ESTIMATION OF LINEAR PANEL DATA MODELS USING GMM

Seung Chan Ahn
Arizona State University, Tempe, AZ 85287, USA

Peter Schmidt
Michigan State University, E. Lansing, MI 48824, USA

August 1997
Revised: October 1998

Abstract

In this chapter we study GMM estimation of linear panel data models. Several different types of models are considered, including the linear regression model with strictly or weakly exogenous regressors, the simultaneous regression model, and a dynamic linear model containing a lagged dependent variable as a regressor. In each case, different assumptions about the exogeneity of the explanatory variables generate different sets of moment conditions that can be used in estimation. This chapter lists the relevant sets of moment conditions and gives some results on simple ways in which they can be imposed. In particular, attention is paid to the question of under what circumstances the efficient GMM estimator takes the form of an instrumental variables estimator.
Use of panel data regression methods has become increasingly popular as the availability of longitudinal data sets has grown. Panel data contain repeated time-series observations (T) for a large number (N) of cross-sectional units (e.g., individuals, households, or firms). An important advantage of using such data is that they allow researchers to control for unobservable heterogeneity, that is, systematic differences across cross-sectional units. Regressions using aggregated time-series and pure cross-section data are likely to be contaminated by these effects, and statistical inferences obtained by ignoring these effects could be seriously biased. When panel data are available, error-components models can be used to control for these individual differences. Such a model typically assumes that the stochastic error term has two components: a time-invariant individual effect which captures the unobservable individual heterogeneity and the usual random noise term. Some explanatory variables (e.g., years of schooling in the earnings equation) are likely to be correlated with the individual effects (e.g., unobservable talent or IQ). A simple treatment to this problem is the within estimator which is equivalent to least squares after transformation of the data to deviations from means.

Unfortunately, the within method has two serious defects. First, the within transformation of a model wipes out time invariant regressors as well as the individual effect, so that it is not possible to estimate the effects of time-invariant regressors on the dependent variable. Second, consistency of the within estimator requires that all the regressors in a given model be strictly
exogenous with respect to the random noise. The within estimator could be inconsistent for models in which regressors are only weakly exogenous, such as dynamic models including lagged dependent variables as regressors. In response to these problems, a number of studies have developed alternative GMM estimation methods.

In this chapter we provide a systematic account of GMM estimation of linear panel data models. Several different types of models are considered, including the linear regression model with strictly or weakly exogenous regressors, the simultaneous regression model, and a dynamic linear model containing a lagged dependent variable as a regressor. In each case, different assumptions about the exogeneity of the explanatory variables generate different sets of moment conditions that can be used in estimation. This chapter lists the relevant sets of moment conditions and gives some results on simple ways in which they can be imposed. In particular, attention is paid to the question of under what circumstances the efficient GMM estimator takes the form of an instrumental variables estimator.¹

### 8.1 Preliminaries

In this section we introduce the general model of interest, some basic assumptions, and some notation. Given linear moment conditions, we consider the efficient GMM estimator and other related instrumental variables estimators. We also examine general conditions under which the efficient GMM estimator is of an instrumental variables form.

#### 8.1.1 General Model

The models we examine in this chapter are of the following common form:

¹Here and throughout this chapter, "efficient" means "asymptotically efficient."
Here $i = 1, \ldots, N$ indexes the cross-sectional unit (individual) and $t = 1, \ldots, T$ indexes time.

The dependent variable is $y_{it}$, while $x_{it}$ is a $1 \times p$ vector of explanatory variables. The $p \times 1$ parameter vector $\beta$ is unknown. The composite error $u_{it}$ contains a time invariant individual effect $\alpha_i$ and random noise $\epsilon_{it}$. We assume that $E(\alpha_i) = 0$ and $E(\epsilon_{it}) = 0$ for any $i$ and $t$; thus, $E(u_{it}) = 0$. The time dimension $T$ is held fixed, so that usual asymptotics apply as $N$ gets large.

We also assume that the data $\{(y_{i1}, y_{i2}, \ldots, y_{iT}, x_{i1}, \ldots, x_{iT})' | i = 1, \ldots, N\}$ are independently and identically distributed (iid) over different $i$ and have finite population moments up to fourth order. Under this assumption, any sample moment (up to fourth order) of the data converges to its counterpart population moment in probability, e.g., $\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} x_{it}' y_{it} = E(x_{it}' y_{it})$.

Some matrix notation is useful throughout this chapter. For any single variable $c_{it}$ or row vector $d_{it}$, we denote $c_i \equiv (c_{i1}, \ldots, c_{iT})'$ and $D_{it} \equiv (d_{i1}', \ldots, d_{iT}')'$. Accordingly, $y_i$ and $X_i$ denote the data matrices of $T$ rows. In addition, for any $T \times 1$ vector $c_i$ or $T \times k$ matrix $D_i$, we denote $c \equiv (c_1', \ldots, c_N')'$ and $D = (D_1', \ldots, D_N')'$. Accordingly, $y$ and $X$ denote the data matrices of $NT$ rows. With this notation, we can rewrite equation (8-1) for individual $i$ as

$$y_i = X_i \beta + u_i; u_i = e_T \otimes \alpha_i + \epsilon_{it}.$$  \hspace{1cm} (8-2)

where $e_T$ is a $T \times 1$ vector of ones, and all $NT$ observations as $y = X\beta + u$.

We treat the individual effects $\alpha_i$ as random, so that we can define $\Sigma \equiv \text{Cov}(u_i) = E(u_i u_i')$. A standard and popular assumption about $\Sigma$ is so-called the random effects structure

$$\Sigma = \sigma_a^2 e_T e_T' + \sigma_e^2 I_T,$$  \hspace{1cm} (8-3)

which arises if the $\epsilon_{it}$ are iid over time and independent of $\alpha_i$. This covariance structure often takes an important role in GMM estimation as we discuss below.
There are two well-known special cases of the model (8-2); the traditional random effects and fixed effects models. Both of these models assume that the regressors $X_i$ are strictly exogenous with respect to the random noise $\varepsilon_i$ (i.e., $E(x_i^t \varepsilon_s^i) = 0$ for any $t$ and $s$). The random effects model (Balestra and Nerlove [1966]) treats the individual effects as random unobservables which are uncorrelated with all of the regressors. Under this assumption, the parameter $\beta$ can be consistently and efficiently estimated by generalized least squares (GLS):

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$

where $\Omega = I_N \otimes \Sigma$.

In contrast, when we treat the $\alpha_i$ as nuisance parameters, the model (8-2) reduces to the traditional fixed effects model. A simple treatment of the fixed effects model is to remove the effects by the (within) transformation of the model (8-2) to deviations from individual means:

$$Q_T' y_i = Q_T' X_i \beta + Q_T' \varepsilon_i = Q_T' X_i \beta + Q_T' \varepsilon_i,$$

where $Q_T = I_T - P_T$, $P_T = T^{-1}e_T e_T'$, and the last equality results from $Q_T e_T = 0$. Least squares on (8-4) yields the familiar within estimator:

$$\hat{\beta}_w = (X'Q_v X)^{-1}X'Q_v y,$$

where $Q_v = I_N \otimes Q_T$.

Although the fixed effects model views the effects $\alpha_i$ as nuisance parameters rather than random variables, the fixed effects treatment (within estimation) is not inconsistent with the random effects assumption. Mundlak [1978] considers an alternative random effects model in which the effects $\alpha_i$ are allowed to be correlated with all of the regressors $x_{i1},...,x_{iT}$. For this model, Mundlak shows that the within estimator is an efficient GLS estimator. This finding implies that the core difference between the random and fixed effects models is not whether the effects are literally random or nuisance parameters, but whether the effects are correlated or uncorrelated with the regressors.

### 8.1.2 GMM and Instrumental Variables
In this subsection we examine GMM and other related instrumental variables estimators for the model (8-2). Our main focus is a general treatment of given moment conditions, so we do not make any specific exogeneity assumption regarding the regressors \(X_i\). We simply begin by assuming that there exists a set of \(T \times k\) instruments \(Z_i\) which satisfies the moment condition

\[E(Z_i'u_i) = 0 \tag{8-5}\]

and the usual identification condition, \(\text{rank}[E(Z_i'X_i)] = p\).

Under (8-5) and other usual regularity conditions, a consistent and efficient estimate of \(\beta\) can be obtained by minimizing the GMM criterion function \(N(y-X\hat{\beta})'Z(V_N)^{-1}Z(y-X\hat{\beta})\), where \(V_N\) is any consistent estimate of \(V \equiv E(Z_i'u_iu_i'Z_i)\). A simple choice of \(V_N\) is

\[N^{-1}\sum_{i=1}^{N} Z_i\hat{u}_i\hat{u}_i'Z_i,\]

where \(\hat{u}_i = y_i - X_i\hat{\beta}\) and \(\hat{\beta}\) is an initial consistent estimator such as two-stage least squares (2SLS). The solution to the minimization leads to the GMM estimator:

\[\hat{\beta}_{GMM} \equiv (X'Z(V_N)^{-1}Z'X)^{-1}X'Z(V_N)^{-1}Z'y.\]

An instrumental variables estimator, which is closely related with this GMM estimator, is three stages least squares (3SLS):

\[\hat{\beta}_{3SLS} \equiv (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'y,\]

where \(\Omega = I_N \otimes \Sigma\). For notational convenience, we assume that \(\Sigma\) is known, although, in practice, it should be replaced by a consistent estimate such as \(\hat{\Sigma} = N^{-1}\sum_{i=1}^{N} \hat{u}_i\hat{u}_i'\). In order to understand the relationship between the GMM and 3SLS estimators, consider the following condition

\[E(Z_i'u_iu_i'Z_i) = E(Z_i'u_i'u_i'Z_i) = V. \tag{8-6}\]

Under this condition, the 3SLS estimator is asymptotically identical to the GMM estimator, because \(\text{plim}_{N \to \infty} N^{-1}Z'\Omega Z = \text{plim}_{N \to \infty} N^{-1}\sum_{i=1}^{N} Z_i\hat{u}_i\hat{u}_i'Z_i = E(Z_i'u_i'u_i'Z_i) = V\). We will refer to (8-6) as
the condition of no conditional heteroskedasticity (NCH). This is a slight misuse of
terminology, since (8-6) is weaker than the condition that \( E(u_i u_i' | Z_i) = \Sigma \). However, (8-6) is
what is necessary for the 3SLS estimator to coincide with the GMM estimator. When (8-6) is violated, \( \hat{\beta}_{GMM} \) is strictly more efficient than \( \hat{\beta}_{3SLS} \).

An alternative to the 3SLS estimator, which is popularly used in the panel data literature, is the 2SLS estimator obtained by premultiplying (8-2) by \( \Sigma^{-\frac{1}{2}} \) to filter \( u_i \), and then applying the instruments \( Z_i \):

\[
\hat{\beta}_{FIV} = \left[ X'\Omega^{-\frac{1}{2}}Z(Z'Z)^{-1}Z'\Omega^{-\frac{1}{2}}X \right]^{-1} X'\Omega^{-\frac{1}{2}}Z(Z'Z)^{-1}Z'\Omega^{-\frac{1}{2}}y .
\]

We refer to this estimator as the filtered instrumental variables (FIV) estimator. This estimator is slightly different from the generalized instrumental variables (GIV) estimator, which is originally proposed by White [1984]. The GIV estimator is also 2SLS applied to the filtered model \( \Sigma^{-\frac{1}{2}}y_i = \Sigma^{-\frac{1}{2}}X_i\beta + \Sigma^{-\frac{1}{2}}u_i \), but it uses the filtered instruments \( \Sigma^{-\frac{1}{2}}X_i \). Thus

\[
\hat{\beta}_{GIV} = \left[ X'\Omega^{-\frac{1}{2}}Z(Z'\Omega^{-\frac{1}{2}}Z)^{-1}Z'\Omega^{-\frac{1}{2}}X \right]^{-1} X'\Omega^{-\frac{1}{2}}Z(Z'\Omega^{-\frac{1}{2}}Z)^{-1}Z'\Omega^{-\frac{1}{2}}y .
\]

Despite this difference, the FIV and GIV estimators are often equivalent in the context of panel data models, especially when \( \Sigma \) is of the random effects structure (8-3). We may also note that the FIV and GIV estimators would be of little interest without the NCH assumption.

A motivation for the FIV (or GIV) estimator is that filtering the error \( u_i \) may improve the efficiency of instrumental variables, as GLS improves upon ordinary least squares (OLS).\(^2\)

However, neither of the 3SLS nor FIV estimators can be shown to generally dominate the other. This is so because the FIV estimator is a GMM estimator based on the different moment condition \( E(Z_i\Sigma^{-\frac{1}{2}}u_i) = 0 \) and the different NCH assumption \( E(Z_i\Sigma^{-\frac{1}{2}}u_i'\Sigma^{-\frac{1}{2}}Z_i) = E(Z_i'Z_i) \).

We now turn to conditions under which the 3SLS and FIV estimators are numerically

\(^2\)White [1984] offers some strong conditions under which the GIV estimator dominates the 3SLS estimator in terms of efficiency.
equivalent, whenever the same estimate of \( \Sigma \) is used.

**Theorem 8.1**

Suppose that there exists a nonsingular, nonstochastic matrix \( B \) such that \( \Sigma^{-\frac{1}{2}} Z_i = Z_i B \) for all \( i \) (that is, \( \Omega^{-\frac{1}{2}} Z = ZB \)). Then, \( \hat{\beta}_{\text{FIV}} = \hat{\beta}_{\text{3SLS}} \).

The proof is omitted because it is straightforward. We note that the numerical equivalence result of Theorem 8.1 holds only if the same estimate of \( \Sigma \) is used for the two estimators \( \hat{\beta}_{\text{FIV}} \) and \( \hat{\beta}_{\text{3SLS}} \). However, even if different (consistent) estimates of \( \Sigma \) are used, the two estimators remain asymptotically identical.

The main point of Theorem 8.1 is that under certain assumption, filtering does not change the efficiency of instrumental variables or GMM. When Theorem 8.1 holds but the instruments \( Z_i \) violate the NCH condition (8-6), both the FIV and 3SLS estimators are strictly dominated by the GMM estimator applied without filtering. Clearly, Theorem 8.1 imposes strong restrictions on the instruments \( Z_i \) and the covariance matrix \( \Sigma \), which do not generally hold. Nonetheless, in the context of some panel data models, the theorem can be used to show that filtering is irrelevant for GMM or 3SLS exploiting all of the moment conditions. We consider a few examples below.

**8.1.2.1 Strictly Exogenous Instruments and Random Effects**

Consider a model in which there exists a \( 1 \times k \) vector of instruments \( h_{it} \) which are strictly exogenous with respect to the \( \epsilon_{it} \) and uncorrelated with \( \alpha_i \). For this model, we have the moment conditions

\[
E(h_{it}' h_{it}) = \mathbf{0}, \; s, t = 1, ..., T. \tag{8-7}
\]
This is a set of $kT^2$ moment conditions. Denote $h^i_t = (h_{1i}, \ldots, h_{ti})$ for any $t = 1, \ldots, T$, and set $Z_{SE,i} \equiv I_T \otimes h^i_{iT}$, so that all of the moment conditions (8-7) can be expressed compactly as $E(Z_{SE,i}' u_i) = 0$.

We now show that filtering the error $u_i$ does not matter in GMM. Observe that for any $T \times T$ nonsingular matrix $A$,

$$AZ_{SE,i} = A(I_N \otimes h^i_{iT}) = A \otimes h^i_{iT} = (I_N \otimes h^i_{IT})(A \otimes I_k) = Z_{SE,i} B,$$

where $B = A \otimes I_k$. If we replace $A$ by $\Sigma^{-\frac{1}{2}}$, Theorem 8.1 holds.

### 8.1.2.2 Strictly Exogenous Instruments and Fixed Effects

We now allow the instruments $h^i_t$ to be correlated with $\alpha_i$ (fixed effects), while they are still assumed to be strictly exogenous with respect to the $\varepsilon^i_t$. For this case, we may first-difference the model (8-2) to remove $\alpha_i$:

$$L_T'y_i = L_T'X_i\beta + L_T'u_i = L_T'X_i\beta + L_T'\varepsilon_i,$$

where $L_T$ is the $T \times (T-1)$ differencing matrix

$$L_T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}.$$

We note that $L_T$ has the same column space as the deviations from means matrix $Q_T$ (in fact, $Q_T = L_T (L_T' L_T)^{-1} L_T'$). This reflects the fact that first differences and deviations from means preserve the same information in the data.

Note that strict exogeneity of the instruments $h^i_t$ with respect to the $\varepsilon^i_t$ implies $E(Z_{SEFE,i}' L_T'u_i) = 0$, where $Z_{SEFE,i} = I_{T-1} \otimes h^i_{iT}$. Thus, the model (8-8) can be estimated by GMM using the instruments $Z_{SEFE,i}$. Once again, given the NCH condition, filtering does not matter for GMM. To see why, define $M = [\text{Cov}(L_T'u_i)]^{-\frac{1}{2}} = (L_T' \Sigma L_T)^{-\frac{1}{2}}$. Then, by essentially the same
algebra as in the previous subsection, we can show the FIV estimator applying the instruments $Z_{SEFE,i}$ and the filter $M$ to (8-8) equals the 3SLS estimator applying the same instruments to (8-8).

### 8.2 Models with Weakly Exogenous Instruments

In this section we consider the GMM estimation of the model (8-2) with weakly exogenous instruments. Suppose that there exists a vector of $1 \times k$ instruments $h_i t$ such that

$$E(h_i^t \epsilon_{it}) = 0, \ t = 1, \ldots, T, \ t \leq s.$$  

(8-9)

There are $T(T+1)k/2$ such moment conditions. These arise if the instruments $h_i t$ are weakly exogenous with respect to the $\epsilon_{it}$ and uncorrelated with the effect $\alpha_i$. If the instruments are weakly exogenous with respect to the $\epsilon_{it}$ but are correlated with the effects, we have a smaller number of moment conditions:

$$E(h_i^t \Delta \epsilon_{it}) = E(h_i^t \Delta \alpha_i) = 0, \ t = 1, \ldots, T-1, \ t \leq s,$$  

(8-10)

where $\Delta u_i s = u_i s - u_{i,s-1}$, and similarly for $\Delta \epsilon_{is}$. In this section, our discussion will be focused on GMM based on (8-9) only. We do this because essentially the same procedures can be used for GMM based on (8-10). The only difference between (8-9) and (8-10) lies in whether GMM applies to the original model (8-2) or the differenced model (8-8).

#### 8.2.1 The Forward Filter Estimator

For the model with weakly exogenous instruments, Keane and Runkle [1992, KR] propose a forward filter (FF) estimator. To be specific, define a $T \times T$ upper-triangular matrix $F = [F_{ij}]$ ($F_{ij}$ = 0 for $i > j$) that satisfies $F \Sigma F' = I_T$, so that Cov($F u_i$) = $I_T$. With this notation, Keane and
Runkle propose a FIV-type estimator, which applies the instruments \( H_i \equiv (h_{i1}',...,h_{iT}')' \) after filtering the error \( u_i \) by \( F \):

\[
\hat{\beta}_{FF} = [X'F.*H(H'H)^{-1}H'F.X]^1X'F.*H(H'H)^{-1}H'F.y ,
\]

where \( F_* = I_N \otimes F \).

The motivation of this FF estimator is that the FIV estimator using the instruments \( H_i \) and the usual filter \( \Sigma^{-\frac{1}{2}} \) is inconsistent unless the instruments \( H_i \) are strictly exogenous with respect to the \( \varepsilon_i \). In contrast, the forward-filtering transformation \( F \) preserves the weak exogeneity of the instruments \( h_{it} \). Hayashi and Sims [1983] provide some efficiency results for this forward filtering in the context of time series data. However, in the context of panel data (with large \( N \) and fixed \( T \)), the FF estimator does not necessarily dominate the GMM or 3SLS estimators using the same instruments \( H_i \).

One technical point is worth noting for the FF estimator. Forward filtering requires that the serial correlations in the error \( u_i \) do not depend on the values of current and lagged values of the instruments \( h_{it} \). (See Hayashi and Sims (1983), pp. 788-789.) This is a slightly weakened version of the usual condition of no conditional heteroskedasticity; it is weakened because conditioning is only on current and lagged values of the instruments. If this condition does not hold, in general,

\[
\text{plim}_{N \to \infty} N^{-1} \Sigma_{i=1}^{N} H_i'u_i'H_i' \neq \text{plim}_{N \to \infty} N^{-1} \Sigma_{i=1}^{N} H_i'F \Sigma F'H_i = \text{plim}_{N \to \infty} N^{-1} \Sigma_{i=1}^{N} H_i'H_i ,
\]

and the rationale for forward filtering is lost. Sufficient conditions under which the autocorrelations in \( u_i \) do not depend on the history of \( h_{it} \) are given by Wooldridge [1996, p. 401].

8.2.2 Irrelevance of Forward Filtering

The FF and 3SLS estimators using the instruments \( H_i \) are inefficient in that they fail to fully
exploit all of the moment conditions implied by (8-10). As Schmidt, Ahn and Wyhowski [1992, SAW] suggest, a more efficient estimator can be obtained by GMM using the $T \times T(T+1)k/2$ instruments

$$Z_{WE,i} \equiv \text{diag}(h_{i1},\ldots,h_{iT}) ,$$

where, as before, $h_{it} = (h_{i1},\ldots,h_{iT})$. When these instruments are used in GMM, filtering $u_i$ by $F$ becomes irrelevant. This result can be seen using Theorem 8.1. SAW show that there exists a $T(T+1)/2 \times T(T+1)/2$ nonsingular, upper-triangular matrix $E$ such that $FZ_{WE,i} = Z_{WE,i}(E \otimes I_q)$ (equation (10), p. 11). Thus, the FF and 3SLS estimators using the instruments $Z_{WE,i}$ are numerically identical (or asymptotically identical if different estimates of $\Sigma$ are used); filtering does not matter. Of course, both of these estimators are dominated by the GMM estimator using the same instruments but an unrestricted weighting matrix, unless the instruments satisfy the NCH condition (8-6).

This irrelevance result does not mean that filtering is meaningless, even practically. In some cases, the (full) GMM estimator utilizing all of the instruments $Z_{WE,i}$ may be practically infeasible. For example, in a model with 10 weakly exogenous instruments and 10 time periods, the total number of instruments is 550. This can cause computational problems for GMM, especially when the cross section dimension is small. Furthermore, GMM with a very large number of moment conditions may have poor finite-sample properties. For example, see Tauchen [1986], Altongi and Segal [1996] and Andersen and Sørensen [1996] for a discussion of the finite-sample bias of GMM in very overidentified problems. In the present context, Ziliak [1997] reports that the full GMM estimator has poor finite-sample properties and is often dominated (in terms of bias and RMSE) by the FF estimator using the same or fewer instruments.
8.2.3 Semiparametric Efficiency Bound

In GMM, imposing more moment conditions never decreases efficiency. An interesting question is whether there is an efficiency bound which GMM estimators cannot improve upon. Once this bound is identified, we may be able to construct an efficient GMM estimator whose asymptotic covariance matrix coincides with the bound. Chamberlain [1992] considers the semiparametric bound for the model in which the assumption (8-9) is replaced by the stronger assumption

\[ E[u_i | h^0_t = (h_{it}, \ldots, h_{iT})] = 0, \text{ for } t = 1, \ldots, T. \]  

(8-11)

For this case, let \( g_{k,it} \) be a \( 1 \times k \) vector of instruments which are some (polynomial) functions of \( h_{it}^0 \). Define \( G^*_k,i = \text{diag}(g_{k,1}, \ldots, g_{k,T}) \); so that under (8-11), \( E(G^*_k,i u_i) = 0 \). Under some suitable regularity conditions, Chamberlain shows that the semiparametric efficiency bound for GMM based on (8-11) is \( B_o^{-1} \), where

\[ B_o = \lim_{k \to \infty} E(X_i'G^*_k) [E(G^*_k u_i u_i') G^*_k]^{-1} E(G^*_k X_i). \]

This bound naturally suggests the GMM estimator based on the moment condition \( E(G^*_k u_i) = 0 \) with large \( k \). However, when \( k \) grows with \( N \) without any restriction, the usual asymptotic GMM inferences obtained by treating \( k \) as fixed would be misleading. In response to this problem, Hahn [1997] rigorously examines conditions under which the usual GMM inferences are valid for large \( k \) and the GMM estimator based on the moment condition \( E(G^*_k u_i) = 0 \) is efficient. Under the assumption that

\[ \lim_{k \to \infty} \frac{k^4}{N} = 0, \]  

(8-12)

Hahn establishes the efficiency of the GMM estimator. A similar result is obtained by Koenker and Machado [1996] for a linear model with heteroskedasticity of general form. They show that the usual GMM inferences, treating the number of moment conditions as fixed, are
asymptotically valid if the number of moment conditions grows more slowly than $N^{\gamma}$. These results do not directly indicate how to choose the number of moment conditions to use, for a given finite value of $N$, but they do provide grounds for suspicion about the desirability of GMM using numbers of moment conditions that are very large.

**8.3 Models with Strictly Exogenous Regressors**

This section considers efficient GMM estimation of linear panel data models with strictly exogenous regressors. The model of interest in this section is the standard panel data regression model

$$y_i = R_i \xi + (e_t \otimes w_i) \gamma + u_i = X_i \beta + u_i; \quad a_t = (e_t \otimes \alpha) + e_t,$$

(8-13)

where $R_i = [r_{i1}, \ldots, r_{iT}]$ is a $T \times k$ matrix of time-varying regressors, $e_t \otimes w_i = [w_i, \ldots, w_i]$ is a $T \times g$ matrix of time-invariant regressors, and $\xi$ and $\gamma$ are $k \times 1$ and $g \times 1$ vectors of unknown parameters, respectively. We assume that the regressors $r_{it}$ and $w_i$ are strictly exogenous to the $\varepsilon_{it}$; that is,

$$E(d_i \otimes \varepsilon_i) = 0,$$

(8-14)

where $d_i = (r_{i1}, \ldots, r_{iT}, w_i)$. We also assume the random-effects covariance structure $\Sigma = \sigma_\alpha^2 e_t e_t^\prime + \sigma_\varepsilon^2 I_T$ as given in equation (8-3). For notational convenience, we treat $\sigma_\alpha^2$ and $\sigma_\varepsilon^2$ as known.

See Hausman and Taylor [1981] for consistent estimation of these variances.

Efficient estimation of the model (8-13) depends crucially on the assumptions about correlations between the regressors $d_i$ and the effects $\alpha_i$. When the regressors are uncorrelated with $\alpha_i$, the traditional GLS estimator is consistent and efficient. If the regressors are suspected to be correlated with the effect, the within estimator can be used. However, a serious drawback
of the within estimator is that it cannot identify $\gamma$ because the within transformation wipes out the time-invariant regressors $w_i$ as well as the individual effects $\alpha_i$.

In response to this problem, Hausman and Taylor [1981, HT] considered the alternative assumption that some but possibly not all of the explanatory variables are uncorrelated with the effect $\alpha_i$. This offers a middle ground between the traditional random effects and fixed effects approaches. Extending their study, Amemiya and MaCurdy [1986, AM] and Breusch, Mizon and Schmidt [1989, BMS] considered stronger assumptions and derived alternative instrumental variables estimators that are more efficient than the HT estimator. A systematic treatment of these estimators can be found in Mátyás and Sevestre [1996, Chapter 6]. In what follows, we will study these estimators and the conditions under which they are efficient GMM estimators.

### 8.3.1 The HT, AM and BMS Estimators

Following HT, we decompose $r_{it}$ and $w_i$ into $r_{it} = (r_{1it}, r_{2it})$ and $w_i = (w_{1i}, w_{2i})$, where $r_{1it}$ and $r_{2it}$ are $1 \times k_1$ and $1 \times k_2$, respectively, and $w_{1i}$ and $w_{2i}$ are $1 \times g_1$ and $1 \times g_2$. With this notation, define:

$$s_{HT,i} = (\bar{r}_{1i}, w_{1i}), s_{AM,i} = (r_{1i1}, \ldots, r_{1iT}, w_{1i}), s_{BMS,i} = (s_{AM,i}, \bar{r}_{2i}).$$

where $\bar{r}_i = (r_{2i1-1}, \ldots, r_{2i,T-1})$. HT, AM and BMS impose the following assumptions, respectively, on the model (8-13):

$$E(s_{HT,i}^t, \alpha_i) = 0; E(s_{AM,i}^t, \alpha_i) = 0; E(s_{BMS,i}^t, \alpha_i) = 0.$$  \hspace{1cm} (8-16)

These assumptions are sequentially stronger. The HT Assumption $E(s_{HT,i}^t, \alpha_i) = 0$ is weaker than the AM assumption $E(s_{AM,i}^t, \alpha_i) = 0$, since it only requires the individual means of $r_{1it}$ to be uncorrelated with the effect, rather than requiring $r_{1it}$ to be uncorrelated with $\alpha_i$ for each $t$. 

14
(However, as AM argue, it is hard to think of cases in which each of the variables $r_{it}$ is correlated with $\alpha_i$ while their individual means are not.) Imposing the AM assumption instead of the HT assumption in GMM would generally lead to a more efficient estimator. The BMS Assumption $E(s_{BMS,i} \alpha_i) = 0$ is based on the stationarity condition,

$$E(r_{2it}^t \alpha_i) \text{ is the same for } t = 1, \ldots, T.$$  \hspace{1cm} (8-17)

This means that, even though the unobserved effect $\alpha_i$ is allowed to be correlated with $r_{2it}$, the covariance does not change with time. Cornwell and Rupert [1988] provide some evidence for the empirical legitimacy of the BMS assumption. They also report that GMM imposing the BMS assumption rather than the HT or AM assumptions would result in significant efficiency gains.

HT, AM and BMS consider FIV estimation of the model (8-13) under the random effects assumption (8-3). The instruments used by HT, AM and BMS are of the common form

$$Z_{A,i} = (Q_T R_i, e_T \otimes s_i),$$  \hspace{1cm} (8-18)

where the form of $s_i$ varies across authors; that is, $s_i = s_{HT,i}, s_{AM,i}$ or $s_{BMS,i}$.

Consistency of the FIV estimator using the instruments $Z_{A,i}$ requires $E(Z_{A,i} \Sigma^{-1} u_i) = 0$. This condition can be easily justified under (8-3), (8-14) and (8-16). Without loss of generality, we set $\sigma^2 = 1$. Then, it can be easily shown that $\Sigma^{-1} = \theta^2 P_T + Q_T$ and $\Sigma^{-1} = \theta P_T + Q_T$, where $\theta^2 = \sigma^2/(\sigma^2 + T \sigma^2)$. With this result, strict exogeneity of the regressors $r_i$ and $w_i$ (8-14) implies

$$E(R_i Q_T \Sigma^{-1} u_i) = E[R_i Q_T (\theta P_T + Q_T) u_i] = E(R_i Q_T e_i) = 0.$$

In addition, both (8-14) and (8-16) imply

$$E[(e_T \otimes s_i) \Sigma^{-1} u_i] = E[(e_T \otimes s_i)(e_T \otimes \alpha_i)] = T \theta \times E(s_i \alpha_i) = 0.$$

Several properties of the FIV estimator are worth noting. First, the usual GMM
identification condition requires that the number of columns in $Z_{A,i}$ should not be smaller than
the number of parameters in $\beta = (\xi', \gamma')'$. This condition is satisfied if the number of variables
in $s_i$ is not less than the number of time-invariant regressors (e.g., $k_1 \geq g_2$ for the HT case).
Second, the FIV estimator is an intermediate case between the traditional GLS and within
estimators. It can be shown that the FIV estimator of $\xi$ equals the within estimator if the model
is exactly identified (e.g, $k_1 = g_2$ for HT), while it is strictly more efficient if the model is
overidentified (e.g., $k_1 > g_2$ for HT). The FIV estimator of $\beta$ is equivalent to the GLS estimator
if $k_2 = g_2 = 0$ (that is, no regressor is correlated with $\alpha_i$). For more details, see HT. Finally,
the FIV estimator is numerically equivalent to the 3SLS estimator applying the instruments $Z_{A,i}$;
thus, filtering does not matter. To see this, observe that
\[ \Sigma^{Z_{A,i}} = (\theta P_T + Q_T)(Q_T R_i e_T \otimes s_i) \]
\[ = (Q_T R_i \theta e_T \otimes s_i) = (Q_T R_i e_T \otimes s_i) \text{diag}(I_{(1)}, \theta I_{(2)}) = Z_{A,i} \text{diag}(I_{(1)}, \theta I_{(2)}) , \]
where $I_{(1)}$ and $I_{(2)}$ are conformable identity matrices. Since the matrix $\text{diag}(I_{(1)}, \theta I_{(2)})$ is
nonsingular, Theorem 8.1 applies: $\hat{\beta}_{FIV} = \hat{\beta}_{3SLS}$. This result also implies that the FIV estimator is
equally efficient as the GMM estimator using the instruments $Z_{A,i}$, if the instruments satisfy the
NCH condition (8-6). When this NCH condition is violated, the GMM estimator using the
instruments $Z_{A,i}$ and an unrestricted weighting matrix is strictly more efficient than the FIV estimator.

8.3.2 Efficient GMM Estimation

We now consider alternative GMM estimators which are potentially more efficient than the HT,
AM or BMS estimators. To begin with, observe that strict exogeneity of the regressors $r_{it}$ and $w_i$
(8-14) implies many more moment conditions than the HT, AM or BMS estimators utilize. The
strict exogeneity condition (8-14) implies
\[ E[(L_T \otimes d_i)'u_i] = E[L_T'(e_i \alpha_i + \epsilon_i) \otimes d_i] = E(L_T' \epsilon_i \otimes d_i) = 0, \]
where \( L_T \otimes d_i \) is \( T \times \{(T-1)(kT+g)\} \). Based on this observation, Arellano and Bover [1995] (and Ahn and Schmidt [1995]) propose the GMM estimator using the instruments
\[ Z_{B,i} = (L_T \otimes d_i, e_i \otimes s_i) , \tag{8-19} \]
which include \( (T-1)(Tk+g) - k \) more instruments than \( Z_{A,i} \). The covariance matrix \( \Sigma \) need not be restricted. Clearly, the instruments \( Z_{B,i} \) subsume \( Z_{A,i} \equiv (Q_T R_i, e_T \otimes s_i) \) which are essentially the HT, AM or BMS instruments. Thus, the GMM estimator utilizing all of the instruments \( Z_{B,i} \) cannot be less efficient than the GMM estimator using the smaller set of instruments \( Z_{A,i} \). In terms of achieving asymptotic efficiency, there is no reason to prefer to use the fewer instruments \( Z_{B,i} \).

However, using all of the instruments \( Z_{A,i} \) may not be practically feasible, even when \( T \) is only moderately large. For example, consider the case in which \( k = g = 5 \) and \( T = 10 \). For this case, the number of the instruments in \( Z_{A,i} \) exceed the number of moment conditions in \( Z_{B,i} \) by 490 (= 495-5). For such cases, the GMM estimator using the HT, AM or BMS instruments would be of more practical use.

In addition, the GMM (or FIV) estimator using the instruments \( Z_{A,i} \) can be shown to be asymptotically as efficient as the GMM estimator using all of the instruments \( Z_{B,i} \), under specific assumptions that are consistent with the motivation for the HT, AM or BMS estimators. Arellano and Bover [1995] provide the foundation for this result.

**Theorem 8.2**

Suppose that \( \Sigma \) has the random effect structure (8-3). Then, the 3SLS estimator using the
instruments $Z_{B,i}$ is numerically identical to the 3SLS estimator using the smaller set of instruments $Z_{A,i}$, if the same estimate of $\Sigma$ is used.\(^3\)

Although Arellano and Bover [1995] provide a detailed proof of the theorem, we provide a shorter alternative proof. In what follows, we use the usual projection notation: For any matrix $B$ of full column rank, we define the projection matrix $P(B) = B(B'B)^{-1}B'$. The following lemma is useful for the proof of Theorem 8.2.

**Lemma 8.1**

Let $L_s = I_N \otimes L_T$ and $D = [(I_{T-1} \otimes d_1)'...(I_{T-1} \otimes d_N)']'$. Define $V = I_N \otimes e_T$ and $W = (w_1',...,w_N')'$, so that $X = (R,VW)$. Then, $P(L_sD)X = P(Q_vR)X$.

**Proof of Lemma 8.1**

Since $Q_v = P(L_s) = L_s(L_s' L_s)^{-1}L_s'$, $Q_vR = L_s[(L_s' L_s)^{-1}L_s' R]$. In addition, since $D$ spans all of the columns in $(L_s' L_s)^{-1}L_s' X$, $L_sD$ must span $Q_vR = L_s[(L_s' L_s)^{-1}L_s' R]$. Finally, since $Q_vL_s = L_s$ and $Q_vV = 0$,

$$P(L_sD)X = P(L_sD)Q_vX = (P(L_sD)Q_vR,0) = (Q_vR,0) = P(Q_vR)X.$$ 

We are now ready to prove Theorem 8.2.

**Proof of Theorem 8.2**

Note that $Z_A = [Q_vR,VW]$ and $Z_B = [L_sD,VS]$, where $S = (s_1',...,s_N')'$. Since $L_s$ and $Q_v$ are in

\(^3\)Even if different estimates of $\Sigma$ are used, the two 3SLS estimators are asymptotically identical.
the same space and orthogonal to both V and P, we have P(Z_A) = P(Q_vR) + P(VS) and P(Z_B) = P(L,D) + P(VS). Using these results and the fact that \( \Omega^{\frac{1}{2}} = \Theta P_v + Q_v \), we can also show that the 3SLS estimators using the instruments \( Z_{A,i} \) and \( Z_{B,i} \), respectively, equal

\[
\hat{\beta}_A = \left( X' \{ \Theta P(Q_v R) + P(VS) \} X \right)^{-1} X' \{ \Theta P(Q_v R) + P(VS) \} y ;
\]

\[
\hat{\beta}_B = \left( X' \{ \Theta P(L,D) + P(VS) \} X \right)^{-1} X' \{ \Theta P(L,D) + P(VS) \} y .
\]

However, Lemma 8.1 implies \( \{ \Theta P(L,D) + P(VS) \} X = \{ \Theta P(Q_v R) + P(VS) \} X \). Thus, \( \hat{\beta}_B = \hat{\beta}_A \).

Theorem 8.2 effectively offers conditions under which the 3SLS (or FIV) estimator using the instruments \( Z_{A,i} \) is an efficient GMM estimator. Under the random-effects structure (8-3), the 3SLS estimator using all of the instruments \( Z_{A,i} \) equals the 3SLS estimator using the full set of instruments \( Z_{B,i} \). Thus if, in addition to (8-3), the instruments \( Z_{B,i} \) satisfy the NCH assumption (8-6), the 3SLS (or FIV) estimator using the instruments \( Z_{A,i} \) should be asymptotically equivalent to the efficient GMM estimator exploiting all of the moment conditions \( E(Z_{B,i}' u_i) = 0 \). Note that both assumptions (8-3) and (8-6) are crucial for this efficiency result. If one of these assumptions is violated, the GMM estimator exploiting all of the instruments \( Z_{B,i} \) is strictly more efficient than the GMM estimator using the instruments \( Z_{A,i} \).

### 8.3.3 GMM with Unrestricted \( \Sigma \)

Im, Ahn, Schmidt and Wooldridge [1996, IASW] examine efficient GMM estimation for the case in which the instruments \( Z_{B,i} \) satisfy the NCH condition (8-6), but \( \Sigma \) is unrestricted. For this case, IASW consider the 3SLS estimator using the instruments \( \Sigma^{-1} Z_{A_i} = \Sigma^{-1} (Q_v e_i \otimes s_i) \), which is essentially the GIV estimator of White [1984]. They show that when \( s_i = s_{BMS,i} \), the 3SLS estimator using the instruments \( \Sigma^{-1} Z_{A,i} \) is numerically equivalent to the 3SLS estimator.
using all of the instruments $Z_{B,i} = (L_T \otimes d_i e_T \otimes s_i)$. However, they also find that this equality does not hold when $s_i = s_{HT,i}$ or $s_{AM,i}$. In fact, without the BMS assumption, the set of instruments $\Sigma^{-1} Z_{A,i}$ is not legitimate in 3SLS. This is true even if $s_i = s_{HT,i}$ or $s_{AM,i}$. To see this, observe that

$$E(R_i Q_T \Sigma^{-1} u_i) = E(R_i Q_T \Sigma^{-1} e_i \alpha_i) + E(R_i Q_T \Sigma^{-1} \varepsilon_i) = E(R_i Q_T \Sigma^{-1} e_i \alpha_i),$$

where the last equality results from given strict exogeneity of $R_i$ with respect to $\varepsilon_i$. However, with unrestricted $\Sigma$ and without the BMS assumption, $E(R_i Q_T \Sigma^{-1} e_i \alpha_i) \neq 0$, and $\Sigma^{-1} Q_T R_i$ is not legitimate.

IASW provide a simple solution to this problem, which is to replace $Q_T$ by a different matrix that removes the effects, $Q_2 = \Sigma^{-1} - \Sigma^{-1} e_T (e_T \Sigma^{-1} e_T)^{-1} e_T \Sigma^{-1}$. Clearly $Q_2 e_T = 0$. Thus $R_i Q_2 e_T \alpha_i = 0$, and $E(R_i Q_2 u_i) = E(R_i Q_2 \varepsilon_i) = 0$ given strict exogeneity of $R_i$ with respect to $\varepsilon_i$. Thus, $Q_2 R_i$ are legitimate instruments for 3SLS.

This discussion motivates modified instruments of the form $(Q_2 R_i, \Sigma^{-1} e_T \otimes s_i)$. IASW show that the 3SLS estimator using these modified instruments is numerically equivalent to the 3SLS estimator using all of the instruments $(L_T \otimes d_i e_T \otimes s_i)$. That is, the modified 3SLS estimator is an efficient GMM estimator, if the instruments $(L_T \otimes d_i e_T \otimes s_i)$ satisfy the NCH condition (8-6).

### 8.4 Simultaneous Equations

In this section we consider GMM estimation of a simultaneous equations model, with panel data and unobservable individual effects in each structural equation. The foundation of this section is the model considered by Cornwell, Schmidt and Wyhowski [1992, CSW] and Mátyás and Sevestre [1996, Chapter 9]:

$$y_{ij} = X_{ij} \beta_j + e_{ij} + w_{ij} + u_{ij}, \quad u_{ij} = e_{ij} \otimes \alpha_{ij} + e_{ij}. \quad (8-20)$$

Here $j = 1, \ldots, J$ indexes the individual structural equation, so that equation (8-20) reflects T
observations for individual i and equation j. \( Y_{j,i} \) denotes the data matrix of included endogenous variables. Other variables are defined similarly to those in (8-13). We denote \( \Sigma_{jh} \equiv E(u_{j,i}u_{h,i}') \) for \( j, h = 1,\ldots,J \).

In order to use the same notation for instrumental variables as in section 8.3, we let \( R_i = (r_{i1},\ldots,r_{iT}') \) and \( w_i \) be the \( T \times k \) and \( 1 \times g \) data matrices of all time-varying and time-invariant exogenous regressors in the system, respectively. With this notation, we can define \( d_i, s_i, Z_{A,i} \) and \( Z_{B,i} \) as in section 8.3. Consistent with CSW, we assume that the variables \( d_i = (r_{i1},\ldots,r_{iT},w_i) \) are strictly exogenous to the \( \epsilon_{j,it} \); that is, \( E(d_i \otimes \epsilon_{j,i}) = 0 \) for all \( j = 1,\ldots,J \). We also assume that a subset \( s_i \) of the exogenous variables \( d_i \) is uncorrelated with the individual effects \( \alpha_{j,i} (j = 1,\ldots,J) \). As in section 8.3, an appropriate choice of \( s_i \) can be made by imposing the HT, AM or BMS assumptions on \( d_i \).

Under these assumptions, CSW consider GMM estimators based on the moment conditions

\[
E(Z_{A,i}'\mu_{j,i}) = 0, \text{ for } j = 1,\ldots,J, \tag{8-21}
\]

where the instruments \( Z_{A,i} = (Q_TR_i,e_T \otimes s_i) \) are of the HT, AM or BMS forms as in (8-18).

Clearly, the model (8-20) implies more moment conditions than those in (8-21). In the same way as in section 8.3.3, we can show that the full set of moment conditions implied by the model (8-20) is

\[
E(Z_{B,i}'u_{j,i}) = 0, \text{ for } j = 1,\ldots,J ,
\]

where \( Z_{B,i} = (L_T \otimes d_i,e_T \otimes s_i) \). We will derive conditions under which the GMM (3SLS) estimator based on (8-21) is asymptotically as efficient as the GMM estimator exploiting the full set of moment conditions.

In (8-21), we implicitly assume that the same instruments \( Z_{A,i} \) are available for each structural equation. This assumption is purely for notational convenience. We can easily allow
the instruments $Z_{A,i}$ to vary over different equations, at the cost of more complex matrix notation (see CSW, section 3.4).

### 8.4.1 Estimation of a Single Equation

We now consider GMM estimation of a particular structural equation in the system (8-20), say the first equation, which we write as

$$y_{1,i} = X_{1,i}\beta_1 + u_{1,i},$$

(8-22)

adopting the notation of (8-20). Using our convention of matrix notation, we can also write this model for all NT observations as $y_i = X_i\beta_1 + u_i$, where $X_i = (Y_i, R_i, VW_i)$.

GMM estimation of the model (8-22) is straightforward. The parameter $\beta_1$ can be consistently estimated by essentially the same GMM or instrumental variables as in section 8.3. Given the assumption (8-21), the GMM estimator using the instruments $Z_{A,i}$ is consistent.

We now consider conditions under which the GMM estimator based on (8-21) is fully efficient (as efficient as GMM based on the full set of moment conditions). The following Lemma provides a clue.

**Lemma 8.2**

$$(L_T \otimes d_i)[E(L_T' L_T \otimes d_i' d_i)]^{-1} E[(L_T \otimes d_i)' Y_{1,i}] = Q_T R_i [E(R_i' Q_i R_i)]^{-1} E(R_i' Q_T Y_{1,i}).$$

**Proof of Lemma 8.2**

Consider the reduced-form equations for the endogenous regressors $Y_{1,i}$:

$$Y_{1,i} = R_{i} \Pi_{11} + (e_i \otimes w_i) \Pi_{12} + (e_i \otimes \alpha_i) \Pi_{13} + v_{1,i},$$

where $\alpha_i = (\alpha_{1,i}, ..., \alpha_{J,i})$, and the error $v_{1,i}$ is a linear function of the structural random errors
εₖᵢ (j = 1,...,J). Since the variables dᵢ = (rᵢ₁,...,rᵢₜ,ωᵢ) are strictly exogenous with respect to the εₖᵢ, so are they with respect to v₁ᵢ; that is, E(dᵢ ⊗ v₁ᵢ) = 0. Note also that since LₜD spans the columns of QᵥR (see the proof of Lemma 8.1), there exists a conformable matrix A such that QᵗᵢRᵢ = (Lₜ ⊗ dᵢ)A for all i. These results imply

\[(Lₜ ⊗ dᵢ)[E(Lᵀₜ'Lᵀₜ ⊗ d'_i)]⁻¹E[(Lₜ ⊗ d')Y₁ᵢ] \]

= \[(Lₜ ⊗ dᵢ)[E(Lᵀₜ'Lᵀₜ ⊗ d'_i)]⁻¹E[(Lₜ ⊗ d')QᵀᵢY₁ᵢ] \]

= \[(Lₜ ⊗ dᵢ)[E(Lᵀₜ'Lᵀₜ ⊗ d'_i)]⁻¹E[(Lₜ ⊗ d')(Lₜ ⊗ d_i)AΠ₁₁] \]

= \[(Lₜ ⊗ d_i)AΠ₁₁ = QᵀᵢRᵢΠ₁₁. \]

However, we also have

\[Qᵀᵢ[R₁[E(R₁'QᵀᵢR₁)]⁻¹E(R₁'QᵀᵢY₁ᵢ)] = Qᵀᵢ[R₁[E(R₁'QᵀᵢR₁)]⁻¹E(R₁'QᵀᵢΠ₁₁)] = QᵀᵢRᵢΠ₁₁. \]

What Lemma 8.2 means is that Proj(Y₁ᵢ|Lₜ ⊗ dᵢ) = Proj(Y₁ᵢ|QᵀᵢRᵢ), where Proj(Bᵢ|Cᵢ) is the population least squares projection of Bᵢ on Cᵢ. Thus, Lemma 8.2 can be viewed as a population (asymptotic) analog of Lemma 8.1. Clearly, P(LₜD)R₁ = P(QᵥR)R₁ and P(LₜD)VW₁ = 0 = P(QᵥR)VW₁. These equalities and Lemma 8.2 imply that for any conformable data matrix Bᵢ of T rows (and B of NT rows),

\[\lim_{N→∞}N⁻¹BᵀP(LₜD)X₁ = E[B₁'(LᵀₜLᵀₜ ⊗ d'_i)]⁻¹E[(Lₜ ⊗ d_i)'X₁ᵢ] \]

\[= E(B₁'QᵀᵢR₁)[E(R₁'QᵀᵢR₁)]⁻¹E(R₁'QᵀᵢX₁ᵢ)] = \lim_{N→∞}N⁻¹BᵀP(QᵥR)X₁ ; \]

\[\lim_{N→∞}N⁻¹BᵀP(LₜD)X₁ = \lim_{N→∞}N⁻¹Σᵢ=₁N⁻¹Bᵢ'(LᵀₜLᵀₜ ⊗ d'_i)]⁻¹E[(Lₜ ⊗ d_i)'X₁ᵢ] \]

\[= \lim_{N→∞}N⁻¹Σᵢ=₁Bᵢ'QᵀᵢR₁[E(R₁'QᵀᵢR₁)]⁻¹E(R₁'QᵀᵢX₁ᵢ)] \]

= \lim_{N→∞}N⁻¹BᵀP(QᵥR)X₁ .

Lemma 8.2 leads to an asymptotic analog of Theorem 8.2.
Theorem 8.3

Suppose that Cov(u_{1,i}) = \Sigma_{11} is of the random effects form (8-3). Then, the 3SLS estimator \( \hat{\beta}_{1,A} \) using the instruments \( Z_{A,i} \) is asymptotically identical to the 3SLS estimator \( \hat{\beta}_{1,B} \) using all of the instruments \( Z_{B,i} \). If, in addition, the instruments \( Z_{B,i} \) satisfy the NCH condition (8-6), \( \hat{\beta}_{1,A} \) is efficient among the class of GMM estimators based on the moment condition \( E(Z_{B,i}'u_{1,i}) = 0 \).

Proof of Theorem 8.3

Let \( \Sigma_{11} = \sigma_{e,1}^2 I_T + \sigma_{\alpha,1}^2 I_T \). Without loss of generality, we set \( \sigma_{e,1}^2 = 1 \). Then, similarly to the proof of Theorem 8.2, we can show

\[
\hat{\beta}_{1,A} = [X_1'(\theta_{11}P(L,D) + P(VS))X_1]^{-1}X_1'(\theta_{11}P(L,D) + P(VS))y_1 ;
\]

\[
\hat{\beta}_{1,B} = [X_1'(\theta_{11}P(Q,R) + P(VS))X_1]^{-1}X_1'(\theta_{11}P(Q,R) + P(VS))y_1 ,
\]

where \( \theta_{11}^2 = \sigma_{e,1}^2 / (\sigma_{e,1}^2 + T\sigma_{\alpha,1}^2) \). But, Lemma 8.2 implies that \( \text{plim}_{N \to \infty} N^{-1/2}(\xi_{1,A} - \xi_{1,B}) = 0 \).

8.4.2 System of Equations Estimation

We now consider the joint estimation of all the equations in the system (8-20). Following our convention for matrix notation, structural equation \( j \) can be written for all NT observations as \( y_j = X_j \beta_j + u_j \). If we stack these equations into the seemingly unrelated regressions (SUR) form, we have

\[
y_* = X_* \beta + u_* ,
\]

where \( y_* = (y_1',...,y_J')' \), \( X_* = \text{diag}(X_1,...,X_J) \), \( u_* = (u_1',...,u_J')' \) and \( \beta = (\beta_1',...\beta_J')' \). We denote \( \Omega_* = \text{Cov}(u_*) \). Straightforward algebra shows that \( \Omega_* = [I_N \otimes \Sigma]_{NTJxNTJ} \); that is, if we partition \( \Omega_* \) evenly into \( J \times J \) blocks, the \( (g,h)' \)th block is \( I_N \otimes \Sigma_{gh} \).
Following CSW, we consider the system GMM estimator based on the moment conditions (8-21). Define \( Z_s^i = I_J \otimes Z_{A,i} \) and \( u_s^i = (u_{1,i}',...,u_{J,i}')' \), so that we can write all of the moment conditions (8-21) compactly as \( E(Z_s^i u_s^i) = 0 \). Use of these moment conditions leads to the system GMM estimator

\[
\hat{\beta}_{SGMM} = \left[ X'^* (V_N^{-1}) Z_s'^* (V_N^{-1}) Z_s'^* y^* \right]^{-1} X'^* (V_N^{-1}) Z_s'^* y^* ,
\]

where \( Z_s^i = I_J \otimes Z_{A,i} \) and \( V_N = N^{-1} \sum_{i=1}^N Z_s^i u_s^i u_s^i' Z_s^i = N^{-1} \sum_{i=1}^N [Z_{A,i} \cdot u_{j,i} u_{h,i} Z_{A,i}] \). The 3SLS version of this system estimator can be obtained if we replace \( V_N \) by

\[
N^{-1} Z_s' \Omega Z_s = N^{-1} \sum_{i=1}^N [Z_{A,i} \cdot \Sigma_{\beta} Z_{A,i}] ;
\]

that is,

\[
\hat{\beta}_{S3SLS} = \left[ X'^* (Z_s'^* \Omega Z_s')^{-1} Z_s' y^* \right]^{-1} X'^* (Z_s'^* \Omega Z_s')^{-1} Z_s'y^* .
\]

In fact, this system 3SLS estimator is a generalization of the 3SLS estimator proposed by CSW. To see this, we make the following assumptions, as in CSW (Assumption 1, p. 157):

**ASSUMPTION 8.1**

(i) The individual effects for person \( i \), \( \alpha_i = (\alpha_{1,i},...\alpha_{J,i})' \), are iid \((0,\Sigma_\alpha)\). (ii) The random errors for person \( i \) at time \( t \), \( (\varepsilon_{1,it},...\varepsilon_{J,it})' \), are iid \((0,\Sigma_\varepsilon)\). (iii) All elements of \( \alpha \) are uncorrelated with all of elements of \( \varepsilon \). (iv) \( \Sigma_\alpha \) and \( \Sigma_\varepsilon \) are nonsingular.

Under Assumption 8.1, we can show

\[
\Omega_s = \Sigma_\varepsilon \otimes I_{NT} + \Sigma_\alpha \otimes (TP_v) = \Sigma_1 \otimes Q_v + \Sigma_2 \otimes P_v,
\]

where \( \Sigma_1 \equiv \Sigma_\varepsilon \) and \( \Sigma_2 \equiv \Sigma_\varepsilon + T \Sigma_\alpha \). Using this result and the facts that \( Z_s^i = I_J \otimes Z_{A,i} = (I_J \otimes Q_v R, I_J \otimes VS) \) and \( R' Q_v VS = 0 \), we can easily show that

\[
Z_s^i (Z_s'^* \Omega Z_s')^{-1} Z_s'^* = \Sigma_1^{-1} \otimes Q_v R + \Sigma_2^{-1} \otimes VS.
\]
Substituting this result into the system 3SLS estimator $\hat{\beta}_{S3SLS}$ yields the 3SLS estimator of CSW (p. 164, equation (34)). Thus $\hat{\beta}_{S3SLS}$ simplifies to the CSW 3SLS estimator when the errors have the random effects covariance structure implied by Assumption 8.1; otherwise, it is the appropriate generalization of the CSW estimator.

8.5 Dynamic Panel Data Models

In this section we consider a regression model for dynamic panel data. The model of interest is given by:

$$y_t = y_{t-1}\delta + R_i\xi + (e_i \otimes \omega)\gamma + u_i = X_i\theta + u_i; u_i = e_i \otimes \alpha_i + \varepsilon_i,$$

where $y_{t-1} = (y_{i0},...,y_{iT-1})'$, $y_{i0}$ is the initial observed value of $y$ (for individual $i$), and other variables are defined exactly as in (8-13).

The basic problem faced in the estimation of this model is that the traditional within estimator is inconsistent, because the within transformation induces a correlation of order $1/T$ between the lagged dependent variable and the random error (see Hsiao [1986]). A popular solution to this problem is to first difference the equation to remove the effects, and then estimate by GMM, using as instruments values of the dependent variable lagged two or more periods as well as other exogenous regressors. Legitimacy of the lagged dependent variables as instruments requires some covariance restrictions on $\varepsilon_i, \alpha_i$ and $y_{i0}$. However, these covariance restrictions imply more moment conditions than are imposed by the GMM estimator based on first differences. In this section, we study the moment conditions implied by a standard set of covariance restrictions and other alternative assumptions. We also examine how these moment conditions can be efficiently imposed in GMM.

A good survey, which emphasizes somewhat different aspects of the estimation problem than
8.5.1 Moment Conditions under Standard Assumptions

In this subsection we count and express the moment conditions implied by a standard set of assumptions about $\alpha_i$, $\epsilon_i$, and $y_{i0}$. For simplicity, and without loss of generality, we do so in the context of the simple dynamic model whose only explanatory variable is the lagged dependent variable:

$$y_i = \delta y_{i, -1} + u_i; \quad u_i = \epsilon_i \otimes \alpha_i + \epsilon_i.$$

(8-25)

Consistently with the previous sections, we assume that $\alpha_i$ and $\epsilon_i$ have mean zero for all $i$ and $t$. (Nonzero mean of $\alpha$ can be handled with an intercept which can be regarded as an exogenous regressor.) We also assume that $E(y_{i0}) = 0$. We make this assumption in order to focus our discussion on the (second-order) moment conditions implied by covariance restrictions on $\epsilon_i$, $\alpha_i$ and $y_{i0}$. If $E(y_{i0}) \neq 0$, the first-order moment conditions $E(u_i) = 0$ ($t = 1, \ldots, T$) are relevant in GMM. Imposing these first-order moment conditions could improve efficiency of GMM estimators, as Crépon, Karamarz and Trognon [1995] suggest. In contrast, if $E(y_{i0}) = 0$ (and $E(\alpha_i) = 0$), the first-order moment conditions become uninformative for the unknown parameter $\delta$ because they cannot identify a unique $\delta$. This is so because for any value of $\delta$,

$$E(u_i) = E(y_{i0} - \delta y_{i, -1}) = E(y_{i0}) - \delta E(y_{i, -1}) = 0.$$

The following assumptions are most commonly adopted in the dynamic panel data literature.

ASSUMPTION 8.2

(i) For all $i$, $\epsilon_{it}$ is uncorrelated with $y_{i0}$ for all $t$.

(ii) For all $i$, $\epsilon_{it}$ is uncorrelated with $\alpha_i$ for all $t$. 

27
(iii) For all $i$, the $\varepsilon_{it}$ are mutually uncorrelated.

Under Assumption 8.2, it is obvious that the following moment conditions hold:

$$E(y_{it}\Delta u_{it}) = 0, \ t = 2, \ldots, T, \ s = 0, \ldots, t-2,$$

where $\Delta u_{it} = u_{it} - u_{i,t-1} = \varepsilon_{it} - \varepsilon_{i,t-1}$. There are $T(T-1)/2$ such conditions. These are the moment conditions that are widely used in the panel data literature (e.g., Anderson and Hsiao [1981], Holtz-Eakin [1988], Holtz-Eakin, Newey and Rosen [1988], Arellano and Bond [1991]). However, as Ahn and Schmidt [1995] find, Assumption 8.2 implies additional moment conditions beyond those in (8-26). In particular, the following $T-2$ moment conditions also hold:

$$E(u_{it}\Delta u_{it}) = 0, \ t = 2, \ldots, T-1,$$

which are nonlinear in terms of $\delta$.

The conditions (8-26) and (8-27) are a set of $T(T-1)/2 + (T-2)$ moment conditions that follow directly from the assumptions that the $\varepsilon_{it}$ are mutually uncorrelated and uncorrelated with $\alpha_i$ and $y_{i0}$. Furthermore, they represent all of the moment conditions implied by these assumptions. A formal proof of the number of restrictions implied by Assumption 8.2 can be given as follows. Define $\sigma_{it} = \text{var}(\varepsilon_{it})$, $\sigma_{i\alpha} = \text{var}(\alpha_i)$ and $\sigma_{00} = \text{var}(y_{i0})$. Then, Assumption 8.2 imposes the following covariance restrictions on the initial value $y_{i0}$ and the composite errors $u_{i1}, \ldots, u_{iT}$:

$$\Delta = \text{Cov} = \begin{bmatrix}
\sigma_{i1} & \sigma_{i2} & \cdots & \sigma_{i2} \\
\sigma_{i2} & \sigma_{i3} & \cdots & \sigma_{i3} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{i(T-1)} & \sigma_{i(T-1)} & \cdots & \sigma_{i(T-1)}
\end{bmatrix},$$

There are $T-1$ restrictions, that $\lambda_{0i}$ is the same for $t = 1, \ldots, T$; and $T(T-1)/2-1$ restrictions, that $\lambda_{ti}$
is the same for t, s = 1, ..., T, t ≠ s. Adding the number of restrictions, we get T(T-1)/2 + (T-2).

Since the moment conditions (8-27) are nonlinear, GMM imposing these conditions requires an iterative procedure. Thus, an important practical question is whether this computational burden is worthwhile. Ahn and Schmidt [1995] provide a partial answer. They compare the asymptotic variances of the GMM estimator based on (8-26) only and the GMM estimator based on both of (8-26) and (8-27). Their computation results show that use of the extra moment condition (8-27) can result in a large efficiency gain, especially when δ is close to one or the variance σ_αα is large.

8.5.2 Some Alternative Assumptions

We now briefly consider some alternative sets of assumptions. The first case we consider is the one in which Assumption 8.2 is augmented by the additional assumption that the ε_it are homoskedastic. That is, suppose that we add the assumption:

ASSUMPTION 8.3

For all i, var(ε_it) is the same for all t.

This assumption, when added to Assumption 8.2, generates the additional (T-1) moment conditions that

\[ \mathbb{E}(u_t^2) \text{ is the same for } t = 1, ..., T. \]  

(8-28)

(In terms of Λ above, λ_t is the same for t = 1, ..., T.) Therefore the total number of moment conditions becomes T(T-1)/2 + (2T-3). These moment conditions can be expressed as (8-26)-(8-28). Alternatively, if we wish to maximize the number of linear moment conditions, these
moment conditions can be expressed as (8-26) plus the additional conditions

$$E(y_{i,t} \Delta u_{i,t+1} - y_{i,t+1} \Delta u_{i,t+2}) = 0, \ t = 1, \ldots, T-2,$$

$$E(\tilde{u}_t \Delta u_{i,t+1}) = 0, \ t = 1, \ldots, T-1,$$

where $\tilde{u}_i = T^{-1} \sum_{t=1}^{T} u_{it}$. Comparing this to the set of moment conditions without homoskedasticity ((8-26)-(8-27)), we see that homoskedasticity adds $T-1$ moment conditions and it allows $T-2$ previously nonlinear moment conditions to be expressed linearly.

Ahn and Schmidt [1995] quantify the asymptotic efficiency gain from imposing the extra moment conditions (8-27) and (8-28) (or equivalently, (8-29) and (8-30)) in addition to (8-26). Their results show that most of efficiency gains come from the moment condition (8-27). That is, we do not gain much efficiency from the assumption of homoskedasticity (Assumption 8.3).

Another possible assumption we may impose on the model (8-25) is the stationarity assumption of Arellano and Bover [1995]:

**ASSUMPTION 8.4**

$$\text{cov}(\alpha_i, y_{it})$$ is the same for all $t$.

This is an assumption of the type made by BMS (see (8-17)); it requires equal covariance between the effects and the variables with which they are correlated. Ahn and Schmidt [1995] show that, given Assumption 8.2, Assumption 8.4 corresponds to the restriction that

$$\sigma_{\alpha \alpha} = \sigma_{uu} / (1 - \delta)$$

and implies one additional moment restriction. Furthermore, they show that it also allows the entire set of available moment conditions to be written linearly; that is, (8-26) plus
This is a set of $T(T-1)/2 + (2T-2)$ moment conditions, all of which are linear in $\delta$. Blundell and Bond [1997] show that the GMM estimator exploiting all of these linear moment conditions has much better asymptotic and finite sample properties than the GMM estimator based on (8-26) only. Thus the stationarity assumption 8.4 may be quite useful.

Finally, we consider an alternative stationarity assumption, which is examined by Ahn and Schmidt [1997]:

**ASSUMPTION 8.5**

In addition to Assumptions 8.2 and 3, the series $y_{i0}, ..., y_{iT}$ is covariance stationary.

To see the connection between the two stationarity assumptions, Assumptions 8.4 and 5, we use the solution

$$y_{it} = \delta y_{i0} + \alpha(1-\delta^t)/(1-\delta) + \sum_{j=0}^{t-1} \delta^j \varepsilon_{i,t-j}$$

to calculate

$$\text{var}(y_{it}) = \sigma_{00}\delta^{2t} + \sigma_{0a}\alpha(1-\delta^t)/(1-\delta) + \sigma_{aa}[(1-\delta^t)/(1-\delta)]^2 + \sigma_{ee}(1-\delta^{2t})/(1-\delta^2),$$

where the calculation assumes Assumptions 8.2 and 3. Assumption 8.5 implies that $\text{var}(y_{it}) = \sigma_{00}$ for all $t$, which occurs if and only if $\sigma_{aa} = (1-\delta)\sigma_{0a}$ and also

$$\sigma_{00} = \sigma_{aa}/(1-\delta) + \sigma_{ee}/(1-\delta^2).$$

Thus Assumption 8.5 implies $\sigma_{aa} = (1-\delta)\sigma_{0a}$, which in turn implies Assumption 8.4. However, it also implies the restriction (8-32) on the variance of the initial observation $y_{i0}$. Imposing (8-
as well as Assumptions 8.2-4 yields one additional, nonlinear moment condition:

\[ E[y_{i0}^2 + y_{i1}u_{i2}/(1-\delta^2) - u_{i2}u_{i1}/(1-\delta)^2] = 0. \]

An interesting question that we do not address here is how many moment conditions we would have if the assumptions discussed above are relaxed. Ahn and Schmidt [1997] give a partial answer by counting the moment conditions implied by many possible combinations of the above assumptions. See that paper for more detail.

### 8.5.3 Estimation

In this subsection we discuss some theoretical details concerning GMM estimation of the dynamic model. We also discuss the relationship between GMM based on the linear moment conditions and 3SLS estimation. Our discussion will proceed under Assumptions 8.2-3, but can easily be modified to accommodate the other cases.

#### 8.5.3.1 Notation and General Results

We now return to the model (8-24) which includes exogenous regressors \( r_{it} \) and \( w_i \). Exogeneity assumptions on \( r_{it} \) and \( w_i \) generate linear moment conditions of the form

\[ E(C_i'u_i) = 0, \]

where \( C_i = Z_{A,i} \) or \( Z_{B,i} \) as defined in section 8.3. In addition, the moment conditions given by (8-26), (8-29) and (8-30) above are valid. The moment conditions in (8-26) above are linear in \( \xi \) and can be written as \( E(A_i'u_i) = 0 \), where \( A_i \) is the \( T \times (T-1)/2 \) matrix

Similarly, the moment conditions in (8-29) above are also linear in \( \beta \) and can be written as \( E(B_i'u_i) = 0 \), where \( B_i \) is the \( T \times (T-2) \) matrix defined by

However, the moment conditions in (8-30) above are quadratic in \( \beta \).

We will discuss GMM estimation based on all of the available moment conditions and GMM
based on a subset (possibly all) of the linear moment conditions. Let $H_i = (C_i, A_i, B_i)$, which represents all of the available linear instruments. The corresponding linear moment conditions are $E[m_i(\beta)] = 0$, with

$$m_i(\beta) \equiv H_i' u_i = m_{ii} + m_{2i} \xi , \quad m_{ii} = H_i' y_i , \quad m_{2i} = -H_i' X_i .$$

The remaining nonlinear moment conditions will be written as $E[g_i(\beta)] = 0$. Since they are at most quadratic, we can write

$$g_i(\beta) \equiv g_{ii} + g_{2i} + (I_q \otimes \beta') g_{3i} \beta ,$$

where $g_{ii}$, $g_{2i}$ and $g_{3i}$ are conformable matrices of functions of data and $q$ is the number of moment conditions in $g_i$. An efficient estimator of $\beta$ can be obtained by GMM based on all of the moment conditions:

$$E[f_i(\beta)] \equiv E[m_i(\beta)' g_i(\beta)']' = 0 .$$
Define \( f_N = N^{-1} \sum_i f_i(\beta) \), with \( m_{1N}, m_{2N}, g_{1N}, g_{2N} \) and \( g_{3N} \) defined similarly; and define

\[ F_N = \frac{\partial f_N}{\partial \beta'} = [\partial m_N'/\partial \beta, \partial g_{1N}'/\partial \beta, \partial g_{2N}'/\partial \beta]' = [M_N', G_N', G_N']', \]

where \( G_N(\beta) = g_{2N} + 2(I \otimes \beta')g_{3N} \) and \( M_N = m_{2N} \). Let \( F = \text{plim} F_N \), with \( M \) and \( G \) defined similarly. Define the optimal weighting matrix:

\[ V = \begin{bmatrix} V_{mm} & V_{mg} \\ V_{gm} & V_{gg} \end{bmatrix} = E(\beta'\beta) \cdot \]

Let \( V_N \) be a consistent estimate of \( V \) of the form

\[ V_N = N^{-1} \sum_{i=1}^N f_i(\hat{\beta})f_i(\hat{\beta})', \]

where \( \hat{\beta} \) is an initial consistent estimate of \( \beta \) (perhaps based on the linear moment conditions \( m_i \), as discussed below); partition it similarly to \( V \).

In this notation, the efficient GMM estimator \( \hat{\beta}_{EGMM} \) minimizes \( Nf_N(\beta)'(V_N)^{-1}f_N(\beta) \). Using standard results, the asymptotic covariance matrix of \( N^{1/2}(\hat{\beta}_{EGMM} - \hat{\beta}) \) is \( [F'V^{-1}F]^{-1} \).

### 8.5.3.2 Linear Moment Conditions and Instrumental Variables

Some interesting questions arise when we consider GMM based on the linear moment conditions \( m_i(\xi) \) only. The optimal GMM estimator based on these conditions is

\[ \hat{\beta}_m = -[m_{2N}'(V_{N,mm})^{-1}m_{2N}]^{-1}m_{2N}'(V_{N,mm})^{-1}m_{1N} = [X'H(V_{N,mm})^{-1}X'H]^1X'H(V_{N,mm})^{-1}X'y \cdot \]

This GMM estimator can be compared to the 3SLS estimator which is obtained by replacing \( V_{N,mm} \) by \( N^{1/2}H_{i}^\prime \Sigma H_i \cdot \) As discussed in section 8.1, they are asymptotically equivalent in the case that \( V_{mm} \equiv E(H_i'u_i'u_i'H_i) = E(H_i'S_iH_i) \). For the case that \( H_i \) consists only of columns of \( C_i \), so that only the moment conditions \( E(C_i'u_i) =0 \) based on exogeneity of \( R_i \) and \( w_i \) are imposed, this equivalence may hold. Arellano and Bond [1991] considered the moment conditions (8-26), so that \( H_i \) also contains \( A_i \), and noted that asymptotic equivalence between the 3SLS and GMM estimates fails if we relax the homoskedasticity assumption, Assumption 8.3,

---

4Note that Assumptions 8.2 and 3 implies the random effects structure (8-3).
even though the moment conditions (8-26) are still valid under only Assumption 8.2. In fact, even the full set of Assumptions 8.2-3 is not sufficient to imply the asymptotic equivalence of the 3SLS and GMM estimates when the moment conditions (8-26) are used. Assumptions 8.2-3 deal only with second moments, whereas asymptotic equivalence of 3SLS and GMM involves restrictions on fourth moments (e.g., \( \text{cov}(y_i^2, \epsilon_{it}^2) = 0 \)). Ahn [1990] proved the asymptotic equivalence of the 3SLS and GMM estimators based on the moment conditions (8-26) for the case that Assumption 8.3 is maintained and Assumption 8.2 is strengthened by replacing uncorrelatedness with independence. Wooldridge [1996] provides a more general treatment of cases in which 3SLS and GMM are asymptotically equivalent. In the present case, his results indicate that asymptotic equivalence would hold if we rewrite Assumptions 8.2-3 in terms of conditional expectations instead of uncorrelatedness; that is, if we assume

\[
E(e_k | y_{10}, \alpha_i, \epsilon_{11}, \ldots, \epsilon_{i-1}) = 0,
\]

\[
E(\epsilon_{it}^2 | y_{10}, \alpha_i, \epsilon_{11}, \ldots, \epsilon_{i-1}) = \sigma_{ee}.
\]

A more novel observation is that the asymptotic equivalence of 3SLS and GMM fails whenever we use the additional linear moment conditions (8-29). This is so even if Assumptions 8.2-3 are strengthened by replacing uncorrelatedness with independence. When uncorrelatedness in Assumptions 8.2-3 is replaced by independence, Ahn [1990, Chapter 3, Appendix 3] shows that, while \( E(A_i'u_i'A_i) = \sigma_{ee}E(A_i'A_i) = E(A_i'\Sigma A_i) \) and \( E(A_i'u_i'B_i) = \sigma_{ee}E(A_i'B_i) = E(A_i'\Sigma B_i) \),

\[
E(B_i'u_i'B_i) = \sigma_{ee}E(B_i'B_i) + (\kappa + \sigma_{ee})L_{T-1}'L_{T-1} = E(B_i'\Sigma B_i) + (\kappa + \sigma_{ee})L_{T-1}'L_{T-1},
\]

where \( \kappa = E(\epsilon^4) - 3\sigma_{ee}^2 \) and \( L_{T-1} \) is the \((T-1)\times(T-2)\) differencing matrix defined similarly to \( L_T \) in section 8.1. Under normality \( \kappa = 0 \) but the term \( \sigma_{ee}L_{T-1}'L_{T-1} \) remains.
8.5.3.3 Linearized GMM

We now consider a linearized GMM estimator. Suppose that $\hat{\beta}$ is any consistent estimator of $\beta$; for example, $\hat{\beta}_m$. Following Newey (1985, p. 238), the linearized GMM estimator is of the form

$$\hat{\beta}_{LGMM} = \hat{\beta} - \left[ F_N(\hat{\beta})' \left( V_N^{-1} - F_N(\hat{\beta}) \right) F_N(\hat{\beta})' \right]^{-1} F_N(\hat{\beta})' \left( V_N^{-1} - F_N(\hat{\beta}) \right) \beta.$$

This estimator is consistent and has the same asymptotic distribution as $\hat{\beta}_{EGMM}$.

When the LGMM estimator is based on the initial estimator $\hat{\beta}_m$, some further simplification is possible. Using the fact that $m_{2N}(V_{N,mm})^{-1}m_N(\hat{\beta}_m) = 0$ and applying the usual matrix inversion rule to $V_N$, we can write the LGMM estimator as follows:

$$\hat{\beta}_{LGMM} = \hat{\beta}_m - \left[ \Gamma_N + B_N \left( V_{N,bb} \right)^{-1} B_N \right]^{-1} B_N \left( V_{N,bb} \right)^{-1} b_N,$$

where $\Gamma_N = m_{2N}(V_{N,bb})^{-1} m_{2N}$, $V_{N,bb} = V_{N,gg} - V_{N,gm}(V_{N,mm})^{-1} V_{N,mg}$, $b_N = g_N(\hat{\beta}_m) - V_{N,gm}(V_{N,mm})^{-1} m_N(\hat{\beta}_m)$, and $B_N = G_N(\hat{\beta}_m) - V_{N,gm}(V_{N,mm})^{-1} m_{2N}$. For more detail, see Ahn and Schmidt [1997].

8.6 Conclusions

In this chapter we have considered the GMM estimation of linear panel data models. We have discussed standard models, including the fixed and random effects linear model, a dynamic model, and the simultaneous equation model. For these models the typical treatment in the literature is some sort of instrumental variables procedure; least squares and generalized least squares are included in the class of such instrumental variables procedures.

It is well known that for linear models the GMM estimator often takes the form of an IV estimator if a no conditional heteroskedasticity condition holds. Therefore we have focused on three related points. First, for each model we seek to identify the complete set of moment conditions (instruments) implied by the assumptions underlying the model. Next, one can
observe that the usual exogeneity assumptions lead to many more moment conditions than standard estimators use, and ask whether some or all of the moment conditions are redundant, in the sense that they are unnecessary to obtain an efficient estimator. Under the no conditional heteroskedasticity assumption, the efficiency of standard estimators can often be established. This implies that the moment conditions which are not utilized by standard estimators are redundant. Finally, we ask whether anything intrinsic to the model makes the assumption of no conditional heteroskedasticity untenable. In some models, such as the dynamic model, this assumption necessarily fails if the full set of moment conditions is used, and correspondingly the efficient GMM estimator is not an instrumental variables estimator.

The set of non-redundant moment conditions can sometimes be very large. For example, this is true in the dynamic model, and also in simpler static models if the assumption of no conditional heteroskedasticity fails. In such cases the finite sample properties of the GMM estimator using the full set of moment conditions may be poor. An important avenue of research is to find estimators which are efficient, or nearly so, and yet have better finite sample properties than the full GMM estimator.
REFERENCES

Ahn, S.C., 1990, Three essays on share contracts, labor supply, and the estimation of models for

of Econometrics 68, 5-27.


Amemiya, T. and T. E. MaCurdy, 1986, Instrumental-variables estimation of an error-
components model, Econometrica 54, 869-880.

Andersen, T. G. and R. E. Sørensen, 1996, GMM estimation of a stochastic volatility model: A

Anderson, T.W. and C. Hsiao, 1981, Estimation of dynamic models with error components,

Arellano, M. and S. Bond, 1991, Tests of specification for panel data: Monte Carlo evidence and
an application to employment equations, Review of Economic Studies 58, 277-297.

Arellano, M. and O. Bover, 1995, Another look at the instrumental variables estimation of error-

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.

Blundell, R. and S. Bond, 1997, Initial conditions and moment restrictions in dynamic panel data
models, Unpublished manuscript, University College London

Breusch, T. S., G. E. Mizon and P. Schmidt, 1989, Efficient estimation using panel data,
Econometrica 57, 695-700.

Chamberlain, G., 1992, Comment: sequential moment restrictions in panel data, Journal of

Cornwell, C., P. Schmidt and D. Wyhowski, 1992, Simultaneous equations and panel data,

Crépon, B., F. Kramarz and A. Trognon, 1995, Parameter of interest, nuisance parameter and
orthogonality conditions, unpublished manuscript, INSEE.

Hahn, J, 1997, Efficient estimation of panel data models with sequential moment restrictions,


Koenker, R. and A. F. Machado, 1997, GMM inferences when the number of moment conditions is large, Unpublished manuscript, University of Illinois at Champaign.


