Chap4 Instrumental Variable Estimation and Dynamic Panel Data Model

Recent panel data applications have relied heavily on the methods of instrumental variables that we are developing here. We will develop this methodology in detail in section 2 where we consider generalized method of moments (GMM) estimation. At this point, we can examine two major building blocks in this set of methods, Huasman and Taylor's (1981) estimator for the random effects model and Bhargava and Sargan's (1983) proposals for estimating a dynamic panel data model. These two tools play a significant role in the GMM estimators of dynamic panel models.

4.1 Instrumental variables estimation for $Cov(X_{it}, u_i) \neq 0$： the Hausman and Taylor estimator

1）Hausman and Taylor estimator

Recall the original specification of the linear model for panel data

$$Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it} \tag{4.1}$$

The random effects model is based on the assumption that the unobserved person specific effects, u_i, are uncorrelated with the included variables, X_{it}. This assumption is a major shortcoming of the model. However, the random effects treatment does allow the model to contain observed time invariant characteristics, such as demographic characteristics, while the fixed effects model does not－if present, they are simply absorbed into the fixed effects. Hausman and Taylor's (1981) estimator for the random effects model suggests a way to overcome the first of these while accommodating the second.

Their model is of the form:

$$Y_{it} = X_{1it}'\beta_1 + X_{2it}'\beta_2 + Z_{1i}'\alpha_1 + Z_{2i}'\alpha_2 + \varepsilon_{it} + u_i \tag{4.2}$$

where $\beta = (\beta_1', \beta_2')'$ and $\alpha = (\alpha_1', \alpha_2')'$. In this formulation, all individual effects denoted Z_i are observed. As before, unobserved individual effects that are contained in $Z_i'\alpha$ in (4.1) are contained in the individual-specific random term, u_i. Hausman and Taylor define four sets of observed variables in the model:

X_{1it} is K_1 variables that are time varying and uncorrelated with u_i,

Z_{1i} is L_1 variables that are time invariant and uncorrelated with u_i,

X_{2it} is K_2 variables that are time varying and are correlated with u_i,

Z_{2i} is L_2 variables that are time invariant and are correlated with u_i.

The assumptions about the random terms in the model are

$$\begin{aligned}
&E\left[u_i \middle| X_{1it}, Z_{1i}\right] = 0 \text{ though } E\left[u_i \middle| X_{2it}, Z_{2i}\right] \neq 0,\\
&Var\left[u_i \middle| X_{1it}, Z_{1i}, X_{2it}, Z_{2i}\right] = \sigma_u^2,\\
&Cov\left[\varepsilon_{it}, u_i \middle| X_{1it}, Z_{1i}, X_{2it}, Z_{2i}\right] = 0,\\
&Var\left[\varepsilon_{it} + u_i \middle| X_{1it}, Z_{1i}, X_{2it}, Z_{2i}\right] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2,\\
&Corr\left[\varepsilon_{it} + u_i, \varepsilon_{is} + u_i \middle| X_{1it}, Z_{1i}, X_{2it}, Z_{2i}\right] = \rho = \sigma_u^2 / \sigma^2
\end{aligned}$$

Note the crucial assumption that one can distinguish sets of variables X_1 and Z_1 that are uncorrelated with u_i from X_2 and Z_2 which are not. The likely presence of X_2 and Z_2 is what complicates specification and estimation of the random effects model in the first place.

By construction, any OLS or GLS estimators of this model are inconsistent when the model contains variables that are correlated with the random effects. Hausman and Taylor have proposed an instrumental variables estimator that uses only the information within the model (i.e., as already stated). The strategy for estimation is based on the following logic: First, by taking deviations from group means, we find that

$$Y_{it} - \overline{Y}_{i.} = \left(X_{1it} - \overline{X}_{1i.}\right)' \beta_1 + \left(X_{2it} - \overline{X}_{2i.}\right)' \beta_2 + \varepsilon_{it} - \overline{\varepsilon}_{i.} \quad (4.3)$$

which implies that β can be consistently estimated by least squares, in spite of the correlation between X_2 and u. Now, in the original model, Hausman and Taylor show that the group mean deviations can be used as (K1 + K2) instrumental variables for estimation of (β, α). That is the implication of (4.3). Because Z_1 is uncorrelated with the disturbances, it can likewise serve as a set of L1 instrumental variables. That leaves a necessity for L2 instrumental variables. The authors show that the group means for X_1 can serve as these remaining instruments, and the model will be identified so long as K1 is greater than or equal to L2. For identification purposes, then, K1 must be at least as large as L2.

The authors propose the following set of steps for consistent and efficient estimation:

Step 1. Obtain the within estimator of $\beta = \left(\beta_1', \beta_2'\right)'$ based on X_1 and X_2. The residual variance estimator from this step is a consistent estimator of σ_ε^2.

Step 2. Form the within model residuals, e_{it}, from the within regression at step 1. Stack the group means of these residuals in a full sample length data vector. Thus, $e_{it}^* = \bar{e}_i = \frac{1}{T}\sum_{t=1}^{T}\left(Y_{it} - X_{it}'\hat{\beta}^{(w)}\right),\ t = 1,\cdots,T, i = 1,\cdots,n$. (The individual constant term, α_i, is not included in e_{it}^*.) These group means are used as the dependent variable in an instrumental variable regression on Z_1 and Z_2 with instrumental variables Z_1 and X_1. (Note the identification requirement that K1, the number of variables in X_1 be at least as large as L2, the number of variables in Z_2.) The time invariant variables are each repeated T times in the data matrices in this regression. This provides a consistent estimator of α.

Step 3. The residual variance in the regression in step 2 is a consistent estimator of $\sigma_{**}^2 = \sigma_u^2 + \sigma_\varepsilon^2 / T$. From this estimator and the estimator of σ_ε^2 in step 1, we deduce an estimator of $\sigma_u^2 = \sigma_{**}^2 - \sigma_\varepsilon^2 / T$. We then form the weight for feasible GLS in this model by forming the estimate of

$$\theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2}}.$$

Step 4. The final step is a weighted instrumental variable estimator. Let the full set of variables in the model be

$$W_{it}' = \left(X_{1it}', X_{2it}', Z_{1i}', Z_{2i}'\right).$$

Collect these nT observations in the rows of data matrix W. The transformed variables for GLS are, as before when we first fit the random effects model,

$$W_{it}^{*\prime} = W_{it}' - \hat{\theta}\overline{W}_{i.}' \quad \text{and} \quad Y_{it}^* = Y_{it} - \hat{\theta}\overline{Y}_i$$

where $\hat{\theta}$ denotes the sample estimate of θ. The transformed data are collected in the rows data matrix W^* and in column vector Y^*. Note in the case of the time invariant variables in W_{it}, the group mean is the original variable, and the transformation just multiplies the variable by $1 - \hat{\theta}$. The instrumental variables are

$$V_{it}' = \left[\left(X_{1it} - \overline{X}_{1i.}\right)', \left(X_{2it} - \overline{X}_{2i.}\right)', Z_{1i}', \overline{X}_{1i}\right].$$

These are stacked in the rows of the nT × (K1 + K2 + L1 + K1) matrix V. Note for the third and fourth sets of instruments, the time invariant variables and group means are repeated for each member of the group. The instrumental variable estimator would be

$$\left(\hat{\beta}', \hat{\alpha}'\right)'_{IV} = \left[\left(W^{*\prime}V\right)\left(V'V\right)^{-1}\left(V'W^{*}\right)\right]^{-1}\left[\left(W^{*\prime}V\right)\left(V'V\right)^{-1}\left(V'Y^{*}\right)\right]. \qquad (4.4)$$

The instrumental variable estimator is consistent if the data are not weighted, that is, if W rather than W^* is used in the computation. But, this is inefficient, in the same way that OLS is consistent but inefficient in estimation of the simpler random effects model.

2)Amemiya-Macurdy Estimator

For the instruments to be valid, Hausman and Taylor's estimatior requires that $\overline{X}_{1i}$ and Z_{1i} be uncorrelated with the rendon-effects u_i .More precisely ,the instruments are valid when $p\lim_{n\to 0}\frac{1}{n}\sum_{i=1}^{n}\overline{X}_{1i.}u_i = 0$ and $p\lim_{n\to 0}\frac{1}{n}\sum_{i=1}^{n}Z_{1i}u_i = 0$. Amemiya-Macurdy(1986) place stricter requirements on the instruments that vary within groups to obtain a more efficient estimator. Specifically , Amemiya-Macurdy(1986) assume that X_{1it} is orthogonal to u_i in every period; i.e. $p\lim_{n\to 0}\frac{1}{n}\sum_{i=1}^{n}X_{1it}u_i = 0$ for $t = 1,\cdots,T$ .With this restriction, they derive the Amemiya-Macurdy estimator as the instrumental-variables regression of (4.2) using instruments $X_{1it} - \overline{X}_{1i.}$, $\breve{X}_{1it}$,and Z_{1i} , where $\breve{X}_{1it} = X_{1i1}, X_{1i2}, \cdots, X_{1iT}$.

4.2 Instrumental variables estimation for $Cov\left(X_{it}, \varepsilon_{it}\right) \neq 0$ ： 2SLS Estimator

Consider an equation of the form

$$y_{it} = Y_{it}\gamma + X_{1it}\beta + u_i + \varepsilon_{it} = Z_{it}\delta + u_i + \varepsilon_{it} \qquad (4.5)$$

where

y_{it} is the dependent variable;

Y_{it} is an $1\times g_2$ vector of observations on g_2 endogenous variables included as covariates, and these variables are followed to be correlated with the ε_{it} ;

X_{1it} is an $1\times k_1$ vector of observations on the exogenous variables included as covariates;

$Z_{it} = \left[Y_{it}, X_{it}\right]$;

γ is a $g_2 \times 1$ vector of coefficients;

β is a $k_1 \times 1$ vector of coefficients;

δ is a $K \times 1$ vector of coefficients, where $K = g_2 + k_1$.

Assume that there is a $1 \times k_2$ vector of observations on the k_2 instruments in X_{2it}. The order condition is satisfied if $k_2 \geq g_2$. Let $X_{it} = \left[X_{1it}, X_{2it} \right]$.Define T_i to be the number of observations on individual i, n to be the number of individuals and N to be the total number of observations; i.e., $N = \sum_{i=1}^{n} T_i$.

1）FD2SLS

As the name implies, this estimator obtains the estimates and conventional VEC from an instrumental-variables regression on the first-differenced data. Specifically, first differencing the data yields

$$y_{it} - y_{it-1} = \left(Z_{it} - Z_{i,t-1} \right) \delta + \varepsilon_{it} - \varepsilon_{i,t-1} \quad (4.6)$$

With the u_i removed by differencing, we can obtain the estimated coefficients and their estimated variance-covariance matrix from a stand two-stage least-squares regression of Δy_{it} on ΔZ_{it} with instruments ΔX_{it}.

2) FE2SLS

At the heart of this model is the within transformation. The within transform of a variable w $\left(for \quad w \in \{ y, Z, \varepsilon \} \right)$ is

$$\tilde{w}_{it} = w_{it} - \overline{w}_{i.} + \overline{\overline{w}}$$

where

$$\overline{w}_{i.} = \frac{1}{T_i} \sum_{t=1}^{T_i} w_{it}$$

$$\overline{\overline{w}} = \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T_i} w_{it}$$

and n is the number of groups and N is the total number of observations on the variable.

The within transform of (4.5) is

$$\tilde{y}_{it} = \tilde{Z}_{it} \beta + \tilde{\varepsilon}_{it}$$

The within transform has removed the u_i. With the u_i gone, the within 2SLS estimator can be obtained from a two-stage least-squares regression of $\tilde{y}_{it}$ on $\tilde{Z}_{it}$ with instruments $\tilde{X}_{it}$.

Suppose that there are K variables in Z_{it}, including the mandatory constant. There are $K+n-1$ parameters estimated in the model, and the conventional VCE for the within estimator is

$$\frac{N-K}{N-n-K+1}V_{IV}$$

where V_{IV} is the VCE from the above two-stage least-squares regression.

From the estimate of $\hat{\delta}$, estimates $\hat{u}_i$ of u_i are obtained as $\hat{u}_i = \bar{y}_i - \bar{Z}_i\hat{\delta}$.

3)BE2SLS

After passing (4.5)through the between transform, we are left with

$$\bar{y}_i = \alpha + \bar{Z}_i\delta + u_i + \bar{\varepsilon}_i \tag{4.6}$$

where

$$\bar{w}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} w_{it} \qquad for \quad w \in \{y, Z, \varepsilon\}$$

Similarly, define $\bar{X}_i$ as the matrix of instruments X_{it} after they have been passed through the between transform.

Through a two-stage least-squares regression of $\bar{y}_i$ on $\bar{Z}_i$ with instruments $\bar{X}_i$ in which each average appears T_i times, we obtain the BE2SLS estimator of (4.6).

4) G2SLS and EC2SLS

Per Baltagi and Chang (2000), let

$$\eta_{it} = u_i + \varepsilon_{it}$$

be the combined errors. Then under the assumptions of the random-effects model,

$$E(\eta\eta') = \sigma_\varepsilon^2 diag\left[I_{T_i} - \frac{1}{T_i} i_{T_i} i'_{T_i}\right] + diag\left[w_i \frac{1}{T_i} i_{T_i} i'_{T_i}\right]$$

where

$$w_i = T_i\sigma_u^2 + \sigma_\varepsilon^2$$

and i_{T_i} is a vector of ones of dimension of T_i.

Since the variance components are unknown, consistent estimates are required to implement feasible GLS.We have two choices. The first choice is a simple extension of the Swamy-Arora method for unbalanced panels.

Let

$$\eta_{it}^w = \tilde{y}_{it} - \tilde{Z}_{it}\hat{\delta}_w$$

be the combined residuals from the within estimator. Let $\tilde{\eta}_{it}$ be the within-transformed η_{it}. Then

$$\hat{\sigma}_{\varepsilon}^{2}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T_i}\tilde{\eta}_{it}^{2}}{N-n-K+1}$$

Let

$$\eta_{it}^{b}=y_{it}-Z_{it}\delta_{b}$$

be the combined residual from the between estimator. Let $\bar{\eta}_{i}^{b}$ be the between residuals after they have been passed through the between transform. Then

$$\hat{\sigma}_{u}^{2}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T_i}\bar{\eta}_{it}^{2}-\left(n-K\right)\hat{\sigma}_{\varepsilon}^{2}}{N-r}$$

where

$$r=trace\left\{\left(\bar{Z}_i'\bar{Z}_i\right)^{-1}\bar{Z}_i'Z_uZ_u'\bar{Z}_i\right\}$$

where

$$Z_u=diag\left(i_{T_i}i_{T_i}'\right)$$

From the second choice , we get the consistent estimators that Baltagi and Chang (2000) are used. These are given by

$$\hat{\sigma}_{\varepsilon}^{2}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T_i}\tilde{\eta}_{it}^{2}}{N-n}$$

and

$$\hat{\sigma}_{u}^{2}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T_i}\bar{\eta}_{it}^{2}-n\hat{\sigma}_{\varepsilon}^{2}}{N}$$

The Swamy-Arora method contains a degree-of-freedom correction to improve its performance in small samples.

Given estimates of the variance components, $\hat{\sigma}_{\varepsilon}^{2}$ and $\hat{\sigma}_{u}^{2}$ the feasible GLS transform of a variable w is

$$w_{it}^{*}=w_{it}-\hat{\theta}_i\bar{w}_{i.} \qquad (4.7)$$

where

$$\bar{w}_{i.}=\frac{1}{T_i}\sum_{t=1}^{T_i}w_{it}$$

$$\hat{\theta}_i=1-\left(\frac{\hat{\sigma}_{\varepsilon}^{2}}{\hat{\omega}_i}\right)^{-\frac{1}{2}}$$

and

$$\hat{\omega}_i = T_i \hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2$$

Using either estimator of the variance components, there are two GLS estimators of the random-effects model. These two estimators differ only in how they construct the GLS instruments from the exogenous and instrumental variables contained in $X_{it} = \left[X_{1it}, X_{2it} \right]$. The first one method, G2SLS, which is from Balestra and Varadharajan-Krishnakumar, uses the exogenous variables after they have been passed through the feasible GLS transform. Mathematically, G2SLS uses X^* for the GLS instruments, where X^* is constructed by passing each variable in X through the GLS transform in (4.7). The G2SLS estimator obtains in coefficient estimates and conventional VCE from an instrumental variable regression of y_{it}^* on Z_{it}^* with instruments X_{it}^*.

For Baltagi's EC2SLS, the instruments are $\tilde{X}_{it}$ and $\overline{X}_{it}$, where $\tilde{X}_{it}$ is constructed by each of the variables in X_{it} throughout the GLS transform in (4.7), and $\overline{X}_{it}$ is made of the group means of each variable in X_{it}. The EC2SLS estimator can be obtained from an instrumental variables regression of y_{it}^* on Z_{it}^* with instruments $\tilde{X}_{it}$ and $\overline{X}_{it}$.

Baltagi and Li (1992) show that although the G2SLS instruments are a subset of those in EC2SLS, the extra instruments in EC2SLS may be redundant in the sense of White (2001).

4.2 Dynamic Panel Data Model

4.2.1 Introduction

Many economic relationships are dynamic in nature and one of the advantages of panel data is that they allow the researcher to better understand the dynamics of adjustment. The dynamic relationship of a panel data model is characterized by the presence of a lagged dependent variable among the regressors, i.e.

$$Y_{it} = \gamma Y_{i,t-1} + X_{it}'\beta + u_i + \varepsilon_{it}, \quad (i = 1,\cdots,n; t = 1,\cdots,T) \tag{4.8}$$

where γ is a scalar, X_{it}' is $1\times K$ and β is $K \times 1$. u_i is the individual effect of ith unit, either fixed or random effects. We have assumptions:

A.4.1.1 $\varepsilon_{it} \sim iid\left(0,\sigma_{\varepsilon}^2\right)$

A.4.1.2 $u_i \sim iid\left(0,\sigma_u^2\right)$

A.4.1.3 u_i and ε_{it} are independent of each other.

The basic problems introduced by the inclusion of a lagged dependent variable.

1) For the fixed effects estimator, we have

$$Y_{it} - \overline{Y}_i = \gamma\left(Y_{i,t-1} - \overline{Y}_{i,-1}\right) + \left(X_{it} - \overline{X}_{i.}\right)'\beta + \varepsilon_{it} - \overline{\varepsilon}_i \tag{4.9}$$

The within transformation wipes out the u_i, but $\left(Y_{i,t-1} - \overline{Y}_{i,-1}\right)$, where $\overline{Y}_{i,-1} = \sum_{t=1}^{T} Y_{i,t-1} \big/ T$ (assuming that $t = 0$ is observed), will be correlated with $\varepsilon_{it} - \overline{\varepsilon}_i$ even if the ε_{it} are not serially correlated. This is because $Y_{i,-1}$ is correlated with $\overline{\varepsilon}_i$, the latter average contains $\varepsilon_{i,t-1}$, which is obviously correlated with $Y_{i,t-1}$.

2) For the random effects GLS estimator, in order to apply GLS, quasi-demeaning is performed and $\left(Y_{i,t-1} - \theta\overline{Y}_{i,-1}\right)$ will be correlated with $\left(\varepsilon_{it} - \theta\overline{\varepsilon}_i\right)$.

So, for a dynamic panel data model, the estimator is biased and inconsistent, whether the effects are treated as fixed or random. This bias is of order $1/T$ and disappears only if $T \to \infty$. The bias can be serious when T is small.

4.2.2 The Difference and System GMM Estimators

1) The differences of Arellano and Bond (1991)

Arellano and Bond (1991) argue that additional instruments can be obtained in a dynamic panel data model if one utilized the orthogonality conditions that exist between lagged values of Y_{it} and the disturbances ε_{it} .

By first differencing model (4.8) as $n \to \infty$ and with T fixed, we have

$$Y_{it} - Y_{i,t-1} = \gamma\left(Y_{i,t-1} - Y_{i,t-2}\right) + \left(X_{it} - X_{i,t-1}\right)' \beta + \varepsilon_{it} - \varepsilon_{i,t-1} \qquad (4.10)$$

i.e.

$$\Delta Y_i = \begin{pmatrix}\Delta Y_{i,-1} & \Delta X_i\end{pmatrix}\begin{pmatrix}\gamma \\ \beta\end{pmatrix} + \Delta\varepsilon_i, \quad \left(i = 1, \cdots, n\right) \qquad (4.11)$$

We note that $\varepsilon_{it} - \varepsilon_{i,t-1}$ is MA(1). In Period $t = 2$, the variable Y_{i0} is a valid instrument, since it is hightly correlated with $\left(Y_{i1} - Y_{i0}\right)$ and not correlated with $\left(\varepsilon_{i2} - \varepsilon_{i1}\right)$ as long as the ε_{it} are not serially correlated. When $t = 3$, Y_{i1} and Y_{i0} are correlated with $\left(Y_{i2} - Y_{i1}\right)$ while, given the assumption of no serial correlation of the ε 's, they are correlated with $\left(\varepsilon_{i3} - \varepsilon_{i2}\right)$. One can continue this fashion, the set of valid instruments becomes $\begin{pmatrix}Y_{i0} & Y_{i2} & \cdots & Y_{i,T-2}\end{pmatrix}$. On the other hand, the GMM estimator depends on the exogeneity of X_{it} .

Case i) $E\left(X_{it}\varepsilon_{is}\right) = 0$ for all $t, s = 0, 1, \cdots, T$

In this case, all the X_{it} are exogenous variables, and they can instrument themselves. So We have IV matrix

$$Z_i = \begin{pmatrix} Y_{i0} & 0 & \cdots & 0 & X_{i0} \\ 0 & Y_{i0}, Y_{i1} & \cdots & 0 & X_{i1} \\ 0 & 0 & \cdots & 0 & X_{i2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & Y_{i0}, Y_{i1}, \cdots, Y_{iT-2} & X_{iT} \end{pmatrix} \qquad (4.12)$$

In (4.12), the instruments X_{it} are in “IV-style”, and the instruments from the second lag of Y are in “GMM-style”, we can substitute zeros for missing observations ,for example, when $t = 0$ is not observed, the first row of the matrix corresponds to $t = 0$. In unbalance panel,one also substitute zeros for other missing values. Althous these instrument sets are part of what defines difference (and system) GMM, researchers are free to incorporate other instruments instead or in addition. According to (4.22), we have orthogonality moment conditions

$$E\left(Z_i'\Delta\varepsilon_i\right) = 0 \qquad (4.13)$$

Its sample analogue

$$\frac{1}{n}\sum_{i=1}^{n} Z_i' \Delta\varepsilon_i = 0 \tag{4.14}$$

The Criterion function

$$q = \left(\frac{1}{n}\sum_{i=1}^{n} \Delta\varepsilon_i' Z_i\right)\hat{\Phi}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i' \Delta\varepsilon_i\right) \tag{4.15}$$

$$\begin{aligned}\Phi &= A\,\mathrm{var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i'\Delta\varepsilon_i\right)\\ &= \frac{1}{n}E\left(\sum_{i=1}^{n} Z_i'\Delta\varepsilon_i\Delta\varepsilon_i' Z_i\right)\\ &= \frac{1}{n}\sum_{i=1}^{n} Z_i' E\left(\Delta\varepsilon_i\Delta\varepsilon'\right)_i Z_i\end{aligned} \tag{4.16}$$

Since $\left(\varepsilon_{it} - \varepsilon_{i,t-1}\right)$ is MA(1), we have

$$E\left(\Delta\varepsilon_i\Delta\varepsilon_i'\right) = \left(\sigma_\varepsilon^2 G\right)$$

Where

$$G = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix} \tag{4.17}$$

Therefore

$$\Phi = \frac{1}{n}\sigma_\varepsilon^2 \sum_{i=1}^{n} Z_i' G Z_i \tag{4.18}$$

$$\frac{\partial q}{\partial \begin{pmatrix}\gamma \\ \beta\end{pmatrix}'} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\Delta Y_{i,-1}, \Delta X_i\right)' Z_i\right]\left[\frac{1}{n}\sum_{i=1}^{n} Z_i' G Z_i\right]^{-1}$$

$$\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i'\left[\Delta Y_i - \left(\Delta Y_{i,-1}, \Delta X_i\right)\begin{pmatrix}\gamma \\ \beta\end{pmatrix}\right]\right\} = 0$$

$$\begin{aligned}\begin{pmatrix}\hat{\gamma} \\ \hat{\beta}\end{pmatrix}_{GMM} &= \left\{\left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\Delta Y_{i,-1} \\ \Delta X_i\end{pmatrix} Z_i\right]\left[\frac{1}{n}\sum_{i=1}^{n} Z_i' G Z_i\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} Z_i'\left(\Delta Y_{i,-1}, \Delta X_i\right)\right]\right\}^{-1}\\ &\quad\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\Delta Y_{i,-1} \\ \Delta X_i\end{pmatrix} Z_i\right]\left[\frac{1}{n}\sum_{i=1}^{n} Z_i' G Z_i\right]^{-1}\frac{1}{n}\sum_{i=1}^{n} Z_i'\Delta Y_i\right\}\end{aligned} \tag{4.19}$$

On the other hand, we write model (4.21) as

$$\Delta Y = \left(\Delta Y_{-1}, \Delta X\right)\begin{pmatrix}\gamma \\ \beta\end{pmatrix} + \Delta\varepsilon \tag{4.20}$$

Where $\left(\Delta Y_{-1}, \Delta X\right) = \begin{pmatrix} \Delta Y_{1,-1} & \Delta X_1 \\ \Delta Y_{2,-1} & \Delta X_2 \\ \vdots & \vdots \\ \Delta Y_{n,-1} & \Delta X_n \end{pmatrix}, \quad \Delta\varepsilon = \begin{pmatrix} \Delta\varepsilon_1 \\ \Delta\varepsilon_2 \\ \vdots \\ \Delta\varepsilon_n \end{pmatrix}.$

Then the matrix of instruments is $Z = \left(Z_1', \cdots, Z_n'\right)'$, premultiplying the differenced equation (4.11) by Z', we get

$$Z'\Delta Y = Z'\left(\Delta Y_{-1}, \Delta X\right)\begin{pmatrix}\gamma \\ \beta\end{pmatrix} + Z'\Delta\varepsilon \tag{4.21}$$

The GLS estimators of (4.21) are

$$\begin{aligned}\begin{pmatrix}\hat{\gamma} \\ \hat{\beta}\end{pmatrix}_{IV,GLS} &= \left\{\left[\left(\Delta Y_{-1}, \Delta X\right)' Z\right]\left[Var\left(Z'\Delta\varepsilon\right)\right]^{-1} Z'\left(\Delta Y_{-1}, \Delta X\right)\right\}^{-1} \\ &\quad \left\{\left[\left(\Delta Y_{-1}, \Delta X\right)' Z\right]\left[Var\left(Z'\Delta\varepsilon\right)\right]^{-1} Z'\Delta Y\right\}\end{aligned} \tag{4.22}$$

Where $Var\left(Z'\Delta\varepsilon\right) = E\left(Z'\Delta\varepsilon\Delta\varepsilon' Z\right) = \frac{\sigma_\varepsilon^2}{n}\sum_{i=1}^{n} Z_i' G Z_i$, thus (4.19) is equivalent to (4.21).

Case ii) $E\left(X_{it}\varepsilon_{is}\right) \neq 0$ for $s < t$, $E\left(X_{it}\varepsilon_{is}\right) = 0$ for $s \geq t$ (4.23)

In this case, X_{it} are predetermined rather than strictly exogenous, then only $\left[X_{i0}', X_{i1}', \cdots, X_{is-1}'\right]$ are valid instruments for the differenced equation at period s, the IV matrix of (4.12) becomes:

$$Z_i = \begin{pmatrix} \left(Y_{i0}, X_{i0}', X_{i1}'\right) & 0 & \cdots & 0 \\ 0 & \left(Y_{i0}, Y_{i1}, X_{i0}', X_{i1}', X_{i2}'\right) & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(Y_{i0}, \cdots, Y_{iT-2}, X_{i0}', \cdots, X_{iT-1}'\right) \end{pmatrix} \tag{4.24}$$

And GMM estimators are again given by (4.19) with this choice of Z_i.

2) The Orthogonal Deviation Method (Arellano and Bover, 1995)

Arellano and Bover (1995) developed a unifying GMM framework for looking at efficient IV estimator for dynamic panel data models. They did it in the context of the Hausman and Taylor (1981) model, which in static form is reproduced here

$$Y_{it} = X'_{it}\beta + Z'_i\alpha + u_i + \varepsilon_{it} \qquad (4.25)$$

Here assumptions are the same as (4.2).In vector form, (425) can be written as

$$Y_i = W_i\delta + \eta_i \qquad (4.26)$$

Where $\delta = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$, $W_i = \left(X_i, i_T Z'_i\right)$, $\eta_i = u_i i_T + \varepsilon_i$, according to the assumptions, we have $\Sigma = E\left(\eta_i\eta'_i \middle| W_i\right) = \sigma_\varepsilon^2 I_T + \sigma_u^2 i_T i'_T$.

Let $X = \left(X_1, X_2\right)$, $Z = \left(Z_1, Z_2\right)$, Z are time-invariant variables, X_1 and Z_1 are exogenous of dimension $nT \times K_1$ and $n \times L_1$, X_2 and Z_2 are correlated with the individual effects and are of dimension $nT \times K_2$ and $n \times L_2$.

Arellano and Bover transform the system of T equations in (4.26) using the nonsingular transformation

$$H = \begin{bmatrix} M^{01} \\ i'_T/T \end{bmatrix} \qquad (4.27)$$

Where M^{01} is any $\left(T-1\right)\times T$ matrix of rank $\left(T-1\right)$ such that $M^{01}i_T = 0$, for example, by taking

$$M^{01} = \left(I_{T-1}, 0\right) - \frac{1}{T} i_{T-1} i'_T$$

M^{01} could be the first $\left(T-1\right)$ rows of the within group operator. Premultiplying (4.25) by matrix H, we have

$$HY_i = HW_i\delta + H\eta_i \qquad (4.28)$$

Where

$$H\eta_i = \begin{pmatrix} M^{01}\eta_i \\ \bar{\eta}_i \end{pmatrix} = \begin{pmatrix} \eta_{it} - \bar{\eta}_i \\ \bar{\eta}_i \end{pmatrix} \begin{matrix} (T-1)\times 1 \\ 1\times 1 \end{matrix} \qquad (4.29)$$

Note that the transformed disturbances $H\eta_i$ have the first $\left(T-1\right)$ transformed errors free of u_i, hence all exogenous variables are valid instruments for the first $\left(T-1\right)$ equations. Arellano and Bover suggest a GMM estimator and show that efficiency gains are available by using a larger set of moment conditions. We now form a matrix of IV. We will form a matrix V_i consisting of $T-1$ rows constructed the same way for $T-1$ observations and a final row that will be different, as discussed latter. The matrix will be of the form

$$V_i = \begin{bmatrix} v_{i1}' & 0' & \cdots & 0' \\ 0' & v_{i2}' & \cdots & 0' \\ \vdots & \vdots & \ddots & \vdots \\ 0' & 0' & \cdots & a_i' \end{bmatrix} \quad (4.30)$$

The instrumental variable sets contained in v_{it}' which have been suggested might include the following from within the model:

X_{it} and $X_{i,t-1}$ (i.e., current and one lag of all the time varying variables),

$X_{i1},\cdots,X_{iT}$ (i.e., all current, past and future values of all the time varying variables),

$X_{i1},\cdots,X_{it}$ (i.e., all current and past values of all the time varying variables).

The time-invariant variables that are uncorrelated with u_i, that is Z_{1i}, are appended at the end of the nonzero part of each of the first $T-1$ rows. It may seem that including X_2 in the instruments would be invalid. However, we will be converting the disturbances to deviations from group means which are free of the latent effects. While the variables are correlated with u_i by construction, they are not correlated with $\varepsilon_{it}-\bar{\varepsilon}_i$. The final row of V_i is important to the construction. Two possibilities have been suggested:

$a_i' = \left[Z_{1i}' \quad \bar{X}_{i1} \right]$ (produces the Hausman and Taylor estimator),

$a_i' = \left[Z_{1i}'..X_{1i1}', X_{1i2}', \cdots, X_{1iT}' \right]$ (produces Amemiya and MaCurdy's estimator).

Note that the a variables are exogenous time-invariant variables, Z_{1i} and the exogenous time-varying variables, either condensed into the single group mean or in the raw form, with the full set of T observations.

The moment conditions are given by

$$E\left(V_i'H\eta_i\right)=0 \quad (4.31)$$

Premultiplying (4.25) by V_i', to get

$$V_i'HY_i = V_i'HW_i\delta + V_i'H\eta_i \quad (4.32)$$

Performing GLS on (4.32), we obtain the Arellano and Bover (1995) estimator:

$$\hat{\delta} = \begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix}$$

$$= \left\{ \left[\frac{1}{n} \sum_{i=1}^{n} W_i' H' V_i \right] \left[\frac{1}{n} \sum_{i=1}^{n} V_i' H \hat{\Sigma} H' V_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} V_i' H W_i \right] \right\}^{-1}$$

$$\left\{ \left[\frac{1}{n} \sum_{i=1}^{n} W_i' H' V_i \right] \left[\frac{1}{n} \sum_{i=1}^{n} V_i' H \hat{\Sigma} H' V_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} V_i' H Y_i \right]_i \right\} \tag{4.33}$$

Where $\hat{\Phi} = \frac{1}{n} \sum_{i=1}^{n} V_i' H \hat{\Sigma} H' V_i$ is the estimator of

$$E\left(V_i' H \eta_i \eta_i' H' V_i\right) = V_i' H E\left(\eta_i \eta_i'\right) H' V = V_i' H \Sigma H' V$$

In practice, the covariance matrix of the transformed system $H \Sigma H'$ is replaced by a consistent estimator usually.

$$H \hat{\Sigma} H' = \sum_{i=1}^{n} \hat{\eta}_i \hat{\eta}_i' \Big/ n \tag{4.34}$$

Where $\hat{\eta}_i$ are residuals based on consistent preliminary estimates, the resulting $\hat{\delta}$ is the optimal GMM estimator of δ based on the above moment restrictions.

let us now introduce a lagged dependent variable into the right-hand side of (4.26), to get

$$Y_{it} = \gamma Y_{i,t-1} + X_{it}' \beta + Z_i' \alpha + u_i + \varepsilon_{it} \tag{4.35}$$

Assuming that $t = 0$ is observed, if ε_{it} are not serially correlated, the transformed error in the equation for period t is independent of u_i and $\left(\varepsilon_{i1}, \cdots, \varepsilon_{i,t-1}\right)$. So that $\left(Y_{i0}, Y_{i1}, \cdots, Y_{i,t-1}\right)$ are additional valid instruments for the equation. Therefore, the matrix of instruments V_i becomes

$$V_i = \begin{bmatrix} \left(V_{i1}', Y_{i0}\right) & 0 & \cdots & 0 & 0 \\ 0 & \left(V_{i2}', Y_{i0}, Y_{i1}\right) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \left(V_{iT-2}', Y_{i0}, \cdots, Y_{i,T-3}\right) & 0 \\ 0 & 0 & \cdots & 0 & a_i' \end{bmatrix} \tag{4.36}$$

Once again, Arellano and Bover (1995) show that the GMM estimator (4.33) that uses (4.36) as the matrix is invariant to the choice of H provided H satisfies the above required conditions.

We have first-difference transform in the above GMM estimation, but the first-difference transform has a weakness. It magnifies gaps in unbalanced panels. If some Y_{it} is missing, for example, then both ΔY_{it} and ΔY_{it+1} are missing in the transformed data. This motivates the second common transformation, call **"forward orthogonal deviations".** Let w_{it} be a variable then the transform is:

$$w_{it}^{*} = c_t \left(w_{it} - \frac{1}{T-t}(w_{it+1} + w_{it+2} + \cdots + w_{iT}) \right)$$

$$for \quad w \in \{Y, X, Z, \varepsilon\} \tag{4.37}$$

where the sum is taken over available future observations, $T-t$ is the number of such observations; since lagged observations do not enter the formula, they are valid as instruments. And the scale c_t is $\sqrt{(T-t)/(T-t+1)}$.

3) The System-GMM(Blundell and Bond,1998)

Blundell and Bond(1998) pointed out, if Y is close to a random walk, then Difference GMM performs poorly because past levels convey little information about future changes, so that untransformed lags are weak instruments for transformed variables.

To increase efficiency, under an additional assumption:

$$E(\Delta Z_{it} u_i) = 0 \text{ for all } i \text{ and } t \tag{4.38}$$

Blundell and Bond(1998) develop an approach outlined in Arellano and Bover(1995).Instead of transforming the regressors to expunge the fixed effects, it transforms-differences-the instruments to make them exogenous to the fixed effects. In a nutshell, where Arellano -Bover instruments differences (or orthogonal deviations) with levels, Blundell –Bond instruments levels with differences. For random walk-like variables ,past changes may indeed be more predictive of current levels than past levels are of current changes, so the new instruments are more relevant. Again, validity depends on the assumption that the ε_{it} are not serially correlated. Otherwise Z_{it-1} and Z_{it-2} correlated with past and contemporary disturbances may correlate with future ones as well. In general, if Z is endogenous, ΔZ_{it-1} is available as an instrument since $\Delta Z_{it-1} = Z_{it-1} - Z_{it-2}$ should not correlated with ε_{it}; earlier realizations of ΔZ can serve as instruments as well. And if Z is predetermined, the contemporaneous $\Delta Z_{it} = Z_{it} - Z_{it-1}$ is also valid, since $E(Z_{it}\varepsilon_{it}) = 0$.

In System GMM, Blundell and Bond build a stacked data set with the observations: in each individual's data, the untransformed observations follow the transformed ones.i.e.

$$X_i^{\perp} = \begin{bmatrix} X_i^* \\ X_i \end{bmatrix}, \qquad Y_i^{\perp} = \begin{bmatrix} Y_i^* \\ Y_i \end{bmatrix}$$

where the $*$ superscript indicates data transformed by differencing or orthogonal deviations. And the instrument matrix

$$Z_i^{\perp} = \begin{pmatrix} Z_i^* & 0 \\ 0 & Z_i \end{pmatrix}$$

For GMM-style instrument, the Arellano -Bover instruments fot the transformed data are set to zero for levels observations, and the new instruments for the levels data are set to zero for transformed observations; for IV-style instrument, a strictly exogenous variable Z_{it} with observation vector as a single-column could be transformed and entered like the regressors above.

In System GMM, one can include time-invariant regressors, which would disappear in Difference GMM. Asymptotically, this does not affect the coefficient estimates for other regressors because all instruments for the levels equation are assumed to be orthogonal to fixed effects indeed to all time-invariant variables.

4.3 Dynamic Panel Data Models with Heterogeneous Slopes

4.3.1 Large Sample Bias of Dynamic Fixed and Random Effects Estimators

Consider the simple dynamic panel data model (ARDL(1,0)):

$$y_{it} = \alpha_i + \gamma_i y_{i,t-1} + \beta_i x_{it} + \varepsilon_{it}, \quad i = 1,2,\ldots,n;\ t = 1,2,\ldots,T, \tag{4.39}$$

where ε_{it}'s are $iid(0,\sigma_i^2)$,the slopes, γ_i and β_i, as well as the intercepts α_i, are allowed to vary across cross-sectional units(groups). Here x_{it} is a scalar random variable but the analysis can be readily extended to the case of more than one regressor. In what follows x_{it} will be treated as strictly exogenous.

Let $\delta_i = \beta_i / (1-\gamma_i)$ be the long-run coefficient of x_{it} for the i-the individual and rewrite (4.39) as:

$$\Delta y_{it} = \alpha_i - (1-\gamma_i)(y_{i,t-1} - \delta_i x_{it}) + \varepsilon_{it} \tag{4.40}$$

or

$$\Delta y_{it} = \alpha_i - \phi_i (y_{i,t-1} - \delta_i x_{it}) + \varepsilon_{it} \tag{4.41}$$

Consider now the random coefficients model

$$\phi_i = \phi + v_{i1}$$

$$\delta_i = \delta + v_{i2}$$

Hence

$$\beta_i = \delta_i \phi_i = \delta\phi + v_{i3} \tag{4.42}$$

where

$$v_{i3} = \phi v_{i2} + \delta v_{i1} + v_{i1} v_{i2} \tag{4.43}$$

$$\begin{pmatrix} v_{i1} \\ v_{i2} \end{pmatrix} \sim iid\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{pmatrix}\right] \tag{4.44}$$

and

$$w_{33} = Var\left(v_{i3}\right) = Var\left(\phi v_{i2} + \delta v_{i1} + v_{i1} v_{i2}\right) \tag{4.45}$$

Letting $\gamma = 1 - \phi$ and $\beta = \delta\phi$, and using the above in (4.46) we have

$$y_{it} = \alpha_i - \gamma y_{i,t-1} + \beta x_{it} + \eta_{it} \tag{4.46}$$

$$\eta_{it} = \varepsilon_{it} - v_{i1} y_{i,t-1} + v_{i3} x_{it} \tag{4.47}$$

It is now clear that η_{it} and $y_{i,t-1}$ are correlated and the fixed effects or random effects estimators will not be consistent.

4.4.2 Mean Group Estimation of Dynamic Heterogeneous Panels

Considering the dynamics panel data model (4.39), We estimate α_i, ϕ_i and δ_i for each individual separately using the observations $t = 1, 2, ..., T$ on T. It is easily seen that the MG estimators are consistent and have asymptotic normal distribution for n and T large. The nonparametric variance-covariance matrix of the MG estimator is given by

$$Var\left(\hat{\hat{\psi}}\right) = \frac{\sum_{i=1}^{n}\left(\hat{\psi}_i - \hat{\hat{\psi}}\right)\left(\hat{\psi}_i - \hat{\hat{\psi}}\right)'}{n\left(n-1\right)} \tag{4.48}$$

where

$$\psi_i = \begin{pmatrix} \alpha_i \\ \gamma_i \\ \beta_i \end{pmatrix} \tag{4.49}$$

Then the Mean Group (MG) estimators are given by

$$\hat{\hat{\alpha}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i, ...\hat{\hat{\gamma}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i ....\hat{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i. \quad .(4.50)$$

notice that the estimates $\hat{\alpha}_i$ and $\hat{\gamma}_i$ and $\hat{\beta}_i$ depend on T.

In general, writing the regression equations for each individual in matrix notation, we have:

$$y_i = X_i h_i + \varepsilon_i \tag{4.51}$$

The OLS estimator of

$$\hat{h}_i = \left(X_i' X_i\right)^{-1} X_i' y_i \quad . \tag{4.52}$$

The MG estimator of h is given by

$$\hat{\bar{h}}_{MG} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i' X_i\right)^{-1} X_i' y_i \quad . \tag{4.53}$$

Under the assumption that

$$\begin{aligned} h_i &= h + v_i \\ v_i &\sim iid\left(0, \Omega_v\right) \end{aligned}$$

we have that

$$\hat{\bar{h}}_{MG} = h + \left(\frac{1}{n}\sum_{i=1}^{n} v_i\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' X_i}{T}\right)^{-1}\frac{X_i'\varepsilon_i}{T} \tag{454}$$

and, asymptotically,

$$\left(\hat{\bar{h}}_{MG} - h\right) \sim N\left[0, Var\left(\hat{\bar{h}}\right)\right], \tag{4.55}$$

as $n \to \infty$ and $T \to \infty$, where $Var\left(\hat{\bar{h}}\right)$ can be obtained as above. When the X_i's are independently distributed across i, $Var\left(\hat{\bar{h}}\right)$ can also be derived analytically:

$$\begin{aligned} \hat{\bar{h}}_{MG} &= h + \left(\frac{1}{n}\sum_{i=1}^{n} v_i\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' X_i}{T}\right)^{-1}\frac{X_i'\varepsilon_i}{T} \\ Var\left(\hat{\bar{h}}_{MG}\right) &= Var\left(\hat{\bar{h}}_{MG} - h\right) \\ &= Var\left[\left(\frac{1}{n}\sum_{i=1}^{n} v_i\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i' X_i}{T}\right)^{-1}\frac{X_i'\varepsilon_i}{T}\right] \\ &= \frac{1}{n}\Omega_v + \frac{1}{n^2}\sum_{i=1}^{n} Var\left(\frac{X_i' X_i}{T}\right)^{-1}\frac{X_i'\varepsilon_i}{T} \\ &= \frac{1}{n}\Omega_v + \frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{X_i' X_i}{T}\right)^{-1}\frac{X_i'}{T}\sigma_i^2 I \frac{X_i}{T}\left(\frac{X_i' X_i}{T}\right)^{-1} \end{aligned}$$

$$Var\left(\hat{\bar{h}}\right) = \frac{1}{n}\Omega_v + \frac{1}{n^2 T}\sum_{i=1}^{n}\sigma_i^2\left(\frac{X_i' X_i}{T}\right)^{-1} \tag{4.56}$$

But this method still required a nonparametric estimation of Ω_v. Furthermore, the assumption that X_i's are distributed independently across i is too restrictive and may not hold in practice. We recommend using the variance formula given by (4.48)

4.3.3 Problem of small sample bias

When T is small, the MG estimator of the dynamic panel data model is biased and can yield misleading results. For finite T, as $n \to \infty$ (under the usual panel assumption of independence across units), the MG estimator still converges to a normal distribution, but with a mean which is not the same as the true value of the parameter underlying equations contain lagged dependent variables or nonexogenous regressors. For a finite T we have:

$$E\left(\hat{\bar{h}}\right) = h + \frac{1}{n}\sum_{i=1}^{n} E\left[\left(X_i'X_i\right)^{-1} X_i'\varepsilon_i\right], \tag{4.57}$$

and $n \to \infty$ will not eliminate the second term. One needs large enough T for the bias to disappear. In practice when the model contains lagged dependent variables we have that

$$E\left[\left(X_i'X_i\right)^{-1} X_i'\varepsilon_i\right] = \frac{K_{iT}}{T} + O\left(T^{-\frac{3}{2}}\right), \tag{4.58}$$

where K_{iT} is bounded in T and a function of the unknown underlying parameters. (See Kiviet and Phillips,1994) Hence

$$E\left(\hat{\bar{h}}\right) = h + \frac{1}{T}\sum_{i=1}^{n} \frac{K_{iT}}{n} + O\left(T^{-\frac{3}{2}}\right). \tag{4.59}$$

Alternative methods of dealing with the small sample bias of the MG estimator are considered in Hsiao, Pesaran, and Tahmiscioglu (1999) and Pesaran and Zhao (1999).

4.3.4 Hypothesis Testing with MG estimators

When using MG estimation in the context of hypothesis testing it is important to note that the validity of the tests crucially depends on the relative size of n and T. When the bias of individual estimates is of order 1/T, then the necessary condition for the validity of tests based on the MG estimator is given by

$$\frac{\sqrt{n}}{T} \to 0 \tag{4.60}$$

as $T \to \infty$ and $n(T) \to \infty$. When the small T bias in the individual estimates is of order $1/\sqrt{T}$, then the above condition becomes

$$\sqrt{\frac{n}{T}} \to 0 \tag{4.61}$$

as $T \to \infty$ and $n(T) \to \infty$.