


REFERENCES


WEBLIOGRAPHY

http://www.bseindia.com
http://www.valueresearchonline.com
http://www.sebi.gov.in
http://www.nseindia.com
http://www.moneycontrol.com
http://www.capitaline.com
http://www.rbi.org.in
http://www.iea.gov
http://www.virtualuniversity.com
http://www.listendata.com
http://www.syedhossain.com
http://www.emeraldinsight.com
http://www.businessballs.com
http://www.indianjournals.com
http://www.oxfordtextbooks.co.uk/orc/dougherty4e/
http://www.oup.com/uk/orc/bin/9780199280964/
NEWSPAPER, MAGAZINES AND OTHER SOURCES:

The Financial Times
Economics Times
Value Research
Business world
Business India
Outlook Money

Company's websites and others.
Various lectures on youtube.com
EBSCO data base
CMIE database/Prowess
Dave Giles Blog
Shishir Shakya Blog
Pablo Fernandez Blog
APPENDIX

Nonstationarity, Unit Roots and Cointegration

The recent literature exhibits an increasing integration of techniques and ideas from time-series analysis, such as unit roots and cointegration, into the area of panel data modelling. The underlying reason for this development is that researchers have increasingly realized that cross-sectional information is a useful additional source of information that should be exploited. To analyse the effect of a certain policy measure, for example adopting a road tax or a pollution tax, it may be more fruitful to compare with other countries than to try to extract information about these effects from the country’s own history. Pooling data from different countries may also help to overcome the problem that sample sizes of time series are fairly small, so that tests regarding long-run properties are not very powerful.

A number of recent articles discuss issues relating to unit roots, spurious regressions and cointegration in panel data. Most of this literature focuses upon the case in which the number of time periods $T$ is fairly large, while the number of cross-sectional units $N$ is small or moderate. As a consequence, it is quite important to deal with potential nonstationarity of the data series, while the presence of a unit root or cointegration may be of specific economic interest. For example, a wide range of applications exist concerning purchasing power parity, including Oh (1996), focusing on (non)stationarity of real exchange rates for a set of countries, or on testing for cointegration between nominal exchange rates and prices (compare Sections 8.5 and 9.3 and Subsection 9.5.4). For ease of discussion, we shall refer below to the cross-sectional units as countries, although they may also correspond to firms, industries or regions.

A crucial issue in analysing the time series on a number of countries simultaneously is that of heterogeneity. Because it is possible to estimate a separate regression for each country, it is natural to think of the possibility that model parameters are different across countries, a case commonly referred to as ‘heterogeneous panels’. Robertson and Symons (1992) and Pesaran and Smith (1995) stress the importance of parameter heterogeneity in dynamic panel data models and analyse the potentially severe biases that may arise from handling it in an inappropriate manner. Such biases are particularly misleading in a nonstationary world as the relationships of the individual series may be completely destroyed.

As long as we consider each time series individually, and the series are of sufficient length, there is nothing wrong with applying the time series techniques from Chapters 8 and 9. However, if we pool different series, we have to be aware of the possibility that their processes not all have the same characteristics or are described by the same parameters. For example, it is conceivable that $y_{it}$ is stationary for country 1 but integrated of order one for country 2. Even when all variables are integrated of order one in each country, heterogeneity in cointegration properties may lead to problems.
For example, if for each country $i$ the variables $y_{it}$ and $x_{it}$ are cointegrated with parameter $\beta_i$, it holds that $y_{it} - \beta_i x_{it}$ is $I(0)$ for each $i$, but in general there does not exist a common cointegrating parameter $\beta$ that makes $y_{it} - \beta x_{it}$ stationary for all $i$. Similarly, there is no guarantee that the cross-sectional averages $\bar{y}_i = (1/N) \sum_i y_{it}$ and $\bar{x}_i$ are cointegrated, even if all underlying individual series are cointegrated.

In Subsections 10.6.1 and 10.6.2, we pay attention to panel data unit root tests and cointegration tests, respectively. Basically, the tests are directed at testing the joint null hypothesis of a unit root (or the absence of cointegration) for each of the countries involved. In comparison to the single time-series case, panel data tests raise a number of additional issues, including cross-sectional dependence, heterogeneity in dynamics and error-term properties, and the type of asymptotics that is employed. While most asymptotic analysis is done with both $N$ and $T$ tending to infinity, there are various ways that this can be done.

### 10.6.1 Panel Data Unit Root Tests

To introduce panel data unit root tests, consider the autoregressive model

$$y_{it} = \alpha_i + \gamma_i y_{i,t-1} + \varepsilon_{it}, \quad (10.68)$$

which we can rewrite as

$$\Delta y_{it} = \alpha_i + \pi_i y_{i,t-1} + \varepsilon_{it}, \quad (10.69)$$

where $\pi_i = \gamma_i - 1$. The null hypothesis that all series have a unit root then becomes $H_0: \pi_i = 0$ for all $i$. A first choice for the alternative hypothesis is that all series are stationary with the same mean-reversion parameter, that is, $H_1: \pi_i = \pi < 0$ for each country $i$, and is used in the approaches of Levin and Lin (1992), Quah (1994) and Harris and Tzavalis (1999). A more general alternative allows the mean-reversion parameters to be potentially different across countries and states that $H_1: \pi_i < 0$ for at least one country $i$. This alternative is used by Maddala and Wu (1999), Choi (2001), Im, Pesaran and Shin (2003) and others. As in the time-series case discussed in Chapter 8, the properties of the test statistics (and their computation) depend crucially upon the deterministic regressors included in the test equation. For example, in (10.69) we have included a dummy for each country, corresponding to the fixed effect. Alternative tests are available in cases where the equation includes a common intercept, or in cases where a deterministic trend is added to the fixed effect.

For all tests, the null hypothesis is that the time series of all individual countries have a unit root. This implies that the null hypothesis can be rejected (in sufficiently large samples) if any one of the $N$ coefficients $\pi_i$ is less than zero. Rejection of the null hypothesis therefore does not indicate that all series are stationary. As Smith and Fuertes (2003) note, if the hypothesis of interest is that all series are stationary (for example, real exchange rates under purchasing power parity), it would be more appropriate to use a panel version of the KPSS test, as discussed in Section 8.4, where
stationarity is the null hypothesis rather than the alternative. However, a test like this may reject if just one series is nonstationary, which may not be interesting either. Because of these issues, Maddala, Wu and Liu (2000) argue that for purchasing power parity panel data unit root tests are the wrong answer to the low power of unit root tests in single time series.

In addition to the choice of deterministic regressors in the test equations, panel data unit root tests offer three additional technical issues in comparison with the single time-series case. First, one has to make assumptions on the cross-sectional dependence between \( e_{it} \)s, noting that virtually all the existing nonstationary panel data literature assume cross-sectional independence. Second, we need to be specific on the properties of \( e_{it} \) and how they are allowed to vary across the different units. This includes serial correlation and the possibility of heteroskedasticity across units. Third, asymptotic properties of estimators and tests depend crucially upon the way in which \( N \), the number of cross-sectional units, and \( T \), the number of time periods, tend to infinity (see Phillips and Moon, 1999). Some tests assume that either \( T \) or \( N \) is fixed and assume that the other dimension tends to infinity. Many tests are based on a sequential limit, where first \( T \) tends to infinity for fixed \( N \), and subsequently \( N \) tends to infinity. Alternatively, some tests assume that both \( N \) and \( T \) tend to infinity along a specific path (e.g. \( T/N \) being fixed). While the type of asymptotics that is applied may seem a theoretical issue, remember that we are using asymptotic theory to approximate the properties of estimators and tests in the finite sample that we happen to have. Although it is hard to make general statements on this matter, some asymptotic approximations are simply better than others. Many papers in this area therefore also contain a Monte Carlo study to analyse the finite sample behaviour of the proposed tests under controlled circumstances. A common finding for many of the tests below is that they tend to be oversized. That is, when the null hypothesis is true, the tests tend to reject more frequently than their nominal size (say, 5%) suggests. Further, many tests do not perform very well when the error terms are cross-sectionally correlated, or in the presence of cross-country cointegration. For example, when real exchange rates are \( I(1) \) and cointegrated across countries, the null hypothesis tends to be rejected too often (see Banerjee, Marcellino and Osbat, 2001, for an illustration).

While it is beyond the scope of this text to discuss alternative panel data unit root tests in great technical detail, a brief discussion of some tests is warranted. More details can be found in Banerjee (1999), Baltagi (2001, Chapter 12) or Enders (2004, Section 4.11). Levin and Lin (1992) and Harris and Tzavalis (1999) base their tests upon the OLS estimator for \( \pi \), assuming that \( e_{it} \) is i.i.d. across countries and time. Depending upon the deterministic regressors included, the OLS estimator may be biased, even asymptotically. When fixed effects are included, the estimator corresponds to the fixed effects estimator for \( \pi \) based on (10.69), which is biased for fixed \( T \) (see Section 10.4). With appropriate correction and standardization factors, test statistics can be derived that are asymptotically normal for \( N \to \infty \) and fixed \( T \) (Harris and Tzavalis) or both \( N, T \to \infty \) (Levin and Lin); see Bal-
tagi (2001, Section 12.2). While the test statistics can be modified to allow for serial correlation in $\varepsilon_t$, they do not allow cross-sectional dependence. This assumption is rather strong, and as stressed by O’Connell (1998) in a panel study on purchasing power parity, allowing for cross-sectional dependence may substantially affect inferences about the presence of a unit root. Because individual observations in a panel typically have no natural ordering, modelling cross-sectional dependence is not obvious.

The above two sets of tests are restrictive because they assume that $\pi_i$ is the same across all countries, also under the alternative hypothesis. The test proposed by Im, Pesaran and Shin (2003) allows $\pi_i$ to be different across individual units. It is based on averaging the augmented Dickey–Fuller (ADF) test statistics (see Section 8.4) over the cross-sectional units, while allowing for different orders of serial correlation. In fact, the alternative hypothesis states that $\pi_i < 0$ for at least one $i$ and thus allows that $\pi_i = 0$ for a subset of the countries. Im, Pesaran and Shin (2003) also propose a test based on the $N$ Lagrange multiplier statistics for $\pi_i = 0$, averaged over all countries. The idea underlying these tests is quite simple: if you have $N$ independent test statistics, their average will be asymptotically normally distributed for $N \to \infty$. Consequently, the tests are based on comparison of appropriately scaled cross-sectional averages with critical values from a standard normal distribution.

An alternative approach to combine information from individual unit root tests is employed by Maddala and Wu (1999) and Choi (2001), who propose panel data unit root tests based on combining the $p$-values of the $N$ cross-sectional tests. Let $p_i$ denote the $p$-value of the (augmented) Dickey–Fuller test for unit $i$. Under the null hypothesis, $p_i$ will have a uniform distribution over the interval $[0, 1]$, small values corresponding to rejection. The combined test statistic is given by

$$P = -2 \sum_{i=1}^{N} \log p_i.$$  \hspace{1cm} (10.70)

For fixed $N$, this test statistic will have a Chi-squared distribution with $2N$ degrees of freedom as $T \to \infty$, so that large values of $P$ lead us to reject the null hypothesis. While this test (sometimes referred to as the Fisher test) is attractive because it allows the use of different ADF tests and different time-series lengths per unit, a disadvantage is that it requires individual $p$-values that have to be derived by Monte Carlo simulations.

While the latter tests may seem attractive and easy to use, a word of caution is appropriate. Before one can apply the individual ADF tests underlying the Maddala and Wu (1999) and Im, Pesaran and Shin (2003) approaches, one has to determine the number of lags and determine whether a trend should be included. It is not obvious how this should be done. For a single time series, a common approach is to perform the ADF test for a range of alternative lag values. For example, in Table 8.2 we presentec
26 different (augmented) Dickey–Fuller test statistics for the log price index. If we were to combine the ADF tests for $N$ different countries, in whatever way, this creates a wide range of possible combinations. Smith and Fuertes (2003) warn for pre-test biases in this context.

### 10.6.2 Panel Data Cointegration Tests

A wide range of alternative tests is available to test for cointegration in a dynamic panel data setting, and research in this area is evolving rapidly. A substantial number of these tests are based on testing for a unit root in the residuals of a panel cointegrating regression. The drawbacks and complexities associated with the panel unit root tests are also relevant in the cointegration case. Several additional issues are of potential importance when testing for cointegration: heterogeneity in the parameters of the cointegrating relationships, heterogeneity in the number of cointegrating relationships across countries and the possibility of cointegration between the series from different countries. A final issue is that of estimating the cointegrating vectors, for which several alternative estimators are available, with different small and large sample properties (depending upon the type of asymptotics that is chosen).

When the cointegrating relationship is unknown, which is almost always the case, most cointegration tests start with estimating the cointegrating regression. Let us focus on the bivariate case and write the panel regression as

$$ y_{it} = \alpha_i + \beta_i x_{it} + \epsilon_{it}, \quad (10.71) $$

where both $y_{it}$ and $x_{it}$ are integrated of order one. Cointegration implies that $\epsilon_{it}$ is stationary for each $i$. Homogeneous cointegration, in addition, requires that $\beta_i = \beta$. If the cointegrating parameter is heterogeneous, and homogeneity is imposed, one estimates

$$ y_{it} = \alpha_i + \beta x_{it} + [(\beta_i - \beta) x_{it} + \epsilon_{it}], \quad (10.72) $$

and in general the composite error term is integrated of order one, even if $\epsilon_{it}$ is stationary. However, the problem of spurious regressions may be less relevant in this situation. This is because a pooled estimator will also average over $i$, so that the noise in the equation will be attenuated. In many circumstances, when $N \to \infty$, the fixed effects estimator for $\beta$ is actually consistent for the long-run average relation parameter, as well as asymptotically normal, despite the absence of cointegration (see Phillips and Moon, 1999). However, the meaning of this long-run relationship, in the absence of cointegration, is open to some interpretation (see Hsiao, 2003, Section 10.2 for some discussion). With heterogeneous cointegration, the long-run average estimated from the pooled regression may differ substantially from the average of the cointegration parameters averaged over countries (see Pesaran and Smith, 1995). Consequently, if there is heterogeneous cointegration, it is much better to estimate the individual cointegrating regressions rather than using a pooled estimator. Obviously, this requires $T \to \infty$. 
To test for cointegration, the panel data unit root tests from the previous section can be applied to the residuals from these regressions, provided that the critical values are appropriately adjusted (see Pedroni, 1999, or Kao, 1999). Recall that these tests assume cross-sectional independence. Some tests assume homogeneity of the cointegrating parameter and use a pooled OLS or dynamic OLS estimator (see Subsection 9.2.2). Additional discussion on these tests can be found in Banerjee (1999), Baltagi (2001, Chapter 4) or Smith and Fuertes (2003).

With more than two variables, an additional complication may arise because more than one cointegrating relationship may exist for one or more of the countries. Further, even with one cointegrating vector per country, the results will be sensitive to the normalization constraint (left-hand side variable) that is chosen. Finally, the existence of between-country cointegration may seriously distort the results of within-country cointegration tests.

10.2 The Static Linear Model

In this section we discuss the static linear model in a panel data setting. We start with two basic models, the fixed effects and the random effects model, and subsequently discuss the choice between the two, as well as alternative procedures that could be considered to be somewhere between a fixed effects and a random effects treatment.

10.2.1 The Fixed Effects Model

The fixed effects model is simply a linear regression model in which the intercept terms vary over the individual units $i$, i.e.

$$ y_{it} = \alpha_i + x_{it}' \beta + \epsilon_{it}, \quad \epsilon_{it} \sim \text{IID}(0, \sigma^2), \quad (10.6) $$

where it is usually assumed that all $x_{it}$ are independent of all $\epsilon_{it}$. We can write this in the usual regression framework by including a dummy variable for each unit $i$ in the model. That is,

$$ y_{it} = \sum_{j=1}^{N} \alpha_j d_{ij} + x_{it}' \beta + \epsilon_{it}, \quad (10.7) $$

where $d_{ij} = 1$ if $i = j$ and 0 elsewhere. We thus have a set of $N$ dummy variables in the model. The parameters $\alpha_1, \ldots, \alpha_N$ and $\beta$ can be estimated by ordinary least squares in (10.7). The implied estimator for $\beta$ is referred to as the least squares dummy variable (LSDV) estimator. It may, however, be numerically unattractive to have a regression model with so many regressors. Fortunately one can compute the estimator for $\beta$ in a simpler way. It can be shown that exactly the same estimator for $\beta$ is obtained if the regression is performed in deviations from individual means. Essentially, this implies that we eliminate the individual effects $\alpha_i$ first by transforming the data. To see this, first note that

$$ \tilde{y}_i = \alpha_i + \tilde{x}_i' \beta + \tilde{\epsilon}_i, \quad (10.8) $$
where \( \bar{y}_i = T^{-1} \sum_i y_{it} \) and similarly for the other variables. Consequently, we can write

\[
y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i). \tag{10.9}
\]

This is a regression model in deviations from individual means and does not include the individual effects \( \alpha_i \). The transformation that produces observations in deviation from individual means, as in (10.9), is called the within transformation. The OLS estimator \( \hat{\beta}_{FE} \) obtained from this transformed model is often called the within estimator or fixed effects estimator, and it is exactly identical to the LSDV estimator described above. It is given by

\[
\hat{\beta}_{FE} = \left( \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i) \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i). \tag{10.10}
\]

If it is assumed that all \( x_{it} \) are independent of all \( \varepsilon_{it} \) (compare assumption (A2) from Chapter 2), the fixed effects estimator can be shown to be unbiased for \( \beta \). If, in addition, normality of \( \varepsilon_{it} \) is imposed, \( \hat{\beta}_{FE} \) also has a normal distribution. For consistency, it is required that

\[
E[(x_{it} - \bar{x}_i)\varepsilon_{it}] = 0 \tag{10.11}
\]

(compare assumption (A7) in Chapters 2 and 5). Sufficient for this is that \( x_{it} \) is uncorrelated with \( \varepsilon_{it} \) and that \( \bar{x}_i \) has no correlation with the error term. These conditions are in turn implied by

\[
E[x_{it}\varepsilon_{it}] = 0 \quad \text{for all } s, t, \tag{10.12}
\]

in which case we call \( x_{it} \) strictly exogenous. A strictly exogenous variable is not allowed to depend upon current, future and past values of the error term. In some applications this may be restrictive. Clearly, it excludes the inclusion of lagged dependent variables in \( x_{it} \), but any \( x_{it} \) variable which depends upon the history of \( y_{it} \) would also violate the condition. For example, if we are explaining labour supply of an individual, we may want to include years of experience in the model, while quite clearly experience depends upon the person’s labour history.

With explanatory variables independent of all errors, the \( N \) intercepts are estimated unbiasedly as

\[
\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta}_{FE}, \quad i = 1, \ldots, N.
\]

Under assumption (10.11) these estimators are consistent for the fixed effects \( \alpha_i \) provided \( T \) goes to infinity. The reason why \( \hat{\alpha}_i \) is inconsistent for fixed \( T \) is clear: when \( T \) is fixed the individual averages \( \bar{y}_i \) and \( \bar{x}_i \) do not converge to anything if the number of individuals increases.
The covariance matrix for the fixed effects estimator $\hat{\beta}_{FE}$, assuming that $\varepsilon_{it}$ is i.i.d. across individuals and time with variance $\sigma^2_v$, is given by

$$
V(\hat{\beta}_{FE}) = \sigma^2_v \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1}.
$$

(10.13)

Unless $T$ is large, using the standard OLS estimate for the covariance matrix based upon the within regression in (10.9) will underestimate the true variance. The reason is that in this transformed regression the error covariance matrix is singular (as the $T$ transformed errors of each individual add up to zero) and the variance of $\varepsilon_{it} - \bar{\varepsilon}_i$ is $(T - 1)/T \sigma^2_v$ rather than $\sigma^2_v$. A consistent estimator for $\sigma^2_v$ is obtained as the within residual sum of squares divided by $N(T - 1)$. That is,

$$
\hat{\sigma}^2_v = \frac{1}{N(T - 1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i - \bar{x}_i \hat{\beta}_{FE})^2
$$

$$
= \frac{1}{N(T - 1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{\tilde{y}}_i - (x_{it} - \bar{x}_i)' \hat{\beta}_{FE})^2.
$$

(10.14)

It is possible to apply the usual degrees of freedom correction in which case $K$ is subtracted from the denominator. Note that using the standard OLS covariance matrix in model (10.7) with $N$ individual dummies is reliable, because the degrees of freedom correction involves $N$ additional unknown parameters corresponding to the individual intercept terms. Under weak regularity conditions, the fixed effects estimator is asymptotically normal, so that the usual inference procedures can be used (like $t$ and Wald tests).

Essentially, the fixed effects model concentrates on differences ‘within’ individuals. That is, it is explaining to what extent $y_{it}$ differs from $\bar{y}_i$ and does not explain why $\bar{y}_i$ is different from $\bar{y}_j$. The parametric assumptions about $\beta$ on the other hand, impose that a change in $x$ has the same (ceteris paribus) effect, whether it is a change from one period to the other or a change from one individual to the other. When interpreting the results, however, from a fixed effects regression, it may be important to realize that the parameters are identified only through the within dimension of the data.

### 10.2.2 The Random Effects Model

It is commonly assumed in regression analysis that all factors that affect the dependent variable, but that have not been included as regressors, can be appropriately summarized by a random error term. In our case, this leads to the assumption that the $\alpha_i$ are random factors, independently and identically distributed over individuals. Thus we write the random effects model as

$$
y_{it} = \mu + x_{it}' \beta + \alpha_i + \varepsilon_{it}, \quad \varepsilon_{it} \sim IID(0, \sigma^2_v); \quad \alpha_i \sim IID(0, \sigma^2_\alpha),
$$

(10.15)

where $\alpha_i + \varepsilon_{it}$ is treated as an error term consisting of two components: an individual specific component, which does not vary over time, and a remainder component, which
is assumed to be uncorrelated over time. That is, all correlation of the error terms over time is attributed to the individual effects $\alpha_i$. It is assumed that $\alpha_i$ and $\varepsilon_{it}$ are mutually independent and independent of $x_{it}$ (for all $j$ and $s$). This implies that the OLS estimator for $\mu$ and $\beta$ from (10.15) is unbiased and consistent. The error components structure implies that the composite error term $\alpha_i + \varepsilon_{it}$ exhibits a particular form of autocorrelation (unless $\sigma^2_\alpha = 0$). Consequently, routinely computed standard errors for the OLS estimator are incorrect and a more efficient (GLS) estimator can be obtained by exploiting the structure of the error covariance matrix.

To derive the GLS estimator, first note that for individual $i$ all error terms can be stacked as $\alpha_i t_T + \varepsilon_i$, where $t_T = (1, 1, \ldots, 1)'$ of dimension $T$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. The covariance matrix of this vector is (see Hsiao, 2003, Section 3.3)

$$V(\alpha_i t_T + \varepsilon_i) = \Omega = \sigma^2_{\alpha} t_T' t_T + \sigma^2_{\varepsilon} I_T,$$

where $I_T$ is the $T$-dimensional identity matrix. This can be used to derive the generalized least squares (GLS) estimator for the parameters in (10.15). For each individual, we can transform the data by premultiplying the vectors $y_i = (y_{i1}, \ldots, y_{iT})'$ etc. by $\Omega^{-1}$, which is given by

$$\Omega^{-1} = \sigma_{\varepsilon}^{-2} \left( I_T - \frac{\sigma^2_{\alpha}}{\sigma^2_{\varepsilon} + T \sigma^2_{\alpha}} t_T' t_T \right),$$

which can also be written as

$$\Omega^{-1} = \sigma_{\varepsilon}^{-2} \left( I_T - \frac{1}{T} t_T' t_T \right) + \psi \frac{1}{T} t_T' t_T,$$

where

$$\psi = \frac{\sigma^2_{\alpha}}{\sigma^2_{\varepsilon} + T \sigma^2_{\alpha}}.$$ 

Noting that $I_T - (1/T) t_T' t_T$ transforms the data in deviations from individual means and $(1/T) t_T' t_T$ takes individual means, the GLS estimator for $\beta$ can be written as

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + \psi T \sum_{i=1}^N (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right)^{-1}$$

$$\times \left( \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + \psi T \sum_{i=1}^N (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}) \right),$$

where $\bar{x} = (1/(NT)) \sum_{i,t} x_{it}$ denotes the overall average of $x_{it}$. It is easy to see that for $\psi = 0$ the fixed effects estimator arises. Because $\psi \to 0$ if $T \to \infty$, it follows that the fixed and random effects estimators are equivalent for large $T$. If $\psi = 1$, the GLS
estimator is just the OLS estimator (and $\Omega$ is diagonal). From the general formula for the GLS estimator it can be derived that

$$\hat{\beta}_{GLS} = \Delta \hat{\beta}_B + (I_k - \Delta) \hat{\beta}_{FE},$$

where

$$\hat{\beta}_B = \left( \sum_{i=1}^N (\tilde{x}_i - \bar{x})(\hat{\tilde{y}} - \bar{y})' \right)^{-1} \sum_{i=1}^N (\tilde{x}_i - \bar{x})(\hat{\tilde{y}} - \bar{y})$$

is the so-called between estimator for $\beta$. It is the OLS estimator in the model for individual means

$$\tilde{y}_i = \mu + \tilde{x}_i' \beta + \alpha_i + \tilde{\epsilon}_i, \quad i = 1, \ldots, N.$$  \hfill (10.18)

The matrix $\Delta$ is a weighting matrix and is proportional to the inverse of the covariance matrix of $\hat{\beta}_B$ (see Hsiang, 2003, Section 3.4, for details). That is, the GLS estimator is a matrix-weighted average of the between estimator and the within estimator, where the weight depends upon the relative variances of the two estimators. (The more accurate one gets the higher the weight.)

The between estimator effectively discards the time series information in the data set. The GLS estimator, under the current assumptions, is the optimal combination of the within estimator and the between estimator, and is therefore more efficient than either of these two estimators. The OLS estimator (with $\psi = 1$) is also a linear combination of the two estimators, but not the efficient one. Thus, GLS will be more efficient than OLS, as usual. If the explanatory variables are independent of all $\varepsilon_{it}$ and all $\alpha_i$, the GLS estimator is unbiased. It is a consistent estimator for $N$ or $T$ or both tending to infinity if in addition to (10.11) it also holds that $E(\tilde{x}_i \varepsilon_{it}) = 0$ and most importantly that

$$E(\tilde{x}_i \alpha_i) = 0.$$  \hfill (10.19)

Note that these conditions are also required for the between estimator to be consistent.

An easy way to compute the GLS estimator is obtained by noting that it can be determined as the OLS estimator in a transformed model (compare Chapter 4), given by

$$(y_{it} - \vartheta \tilde{y}_i) = \mu (1 - \vartheta) + (x_{it} - \bar{x}_i)' \beta + u_{it},$$  \hfill (10.20)

where $\vartheta = 1 - \psi^{1/2}$. The error term in this transformed regression is i.i.d. over individuals and time. Note again that $\psi = 0$ corresponds to the within estimator ($\vartheta = 1$). In general, a fixed proportion $\vartheta$ of the individual means is subtracted from the data to obtain this transformed model ($0 \leq \vartheta \leq 1$).

Of course, the variance components $\sigma^2_e$ and $\sigma^2_e$ are unknown in practice. In this case we can use the feasible GLS estimator (EGLS), where the unknown variances are consistently estimated in a first step. The estimator for $\sigma^2_e$ is easily obtained from the within residuals, as given in (10.14). For the between regression the error variance is $\sigma^2_e + (1/T)\sigma^2_e$, which we can estimate consistently by

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{1}{N} \sum_{i=1}^N (\tilde{y}_i - \bar{y}_i) (\tilde{y}_i - \bar{y}_i)' (\hat{\beta}_B)' (\hat{\beta}_B),$$  \hfill (10.21)
where \( \hat{\mu}_B \) is the between estimator for \( \mu \). From this, a consistent estimator for \( \sigma^2 \) follows as
\[
\hat{\sigma}^2 = \hat{\sigma}^2_B - \frac{1}{T} \hat{\sigma}^2_e.
\] (10.22)

Again, it is possible to adjust this estimator by applying a degrees of freedom correction, implying that the number of regressors \( K+1 \) is subtracted in the denominator of (10.21) (see Hsiao, 2003, Section 3.3). The resulting EGLS estimator is referred to as the random effects estimator for \( \beta \) (and \( \mu \)), denoted below as \( \hat{\beta}_{RE} \). It is also known as the Balestra–Nerlove estimator.

Under weak regularity conditions, the random effects estimator is asymptotically normal. Its covariance matrix is given by
\[
V(\hat{\beta}_{RE}) = \sigma^2_e \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + \psi T \sum_{i=1}^{N} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right)^{-1},
\] (10.23)

which shows that the random effects estimator is more efficient than the fixed effects estimator as long as \( \psi > 0 \). The gain in efficiency is due to the use of the between variation in the data \( (\bar{x}_i - \bar{x}) \). The covariance matrix in (10.23) is routinely estimated by the OLS expressions in the transformed model (10.20).

In summary, we have seen a range of estimators for the parameter vector \( \beta \). The basic two are:

1. The between estimator, exploiting the between dimension of the data (differences between individuals), determined as the OLS estimator in a regression of individual averages of \( y \) on individual averages of \( x \) (and a constant). Consistency, for \( N \to \infty \), requires that \( E(\bar{x}_i \alpha_i) = 0 \) and \( E(\bar{x}_i \bar{e_i}) = 0 \). Typically this means that the explanatory variables are strictly exogenous and uncorrelated with the individual specific effect \( \alpha_i \).

2. The fixed effects (within) estimator, exploiting the within dimension of the data (differences within individuals), determined as the OLS estimator in a regression in deviations from individual means. It is consistent for \( \beta \) for \( T \to \infty \) or \( N \to \infty \), provided that \( E((x_{it} - \bar{x}_i)e_{it}) = 0 \). Again this requires the \( x \)-variables to be strictly exogenous, but it does not impose any restrictions upon the relationship between \( \alpha_i \) and \( x_{it} \).

The other two estimators are:

3. The OLS estimator, exploiting both dimensions (within and between) but not efficiently. Determined (of course) as OLS in the original model given in (10.15). Consistency for \( T \to \infty \) or \( N \to \infty \) requires that \( E(x_{it} e_{it}) = 0 \). This requires the explanatory variables to be uncorrelated with \( \alpha_i \) but does not impose that they are strictly exogenous. It suffices that \( x_{it} \) and \( e_{it} \) are contemporaneously uncorrelated.

4. The random effects (EGLS) estimator, combining the information from the between and within dimensions in an efficient way. It is consistent for \( T \to \infty \) or \( N \to \infty \) under the combined conditions of 1 and 2. It can be determined as a
weighted average of the between and within estimator or as the OLS estimator in a regression where the variables are transformed as \( \hat{y}_i - \hat{\theta} \tilde{y}_i \), where \( \hat{\theta} \) is an estimate for \( \theta = 1 - \psi^{1/2} \) with \( \psi = \sigma_e^2 / (\sigma_e^2 + T \sigma_\alpha^2) \).

**Fixed Effects or Random Effects?**

**HAUSMAN’S SPECIFICATION TEST FOR THE RANDOM EFFECTS MODEL**

At various points, we have made the distinction between fixed and random effects models. An inevitable question is, Which should be used? From a purely practical standpoint, the dummy variable approach is costly in terms of degrees of freedom lost. On the other hand, the fixed effects approach has one considerable virtue. There is little justification for treating the individual effects as uncorrelated with the other regressors, as is assumed in the random effects model. The random effects treatment, therefore, may suffer from the inconsistency due to this correlation between the included variables and the random effect.22

The specification test devised by Hausman (1978)23 is used to test for orthogonality of the random effects and the regressors. The test is based on the idea that under the hypothesis of no correlation, both OLS in the LSDV model and GLS are consistent, but OLS is inefficient,24 whereas under the alternative, OLS is consistent, but GLS is not. Therefore, under the null hypothesis, the two estimates should not differ systematically, and a test can be based on the difference. The other essential ingredient for the test is the covariance matrix of the difference vector, \([\mathbf{b} - \hat{\mathbf{b}}]\):

\[
\text{Var}[\mathbf{b} - \hat{\mathbf{b}}] = \text{Var}[\mathbf{b}] + \text{Var}[\hat{\mathbf{b}}] - \text{Cov}[\mathbf{b}, \hat{\mathbf{b}}] - \text{Cov}[\mathbf{b}, \hat{\mathbf{b}}].
\]  

Hausman’s essential result is that the covariance of an efficient estimator with its difference from an inefficient estimator is zero, which implies that

\[
\text{Cov}([\mathbf{b} - \hat{\mathbf{b}}], \hat{\mathbf{b}}) = \text{Cov}[\mathbf{b}, \hat{\mathbf{b}}] - \text{Var}[\hat{\mathbf{b}}] = 0
\]
or that

\[
\text{Cov}[\mathbf{b}, \hat{\mathbf{b}}] = \text{Var}[\hat{\mathbf{b}}].
\]

Inserting this result in (13.32) produces the required covariance matrix for the test,

\[
\text{Var}[\mathbf{b} - \hat{\mathbf{b}}] = \text{Var}[\mathbf{b}] - \text{Var}[\hat{\mathbf{b}}] = \Psi.
\]  

The chi-squared test is based on the Wald criterion:

\[
W = \chi^2[K - 1] = [\mathbf{b} - \hat{\mathbf{b}}]^T \Psi^{-1} [\mathbf{b} - \hat{\mathbf{b}}].
\]  

For \( \Psi \), we use the estimated covariance matrices of the slope estimator in the LSDV model and the estimated covariance matrix in the random effects model, excluding the constant term. Under the null hypothesis, \( W \) has a limiting chi-squared distribution with \( K - 1 \) degrees of freedom.
GMM ESTIMATION OF DYNAMIC PANEL DATA MODELS

Panel data are well suited for examining dynamic effects, as in the first-order model,

\[ y_{it} = x_{it}' \beta + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it} = w_{it}' \delta + \alpha_i + \varepsilon_{it}, \]

where the set of right hand side variables, \( w_{it} \), now includes the lagged dependent variable, \( y_{i,t-1} \). Adding dynamics to a model in this fashion is a major change in the interpretation of the equation. Without the lagged variable, the “independent variables” represent the full set of information that produce observed outcome \( y_{it} \). With the lagged variable, we now have in the equation, the entire history of the right hand side variables, so that any measured influence is conditioned on this history; in this case, any impact of \( x_{it} \) represents the effect of new information. Substantial complications arise in estimation of such a model. In both the fixed and random effects settings, the difficulty is that the lagged dependent variable is correlated with the disturbance, even if it is assumed that \( \varepsilon_{it} \) is not itself autocorrelated. For the moment, consider the fixed effects model as an ordinary regression with a lagged dependent variable. We considered this case in Section 5.3.2 as a regression with a stochastic regressor that is dependent across observations. In that dynamic regression model, the estimator based on \( T \) observations is biased in finite samples, but it is consistent in \( T \). That conclusion was the main result of Section 5.3.2. The finite sample bias is of order \( 1/T \). The same result applies here, but the difference is that whereas before we obtained our large sample results by allowing \( T \) to grow large, in this setting, \( T \) is assumed to be small and fixed, and large-sample results are obtained with respect to \( n \) growing large, not \( T \). The fixed effects estimator of \( \delta = [\beta, \gamma] \) can be viewed as an average of \( n \) such estimators. Assume for now that \( T \geq K + 1 \) where \( K \) is the number of variables in \( x_{it} \). Then, from (13-4),

\[ \hat{\delta} = \left[ \sum_{i=1}^{n} W_i' M^{-1} W_i \right]^{-1} \left[ \sum_{i=1}^{n} W_i' M^{-1} y_i \right] = \left[ \sum_{i=1}^{n} W_i' M^{-1} W_i \right]^{-1} \left[ \sum_{i=1}^{n} W_i' M^{-1} W_i \right] d_i \]

where the rows of the \( T \times (K + 1) \) matrix \( W_i \) are \( w_{it} \) and \( M^{-1} \) is the \( T \times T \) matrix that creates deviations from group means [see (13-5)]. Each group specific estimator, \( d_i \) is inconsistent, as it is biased in finite samples and its variance does not go to zero as \( n \) increases. This matrix weighted average of \( n \) inconsistent estimators will also be inconsistent. (This analysis is only heuristic. If \( T < K + 1 \), then the individual coefficient vectors cannot be computed.26)
The problem is more transparent in the random effects model. In the model
\[ y_{it} = \gamma y_{i,t-1} + x_{it}' \beta + u_i + \varepsilon_{it}, \]
the lagged dependent variable is correlated with the compound disturbance in the model, since the same \( u_i \) enters the equation for every observation in group \( i \).

Neither of these results renders the model inestimable, but they do make necessary some technique other than our familiar LSDV or FGLS estimators. The general approach, which has been developed in several stages in the literature,\textsuperscript{27} relies on instrumental variables estimators and, most recently [by Arellano and Bond (1991) and Arellano and Bover (1995)] on a GMM estimator. For example, in either the fixed or random effects cases, the heterogeneity can be swept from the model by taking first differences, which produces
\[ y_{it} - y_{i,t-1} = \delta (y_{i,t-1} - y_{i,t-2}) + (x_{it} - x_{i,t-1})' \beta + (\varepsilon_{it} - \varepsilon_{i,t-1}). \]

This model is still complicated by correlation between the lagged dependent variable and the disturbance (and by its first-order moving average disturbance). But without the group effects, there is a simple instrumental variables estimator available. Assuming that the time series is long enough, one could use the lagged differences, \((y_{i,t-2} - y_{i,t-3})\), or the lagged levels, \(y_{i,t-2}\) and \(y_{i,t-3}\), as one or two instrumental variables for \((y_{i,t-1} - y_{i,t-2})\). (The other variables can serve as their own instruments.) By this construction, then, the treatment of this model is a standard application of the instrumental variables technique that we developed in Section 5.4.\textsuperscript{28} This illustrates the flavor of an instrumental variable approach to estimation. But, as Arellano et al. and Ahn and Schmidt (1995) have shown, there is still more information in the sample which can be brought to bear on estimation, in the context of a GMM estimator, which we now consider.

We extend the Hausman and Taylor (HT) formulation of the random effects model to include the lagged dependent variable;
\[ y_{it} = \delta (y_{i,t-1} - y_{i,t-2}) + (x_{it} - x_{i,t-1})' \beta + (\varepsilon_{it} - \varepsilon_{i,t-1}). \]

This model is still complicated by correlation between the lagged dependent variable and the disturbance (and by its first-order moving average disturbance). But without the group effects, there is a simple instrumental variables estimator available. Assuming that the time series is long enough, one could use the lagged differences, \((y_{i,t-2} - y_{i,t-3})\), or the lagged levels, \(y_{i,t-2}\) and \(y_{i,t-3}\), as one or two instrumental variables for \((y_{i,t-1} - y_{i,t-2})\). (The other variables can serve as their own instruments.) By this construction, then, the treatment of this model is a standard application of the instrumental variables technique that we developed in Section 5.4.\textsuperscript{28} This illustrates the flavor of an instrumental variable approach to estimation. But, as Arellano et al. and Ahn and Schmidt (1995) have shown, there is still more information in the sample which can be brought to bear on estimation, in the context of a GMM estimator, which we now consider.

We extend the Hausman and Taylor (HT) formulation of the random effects model to include the lagged dependent variable:
\[ y_{it} = \gamma y_{i,t-1} + x_{it}' \beta_1 + x_{i2t}' \beta_2 + \varepsilon_{i1}' \alpha_1 + \varepsilon_{i2}' \alpha_2 + \varepsilon_{it} + u_t \]
\[ = \delta' w_{it} + \varepsilon_{it} + u_t \]
\[ = \delta' w_{it} + \eta_{it} \]

where

\[ w_{it} = [y_{i,t-1}, x'_{1it}, x'_{2it}, z'_{1i}, z'_{2i}] \]

is now a \((1 + K_1 + K_2 + L_1 + L_2) \times 1\) vector. The terms in the equation are the same as in the Hausman and Taylor model. Instrumental variables estimation of the model without the lagged dependent variable is discussed in the previous section on the HT estimator. Moreover, by just including \(y_{i,t-1}\) in \(x_{2it}\), we see that the HT approach extends to this setting as well, essentially without modification. Arellano et al. suggest a GMM estimator, and show that efficiency gains are available by using a larger set of moment conditions. In the previous treatment, we used a GMM estimator constructed as follows: The set of moment conditions we used to formulate the instrumental variables were

\[ E \left[ \begin{pmatrix} x'_{1it} \\ x'_{2it} \\ z'_{1i} \\ z'_{2i} \end{pmatrix} (\eta_{it} - \tilde{\eta}_{it}) \right] = E \left[ \begin{pmatrix} x'_{1it} \\ x'_{2it} \\ z'_{1i} \\ z'_{2i} \end{pmatrix} (\epsilon_{it} - \tilde{\epsilon}_{it}) \right] = 0. \]

This moment condition is used to produce the instrumental variable estimator. We could ignore the nonscalar variance of \(\eta_{it}\) and use simple instrumental variables at this point. However, by accounting for the random effects formulation and using the counterpart to feasible GLS, we obtain the more efficient estimator in (13-37). As usual, this can be done in two steps. The inefficient estimator is computed in order to obtain the residuals needed to estimate the variance components. This is Hausman and Taylor’s steps 1 and 2. Steps 3 and 4 are the GMM estimator based on these estimated variance components.

Arellano et al. suggest that the preceding does not exploit all the information in the sample. In simple terms, within the \(T\) observations in group \(i\), we have not used the fact that

\[ E \left[ \begin{pmatrix} x'_{1it} \\ x'_{2it} \\ z'_{1i} \\ z'_{2i} \end{pmatrix} (\eta_{is} - \tilde{\eta}_{is}) \right] = 0 \text{ for some } s \neq t. \]

Thus, for example, not only are disturbances at time \(t\) uncorrelated with these variables at time \(t\), arguably, they are uncorrelated with the same variables at time \(t - 1, t - 2, \) possibly \(t + 1\), and so on. In principle, the number of valid instruments is potentially enormous. Suppose, for example, that the set of instruments listed above is strictly exogenous with respect to \(\eta_{it}\) in every period including current, lagged and future. Then, there are a total of \([T(K_1 + K_2) + L_1 + L_2]\) moment conditions for every observation on this basis alone. Consider, for example, a panel with two periods. We would have for the two periods,
\[ E \begin{bmatrix} x_{1i1} \\ x_{2i1} \\ x_{1i2} \\ x_{2i2} \\ z_{i1} \\ z_{i2} \\ \tilde{x}_{i1} \\ \tilde{x}_{i2} \end{bmatrix} (\eta_{i1} - \tilde{\eta}_{i}) = E \begin{bmatrix} x_{1i1} \\ x_{2i1} \\ x_{1i2} \\ x_{2i2} \\ z_{i1} \\ z_{i2} \\ \tilde{x}_{i1} \\ \tilde{x}_{i2} \end{bmatrix} (\eta_{i2} - \tilde{\eta}_{i}) = 0. \quad (13-38) \]

How much useful information is brought to bear on estimation of the parameters is uncertain, as it depends on the correlation of the instruments with the included exogenous variables in the equation. The farther apart in time these sets of variables become the less information is likely to be present. (The literature on this subject contains references to “strong” versus “weak” instrumental variables.) In order to proceed, as noted, we can include the lagged dependent variable in \( x_{2i} \). This set of instrumental variables can be used to construct the estimator, actually whether the lagged variable is present or not. We note, at this point, that on this basis, Hausman and Taylor’s estimator did not actually use all the information available in the sample. We now have the elements of the Arellano et al. estimator in hand; what remains is essentially the (unfortunately, fairly involved) algebra, which we now develop.

Let

\[ W_i = \begin{bmatrix} w_{i1}' \\ w_{i2}' \\ \vdots \\ w_{1i}' \end{bmatrix} \]  

the full set of rhs data for group \( i \), and \( y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} \).

Note that \( W_i \) is assumed to be, a \( T \times (1 + K_1 + K_2 + L_1 + L_2) \) matrix. Since there is a lagged dependent variable in the model, it must be assumed that there are actually \( T + 1 \) observations available on \( y_{it} \). To avoid a cumbersome, cluttered notation, we will leave this distinction embedded in the notation for the moment. Later, when necessary, we will make it explicit. It will reappear in the formulation of the instrumental variables. A total of \( T \) observations will be available for constructing the IV estimators. We now form a matrix of instrumental variables. Different approaches to this have been considered by Hausman and Taylor (1981), Arellano et al. (1991, 1995, 1999), Ahn and Schmidt (1995) and Amemiya and MaCurdy (1986), among others. We will form a matrix \( V_i \) consisting of \( T_i - 1 \) rows constructed the same way for \( T_i - 1 \) observations and a final row that will be different, as discussed below. [This is to exploit a useful algebraic result discussed by Arellano and Bover (1995).] The matrix will be of the form

\[ V_i = \begin{bmatrix} v_{1i}' & 0' & \cdots & 0' \\ 0' & v_{2i}' & \cdots & 0' \\ \vdots & \vdots & \ddots & \vdots \\ 0' & 0' & \cdots & a_i' \end{bmatrix}. \quad (13-39) \]

The instrumental variable sets contained in \( V_i' \) which have been suggested might include the following from within the model:
\( x_t \) and \( x_{t-1} \) (i.e., current and one lag of all the time varying variables)
\( x_{1t}, \ldots, x_{Tt} \) (i.e., all current, past and future values of all the time varying variables)
\( x_{1t}, \ldots, x_{rt} \) (i.e., all current and past values of all the time varying variables)

The time invariant variables that are uncorrelated with \( u_t \), that is \( z_{it} \), 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—that is, this set of moment conditions will ultimately be converted to what appears in (13-38). While the variables are correlated with \( u_t \) by construction, they are not correlated with \( e_{it} - \bar{e}_t \). The final row of \( V_t \) is important to the construction. Two possibilities have been suggested:

\[
\begin{align*}
\mathbf{a}_t' &= [ z_{ii} \ x_{1t} ] \text{ (produces the Hausman and Taylor estimator) } \\
\mathbf{a}_t' &= [ z_{ii} \ x_{1t}, x_{1t2}, \ldots, x_{HT} ] \text{ (produces Amemiya and MaCurdy's estimator). }
\end{align*}
\]

Note that the \( m \) variables are exogenous time invariant variables, \( z_{it} \) 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.

To construct the estimator, we will require a transformation matrix, \( \mathbf{H} \) constructed as follows. Let \( \mathbf{M}^{01} \) denote the first \( T - 1 \) rows of \( \mathbf{M}^{0} \), the matrix that creates deviations from group means. Then,

\[
\mathbf{H} = \begin{bmatrix} \mathbf{M}^{01} \\
1/T' 
\end{bmatrix}.
\]

Thus, \( \mathbf{H} \) replaces the last row of \( \mathbf{M}^{0} \) with a row of \( 1/T \). The effect is as follows: if \( q \) is \( T \) observations on a variable, then \( \mathbf{H}q \) produces \( q' \) in which the first \( T - 1 \) observations are converted to deviations from group means and the last observation is the group mean. In particular, let the \( T \times 1 \) column vector of disturbances

\[
\eta_t = [ \eta_{1t}, \eta_{2t}, \ldots, \eta_{Tt} ] = [ (e_{1t} + u_t), (e_{2t} + u_t), \ldots, (e_{Tt} + u_t) ],
\]

then

\[
\mathbf{H}\eta = \begin{bmatrix}
\eta_{1t} - \bar{\eta}_t \\
\vdots \\
\eta_{Tt-1} - \bar{\eta}_t \\
\eta_{Tt} - \bar{\eta}_t
\end{bmatrix}.
\]

We can now construct the moment conditions. With all this machinery in place, we have the result that appears in (13-40), that is

\[
E[V_t' \mathbf{H} \eta_t] = E[\mathbf{g}_t] = 0.
\]
It is useful to expand this for a particular case. Suppose $T = 3$ and we use as instruments the current values in Period 1, and the current and previous values in Period 2 and the Hausman and Taylor form for the invariant variables. Then the preceding is

$$
E \begin{bmatrix}
    x_{1t} & 0 & 0 \\
    x_{2t} & 0 & 0 \\
    z_{t+1} & 0 & 0 \\
    0 & x_{1t+1} & 0 \\
    0 & x_{2t+1} & 0 \\
    0 & x_{3t+2} & 0 \\
    0 & z_{t+1} & 0 \\
    0 & 0 & z_{t+1}
\end{bmatrix}
\begin{bmatrix}
    \eta_{t+1} - \tilde{\eta}_t \\
    \eta_{t+2} - \tilde{\eta}_t \\
    \eta_{t+3} - \tilde{\eta}_t \\
    \tilde{\eta}_{t+1} - \tilde{\eta}_t \\
    \tilde{\eta}_{t+2} - \tilde{\eta}_t \\
    \tilde{\eta}_{t+3} - \tilde{\eta}_t
\end{bmatrix} = 0.
$$

(13-40)

This is the same as (13-38). The empirical moment condition that follows from this is

$$
p\lim \frac{1}{n} \sum_{t=1}^{n} V_t' H \eta_t = p\lim \frac{1}{n} \sum_{t=1}^{n} V_t' H \left( \begin{array}{c}
    y_{t+1} - y_{t+1}' \theta_0 - x_{1t+1}' \beta_1 - x_{2t+1}' \beta_2 - z_{t+1}' \alpha_1 - z_{t+2}' \alpha_2 \\
    y_{t+2} - y_{t+2}' \theta_0 - x_{1t+2}' \beta_1 - x_{2t+2}' \beta_2 - z_{t+1}' \alpha_1 - z_{t+2}' \alpha_2 \\
    \vdots \\
    y_{t+3} - y_{t+3}' \theta_0 - x_{1t+3}' \beta_1 - x_{2t+3}' \beta_2 - z_{t+1}' \alpha_1 - z_{t+2}' \alpha_2
\end{array} \right) = 0.
$$

Write this as

$$
p\lim \frac{1}{n} \sum_{t=1}^{n} m_t = p\lim \hat{m} = 0.
$$

The GMM estimator $\hat{\delta}$ is then obtained by minimizing

$$
q = \hat{m}' A \hat{m}
$$

with an appropriate choice of the weighting matrix, $A$. The optimal weighting matrix will be the inverse of the asymptotic covariance matrix of $\sqrt{n} \hat{m}$. With a consistent estimator of $\delta$ in hand, this can be estimated empirically using

$$
\text{Est. Asy. Var}[\sqrt{n} \hat{m}] = \frac{1}{n} \sum_{t=1}^{n} \hat{m}_t \hat{m}_t' = \frac{1}{n} \sum_{t=1}^{n} V_t' H \hat{\eta}_t \hat{\eta}_t' H' V_t.
$$

This is a robust estimator that allows an unrestricted $T \times T$ covariance matrix for the $T$ disturbances, $\epsilon_t + \mu_t$. But, we have assumed that this covariance matrix is the $\Sigma$ defined in (13-20) for the random effects model. To use this information we would, instead, use the residuals in

$$
\hat{\eta}_t = y_t - W_t \hat{\delta}
$$

to estimate $\sigma_\epsilon^2$ and $\sigma_\mu^2$ and then $\Sigma$, which produces

$$
\text{Est. Asy. Var}[\sqrt{n} \hat{m}] = \frac{1}{n} \sum_{t=1}^{n} V_t' \hat{\Sigma} H' V_t.
$$
We now have the full set of results needed to compute the GMM estimator. The solution to the optimization problem of minimizing \( q \) with respect to the parameter vector \( \delta \) is

\[
\hat{\delta}_{\text{GMM}} = \left[ \left( \sum_{i=1}^{n} W_i' H V_i \right) \left( \sum_{i=1}^{n} V_i' H' \hat{\Sigma} H V_i \right)^{-1} \left( \sum_{i=1}^{n} V_i' H' W_i \right) \right]^{-1} \times \left( \sum_{i=1}^{n} W_i' H V_i \right) \left( \sum_{i=1}^{n} V_i' H' \hat{\Sigma} H V_i \right)^{-1} \left( \sum_{i=1}^{n} V_i' H' \hat{\Sigma} H V_i \right) .
\]

(13-41)

The estimator of the asymptotic covariance matrix for \( \hat{\delta} \) is the inverse matrix in brackets.

The remaining loose end is how to obtain the consistent estimator of \( \delta \) to compute \( \Sigma \). Recall that the GMM estimator is consistent with any positive definite weighting matrix, \( A \) in our expression above. Therefore, for an initial estimator, we could set \( A = I \) and use the simple instrumental variables estimator,

\[
\hat{\delta}_{\text{IV}} = \left[ \left( \sum_{i=1}^{N} W_i' H V_i \right) \left( \sum_{i=1}^{N} V_i' H W_i \right) \right]^{-1} \left( \sum_{i=1}^{N} W_i' H V_i \right) \left( \sum_{i=1}^{N} V_i' H W_i \right) .
\]

It is more common to proceed directly to the “two stage least squares” estimator (see Chapter 15) which uses

\[
A = \left( \frac{1}{n} \sum_{i=1}^{N} V_i' H' H V_i \right)^{-1} .
\]

The estimator is, then, the one given earlier in (13-41) with \( \hat{\Sigma} \) replace by \( \hat{\Sigma} \). Either estimator is a function of the sample data only and provides the initial estimator we need.

Ahn and Schmidt (among others) observed that the IV estimator proposed here, as extensive as it is, still neglects quite a lot of information and is therefore (relatively) inefficient. For example, in the first differenced model,

\[
E[y_{is}(\varepsilon_{it} - \varepsilon_{it-1})] = 0, \quad s = 0, \ldots, t - 2, \quad t = 2, \ldots, T.
\]

That is, the level of \( y_{is} \) is uncorrelated with the differences of disturbances that are at least two periods subsequent.\(^3\) (The differencing transformation, as the transformation to deviations from group means, removes the individual effect.) The corresponding moment equations that can enter the construction of a GMM estimator are

\[
\frac{1}{n} \sum_{i=1}^{n} y_{is} [(\delta y_{is} - \delta y_{is-1}) - \hat{\delta} (\hat{y}_{is} - \hat{y}_{is-2}) - (\hat{s}_{it} - \hat{s}_{it-1})] = 0
\]

\[
s = 0, \ldots, t - 2, \quad t = 2, \ldots, T.
\]

Altogether, Ahn and Schmidt identify \( T(T - 1)/2 + T - 2 \) such equations that involve mixtures of the levels and differences of the variables. The main conclusion that they demonstrate is that in the dynamic model, there is a large amount of information to be gleaned not only from the familiar relationships among the levels of the variables but also from the implied relationships between the levels and the first differences. The issue of correlation between the transformed \( y_{it} \) and the deviations of \( \varepsilon_{it} \) is discussed in the papers cited. (As Ahn and Schmidt show, there are potentially huge numbers of additional orthogonality conditions in this model owing to the relationship between
Arellano-Bond Model

Present literature in Economics identifies three major sources: labors, capital, and technological progress as determinants of the level of GDP per capita (GDPPC) in any economy. In order to establish link between the financial markets development and the real economy, let us assume the economies featured by an aggregate production function; Where output $y_{it}$, is produced during period $t$ by country $i$ can be defined as:

$$y_{it} = f( k_{it}, X_{it}, FMD_{it} )$$  

Where $k_{it}$ is the capital per unit of labor of country $i$ in period $t$; $x_{it}$ represent control variables, and $FMD_{it}$ is the level of financial market development of country $i$ in period $t$.

The $FMD_{it}$ represents banking sector development indicators ($BSDI_{it}$) and stock market development indicators ($SMDI_{it}$). Because of the dependency of GDP per capita on its previous values, Dynamic Panel Data Model (DMPD) is used for the analysis.

$$Y_{it} = \alpha_i + \gamma_{i,t-1} + \beta X_{i,t} + \varnothing FMD_{i,t} + \epsilon_{i,t}$$  

Where $\alpha_i$ are fixed effects, $x_{it}$ is a vector of control variables, and $\epsilon_{it}$ is the random disturbance. A problem with estimating equation (2.4) is that time-invariant country characteristics $\alpha_i$ may be correlated with the explanatory variables. Moreover, $FMD_{it}$ might also be endogenous because causality may run in both directions; from the financial markets development to economic growth and vice versa. Another problem of the autocorrelation can arise because of the presence of the lagged dependent variable $Y_{it}$.

Estimating a fixed effect model by Least Square Dummy Variable (LSDV) in the presence of lagged dependent variable generate biased estimate of the coefficients. Several estimators have been proposed to estimate equation (2.4) without bias. To cope with the problems of endogeneity between the financial markets development indicators and economic growth study employ Arellano and Bond (1991) difference GMM estimator first proposed by Holmstrom and Tirole (1993). Arellano-Bond model along with exogenous instruments, lagged levels of the endogenous regressors, and the first differenced lagged dependent variable are also added as instruments. Lagged levels of endogenous regressors
make the endogenous variables pre-determined and therefore, not correlated with the error term. The first differenced lagged dependent variable as instrument help to remove the autocorrelation. Another incentive of using Arellano-Bond model is that estimator is designed for small-T large-N panels.