σ_η^2	.0097	$\sigma_{\eta*}^2 = .1404$	.0079	$\sigma_{\eta*}^2 = .1432$

[a] Calculated as $\bar{\sigma}^2 = \sigma_\eta^2/\bar{\rho}^2$

TABLE 7

Asymptotic Tests of Random Effects Constraints (Model 1)

	Criteria	χ^2 Prob. Limit	Imhof Prob. Limit
	Heteroscedastic Moving Average Scheme (D.F.=34)		
Likelihood Ratio	82.7	1.00	<0.57
Normal-Wald	80.8	1.00	0.53
Wald	32.1	0.44	-
	Homoscedastic Moving Average Scheme (D.F.=42)		
Likelihood Ratio	138.5	1.00	<0.78
Normal-Wald	142.5	1.00	0.79
Wald	51.5	0.85	-

TABLE 8

Standardised Fourth Order Cumulants of Log Earnings Errors

t	Value[a]
1	7.01
2	5.83
3	14.20
4	10.79
5	12.25
6	5.33
7	8.84
8	9.29
9	10.74
Average Value[b]	9.74

[a] Calculated as $\left\{(1/N)\sum_{i=1}^{N} \hat{u}_{it}^{4}/\hat{\omega}_{tt}^{2}\right\} - 3.$

[b] Calcualted as $\left\{\left[(1/N)\sum_{t=1}^{T}\sum_{i=1}^{N} \hat{u}_{it}^{4}\right] / \left(\sum_{t=1}^{T} \hat{\omega}_{tt}^{2}\right)\right\} - 3.$

NOTES

1 See Arellano (1984) for an alternative derivation of the AVM of $\hat{\omega}^*$ based on the reduced form of the model and which makes explicit use of unconstrained third order moments throughout.

2 Note that by making use of a second order Taylor expansion of $L(\psi)$ about $\hat{\psi}_u = (\hat{\theta}_u \hat{\omega}^*)'$, since $\partial L/\partial\psi|\hat{\psi}_u = 0$, we have

$$QLR = 2[L(\hat{\psi}_u) - L(\hat{\psi}_r)] = N(\hat{\psi}_u - \hat{\psi}_r)'\left(-\frac{1}{N}\frac{\partial^2 L}{\partial\psi\partial\psi'}\bigg|_{\psi^*}\right)(\hat{\psi}_u - \hat{\psi}_r)$$

where ψ^* lies between $\hat{\psi}_u$ and $\hat{\psi}_r$, and we are using $\psi' = (\theta'\omega^{*\prime})$ and $\hat{\psi}_r' = (\hat{\theta}_r'\omega^*(\hat{\tau})')$.

3 The computations were carried out on a Cray-1S computer at the University of London Computer Centre. Each experiment took between 12 and 13 CPU minutes.

4 The transformation of the data instead of the inclusion of the time dummies reduces the number of coefficients to be estimated, thus considerably lowering the computer costs.

5 This calculation is made using the results of the third and fourth columns in Table 6. In that case, the stationary solution of the model has an ARMA(1,1) random error whose variance is given by

$$\bar{\sigma}_w^2 = \left(\frac{1 + \lambda^2 + 2\alpha\lambda}{1 - \alpha^2}\right)\bar{\sigma}^2 = .09,$$

therefore, total error variance for the stationary solution is $\sigma^2_{\eta^*} + \bar{\sigma}^2_w = .23$.

6 These results correspond to the model that excludes the occupational dummy, but the inclusion of this variable leaves the value of the test criteria almost unchanged.

CHAPTER 5

MINIMUM DISTANCE AND GLS ESTIMATION OF TRIANGULAR MODELS WITH COVARIANCE RESTRICTIONS

5.1 Introduction

Having discussed the various aspects of estimation and testing of dynamic models in a quasi-maximum likelihood framework, now we turn to consider methods of estimation based on the minimum distance or minimum chi-square principle. A convenient level of generality for our purposes is provided by the triangular system with covariance restrictions introduced in Chapter 3. Thus, the present discussion will be conducted on the basis of these previous results and the same notation will also be used here.

Let $\hat{p}$ be an unconstrained estimator of the coefficient vector $\bar{p}$ which is asymptotically normal with asymptotic covariance matrix equal to V_p. Assume that $\bar{p}$ depends on a set of constraint parameters $\bar{\delta}$, $\bar{p} = p(\bar{\delta})$. The problem of estimating $\bar{\delta}$ is that of finding a value of p satisfying the constraints at a minimum distance from the value $\hat{p}$ indicated by the sample. The minimum distance estimator (MDE), $\tilde{\delta}$, minimises the distance function

$$(5.1.1) \qquad s(\delta) = [\hat{p} - p(\delta)]' \, Q[\hat{p} - p(\delta)]$$

where any consistent estimator of V_p^{-1} is an optimal choice for the norm Q (see Appendix A.5). If V_p^{-1} is the information matrix when there are no constraints, then the basic theorem of the minimum distance method establishes that $\tilde{\delta}$ is asymptotically efficient and therefore asymptotically equivalent to the maximum likelihood estimator. More generally, as discussed by Chamberlain (1982), the quasi-maximum likelihood estimator has the same limiting distribution as a certain minimum distance estimator; but in general that minimum distance estimator is not using the optimal norm. The cases where the AVM of the unconstrained QML estimator of $\bar{p}$ remains the same under non-normality constitute a relevant exception. We know that this is not the case in a simultaneous equation model if covariance constraints are available. Thus, in our model we may expect to obtain estimators of both slope and covariance parameters that are efficient relative to the QML estimator by application of the minimum distance method when the errors are non-normal. Most of the basic discussion of the general principle is contained in Chiang (1956) and Ferguson (1958). Malinvaud (1970) considers the minimum distance estimation of multivariate nonlinear regression models with unrestricted covariance. The non-normal case and its relation to QML estimators of covariance parameters is discussed in Chamberlain (1982).

The order of presentation in this Chapter is as follows. Section 5.2 defines the joint MDE of slope and covariance parameters that makes use of the optimal norm. In Section 5.3 an expression of the AVM of this estimator is derived which can be used as a bound to characterise efficient estimators. Section 5.4 deals with separate MDE of covariance parameters based on 3SLS, MD or QML estimates of unrestricted Ω, which are shown to be efficient. Section 5.5 discusses various generalised least squares estimators of the slope coefficients under linear constraints, and an efficient GLS estimator robust to non-normality is presented.[1] Finally, Section 5.6 examines the problem of estimating subsets of equations.

5.2 Joint Minimum Distance Estimation of Slope and Covariance Parameters

The reduced form formulation of model (3.2.1) is given by

$$\text{(5.2.1)} \qquad \underset{n\times 1}{y_i} = \underset{n\times k}{P(\bar{\theta})}\ \underset{k\times 1}{z_i} + \underset{n\times 1}{\bar{v}_i} \qquad (i=1,\dots,N)$$

$$E(\bar{v}_i\ \bar{v}_i') = \Omega_v(\bar{\theta},\ \bar{\tau})$$

where $\bar{\theta}$ is px1, $\bar{\tau}$ is qx1, $\bar{v}_i = \bar{B}^{-1}\ \bar{u}_i$ and

$$P(\bar{\theta}) = \bar{P} = -\ \bar{B}^{-1}\ \bar{C},$$

$$\Omega_v(\bar{\theta},\ \bar{\tau}) = \bar{\Omega}_v = \bar{B}^{-1}\ \bar{\Omega}\ \bar{B}'^{-1}\ .$$

Let us consider the statistics

$$\text{(5.2.2)} \qquad \hat{P} = \left(\frac{Y'Z}{N}\right)\left(\frac{Z'Z}{N}\right)^{-1}$$

$$\text{(5.2.3)} \qquad \hat{\Omega}_v = \left(\frac{Y'Y}{N}\right) - \left(\frac{Y'Z}{N}\right)\left(\frac{Z'Z}{N}\right)^{-1}\left(\frac{Z'Y}{N}\right)$$

where $Z'Z = \sum_{i=1}^{N} z_i z_i'$, $Y'Z = \sum_{i=1}^{N} y_i z_i'$ and $Y'Y = \sum_{i=1}^{N} y_i y_i'$. $\hat{P}$ and $\hat{\Omega}_v$ are the unconstrained least-squares estimators of $\bar{P}$ and $\bar{\Omega}_v$, and they can be written in vector form as

$$\hat{w}_1 = \text{vec}(\hat{P}) \quad , \qquad \hat{w}_2 = \nu(\hat{\Omega}_v) \quad , \qquad \hat{w} = \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \end{bmatrix} .$$

Accordingly, we set $\bar{w}_1 = \text{vec}(\bar{P})$, $\bar{w}_2 = \nu(\bar{\Omega}_v)$ and $\bar{w}' = (\bar{w}_1' \; \bar{w}_2')$. Moreover, let $\Delta_{v,3}$ and $\Delta_{v,4}$ be matrices of reduced form third order and fourth order moments respectively given by $\Delta_{v,3} = E\{\bar{v}_i [\nu(\bar{v}_i \bar{v}_i')]'\}$ and $\Delta_{v,4} = E\{\nu(\bar{v}_i \bar{v}_i') [\nu(\bar{v}_i \bar{v}_i')]'\}$ for all i. $\hat{P}$ and $\hat{\Omega}_v$ are consistent and asymptotically normal, i.e. we have (e.g. see Arellano (1984)):

$$\sqrt{N}(\hat{w} - \bar{w}) \xrightarrow{d} N(O, V) \quad , \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} .$$

where

$$(5.2.4) \qquad V_{11} = \bar{\Omega}_v \otimes M^{-1} ,$$

$$(5.2.5) \qquad V_{12} = (I \otimes M^{-1} m) \, \Delta_{v,3} ,$$

$$(5.2.6) \qquad V_{22} = \Delta_{v,4} - \bar{w}_2 \bar{w}_2' ;$$

the partition in V corresponds to that in $\bar{w}$, $M = \lim_{N\to\infty} (Z'Z/N)$ and $m = \lim_{N\to\infty} (1/N) \sum_{i=1}^{N} z_i$. We shall assume that $V_{12} = O$; note that this can happen either because $\Delta_{v,3} = O$ or as a consequence of having $m = O$.

The joint minimum distance estimators of $\bar{\theta}$ and $\bar{\tau}$, $\tilde{\theta}_r$ and $\tilde{\tau}$, minimise the function

$$(5.2.7) \qquad s(\theta,\tau) = [\hat{w} - w(\theta,\tau)]' \begin{pmatrix} \hat{V}_{11}^{-1} & O \\ O & \hat{V}_{22}^{-1} \end{pmatrix} [\hat{w} - w(\theta,\tau)]$$

or equivalently

$$(5.2.7a) \quad s(\theta,\tau) = s_1(\theta) + s_2(\theta,\tau)$$

where

$$(5.2.8) \quad s_1(\theta) = [\hat{w}_1 - w_1(\theta)]' \hat{V}_{11}^{-1}[\hat{w}_1 - w_1(\theta)] ,$$

$$(5.2.9) \quad s_2(\theta,\tau) = [\hat{w}_2 - w_2(\theta,\tau)]' \hat{V}_{22}^{-1}[\hat{w}_2 - w_2(\theta,\tau)] .$$

We assume that the optimal weighting matrices have been chosen so that plim $\hat{V}_{11} = V_{11}$ and plim $\hat{V}_{22} = V_{22}$. A distance function similar to (5.2.7) is discussed by Rothenberg (1973); however, he sets $\hat{V}_{22}$ equal to $2\ D^+(\hat{\Omega}_v \otimes \hat{\Omega}_v)D^{+'}$, what only leads to the optimal MDE of $\bar{\theta}$ and $\bar{\tau}$ if the fourth order moments attain their gaussian values.

Letting $\tilde{\psi}_r' = (\tilde{\theta}_r'\ \tilde{\tau}')$ and $\bar{\psi}' = (\bar{\theta}'\ \bar{\tau}')$ we have the asymptotic normality result $\sqrt{N}(\tilde{\psi}_r - \bar{\psi}) \overset{d}{\rightarrow} N(0,\ \Psi_r^{-1})$ where

$$(5.2.10) \quad \Psi_r = \left(\frac{\partial w(\psi)}{\partial \psi'}\Big|_{\bar{\psi}} \right)' V^{-1} \left(\frac{\partial w(\psi)}{\partial \psi'}\Big|_{\bar{\psi}} \right) \text{ (see Appendix 5.A)}.$$

Corresponding to θ and τ we define partitions in Ψ_r and Ψ_r^{-1}

$$\Psi_r = \begin{pmatrix} \Psi_{r,11} & \Psi_{r,12} \\ \Psi_{r,21} & \Psi_{r,22} \end{pmatrix}, \quad \Psi_r^{-1} = \begin{pmatrix} \Psi_r^{11} & \Psi_r^{12} \\ \Psi_r^{21} & \Psi_r^{22} \end{pmatrix} .$$

In this context, we shall say that particular estimators of $\bar{\theta}$ and $\bar{\tau}$ are efficient if they attain the same AVM as $\tilde{\theta}_r$ and $\tilde{\tau}$, Ψ_r^{11} and Ψ_r^{22} respectively. We have assumed $\Delta_{v,4}$ to be finite and unrestricted; alternatively we could consider a set of constraints in Δ_4 (e.g. in the random effects model this can be done by assuming independence between error components and that the fourth order moments are homoscedastic over time) in which case more efficient estimators could be obtained by including a further set of statistics in the definition of (5.2.7). Thus, the estimators we propose, represent a feasible compromise: if constraints are enforced in fourth order moments, determining their sampling variances would require the evaluation of eighth order moments. The same reasoning could apply to the latter and the process would have no end.

If the errors are normally distributed Ψ_r equals $\Phi_r = \text{plim}_{N\to\infty}[-(1/N)\partial^2 L_r/\partial\psi\partial\psi'|_{\bar{\psi}}]$ where L_r is the log-likelihood function associated to our model. But in general, $\tilde{\psi}_r$ will be efficient relative to the QML estimator since the latter is asymptotically equivalent to the MDE whose norm converges in probability to

$$\begin{bmatrix} \bar{\Omega}_v^{-1} \otimes M & 0 \\ 0 & \tfrac{1}{2} D'(\bar{\Omega}_v^{-1} \otimes \bar{\Omega}_v^{-1})D \end{bmatrix}$$

which is not optimal.

The relation between the reduced form and the structural form covariance matrices can be written in vector form as

(5.2.11) $w_2(\theta,\tau) = F(\theta)\ \omega(\tau)$

where $\omega(\tau) = \nu[\Omega(\tau)]$ and

(5.2.12) $F(\theta) = D^+[B^{-1}(\theta) \otimes B^{-1}(\theta)]D.$

Notice that this case where structural covariances are not functionally related to slope parameters must be distinguished from the more general situation where

$$w_2(\theta,\tau) = F(\theta)\ \omega(\tau,\theta).$$

We are primarily concerned with the former case and in that context it will be shown that there exist two stage methods that are asymptotically equivalent to the estimators found by minimising (5.2.7). Nevertheless, the distance function (5.2.7) is equally appropriate for the latter case (an example of which, in panel data, are models b discussed in Chapter 2).

As a final remark, it is worth considering the MDE when Ω is unrestricted. In this case $\omega = \tau$ so that $w_2(\theta,\omega) = F(\theta)\omega$, which does not restrict w_2. Hence the joint criterion function (5.2.7a) becomes

(5.2.13) $s(\theta,\omega) = s_1(\theta) + [\hat{w}_2 - F(\theta)\omega]'\ V_{22}^{-1}[\hat{w}_2 - F(\theta)\omega].$

Differentiating with respect to ω yields

$$\frac{\partial s(\theta,\omega)}{\partial\omega} = -\ 2\ F'(\theta)\ \hat{V}_{22}^{-1}[\hat{w}_2 - F(\theta)\omega] = 0$$

but since $F'(\theta)\ \hat{V}_{22}^{-1}$ is an $\frac{1}{2}\,n(n+1) \times \frac{1}{2}\,n(n+1)$ nonsingular matrix, the MDE of ω is just

$$(5.2.14) \quad \tilde{\omega} = F^{-1}(\theta)\ \hat{w}_2$$

or equivalently

$$(5.2.15) \quad \tilde{\Omega} = B(\theta)\ \hat{\Omega}_v\ B'(\theta)\ .$$

Substitution of (5.2.14) in (5.2.13) immediately reveals that the concentrated distance function is simply $s_1(\theta)$. The minimiser of $s_1(\theta)$, $\tilde{\theta}_u$ say, is the standard Malinvaud's minimum distance estimator of a simultaneous equations model without covariance constraints, which is well known to be asymptotically equivalent to the 3SLS and Ω-unrestricted QML estimators of $\bar{\theta}$, $\hat{\theta}_{3SLS}$ and $\hat{\theta}_u$. Alternatively, the 'QML-type' of estimator of Ω unconstrained takes the form

$$(5.2.16) \quad \hat{\Omega} = A(\theta)\ (\frac{X'X}{N})\ A'(\theta).$$

Both $\tilde{\Omega}$ and $\hat{\Omega}$ can be regarded as functions of θ which provide a range of asymptotically equivalent estimators of $\bar{\Omega}$ when evaluated indistinctly at $\tilde{\theta}_u$, $\hat{\theta}_u$ or $\hat{\theta}_{3SLS}$. The AVM of these estimators is given by $W_{u,22}$ in (3.4.3).

5.3 The Asymptotic Variance Matrix of the Optimal Joint MDE

In what follows, we derive explicit expressions of the partitions of Ψ_r and Ψ_r^{-1} that relate these matrices to the results given in Chapter 3. First, note that since

$$\left.\frac{\partial w(\psi)}{\partial \psi'}\right|_{\bar{\psi}} = \begin{pmatrix} \frac{\partial \bar{w}_1}{\partial \theta'} & O \\ \frac{\partial \bar{w}_2}{\partial \theta'} & \frac{\partial \bar{w}_2}{\partial \tau'} \end{pmatrix} , \quad V^{-1} = \begin{pmatrix} V_{11}^{-1} & O \\ O & V_{22}^{-1} \end{pmatrix} ,$$

from (5.2.10) we have

$$\Psi_{r,11} = \left(\frac{\partial \bar{w}_1}{\partial \theta'}\right)' V_{11}^{-1} \left(\frac{\partial \bar{w}_1}{\partial \theta'}\right) + \left(\frac{\partial \bar{w}_2}{\partial \theta'}\right)' V_{22}^{-1} \left(\frac{\partial \bar{w}_2}{\partial \theta'}\right),$$

$$\Psi_{r,12} = \left(\frac{\partial \bar{w}_2}{\partial \theta'}\right)' V_{22}^{-1} \left(\frac{\partial \bar{w}_2}{\partial \tau'}\right) ,$$

$$\Psi_{r,22} = \left(\frac{\partial \bar{w}_2}{\partial \tau'}\right)' V_{22}^{-1} \left(\frac{\partial \bar{w}_2}{\partial \tau'}\right) .$$

The matrices of coefficient partial derivatives evaluated at the true values are given by

(5.3.1) $$\frac{\partial \bar{w}_1}{\partial \theta'} = - (\bar{B}^{-1} \otimes P^{*\prime}) R,$$

(5.3.2) $$\frac{\partial \bar{w}_2}{\partial \theta'} = \bar{F} \Phi_{u,22}^{-1} \Phi_{u,21} ,$$

(5.3.3) $$\frac{\partial \bar{w}_2}{\partial \tau'} = \bar{F} G ,$$

where $\bar{F} = F(\bar{\theta})$, $\Phi_{u,22}^{-1}\, \Phi_{u,21}$ is given in (3.3.4), an P^*, R and G are defined in Section 3.2.[2] It is convenient to re-write V_{22} as a function of the structural form fourth order moments, thus we have

$$(5.3.4) \qquad V_{22} = \Delta_{v,4} - \bar{w}_2\, \bar{w}_2' = \bar{F}(\Delta_4 - \bar{\omega}\, \bar{\omega}')\bar{F}' \quad .$$

Hence substituting (5.2.4) and (5.3.1) to (5.3.4) in our previous expressions, since all terms $\bar{F}$ cancel, we have

$$(5.3.5) \qquad \Psi_{r,11} = (\Phi_u^{11})^{-1} + \Phi_{u,12}\, \Phi_{u,22}^{-1}(\Delta_4 - \bar{\omega}\, \bar{\omega}')^{-1}\, \Phi_{u,22}^{-1}\, \Phi_{u,21} \, ,$$

$$(5.3.6) \qquad \Psi_{r,12} = \Phi_{u,12}\, \Phi_{u,22}^{-1}(\Delta_4 - \bar{\omega}\, \bar{\omega}')^{-1}\, G \quad ,$$

$$(5.3.7) \qquad \Psi_{r,22} = G'(\Delta_4 - \bar{\omega}\, \bar{\omega}')^{-1}\, G \quad .$$

Note that under normality $(\Delta_4 - \bar{\omega}\, \bar{\omega}')^{-1} = \Phi_{u,22}$, so that $\Psi_r = \Phi_r$.

Next, let us define the matrix Ψ_u such that

$$\Psi_{u,11} = \Psi_{r,11} \quad ,$$

$$\Psi_{u,12} = \Phi_{u,12}\, \Phi_{u,22}^{-1}(\Delta_4 - \bar{\omega}\, \bar{\omega}')^{-1} \, ,$$

$$\Psi_{u,22} = (\Delta_4 - \bar{\omega}\, \bar{\omega}')^{-1} \, ,$$

thus

$$(5.3.8) \qquad \Psi_r = \begin{bmatrix} \Psi_{u,11} & \Psi_{u,12}G \\ G'\Psi_{u,21} & G'\Psi_{u,22}G \end{bmatrix} .$$

Using the formulae for partitioned inverses it is straightforward to check that Ψ_u is the inverse of the matrix W_u in Section 3.4. Then

$$(5.3.9) \qquad \Psi_u^{-1} = \begin{bmatrix} W_{u,11} & W_{u,12} \\ W_{u,21} & W_{u,22} \end{bmatrix} .$$

Finally, in view of (5.3.8) and (5.3.9) we may apply the Lemma in Section 3.3 to obtain

$$(5.3.10) \qquad \Psi_r^{11} = W_{u,11} - W_{u,12}\, W_{u,22}^{-1}(I-P_W)\, W_{u,22}(I-P_W')\, W_{u,22}^{-1}\, W_{u,21} ,$$

$$(5.3.11) \qquad \Psi_r^{22} = (G'\, W_{u,22}^{-1}\, G)^{-1}$$

with $\qquad P_W = G(G'\, W_{u,22}^{-1}\, G)^{-1}\, G'\, W_{u,22}^{-1}$.

Ψ_r^{11} and Ψ_r^{22} are the asymptotic covariance matrices of $\tilde{\theta}_r$ and $\tilde{\tau}$, respectively. Remark that $W_{u,11}$ is the AVM of the Ω-unrestricted QML and MD estimator of $\bar{\theta}$, $\hat{\theta}_u$ and $\tilde{\theta}_u$. Thus, since $W_{u,22}$ is a positive definite matrix, from (5.3.10) we always have

$$AVM(\tilde{\theta}_r) < AVM(\tilde{\theta}_u)$$

irrespective of non-normality.

5.4 Efficient MDE of Covariance Parameters and Minimum Chi-square Specification Tests of Covariance Constraints

Chamberlain (1982) advocates the use of MD estimators to impose restrictions on covariance matrices in the context of i.i.d. random vectors with unrestricted mean. Chamberlain shows that in general the QMLE of the constraint parameters is less efficient than the optimal MDE. We show that these results hold true for structural covariance matrices. Furthermore, separate (optimal) MDE of covariance parameters based on efficient unrestricted estimators of Ω are efficient in the sense of attaining the same limiting distribution as the joint estimators defined in Section 5.2.

Thus let us consider the following criterion function

$$(5.4.1) \qquad s(\tau) = [\hat{\omega} - \omega(\tau)]' \, \hat{V}_{\omega}^{-1} [\hat{\omega} - \omega(\tau)]$$

where $\hat{\omega} = \nu(\hat{\Omega})$ and $\hat{\Omega}$ is indistinctly the 3SLS, the MD or the QML estimator of $\bar{\Omega}$ unrestricted. We assume that $\text{plim}_{N\to\infty} \hat{V}_{\omega} = \bar{V}_{\omega}$ is positive definite. Let $\tilde{\tau}_{MD}$ be the minimiser of $s(\tau)$, so that $\tilde{\tau}_{MD}$ solves the following system of equations

$$(5.4.2) \qquad \left(\frac{\partial \omega(\tau)}{\partial \tau'}\right)' \, \hat{V}_{\omega}^{-1} [\hat{\omega} - \omega(\tau)] = 0 \, .$$

In particular, if the restrictions in $\bar{\Omega}$ are linear and homogeneous, i.e. $\omega(\tau) = G\,\tau$ where G is a matrix of known constants, an explicit solution for $\tilde{\tau}_{MD}$ is available

(5.4.3) $\tilde{\tau}_{MD} = (G' \hat{V}_{\omega}^{-1} G)^{-1} G' \hat{V}_{\omega}^{-1} \hat{\omega}$.

Note that since (5.4.3) is linear, $\Omega(\tilde{\tau}_{MD})$ may not be positive definite. Recalling that $\sqrt{N}(\hat{\omega} - \bar{\omega}) \xrightarrow{d} N(0, W_{u,22})$ and in particular, under normality, $W_{u,22} = \Phi_u^{22}$, the asymptotic distribution of $\tilde{\tau}_{MD}$ follows as an application of the results in Appendix 5.A. Thus, we have that $\sqrt{N}(\tilde{\tau}_{MD}-\bar{\tau}) \xrightarrow{d} N(0,W_{\tau}(\bar{V}_{\omega}))$ where

(5.4.4) $W_{\tau}(\bar{V}_{\omega}) = (G' \bar{V}_{\omega}^{-1} G)^{-1} (G' \bar{V}_{\omega}^{-1} W_{u,22} \bar{V}_{\omega}^{-1} G)(G' \bar{V}_{\omega}^{-1} G)^{-1}$

This result applies to general constraints and so G is the matrix of partial derivatives $\partial\omega(\tau)/\partial\tau'$ evaluated at the true values. (5.4.4) makes clear that the optimal choice for $\bar{V}_{\omega}$ is k $W_{u,22}$, where k is an arbitrary real number, in which case W_{τ} reduces to $(G' W_{u,22}^{-1} G)^{-1} = \Psi_r^{22}$, further establishing that the optimal $\tilde{\tau}_{MD}$ is fully efficient.

Now we can refer to our results in (3.3.10) and (3.4.6a). It turns out that the QML estimator of $\bar{\tau}$ is asymptotically equivalent to the MDE that sets $\bar{V}_{\omega}$ equal to k Φ_u^{22}, and therefore it is generally inefficient relative to the optimal MDE. Only under gaussian kurtosis (or in the special case where $W_{u,22}$ is proportional to Φ_u^{22}) QML and optimal MD are asymptotically equivalent. Thus,

(5.4.5) $W_{\tau}(\Phi_u^{22}) = W_{r,22}$,

(5.4.6) $W_{\tau}(W_{u,22}) = \Psi_r^{22}$,

and

$$W_{r,22} \geq \Psi_r^{22} .$$

Minimum Chi-square Specification Tests of Covariance Constraints

The assumption that ω depends on a q x 1 vector of parameters τ imposes $r = \frac{1}{2}\, n(n+1) - q$ restrictions on ω. Suppose that we wish to test this set of constraints. In Chapter 4 we discussed Wald and quasi-likelihood ratio tests; alternatively we can rely on the statistic $N \cdot s(\tilde{\tau}_{MD})$, since we have the following result

$$(5.4.7) \qquad MCS = N[\hat{\omega} - \omega(\tilde{\tau}_{MD})]' \, \hat{W}_{u,22}^{-1} [\hat{\omega} - \omega(\tilde{\tau}_{MD})] \overset{d}{\to} \chi_r^2 \, .$$

Proof: See Chamberlain (1982, Appendix B, Proposition 8).

That is, under the null hypothesis $N \cdot s(\tilde{\tau}_{MD})$ is asymptotically distributed as a χ^2 variate with r degrees of freedom if $\tilde{\tau}_{MD}$ is the optimal MDE and an optimal norm is used in setting up the statistic, i.e. plim $\hat{W}_{u,22} = W_{u,22}$.

The advantage of the statistic MCS is that it does not require explicit expressions of the constraint equations. This feature makes the minimum chi-square tests specially attractive in panel data, where serial covariance matrices are initially expressed in terms of constraint parameters. Moreover, notice that since separate MDE of $\bar{\tau}$ based on $\hat{\Omega}$ are efficient, we do not require Ω-restricted estimates of the slope coefficients in setting up the minimum chi-square statistics.

Finally, suppose that we consider testing an additional set of constraints $\tau = \tau(\kappa)$ where κ is sx1($s \leq q$). Then $\tilde{\kappa}_{MD}$ minimises

$$(5.4.8) \qquad s(\kappa) = [\hat{\omega} - \omega^{+}(\kappa)]' \, \hat{W}^{-1}_{u,22}[\hat{\omega} - \omega^{+}(\kappa)]$$

where $\omega^{+}(\kappa) = \omega[\tau(\kappa)]$. Then if we consider the statistic

$$(5.4.9) \qquad MCS_1 = N[\hat{\omega} - \omega^{+}(\tilde{\kappa}_{MD})]' \, \hat{W}^{-1}_{u,22}[\hat{\omega} - \omega^{+}(\tilde{\kappa}_{MD})]$$

we can show that MCS_1 - MCS is asymptotically distributed as a χ^2 with q-s degrees of freedom independent of MCS. (cf. Chamberlain (1982, Proposition 8')).

5.5 Generalised Least Squares Estimation of Regression Coefficients

In Section 3.2 we noticed that when the restrictions in $A(\theta)$ are linear the QML estimator of $\bar{\theta}$ takes the form of a GLS estimator. This observation, coupled with the results of the previous Section on separate estimators of restricted covariances, suggests to consider GLS estimators of $\bar{\theta}$ based on MDE of $\bar{\tau}$. From the work of Lahiri and Schmidt (1978) we know that GLS estimators based on efficient but unrestricted estimates of $\bar{\Omega}$ are asymptotically equivalent to full information simultaneous equations estimators (e.g. QML, 3SLS); thus if no a priori information on $\bar{\Omega}$ is available, GLS estimators of triangular systems are not too interesting, except perhaps as an algorithm for the computation of the QML estimates. On the other hand, unlike other GLS applications, GLS estimators of triangular systems are only consistent if they use consistent estimators of $\bar{\Omega}$.

Thus let us consider estimators of $\bar{\theta}$ that solve

$$(5.5.1) \qquad R'(\tilde{\Omega}^{-1} \otimes X'X)R\,\tilde{\theta}_{GLS} = R'(\tilde{\Omega}^{-1} \otimes X'X)r$$

where $\tilde{\Omega}$ is such that $\text{plim}_{N\to\infty}\,\tilde{\Omega} = \bar{\Omega}$, R and r are now a matrix and a vector of known constants such that $\text{vec}(\bar{A}) = R\,\bar{\theta} - r$, and let $\tilde{\omega} = \nu(\tilde{\Omega})$. In what follows we discuss the asymptotic properties of $\tilde{\theta}_{GLS}$.

Consistency

We may re-write (5.5.1) as

$$(5.5.2) \qquad R'(\tilde{\Omega}^{-1} \otimes (\frac{X'X}{N}))R(\tilde{\theta}_{GLS} - \bar{\theta}) = -\,R'\,\text{vec}(\tilde{\Omega}^{-1}\,\frac{\bar{U}'X}{N}) \;.$$

where $\bar{U}' = \bar{A}\,X'$. In general, for a simultaneous system the limit in probability of the left hand side of (5.5.2) does not vanish, and thus $\tilde{\theta}_{GLS}$ is not consistent for $\bar{\theta}$. However, this is not the case for a triangular system: using that $\text{plim}(\bar{U}'X/N) = \bar{\Omega}\,\tilde{B}'$ we have

$$\underset{N\to\infty}{\text{plim}}[R'\,\text{vec}(\tilde{\Omega}^{-1}\,\frac{\bar{U}'X}{N})] = R'\,\text{vec}(\tilde{B}') = 0$$ (see note 2 in Chapter 3).

Finally, since $\text{plim}[R'(\tilde{\Omega}^{-1} \otimes (\frac{X'X}{N}))R] = \Phi_{u,11}$ as in (3.2.12), the consistency of $\tilde{\theta}_{GLS}$ is established.

Asymptotic Normality

If we regard (5.5.2) as a vector valued function of ω, using a first order expansion about $\bar{\omega}$ and rescaling we can write

$$(5.5.3) \qquad [R'(\tilde{\Omega}^{-1} \otimes \frac{X'X}{N})R] \sqrt{N}(\tilde{\theta}_{GLS} - \theta) =$$

$$- \frac{1}{\sqrt{N}} R' \operatorname{vec}(\bar{\Omega}^{-1} \bar{U}'X) + [R'(I \otimes \frac{X'\bar{U}}{N})(\Omega_*^{-1} \otimes \Omega_*^{-1})D] \sqrt{N}(\tilde{\omega} - \bar{\omega})$$

where $\omega_* = \nu(\Omega_*)$ lies between $\tilde{\omega}$ and $\bar{\omega}$. Now notice that since $\operatorname{plim} \Omega_* = \bar{\Omega}$, in view of (3.2.13) we have $\operatorname{plim}[R'(I \otimes \frac{X'\bar{U}}{N})(\Omega_*^{-1} \otimes \Omega_*^{-1})D] = -\Phi_{u,12}$ so that

$$(5.5.4) \qquad \Phi_{u,11} \sqrt{N}(\tilde{\theta}_{GLS} - \bar{\theta}) = - \frac{1}{\sqrt{N}} R' \operatorname{vec}(\bar{\Omega}^{-1} \bar{U}'X) - \Phi_{u,12} \sqrt{N}(\tilde{\omega} - \bar{\omega}) + o_p(1).$$

Next, assuming that $\tilde{\Omega} = \Omega(\tilde{\tau}_{MD})$ where $\tilde{\tau}_{MD}$ minimises a distance function of the type (5.4.1), if we define the indempotent matrix $P_{\bar{V}} = G(G' \bar{V}_{\omega}^{-1} G)^{-1} G' \bar{V}_{\omega}^{-1}$, we have

$$(5.5.5) \qquad \sqrt{N}(\omega(\tilde{\tau}_{MD}) - \bar{\omega}) = P_{\bar{V}} \sqrt{N}(\hat{\omega} - \bar{\omega}) + o_p(1).$$

Thus, (5.5.4) becomes

$$(5.5.6) \qquad \Phi_{u,11} \sqrt{N}(\tilde{\theta}_{GLS} - \bar{\theta}) = - \frac{1}{\sqrt{N}} R' \operatorname{vec}(\bar{\Omega}^{-1} \bar{U}'X) - \Phi_{u,12} P_{\bar{V}} \sqrt{N}(\hat{\omega} - \bar{\omega}) + o_p(1)$$

Moreover, it can be proved that[3]

$$-\frac{1}{\sqrt{N}} R' \operatorname{vec}(\bar{\Omega}^{-1} \bar{U}'X) = \Phi_{u,11} \sqrt{N}(\hat{\theta}_u - \bar{\theta}) + \Phi_{u,12} \sqrt{N}(\hat{\omega} - \bar{\omega}) + o_p(1)$$

where $\hat{\theta}_u$ is the Ω-unrestricted QML estimator of $\bar{\theta}$. Hence

$$(5.5.7) \qquad \sqrt{N}(\tilde{\theta}_{GLS} - \bar{\theta}) = \sqrt{N}(\hat{\theta}_u - \bar{\theta}) + \Phi_{u,11}^{-1} \Phi_{u,12}(I-P_{\bar{V}})\sqrt{N}(\hat{\omega} - \bar{\omega}) + o_p(1).$$

Therefore, since we know that $\sqrt{N}\begin{pmatrix} \hat{\theta}_u - \bar{\theta} \\ \hat{\omega} - \bar{\omega} \end{pmatrix} \xrightarrow{d} N(0, W_u)$, $\sqrt{N}(\tilde{\theta}_{GLS} - \bar{\theta})$ is also asymptotically normal and its variance matrix is given by

$$W_\theta(\bar{V}_\omega) = W_{u,11} + \Phi_{u,11}^{-1} \Phi_{u,12}(I-P_{\bar{V}}) W_{u,22}(I-P_{\bar{V}})' \Phi_{u,21} \Phi_{u,11}^{-1}$$

$$+ \Phi_{u,11}^{-1} \Phi_{u,12}(I-P_{\bar{V}}) W_{u,21} + W_{u,12}(I-P_{\bar{V}})' \Phi_{u,21} \Phi_{u,11}^{-1}.$$

Equivalently, in view of (3.4.1) and (3.4.2), $W_{u,11} = \Phi_u^{11}$ and $W_{u,12} = \Phi_u^{12} = -\Phi_{u,11}^{-1} \Phi_{u,12} \Phi_u^{22}$, thus

$$(5.5.8) \qquad W_\theta(\bar{V}_\omega) = \Phi_u^{11} +$$

$$\Phi_{u,11}^{-1} \Phi_{u,12}[(I-P_{\bar{V}}) W_{u,22}(I-P_{\bar{V}}') - \Phi_u^{22}(I-P_{\bar{V}}') - (I-P_{\bar{V}})\Phi_u^{22}]\Phi_{u,21} \Phi_{u,11}^{-1}.$$

(5.5.8) can be used to compute the AVM of different GLS estimators based on particular choices of $\bar{V}_\omega$. In particular, the GLSE that sets $\bar{V}_\omega$ equal to Φ_u^{22} is asymptotically equivalent to $\hat{\theta}_r$ (i.e. the Ω-restricted QMLE of $\bar{\theta}$). This can be easily seen by re-writing (5.5.8) in a way more comparable to (3.4.5); after some manipulation we have

$$(5.5.8a)\quad W_\theta(\bar{V}_\omega) = \Phi_u^{11} - \Phi_{u,11}^{-1}\,\Phi_{u,12}\{(I-P_{\bar{V}})[\Phi_u^{22} - (\Delta_4-\Delta_4^{n})](I-P_{\bar{V}}')$$

$$+ P_{\bar{V}}\,\Phi_u^{22}(I-P_{\bar{V}}') + (I-P_{\bar{V}})\,\Phi_u^{22}\,P_{\bar{V}}'\}\;\Phi_{u,21}\,\Phi_{u,11}^{-1}.$$

Simply noting that $P_\Phi\,\Phi_u^{22}(I-P_{\bar{V}}') = 0$ it follows that

$$(5.5.9)\quad W_\theta(\Phi_u^{22}) = W_{r,11}\quad .$$

Another intuitively relevant choice for $\bar{V}_\omega$ is $W_{u,22}$, that is, the GLSE of $\bar{\theta}$ that uses the optimal MDE of $\bar{\tau}$. Under normality $W_\theta(W_{u,22}) = W_\theta(\Phi_u^{22})$, but in general the matrix $W_\theta(\Phi_u^{22}) - W_\theta(W_{u,22})$ is indefinite. Making use of (5.5.8) after some reductions we have

$$(5.5.10)\quad W_\theta(\Phi_u^{22}) - W_\theta(W_{u,22}) = \Phi_{u,11}^{-1}\,\Phi_{u,12}[(P_W-P_\Phi)\;W_{u,22}(P_W' - P_\Phi')$$

$$+ (P_W-P_\Phi)(W_{u,22}-\Phi_u^{22}) + (W_{u,22}-\Phi_u^{22})(P_W'-P_\Phi')]\Phi_{u,21}\,\Phi_{u,11}^{-1}$$

and although $(P_W-P_\Phi)\;W_{u,22}(P_W'-P_\Phi')$ is positive semi-definite, the matrix in square brackets on the right hand side of (5.5.10) is indefinite. Therefore, none of these two estimators is generally efficient in the sense introduced in Section 5.2.

<u>An Efficient GLS Estimator</u>

Let us consider the GLS estimator $\bar{\bar{\theta}}_{GLS}$ of the type (5.5.1) which is based in the following choice of $\tilde{\omega}$:

$$(5.5.11)\quad \bar{\bar{\omega}} = (I - \hat{\Phi}_u^{22}\,\hat{W}_{u,22}^{-1})\hat{\omega} + (\hat{\Phi}_u^{22}\,\hat{W}_{u,22}^{-1})\omega(\tilde{\tau}_{OMD})$$

where plim $\hat{W}_{u,22} = W_{u,22}$, plim $\hat{\Phi}_u^{22} = \Phi_u^{22}$ and $\tilde{\tau}_{OMD}$ is the optimal MDE of $\bar{\tau}$. Thus $\bar{\bar{\omega}}$ is defined as a matrix-weighted average of $\hat{\omega}$ and $\omega(\tilde{\tau}_{OMD})$. Under normality, $W_{u,22} = \Phi_u^{22}$ and then $\bar{\bar{\omega}} = \omega(\tilde{\tau}_{OMD})$, but as $W_{u,22}$ becomes large relative to Φ_u^{22} an increasing weight is being put on $\hat{\omega}$ relative to $\omega(\tilde{\tau}_{OMD})$. We prove below that the GLS estimator so defined attains the same AVM as the joint MDE $\tilde{\theta}_r$ and is therefore efficient. Note that since $\bar{\bar{\omega}}$ is a linear combination of consistent estimators of $\bar{\omega}$, it is itself a consistent estimator of $\bar{\omega}$; though since $\hat{\omega}$ is an unrestricted estimator of $\bar{\omega}$, $\bar{\bar{\omega}}$ will not satisfy the covariance restrictions.

To obtain the limiting distribution of $\sqrt{N}(\bar{\bar{\theta}}_{GLS} - \bar{\theta})$, we begin by re-writing (5.5.11) as

$$\sqrt{N}(\bar{\bar{\omega}}-\bar{\omega}) = (I-\hat{\Phi}_u^{22}\hat{W}_{u,22}^{-1})\sqrt{N}(\hat{\omega}-\bar{\omega}) + (\hat{\Phi}_u^{22}\hat{W}_{u,22}^{-1})\sqrt{N}[\omega(\tilde{\tau}_{OMD})-\bar{\omega}],$$

furthermore, using (5.5.5) with P_W in place of $P_{\bar{V}}$ we have

$$\text{(5.5.12)} \quad \sqrt{N}(\bar{\bar{\omega}}-\bar{\omega}) = [I-\Phi_u^{22}W_{u,22}^{-1}(I-P_W)]\sqrt{N}(\hat{\omega}-\bar{\omega}) + o_p(1).$$

Now applying (5.5.4) to the present case it turns out that

$$\text{(5.5.13)} \quad \sqrt{N}(\bar{\bar{\theta}}_{GLS}-\bar{\theta}) = \sqrt{N}(\hat{\theta}_u-\bar{\theta}) - W_{u,12}\, W_{u,22}^{-1}(I-P_W)\, \sqrt{N}(\hat{\omega}-\bar{\omega}) + o_p(1) ,$$

where we have used the fact that $W_{u,12} = \Phi_u^{12} = - \Phi_{u,11}^{-1}\, \Phi_{u,12}\, \Phi_u^{22}$.

Finally, from (5.5.13) the AVM of $\bar{\bar{\theta}}_{GLS}$ is given by

$$\text{(5.5.14)} \quad AVM(\bar{\bar{\theta}}_{GLS}) = (I \vdots K_o) \begin{pmatrix} W_{u,11} & W_{u,12} \\ W_{u,21} & W_{u,22} \end{pmatrix} \begin{pmatrix} I \\ K_o' \end{pmatrix}$$

with

$$K_o = - W_{u,12} W_{u,22}^{-1} (I-P_w);$$

after some manipulation (5.5.14) reduces to

$$\text{(5.5.15)} \quad AVM(\bar{\bar{\theta}}_{GLS}) = W_{u,11} - W_{u,12} W_{u,22}^{-1} (I-P_w) W_{u,22} (I-P_w') W_{u,22}^{-1} W_{u,21}$$

which is equivalent to Ψ_r^{11} in (5.3.10).

5.6 Subsystem Estimation

We now consider the case in which a subset of n_1 equations of the complete model (3.2.1) are unrestricted reduced form equations. Thus, defining the partitions $\theta' = (\theta_1' \; \theta_2')$, $\tau' = (\tau_1' \; \tau_2')$, we have

$$\bar{B} = \begin{pmatrix} I & O \\ B_{21}(\bar{\theta}_2) & B_{22}(\bar{\theta}_2) \end{pmatrix}, \quad \bar{C} = \begin{pmatrix} \bar{C}_1 \\ C_2(\bar{\theta}_2) \end{pmatrix},$$

$$\bar{\Omega} = \begin{pmatrix} \bar{\Omega}_{11} & \bar{\Omega}_{12} \\ \bar{\Omega}_{21} & \bar{\Omega}_{22}(\bar{\tau}_2) \end{pmatrix}.$$

$\bar{\Omega}_{11}$, $\bar{\Omega}_{12}$ and $\bar{C}_1$ are unrestricted matrices whose coefficients can be arranged in the vectors $\bar{\tau}_1$ and $\bar{\theta}_1$, and $B_{22}(\bar{\theta}_2)$ is lower trianguler. (3.2.1) then becomes

$$(5.6.1) \qquad \underset{}{y_{1i}} + \underset{n_1\times k}{\bar{C}_1} z_i = \underset{n_1\times 1}{\bar{u}_{1i}} \quad ,$$

$$(5.6.2) \qquad \underset{n_2\times n_1}{\bar{B}_{21}} y_{1i} + \underset{n_2\times n_2}{\bar{B}_{22}} y_{2i} + \underset{n_2\times k}{\bar{C}_2} z_i = A_2(\bar{\theta}_2)\, x_i = \underset{n_2\times 1}{\bar{u}_{2i}}$$

where $A_2(\bar{\theta}_2) = (\bar{B}_{21}\ \bar{B}_{22}\ \bar{C}_2)$ and $n_1+n_2=n$. Moreover let us assume that the restrictions in A_2 are linear so that

$$R = \begin{pmatrix} R_1 & O \\ O & R_2 \end{pmatrix}, \; r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}, \; \mathrm{vec}(A_2) = R_2\,\theta_2 - r_2 \;;$$

in particular note that $R_1 = K\begin{pmatrix} O \\ I \end{pmatrix}$ where K is the commutation matrix of order $n_1(n_1+k)$, O is a $n_1^2 \times n_1k$ matrix of zeroes and I is a n_1k unit matrix. Also let us introduce a notation for the partition of Ω^{-1}

$$\Omega^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} \\ \Omega^{21} & \Omega^{22} \end{pmatrix}.$$

Clearly the estimator that solves

$$(5.6.3) \qquad R_2'(\tilde{\Omega}_{22}^{-1} \otimes X'X)R_2\,\tilde{\theta}_2 = R_2'(\tilde{\Omega}_{22}^{-1} \otimes X'X)r_2$$

(with plim $\tilde{\Omega}_{22} = \bar{\Omega}_{22}$) is inconsistent for $\bar{\theta}_2$ unless $\bar{\Omega}_{12} = O$.[4]

On the other hand since $\bar{\Omega}$ is positive definite we can write $\bar{\Omega} = P^{-1} P'^{-1}$ where P is lower triangular. Transforming the original complete model by mean of P leads to a recursive system,

$$P \begin{pmatrix} I & O \\ \bar{B}_{21} & \bar{B}_{22} \end{pmatrix} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} + \begin{pmatrix} P_{11} \bar{C}_1 \\ P_{21} \bar{C}_1 + P_{22} \bar{C}_2 \end{pmatrix} z_i = P \begin{pmatrix} \bar{u}_{1i} \\ \bar{u}_{2i} \end{pmatrix}$$

where $P = \begin{pmatrix} P_{11} & O \\ P_{21} & P_{22} \end{pmatrix}$.

Therefore, in the transformed recursive system the coefficients in the two blocks of equations are functionally related. This makes clear the impossibility of obtaining subsystem least squares estimates from the transformed model that take into account all the restrictions in the original system.

However we still can solve the complete system of GLS equations for the subset of estimates of $\bar{\theta}_2$ as follows. Let us consider the partitions

$$(5.6.4) \quad \begin{pmatrix} R_1'(\tilde{\Omega}^{11} \otimes X'X)R_1 & R_1'(\tilde{\Omega}^{12} \otimes X'X)R_2 \\ R_2'(\tilde{\Omega}^{21} \otimes X'X)R_1 & R_2'(\tilde{\Omega}^{22} \otimes X'X)R_2 \end{pmatrix}^{-1} = \begin{pmatrix} H^{11} & H^{12} \\ H^{21} & H^{22} \end{pmatrix},$$

$$(5.6.5) \quad \begin{pmatrix} R_1'(\tilde{\Omega}^{11} \otimes X'X)r_1 + R_1'(\tilde{\Omega}^{12} \otimes X'X)r_2 \\ R_2'(\tilde{\Omega}^{21} \otimes X'X)r_1 + R_2'(\tilde{\Omega}^{22} \otimes X'X)r_2 \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

Then, using $H^{21} = - H^{22} H_{21} H_{11}^{-1}$, the GLS estimator of $\bar{\theta}_2$ is given by

$$(5.6.6) \qquad \tilde{\theta}_{2,GLS} = H^{22}(h_2 - H_{21} H_{11}^{-1} h_1) \ .$$

Furthermore, since $H^{22} = (H_{22} - H_{21} H_{11}^{-1} H_{12})^{-1}$, and noting that in view of the form of R_1 we have $R_1'(\tilde{\Omega}^{11} \otimes X'X)R_1 = Z'Z \otimes \tilde{\Omega}^{11}$, we end up with

$$(H^{22})^{-1} = R_2'\{(\tilde{\Omega}^{22} \otimes X'X) - [\tilde{\Omega}^{21}(\tilde{\Omega}^{11})^{-1}\tilde{\Omega}^{12} \otimes (X'Z)(Z'Z)^{-1}(Z'X)]\}R_2 \ .$$

Thus, after some manipulation we obtain

$$(5.6.7) \qquad \tilde{\theta}_{2,GLS} = (R_2' \, \Delta \, R_2)^{-1}[R_2' \, \Delta \, r_2 + R_2'(\tilde{\Omega}^{21} \otimes S_{11})]$$

where

$$(5.6.8) \qquad \Delta = (\tilde{\Omega}^{22} \otimes \frac{X'X}{N}) - [\tilde{\Omega}^{21}(\tilde{\Omega}^{11})^{-1} \tilde{\Omega}^{12} \otimes (\frac{X'Z}{N})(\frac{Z'Z}{N})^{-1}(\frac{Z'X}{N})]$$

and

$$(5.6.9) \qquad S_{11} = \left(\frac{Y_1'Y_1}{N}\right) - \left(\frac{Y_1'Z}{N}\right)\left(\frac{Z'Z}{N}\right)^{-1}\left(\frac{Z'Y_1}{N}\right).$$

Finally, we can illustrate the general expression in (5.6.7) by considering the GLS estimator of the second equation in the simple model of Section 3.5. This turns out to be

$$(5.6.10) \quad \tilde{\delta}_{GLS} = [\kappa(X'X) + (1-\kappa)(X'Z)(Z'Z)^{-1}(Z'X)]^{-1}$$

$$[\kappa(X'y_2) + (1-\kappa)(X'Z)(Z'Z)^{-1}(Z'y_2) + \tilde{\omega}_{22}\,\tilde{\omega}^{21}\,N\,s_{11}\,d_1]$$

where

$$\kappa = \frac{1}{1-\rho^2}, \quad \rho^2 = \frac{\tilde{\omega}_{12}^2}{\tilde{\omega}_{11}\tilde{\omega}_{22}}, \quad d_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

and

$$s_{11} = \frac{1}{N}[y_1'y_1 - y_1'Z(Z'Z)^{-1}\,Z'y_1] \ .$$

Note that if we are using a restricted estimator of $\tilde{\Omega}$, $\tilde{\omega}_{12}=0$, and thus $\kappa=1$ so that (5.6.10) reduces to the O.L.S. estimator.

NOTES

1 The results given in Sections 5.4 and 5.5 are a revised version of those given in Arellano (1985).

2 In the derivation of (5.3.1), (5.3.2) and (5.3.3) we have made use of the result $d\, vec(A^{-1}) = - (A \otimes A')^{-1}\, d\, vec(A)$, for non-singular A. The matrices of partial derivatives are immediately determined from the differential by noting that for an n x 1 vector valued function ϕ of an m x 1 vector θ we have

$$d\, \phi_j(\theta) = \sum_{k=1}^{m} \frac{\partial \phi_j(\theta)}{\partial \theta_k} d\, \theta_k \qquad (j=1,\ldots,n)$$

or equivalently $d\, \phi(\theta) = \frac{\partial \phi(\theta)}{\partial \theta'} d\, \theta$.

3 Using $X = Z\, P^{*\prime} + \bar{\tilde{U}}B'$ and (3.2.13) we have

$$R'\, vec(\bar{\Omega}^{-1}\, \bar{U}'X) = R'\, vec(\bar{\Omega}^{-1}\, \bar{U}'ZP^{*\prime}) - \Phi_{u,12}\, \nu(\bar{U}'\bar{U}) \; ;$$

moreover since $\Phi_{u,12}\, \bar{\omega} = 0$ for a triangular model (see note 2, Chapter 3), this is equivalent to

$$\frac{1}{\sqrt{N}} R'vec(\bar{\Omega}^{-1}\bar{U}'X) = \frac{1}{\sqrt{N}} R'vec(\bar{\Omega}^{-1}\bar{U}'ZP^{*\prime}) - \Phi_{u,12}\, \sqrt{N}[\frac{1}{N}\, \nu(\bar{U}'\bar{U}) - \bar{\omega}] .$$

Now using

$$\sqrt{N}(\hat{\theta}_u-\bar{\theta}) = -\Phi_u^{11}\frac{1}{\sqrt{N}}R'\mathrm{vec}(\bar{\Omega}^{-1}\bar{U}'ZP*') + o_p(1)$$

and noting that a first order expansion of $\sqrt{N}(\hat{\omega}-\bar{\omega})$ about $\bar{\theta}$ yields

$$\sqrt{N}(\hat{\omega}-\bar{\omega}) = \sqrt{N}[\frac{1}{N}\nu(\bar{U}'\bar{U}) - \bar{\omega}] - \Phi_{u,22}^{-1}\Phi_{u,21}\sqrt{N}(\hat{\theta}_u-\bar{\theta}) + o_p(1) ,$$

we have

$$-\frac{1}{\sqrt{N}}R'\mathrm{vec}(\bar{\Omega}^{-1}\bar{U}'X) = [(\Phi_u^{11})^{-1} + \Phi_{u,12}\Phi_{u,22}^{-1}\Phi_{u,21}]\sqrt{N}(\hat{\theta}_u-\bar{\theta})$$

$$+ \Phi_{u,12}\sqrt{N}(\hat{\omega}-\bar{\omega}) + o_p(1)$$

from which the result claimed follows.

4 (5.6.3) can be re-written as

$$R_2'(\tilde{\Omega}_{22}^{-1}\otimes\frac{X'X}{N})R_2(\tilde{\theta}_2-\bar{\theta}) = -R_2'(\tilde{\Omega}_{22}^{-1}\otimes\frac{X'X}{N})\mathrm{vec}(\bar{A}_2)$$

$$= -R_2'\mathrm{vec}(\tilde{\Omega}_{22}^{-1}\frac{\bar{U}_2'X}{N}) .$$

Then noting that plim $(\bar{U}_2'X/N) = \mathrm{plim}(\bar{U}_2'\bar{U}\tilde{B}'/N)$ we have that

$$\underset{N\to\infty}{\mathrm{plim}}\, R_2'\mathrm{vec}(\tilde{\Omega}_{22}^{-1}\frac{\bar{U}_2'X}{N}) = R_2'\mathrm{vec}[(\bar{\Omega}_{22}^{-1}\bar{\Omega}_{21} \vdots I)\tilde{B}'] .$$

After some reductions it turns out that a typical element of this vector takes the form

$$\mathrm{tr}\left[\left(\frac{\partial \bar{B}_{21}}{\partial \theta_k} - \frac{\partial \bar{B}_{22}}{\partial \theta_k} \bar{B}_{22}^{-1} \bar{B}_{21}\right) \bar{\Omega}_{12} \bar{\Omega}_{22}^{-1}\right]$$

where θ_k is the kth element of θ_2; despite the triangularity of $\bar{B}_{22}^{-1}$ and $\partial \bar{B}_{22}/\partial \theta_k$, in general these elements do not vanish unless $\bar{\Omega}_{12} = 0$.

APPENDIX 5.A

The Limiting Distribution of the Minimum Distance Estimator

Let $\hat{p}$ be an unconstrained estimator of the s x 1 coefficient vector $\bar{p}$, such that

(5.A.1) $$\operatorname*{plim}_{N\to\infty} \hat{p} = \bar{p} \,,$$

(5.A.2) $$\sqrt{N}(\hat{p} - \bar{p}) \overset{d}{\to} N(0,V_p) \,.$$

Assume that $\bar{p}$ depends on a set of constraint parameters $\bar{\delta}$, $\bar{p} = p(\bar{\delta})$. We further assume that $p(\delta) = p(\bar{\delta})$ for some δ in the parameter space implies that $\delta = \bar{\delta}$, and that $p(\delta)$ has continuous second partial derivatives in a neighborhood of $\bar{\delta}$. It is also assumed that $\bar{D} = D(\bar{\delta}) = \partial p(\delta)/\partial\delta'|_{\bar{\delta}}$ has full column rank.

Let $\tilde{\delta}$ be the minimiser of the distance function

(5.A.3) $$s(\delta) = [\hat{p} - p(\delta)]' \, Q[\hat{p} - p(\delta)]$$

where Q is an sxs matrix such that plim $Q = \bar{Q}$ exists and is positive definite. (5.A.1) and our identification assumption ensure the consistency of $\tilde{\delta}$ for $\bar{\delta}$. By the definition of $\tilde{\delta}$, $\partial\, s(\delta)/\partial\delta|\tilde{\delta} = 0$ so that a first order expansion of $\partial\, s(\delta)/\partial s$ about $\bar{\delta}$ yields

(5.A.4) $$-\left(\frac{\partial^2\, s(\delta)}{\partial\delta\partial\delta'}\bigg|_{\delta*}\right) \sqrt{N}(\tilde{\delta}-\bar{\delta}) = \sqrt{N}\,\frac{\partial\, s(\delta)}{\partial\delta}\bigg|_{\bar{\delta}} \,.$$

where δ^* lies between $\tilde{\delta}$ and $\bar{\delta}$. Now since

$$(5.A.5) \qquad \left.\frac{\partial\, s(\delta)}{\partial \delta}\right|_{\bar{\delta}} = -\,2\,\bar{D}'\,Q[\hat{p} - p(\bar{\delta})]$$

in view of (5.A.2) we have

$$(5.A.6) \qquad \sqrt{N}\,\left.\frac{\partial\, s(\delta)}{\partial \delta}\right|_{\bar{\delta}} \overset{d}{\rightarrow} N(0,\ 4\,\bar{D}'\bar{Q}\,V_p\,\bar{Q}\bar{D}).$$

On the other hand, since plim $\delta^* = \bar{\delta}$, direct evaluation shows that

$$(5.A.7) \qquad \underset{N\to\infty}{\text{plim}} \left\{ \left.\frac{\partial^2\, s(\delta)}{\partial\delta\,\partial\delta'}\right|_{\delta^*} \right\} = 2\,\bar{D}'\,\bar{Q}\,\bar{D}$$

Hence, using the Cramer linear transformation theorem

$$(5.A.8) \qquad \sqrt{N}(\tilde{\delta} - \bar{\delta}) \overset{d}{\rightarrow} N(0,\ V_\delta)$$

where

$$(5.A.9) \qquad V_\delta = (\bar{D}'\,\bar{Q}\,\bar{D})^{-1}(\bar{D}'\,\bar{Q}\,V_p\,\bar{Q}\,\bar{D})(\bar{D}'\,\bar{Q}\,\bar{D})^{-1}\ .$$

Clearly, an optimal choice for $\bar{Q}$ is V_p^{-1}, in which case the asymptotic covariance matrix of $\tilde{\delta}$ reduces to

$$(5.A.10) \qquad V_\delta = (\bar{D}'\,V_p^{-1}\,\bar{D})^{-1}\ .$$

CHAPTER 6

EFFICIENT MD AND GLS ESTIMATORS APPLIED TO PANEL DATA

6.1 Introduction

The methods developed in Chapter 5 suggest an estimation and modelling strategy for dynamic models from panel data. Regression specification analyses can be based on Ω-unrestricted 3SLS or QML estimates of alternative versions of the model under consideration (though 3SLS estimates have the advantage of not requiring iterative optimisation). Once a particular specification has been chosen we can proceed to estimate and test different structures for Ω using minimum distance estimators and minimum chi-square or Wald tests. If eventually a particular covariance specification is not rejected, this information can be used to obtain more efficient estimates of the regression parameters by mean of the GLS procedure discussed in Sections 5.5 and 5.6. A computer program has been written in Fortran 77 to perform the calculations involved in this sequence.

This Chapter discusses the application of the results in Chapter 5 to panel data, and also various calculations are performed to assess the practical performance of the proposed methods. Section 6.2 presents the analytical results and in Section 6.3 a simulation is carried out and the Michigan earnings function is re-estimated.

6.2 The Estimators

We begin by considering in some detail the MD estimation of first order moving average covariance structures. We use the same notation as in Chapters 2 and 4. The T x T random effects MA(1) covariance matrix may alternatively be parameterised as

$$(6.2.1) \qquad \Omega = \begin{pmatrix} g_1 & g_2 & g_3 & \cdots & g_3 \\ g_2 & g_1 & g_2 & \cdots & g_3 \\ g_3 & g_2 & g_1 & \cdots & g_3 \\ \vdots & & & & \vdots \\ g_3 & g_3 & g_3 & \cdots & g_1 \end{pmatrix}$$

where

$$g_1 = \sigma^2(1+\lambda^2) + \sigma_\eta^2 ,$$

$$g_2 = \sigma^2 \lambda + \sigma_\eta^2 ,$$

$$g_3 = \sigma_\eta^2 .$$

λ, σ^2 and σ_η^2 can be easily retrieved from g_1, g_2 and g_3 by noting that λ solves $\lambda^2 - c\lambda + 1 = 0$, with $c = (g_1-g_3)/(g_2-g_3)$, so that

$$(6.2.2) \qquad \lambda = \tfrac{1}{2}(c \pm\sqrt{c^2-4}) ,$$

$$(6.2.3) \qquad \sigma^2 = \frac{g_2-g_3}{\lambda} ,$$

(6.2.4) $\sigma_\eta^2 = g_3$.

The indeterminacy of λ and σ^2 is eliminated by choosing the solution for λ that lies inside the unit circle.

Since the restrictions implicit in (6.2.1) are linear so is the MD estimator of $g' = (g_1\ g_2\ g_3)$. On the other hand, since we are assuming that the elements $\omega_{00}, \omega_{01}, \ldots, \omega_{0T}$ of the top row of Ω^* are unrestricted, the MD estimator of g, $\tilde{g}_{MD}$, remains unaffected if we drop these elements from the distance function. In order to obtain an explicit expression for $\tilde{g}_{MD}$ it is convenient to introduce a permutation of $\omega^* = \nu(\Omega^*)$ as follows; let Π_p be a permutation of the rows of the $\frac{1}{2}(T+1)(T+2)$ unit matrix such that

$$\omega_p^* = \Pi_p\ \omega^*$$

where

$$\omega_p^* = \begin{bmatrix} \omega_p \\ \\ \omega_0 \end{bmatrix}, \quad \omega_0' = (\omega_{00}, \omega_{10}, \ldots, \omega_{T0})$$

and

$$\omega_p' = (\omega_{11}, \ldots, \omega_{TT}, \omega_{21}, \ldots, \omega_{T(T-1)}, \omega_{31}, \ldots, \omega_{T(T-2)}, \ldots, \omega_{T1}).$$

ω_p contains the same coefficients as $\nu(\Omega)$ but now they are ordered by diagonals.

In Section 4.2 we established that the AVM of any asymptotically efficient estimator of ω^* unrestricted is given by

$$W_{\omega\omega} = \text{Avar}(\hat{\alpha}_u)\, qq' + \Delta_4 - \omega^* \omega^{*'}$$

where

$$q = \nu(B_\alpha B^{-1} \Omega^* + \Omega^* B'^{-1} B'_\alpha) \;, \; B_\alpha = - \frac{\partial B(\alpha)}{\partial \alpha} \;.$$

Accordingly, the AVM of $\hat{\omega}^*_p$ is given by $\Pi_p W_{\omega\omega} \Pi'_p = W^*_{pp}$, say. Moreover, let us introduce the partition

$$W^*_{pp} = \begin{pmatrix} W_{pp} & W_{pO} \\ W_{Op} & W_{OO} \end{pmatrix}$$

where W_{pp} is $\frac{1}{2}T(T+1) \times \frac{1}{2}T(T+1)$ and W_{OO} is $(T+1) \times (T+1)$ corresponding, respectively, to the AVM's of $\hat{\omega}_p$ and $\hat{\omega}_O$. Now we can write (6.2.1) as

$$\text{(6.2.5)} \qquad \omega_p = \begin{pmatrix} \omega_{11} \\ \vdots \\ \omega_{TT} \\ \hline \omega_{21} \\ \vdots \\ \omega_{T(T-1)} \\ \hline \omega_{31} \\ \vdots \\ \omega_{T1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = G\, g.$$

$\tilde{g}_{MD}$ minimises

$$s(g) = (\hat{\omega}_p - G\ g)'\ \hat{W}_{pp}^{-1}(\hat{\omega}_p - G\ g)$$

thus obtaining

$$(6.2.6) \qquad \tilde{g}_{MD} = (G'\ \hat{W}_{pp}^{-1}\ G)^{-1}\ G\ \ \hat{W}_{pp}^{-1}\ \hat{\omega}_p$$

where $\hat{W}_{pp}$ is a consistent estimator of W_{pp} and $\hat{\omega}_p$ is an asymptotically efficient estimator of $\hat{\omega}_p$ unrestricted. Incidentally, note that if we replace $\hat{W}_{pp}^{-1}$ by an unit matrix we obtain a 'crude minimum distance' estimator; this estimator is consistent although inefficient, and in view of the form of the matrix G, crude MD estimators of g_1, g_2 and g_3 are simple arithmetic means given by

$$\tilde{g}_{1,CMD} = \frac{1}{T} \sum_{t=1}^{T} \hat{\omega}_{tt}\ ,\ \tilde{g}_{2,CMD} = \frac{1}{(T-1)} \sum_{t=2}^{T} \hat{\omega}_{t(t-1)}\ ,$$

$$\tilde{g}_{3,CMD} = \frac{1}{T^*} (\omega_{31} + \ldots + \omega_{T1})$$

where $T^* = \frac{1}{2}\,T(T+1) - T - (T-1)$. We further remark that (6.2.6) does not impose the restrictions $g_3 > 0$ and $(g_1 - g_3) > 0$ and thus it is not guaranteed that the implied estimates for σ^2 and σ_η^2 will be positive. However, significantly negative values of $\tilde{\sigma}^2$ and/or $\tilde{\sigma}_\eta^2$ will commonly be an indication of misspecification. A minimum chi-square test of the structure (6.2.1) is given by

$$(6.2.7) \qquad MCS = N(\hat{\omega}_p - G\,\tilde{g}_{MD})'\,\hat{W}_{pp}^{-1}(\hat{\omega}_p - G\,\tilde{g}_{MD}) \;.$$

MCS is distributed as a χ^2 with ½ T(T+1)-3 degrees of freedom under the null hypothesis.

Once $\tilde{g}_{MD}$ has been obtained, we may proceed to calculate the MD estimator of ω_O, $\tilde{\omega}_O$. This is only necessary if the slope parameters are going to be re-estimated by GLS; $\tilde{\omega}_O$ is given by

$$(6.2.8) \qquad \tilde{\omega}_O = \hat{\omega}_O - \hat{W}_{Op}\,\hat{W}_{pp}^{-1}(\hat{\omega}_p - G\,\tilde{g}_{MD})$$

where $\hat{W}_{Op}$ is a consistent estimator of W_{Op} and $\hat{\omega}_O$ is an asymptotically efficient Ω^*-unrestricted estimator of ω_O.

The previous results can be generalised to higher order moving average schemes. If we consider an sth order moving average case (with $s < T - 1$), Ω will depend on $s + 2$ constraint parameters

$$(6.2.9) \qquad \begin{aligned} g_1 &= \sigma^2(1 + \lambda_1^2 + \ldots + \lambda_s^2) + \sigma_\eta^2 \;, \\ g_2 &= \sigma^2(\lambda_1 + \lambda_1\lambda_2 + \ldots + \lambda_{s-1}\lambda_s) + \sigma_\eta^2 \;, \\ &\;\vdots \\ g_{s+1} &= \sigma^2\,\lambda_s + \sigma_\eta^2 \\ g_{s+2} &= \sigma_\eta^2 \end{aligned}$$

where $v_{it} = \varepsilon_{it} + \lambda_1\,\varepsilon_{i(t-1)} + \ldots + \lambda_s\,\varepsilon_{i(t-s)}$. Again, $\lambda_1, \ldots, \lambda_s$, σ^2 and σ_η^2 can be retrieved from $g_1, \ldots, g_{s+2}$ by solving the nonlinear system (6.2.9).

However, note that if the only purpose in enforcing the restrictions in Ω is to obtain GLS estimates of the slope parameters, there is no need to retrieve the moving average coefficients.

Finally, if autoregressive or ARMA schemes have to be estimated, the corresponding distance function will have to be minimised by mean of iterative techniques.

GLS Estimation of α, β and γ

This is a particular case of the problem of subsystem estimation studied in Section 5.6; there is only one unrestricted reduced form equation and it corresponds to the prediction equation for y_0. Let

$$(6.2.10) \quad \bar{\bar{\omega}}^* = \nu(\bar{\bar{\Omega}}^*) = (I - \hat{\Xi}_{\omega\omega} \hat{W}_{\omega\omega}^{-1})\hat{\omega}^* + \hat{\Xi}_{\omega\omega} \hat{W}_{\omega\omega}^{-1} \omega^*(\tilde{g}_{MD}, \tilde{\omega}_0)$$

where $\hat{\Xi}_{\omega\omega}$ is a consistent estimate of $\Xi_{\omega\omega}$ as given in (4.2.9). $\bar{\bar{\Omega}}^*$ is the estimate of Ω^* that leads to the optimal GLS of $\delta' = (\alpha\ \beta'\gamma')$ robust to non-normality (NNGLS). Moreover, let us introduce the partition

$$\bar{\bar{\Omega}}^{*-1} = \begin{pmatrix} \bar{\bar{\omega}}^{00} & \bar{\bar{\omega}}^{01} \\ \bar{\bar{\omega}}^{10} & \bar{\bar{\Omega}}^{11} \end{pmatrix}$$

and let $\bar{\bar{\Omega}}^{11} = \{\bar{\bar{\omega}}^{ts}\}$ and $\bar{\bar{\omega}}^{01} = (\bar{\bar{\omega}}^{01}, \ldots, \bar{\bar{\omega}}^{0T})$. Then, direct application of the results in (5.6.7), (5.6.8) and (5.6.9) gives

$$(6.2.11) \quad \tilde{\delta}_{GLS} = (X^{+'}\ \Psi\ X^{+})^{-1}(X^{+'}\ \Psi\ y + \psi\ d_1)$$

where

$$\Psi = (I_N \otimes \bar{\bar{\Omega}}^{11}) - [Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} \otimes \bar{\bar{\omega}}^{10}\,\bar{\bar{\omega}}^{01}/\bar{\bar{\omega}}^{00}],$$

$$\psi = \sum_{t=1}^{T} \bar{\bar{\omega}}^{0t}\, y'_{t-1}(I - Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime})y_o \;,$$

and d_1 is an (n+m+1) vector with one in the first position and zero elsewhere. Finally, in view of the discussion in Section 2.4, a computationally more convenient expression for $\tilde{\delta}_{GLS}$ is given by

$$\text{(6.2.12)} \quad \tilde{\delta}_{GLS} = \left[\sum_{t=1}^{T}\sum_{s=1}^{T} (\bar{\bar{\omega}}^{ts}\, x_t^{+\prime}\, x_s^{+} - \frac{\bar{\bar{\omega}}^{t0}\,\bar{\bar{\omega}}^{s0}}{\bar{\bar{\omega}}^{00}}\, \hat{x}_t^{+\prime}\, \hat{x}_s^{+}) \right]^{-1}$$

$$\left[\sum_{t=1}^{T}\sum_{s=1}^{T} (\bar{\bar{\omega}}^{ts}\, x_t^{+\prime}\, y_s - \frac{\bar{\bar{\omega}}^{t0}\,\bar{\bar{\omega}}^{s0}}{\bar{\bar{\omega}}^{00}}\, x_t^{+\prime}\, y_s) + \psi\, d_1 \right]$$

6.3 Numerical Results

Two Monte Carlo experiments were conducted in order to investigate the performance of GLS and MD estimators, particularly the magnitude of the finite sample efficiency increase that results from covariance restrictions. The performance of minimum chi-square tests was also examined. The present experiments were based on the same model we used in Section 4.4. However, a shorter number of replications (30 samples) were generated in this case due to CPU time limitations. Tables 1, 2 and 3 summarise the results. For the normal model, the biases in the 3SLS estimates are of a similar magnitude to the biases in the QML estimates corresponding to the moving average data reported in Chapter 2; however, the finite sample variances appear to be smaller in the case of the 3SLS estimator. On the other hand, GLS estimators that use the a priori information on the covariance matrix have in general a smaller variance than the 3SLS estimator, although the reduction in variance varies considerably from one parameter to another. For example, considering the non-normal model, the variance of the NNGLS estimate of α is cut by an amount of 22 percent of its 3SLS value, but the variance reduction relative to NGLS is of only 2 percent. In the case of the intercept, the NNGLS variance is 14 percent less than its 3SLS variance and 6.5 percent less than the NGLS variance, which is also the case for γ_1. In the case of β there is no reduction at all.

Turning to MD estimates of covariance parameters (Table 2), in the nonnormal experiment the NNMD variances of σ_η^2, σ^2 and λ are reduced by an amount of 49.5, 1 and 24 percent of their NMD variances, respectively. However, the biases in the NNMD estimates of σ_η^2 and σ^2 are respectively 4 times and 2 times larger than the corresponding NMD biases, whereas in the normal experiment they are roughly the same.

Table 3 reports the results concerning the minimum chi-square tests. The performance of the MCS and the NMCS tests turns out to be rather similar to that of the Wald and normal-Wald tests studied in Chapter 4. MCS is slightly upward biased, particularly in the non-normal case, while NMCS shows a bias in the opposite direction in the experiment with normal data. On the other hand, when the errors are long-tailed the mean and variance of NMCS are far beyond the χ^2 values, as expected.

Finally, as an illustration we have re-estimated the earnings function for the US discussed in Section 4.5. Table 4 presents Ω-unrestricted 3SLS estimates, and Tables 5 and 6 present the NNGLS and NGLS estimates of the slope coefficients and the NNMD and NMD estimates of the covariance parameters. The mean effects of the explanatory variables are rather stable for the different methods of estimation and there are no noticeable differences in relation with our previous QML estimates. However, the estimated coefficient of the lagged dependent variable is smaller in the present case, what reflects the lack of identification to which we referred in Section 4.5. Turning to minimum chi-square tests, the values of MCS and NMCS for Model 1 are rather similar to those found earlier for the Wald and the normal-Wald statistics and therefore the conclusions are also the same; namely, that if proper account is taken of the non-normality of the errors the first order moving average restrictions are not rejected at the 90 percent level.

TABLE 1

Biases in the Estimates of the Slope Parameters[a]

Parameter	Model with Normal Errors CIV	3SLS	NGLS[b]	NNGLS[c]	Model with Long-tailed Errors CIV	3SLS	NGLS	NNGLS
γ_0	-.0881 (.0132)[d]	-.0247 (.0115)	-.0170 (.0098)	-.0218 (.0108)	-.0805 (.0116)	-.0177 (.0100)	-.0006 (.0096)	-.0158 (.0093)
γ_1	-.0050 (.0034)	.0018 (.0036)	.0024 (.0035)	.0014 (.0035)	-.0018 (.0033)	.0055 (.0036)	.0067 (.0036)	.0048 (.0034)
β	-.0075 (.0017)	.0004 (.0017)	.0007 (.0017)	.0006 (.0018)	-.0072 (.0016)	.0001 (.0017)	.0009 (.0017)	.0000 (.0017)
α	.0294 (.0036)	.0065 (.0031)	.0043 (.0025)	.0057 (.0028)	.0266 (.0034)	.0043 (.0028)	-.0007 (.0025)	.0038 (.0025)

[a] N = 500, T = 9, 30 replications

[b] Efficient GLS estimator under normality

[c] Efficient GLS estimator robust to non-normality

[d] Standard errors of bias.

TABLE 2

Biases in the Estimates of the Covariance Parameters[a]

Parameter	Model with Normal Errors NMD[b]	Model with Normal Errors NNMD[c]	Model with Long-Tailed Errors NMD	Model with Long-Tailed Errors NNMD
σ_η^2	-.0074 (.0033)[d]	-.0086 (.0035)	-.0061 (.0066)	-.0244 (.0047)
σ^2	-.0095 (.0012)	-.0094 (.0014)	-.0330 (.0024)	-.0680 (.0024)
λ	-.0033 (.0038)	-.0028 (.0040)	.0036 (.0039)	-.0001 (.0034)

a N = 500, T = 9, 30 replications

b Efficient MD estimator under normality

c Efficient MD estimator robust to non-normality

d Standard errors of bias.

TABLE 3

Simulation Results for Minimum Chi-Square Tests

	Model with Normal Errors		Model with Long-Tailed Errors	
	NMCS[a]	MCS[b]	NMCS	MCS
Size	Number of Rejections out of 30 cases[c]			
0.10	2	8	30	8
0.05	1	3	30	7
0.01	0	2	29	2
Mean	40.176	45.467	102.723	48.876
Variance	77.115	109.175	629.642	124.993

a Minimum Chi-square Test under normality

b Robust Minimum Chi-square Test

c According to a chi-square with 42 degrees of freedom.

TABLE 4

Three Stage Least Squares Estimates

Dependent Variable: Log Hourly Earnings[a]

	Model 1		Model 2	
	Estimates	Derived Mean Effects[e]	Estimates	Derived Mean Effects
Years of Education	.0168 (.0039)[b]	.0755	.0167 (.0037)	.0532
Age	.0076 (.0035)	.0340	.0107 (.0038)	.0341
Age squared	-.000072 (.000036)	-.0003	-.000103 (.000040)	-.0003
Race[c]	.0332 (.0160)	.1497	.0470 (.0193)	.1494
Occupation[d]	-	-	.0768 (.0183)	.2445
Lagged dependent variable	.7779 (.0460)		.6857 (.0549)	

a Data in mean deviation form (N=742, T=9, period 1967-1976)

b Standard errors in parentheses

c Dummy variable: 1 if individual is white

d Dummy variable: 1 if individual belongs to professional or managerial groups in 1967

e Calculated as $\hat{\gamma}_k^* = \hat{\gamma}_k/(1 - \hat{\alpha})$.

TABLE 5

NNGLS Estimates of Slope Parameters and NNMD Estimates of Covariance Parameters

Dependent Variable: Log Hourly Earnings

	Model 1		Model 2	
	Estimates	Derived Mean Effects	Estimates	Derived Mean Effects
Years of Education	.0166[a]	.0751	.0166	.0530
Age	.0075	.0342	.0104	.0332
Age squared	-.000072	-.0003	-.00010	-.0003
Race	.0333	.1506	.0470	.1502
Occupation	-	-	.0759	.2428
Lagged dependent variable	.7789		.6873	
λ	-.3983 (.0381)		-.3661 (.0433)	
ρ^2 [b]	.1303			
σ^2	.0478 (.0026)		.0463 (.0025)	
σ^2_η	.0062 (.0020)	$\sigma^2_{\eta*}$[c]=.1275	.0094 (.0028)	$\sigma^2_{\eta*}$=.0958
MCS[d]	51.1	(D.F.=42)	54.4	

a S.E. of NNGLS Estimates have not been calculated; however, 3SLS S.E. in Table 4 provide an upper bound for NNGLS S.E.

b $\rho^2 = \sigma^2_\eta/\sigma^2$

c $\sigma^2_{\eta*} = \sigma^2_\eta/(1-\alpha)^2$

d Robust Minimum Chi-Square Tests.

TABLE 6

NGLS Estimates of Slope Parameters and NMD Estimates of Covariance Parameters

Dependent Variable: Log Hourly Earnings

	Model 1		Model 2	
	Estimates	Derived Mean Effects	Estimates	Derived Mean Effects
Years of Education	.0153	.0745	.0133	.0518
Age	.0073	.0356	.0087	.0336
Age squared	-.000071	-.0003	-.000085	-.0003
Race	.0294	.1428	.0366	.1421
Occupation	-	-	.0621	.2413
Lagged dependent variable	.7940		.7425	
λ	-.3966 (.0250)[a]		-.3663 (.0285)	
ρ^2	.0931		.1450	
σ^2	.0612 (.0014)		.0594 (.0013)	
σ^2_η	.0057 (.0017)	$\sigma^2_{\eta*}$=.1341	.0086 (.0024)	$\sigma^2_{\eta*}$=.1299
NMCS[b]	140.5	(D.F.=42)	143.6	

a Reported S.E. of NMD estimates are only consistent under normality

b Minimum Chi-square Test under normality.

CONCLUSION

This thesis has presented methods of estimation and tests of specification for dynamic econometric models from panel data when the errors are serially correlated and the number of time periods is small. We have derived the asymptotic properties of such estimators in the context of general triangular systems with covariance restrictions when normality holds and also when the errors are non-normal. Throughout, the quasi-maximum likelihood framework has proved useful from a theoretical point of view in organising the relevant discussion.

QML estimators are also of interest in view of their satisfactory performance in Monte Carlo experiments, which is further supported by the results in our empirical application using the Michigan data. Nevertheless, there exists a full information minimum distance estimator that is never less efficient than the QML and is strictly better when the assumption of normality is false. Moreover, we have developed separate MD estimators of covariance parameters and GLS estimators of slope coefficients that are asymptotically equivalent to the full information MD, thus providing a computationally simpler alternative for the efficient estimation of dynamic models from panel data.

Tests of covariance specification ought not to be based upon the assumption of normality. Robust Wald and minimum chi-square tests as well as appropriate probability limits for the quasi-likelihood ratio test have been proposed and successfully applied in testing the serial correlation structure to earnings equations for the US.

A comprehensive treatment of the statistical problems posed by models with non-exogenous explanatory variables or variables with measurement errors still has to be done. But this will be the purpose of future research.

REFERENCES

Ali, M.M. (1974): "Stochastic Ordering and Kurtosis Measure", Journal of the American Statistical Association, 69, 543-545.

Anderson, T.W. and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components", Journal of the American Statistical Association, 76, 598-606.

Anderson, T.W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data", Journal of Econometrics, 18, 47-82.

Arellano, M. (1983): "Estimation of Dynamic Random Effects Models with Serially Correlated Time-Varying Errors", Paper presented at the European Meeting of the Econometric Society, Pisa.

Arellano, M. (1984): "Testing Random Effects Specifications in Dynamic Models for Panel Data", DEMEIC Paper, London School of Economics.

Arellano, M. (1985): "GLS and QML Estimation of Triangular Models with Covariance Restrictions", DEMEIC Paper, London School of Economics.

Balestra, P. and M. Nerlove (1966): "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", Econometrica, 34, 585-612.

Bhargava, A., L. Franzini and W.Narendranathan (1982): "Serial Correlation and the Fixed Effects Model", Review of Economic Studies, 49, 533-549.

Bhargava, A. and J.D. Sargan (1982): "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods", Discussion Paper no.71, University of Florida.

Bhargava, A. and J.D. Sargan (1983): "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods", Econometrica, 51, 1635-1659.

Box, G.E.P. (1953): "Non-normality and Tests on Variances", Biometrika, 40, 318-335.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data", Review of Economic Studies, 47, 225-238.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data", Journal of Econometrics, 18, 5-46.

Chamberlain, G. (1984): "Panel Data", in Z. Griliches and M.D. Intriligator (eds.), Handbook of Econometrics, Volume II, Elsevier Science Publishers.

Chiang, C.L. (1956): "On Regular Best Asymptotically Normal Estimates", Annals of Mathematical Statistics, 27, 336-351.

Chowdhury, G. and S. Nickell (1982): "Individual Earnings in the U.S. Another Look at Unionization, Schooling, Sickness and Unemployment Using Panel Data", Discussion Paper No.141, Centre for Labour Economics, London School of Economics.

Ferguson, T.S. (1958): "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities", Annals of Mathematical Statistics, 29, 1046-1062.

Foutz, R.V. and R.C. Srivastava (1977): "The Performance of the Likelihood Ratio Test when the Model is Incorrect", The Annals of Statistics, 5, 1183-1194.

Hammersley, J.M. and D.C. Handscomb (1964): Monte Carlo Methods, Methuen, London.

Hausman, J.A. and W.E. Taylor (1981): "Panel Data and Unobservable Individual Effects", Econometrica, 49, 1377-1398.

Hausman, J.A., W. Newey and W.E. Taylor (1983): "Efficient Estimation and Identification of Simultaneous Equation Models with Covariance Restrictions", mimeo.

Hendry, D.F. (1984): "Monte Carlo Experimentation in Econometrics", in Z. Griliches and M.D. Intriligator (eds.), Handbook of Econometrics, Volume II, Elsevier Science Publishers.

Hendry, D.F. and R.W. Harrison (1974): "Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-Stage Least Squares", Journal of Econometrics, 2, 151-174.

Imhof, J.P. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables", Biometrika, 48, 419-426.

Kiefer, N.M. (1980): "Estimation of Fixed Effect Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance", Journal of Econometrics, 14, 195-202.

Koerts, J. and A.P.J. Abrahamse (1969): On the Theory and Application of the General Linear Model, University Press, Rotterdam.

Koopmans, T.C., H. Rubin and R.B. Leipnik (1950): "Measuring the Equation Systems of Dynamic Economics", in T.C. Koopmans (ed.), Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10, John Wiley, New York.

Lahiri, K. and P. Schmidt (1978): "On the Estimation of Triangular Structural Systems", Econometrica, 46, 1217-1221.

Lillard, L.A. and R.J. Willis (1978): "Dynamic Aspects of Earnings Mobility", Econometrica, 46, 985-1012.

MaCurdy, T.E. (1981): "Asymptotic Properties of Quasi-Maximum Likelihood Estimators and Test Statistics", National Bureau of Economic Research Technical Paper no.14.

MaCurdy, T.E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", Journal of Econometrics, 18, 83-114.

Maddala, G.S. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data", Econometrica, 39, 341-358.

Magnus, J.R. and H. Neudecker (1980): "The Elimination Matrix: Some Lemmas and Applications", SIAM Journal of Algebraic and Discrete Methods, 1, 422-449.

Malinvaud, E. (1970): Statistical Methods of Econometrics, North-Holland, Amsterdam.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data", Econometrica, 46, 69-85.

Nag Library of Computer Programs, Mark IX, Oxford University Computing Library, Oxford.

Nerlove, M. (1967): "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections", Economic Studies Quarterly, 18, 42-74.

Nerlove, M. (1971): "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections", Econometrica, 39, 359-382.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects", Econometrica, 49, 1417-1426.

Pudney, S.E. (1979): "The Estimation and Testing of Some Error Components Models", Discussion Paper, London School of Economics, 1979.

Rao, C.R. (1973): Linear Statistical Inference and its Applications, Second Edition, John Wiley, New York.

Richard, J.-F. (1975): "A Note on the Information Matrix of the Multivariate Normal Distribution", Journal of Econometrics, 3, 57-60.

Rothenberg, T.J. (1973): Efficient Estimation with A Priori Information, Cowles Foundation Monograph 23, Yale University Press, New Haven.

Rothenberg, T.J. and C.T. Leenders (1964): "Efficient Estimation of Simultaneous Equation Systems", Econometrica, 32, 57-76.

Sargan, J.D. (1964): "Three-Stage Least-Squares and Full Maximum Likelihood Estimates", Econometrica, 32, 77-81.

Sargan, J.D. (1975): "The Identification and Estimation of Sets of Simultaneous Stochastic Equations", London School of Economics Econometrics Programme Discussion Paper A.1.

Sargan, J.D. and A. Bhargava (1983): "Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root Lies on the Unit Circle", Econometrica, 51, 799-820.

Tiao, G.C. and M.M. Ali (1971): "Analysis of Correlated Random Effects: Linear Model with Two Random Components", Biometrika, 58, 37-51.

White, H. (1980): "Using Least Squares to Approximate Unknown Regression Functions", International Economic Review, 21, 149-170.

White, H. (1980a): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", Econometrica, 48, 817-838.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1-25.