Panel Data Analysis

Mei-Yuan Chen

Department of Finance
National Chung Hsing University

April, 22, 2003
1 Introduction

Among panel data analyzes, dynamic panel data and nonstationary panel time series models have received tremendous attention recently. Using cross-country time series, purchasing power parity, growth convergence and international R&D spillovers have been studied empirically. Instead of the usual asymptotics of micro panels with large $N$ (number of countries) and small $T$ (length of time series) having been studied, discussions of the asymptotics of macro panels with large $N$ and large $T$. The limiting distributions of double indexed integrated processes have been developed by Phillips and Moon (1999a). As the univariate unit root tests suffer from the drawback of low power due to the short length of time series, panel unit root tests have been proposed to increase the testing power by data increase via adding the cross-section dimension. The addition of the cross-section dimension can be viewed as repeated sampling from the same distribution, under certain assumptions.

In spite of power increment, unit root tests with panel data are still criticized. Madala, Wu, and Liu (2000) argue that panel unit root tests do not rescue purchasing power parity (PPP). Besides, the panel unit root tests are the wrong answer to the low power of univariate unit root tests since the null hypothesis of a single unit root is different from the one of a panel unit root for the PPP hypothesis. Madala (1999) also argued that panel unit root test did not help settle the question of growth convergence among countries. A survey on nonstationary panels, cointegration in panels and dynamic panels has been provided by Phillips and Moon (1999b), Banerjee (1999), and Baltagi and Kao (2000).

2 Panel Data Models

Consider the linear regression model for panel data:

$$ y_{it} = \beta_0 + x_{it}' \beta + \epsilon_{it}, $$

(i = 1, \ldots, N (size of the cross-section), and t = 1, \ldots, T (number of time periods). The usefulness of panel data is more than its large sample size. The very nature of repeated observations on the same individual offers the opportunity to infer behavioral changes of an individual over time while controlling the differences across individuals. To see this, consider
the regression equation for the estimation of income ($m_i$) elasticity of health care ($c_i$):

$$\ln c_i = \beta_0 + \beta_1 \ln m_i + \epsilon_i, \ i = 1, \ldots, N.$$  

Obviously, the above regression equation is too simple since there are many factors other than income that can influence health care consumption decision, such as age, gender, marital status, education, etc. Moreover, number of individual characteristics, such as current health condition, attitude toward exercises, etc., are usually not covered by cross sectional survey data but can greatly affect the health care consumption decision. As a result, no matter how we expand the above regression model, there may always be some important variables left out from the estimation.

Now, let’s consider the linear regression model for panel data:

$$\ln c_{it} = \beta_0 + \beta_1 \ln m_{it} + \epsilon_{it}.$$  

It is not hard to see that while health care consumption $c_{it}$ and income $m_{it}$ change from person to person (over $i$) as well as from time to time (over $t$), most of those personal characteristics, either observed or unobserved, differ among individuals but remain the same over time; that is, they are time-invariant. Hence, it seems possible to control all those time-invariant inter-personal (i.e., cross-sectional) differences by employing the information only from intertemporal changes in each individuals behavior. In other words, by comparing individuals’ behavior changes over time only, we might be able to eliminate the effects of all those observed and unobserved inter-personal differences when estimating the income elasticity of health care. This possibility offered by the panel structure is the main advantage of the panel data. Certainly, in order to improve the estimation efficiency, we may also consider some information from the inter-personal behavioral differences and include it in a systematic but controlled manner.

2.1 Individual Effects

Given a panel data set with $N$ cross-section units and $T$ observations, the linear specification allowing for individual effects is

$$y_{it} = x_{it}'\beta_i + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$  


where \( x_{it} \) is \( k \times 1 \) and \( \beta_i \) is the parameter vector depending only on \( i \) but not on \( t \). In this specification, individual effects are characterized by \( \beta_i \), and there is no time-specific effect. This may be reasonable when a short time series is observed for each individual units. This specification can also be written as

\[
y_i = X_i \beta_i + e_i, \quad i = 1, \ldots, N,
\]

where \( y_i \) is \( T \times 1 \), \( X_i \) is \( T \times k \), and \( e_i \) is \( T \times 1 \).

When \( T \) is small, estimating (2) is not feasible. A simpler form of (2) is such that only the intercept changes with \( i \) and the other parameters remain constant across \( i \):

\[
y_i = \ell_T a_i + Z_i b + e_i, \quad i = 1, \ldots, N,
\]

where \( \ell_T \) is the \( T \)-dimensional vector of ones, \( [\ell_T \ Z_i] = X_i \) and \( [a_i \ b]' = \beta_i \). In (3), individual effects are completely captured by the intercept \( a_i \). This specification simplifies (2) from \( kN \) to \( N + k - 1 \) parameters and is known as the fixed-effects model. Stacking \( N \) equations in (3) together we obtain

\[
y = D a + Z b + e.
\]

Note that each column of \( D \) is in effect a dummy variable for the \( i \)th individual unit. In what follows, an individual unit will be referred to as a “group”.

Denote the \( i \)th group average over time is

\[
\bar{z}_i = \frac{1}{T} \sum_{t=1}^{T} z_{it} = \frac{1}{T} Z_i \ell_T;
\]

the \( i \)th group average of \( y_{it} \) over time is

\[
\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it} = \frac{1}{T} Y_i \ell_T;
\]

the overall sample average of \( z_{it} \) (average over time and group) is

\[
\bar{z} = \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} = \frac{1}{TN} Z' \ell_T N;
\]
and the overall sample average of $z_{it}$ is
\[
\bar{y} = \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} = \frac{1}{TN} y' T N.
\]
Observe that the overall sample averages are
\[
\bar{z} = \frac{1}{N} \sum_{i=1}^{N} \bar{z}_i, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} \bar{y}_i,
\]
which are the sample averages of group averages. By Frisch-Waugh-Lovell Theorem, the OLS estimator for $b$ is
\[
\hat{b}_{TN} = [Z'(I_{TN} - P_D)Z]^{-1} Z'(I_{TN} - P_D)y, \quad (4)
\]
where $P_D = D(D' D)^{-1} D'$ is a projection matrix. Thus, $\hat{b}_{TN}$ can be obtained by regressing $(I_{TN} - P_D)y$ on $(I_{TN} - P_D)Z$. Let $\hat{a}_{TN}$ denote the OLS estimator of the vector $a$ of individual effects. By the facts that
\[
D' \dot{y} = D' D \hat{a}_{TN} + D' Z \hat{b}_{TN},
\]
and that the OLS residual vector is orthogonal to $D$, $\hat{a}_{TN}$ can be computed as
\[
\hat{a}_{TN} = (D' D)^{-1} D'( \hat{y} - Z \hat{b}_{TN}) = (D' D)^{-1} D'( \hat{y} - \hat{e} - Z \hat{b}_{TN}) = (D' D)^{-1} D'( y - Z \hat{b}_{TN}).
\]
Writing $D = I_N \otimes \ell_T$, we have
\[
P_D = (I_N \otimes \ell_T)(I_N \otimes \ell_T')^{-1}(I_N \otimes \ell_T) = (I_N \otimes \ell_T)[I_N \otimes (\ell_T' \ell_T)^{-1}](I_N \otimes \ell_T') = I_N \otimes [\ell_T (\ell_T' \ell_T)^{-1} \ell_T'] = I_N \otimes \ell_T \ell_T' / T,
\]
where $\ell_T \ell_T' / T$ is also a projection matrix. Thus,
\[
I_{TN} - P_D = I_N \otimes (I_T - \ell_T \ell_T' / T),
\]
and \((I_T - \ell_T\ell_T' / T)y_i = y_i - \ell_T\bar{y}_i\) with the \(i\)th element being \(y_{it} - \bar{y}_i\). It follows that

\[
(I_{TN} - P_D)y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} \ell_T\bar{y}_1 \\ \ell_T\bar{y}_2 \\ \vdots \\ \ell_T\bar{y}_N \end{bmatrix}.
\]

Similarly,

\[
(I_{TN} - P_D)Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix} - \begin{bmatrix} \ell_T\bar{z}_1 \\ \ell_T\bar{z}_2 \\ \vdots \\ \ell_T\bar{z}_N \end{bmatrix}.
\]

This shows that the OLS estimator (4) can be obtained by regressing \(y_{it} - \bar{y}_i\) on \(z_{it} - \bar{z}_i\) for \(i = 1, \ldots, N\), and \(t = 1, \ldots, T\). That is,

\[
\hat{b}_{TN} = \left( \sum_{i=1}^{N} (Z'_i - \bar{z}_i\ell_T')(Z_i - \ell_Tz'_i) \right)^{-1} \left( \sum_{i=1}^{N} (Z'_i - \bar{z}_i\ell_T')(y_i - \ell_T\bar{y}_i) \right) \tag{5}
\]

\[
= \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (z_{it} - \bar{z}_i)(z_{it} - \bar{z}_i)' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (z_{it} - \bar{z}_i)(y_{it} - \bar{y}_i) \right).
\]

The estimator \(\hat{b}_{TN}\) will be referred to as the \textit{within-groups estimator} because it is based on the observations that are deviations from their group averages. It is also easily seen that the \(i\)th element of \(\hat{a}_{TN}\) is

\[
\hat{a}_{TN,i} = \frac{1}{T}(\ell_T'\bar{y}_i - \ell_T'Z_i\hat{b}_{TN}) = \bar{y}_i - \bar{z}_i\hat{b}_{TN}.
\]

To distinguish \(\hat{a}_{TN,i}\) and \(\hat{b}_{TN}\) from other estimators, we will suppress their subscript \(TN\) and denote them as \(\hat{a}_w\) and \(\hat{b}_w\).

Suppose that the classical conditions: \(Z\) are non-stochastic and \(E(y) = Da_0 + Zb_0\), which is equivalent to

\[
E(y_i) = \ell_Ta_0 + Z_i b_0, i = 1, \ldots, N,
\]

then, \(\hat{a}_w\) and \(\hat{b}_w\) are unbiased for \(a_0\) and \(b_0\), respectively. Suppose also that \(\text{var}(y_i) = \sigma_Y^2 I_T\) for every equation \(i\) and that \(\text{cov}(y_i, y_j) = 0\) for every \(i \neq j\), then \(\text{var}(y) = \sigma_Y^2 I_{TN}\). As \(\hat{a}_w\)
and $\hat{b}_w$ are linear $y$ and unbiased, by the Gauss-Markov theorem, $\hat{a}_w$ and $\hat{b}_w$ are the BLUEs for $a_0$ and $b_0$. The variance-covariance matrix of $\hat{a}_w$ and $\hat{b}_w$ are

$$\text{var}(\hat{b}_w) = \sigma_0^2 [Z' (I_{TN} - PD) Z]^{-1}$$

and

$$\text{var}(\hat{a}_{w,i}) = \frac{1}{T} \sigma_0^2 + \bar{z}_i' [\text{var}(\hat{b}_w)] \bar{z}_i.$$

The OLS estimator for the regression variance $\sigma_0^2$ in this case is

$$\hat{\sigma}_w^2 = \frac{1}{TN - N - k + 1} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \hat{a}_w, i - \hat{z}_i')^2,$$

which can be used to compute the estimators of $\text{var}(\hat{a}_{w,i})$ and $\text{var}(\hat{b}_w)$.

Note that the conditions $\text{var}(y_i) = \sigma_0^2 I_T$ for all $i$ and $\text{cov}(y_i, y_j) = 0$ for every $i \neq j$ may be too restrictive in applications. When any one of these conditions fails, $\text{var}(y_i) \neq \sigma_0^2 I_{TN}$ and $\hat{a}_w$ and $\hat{b}_w$ are no longer the BLUEs. Despite that $\text{var}(y)$ may not be a scalar variance-covariance matrix in practice, the fixed-effects model is typically estimated by the OLS method and hence also known as the least squares dummy variable model.

Suppose that

$$y \sim N(Da_0 + Zb_0, \sigma_0^2 I_{TN}),$$

the conventional $t$- and $F$-tests remain applicable. An interesting hypothesis for the fixed-effects model is whether fixed (individual) effects indeed exist. This amounts to applying an $F$-test to the hypothesis

$$H_0 : a_{1,0} = a_{2,0} = \cdots = a_{N,0}.$$

The null distribution of the $F$-test is $F(N - 1, TN - N - k + 1)$. In practice, it may be more convenient to estimate the following specification for the fixed-effects model:

$$y_1 = \ell_T a_1 + e_1$$
$$y_2 = \ell_T a_2 + e_2$$
$$\vdots$$
$$y_N = \ell_T a_N + e_N.$$
In this specification, the parameters \( a_i, i = 1, \ldots, N \) denote the differences between the \( i \)th and the first group effects. Testing the existence of fixed-effects is then equivalent to testing

\[
H_0 : a_{2,0} = \cdots = a_{N,0} = 0.
\]

2.2 Random-Effects Model

Given the specification (3) that allows for individual effects:

\[
y_i = \ell_T a_i + Z_i b + e_i, \quad i = 1, \ldots, N,
\]

we now treat \( a_i \) as random variables rather than constant parameters. Writing \( a_i = a + u_i \) with \( a = E(a_i) \), the specification above can be expressed as

\[
y_i = \ell_T a + Z_i b + (\ell_T u_i + e_i), \quad i = 1, \ldots, N.
\]

(6)

where \( \ell_T u_i \) and \( e_i \) form the error term. This specification differs from the fixed-effects model in that the intercept does not vary across units \( i \). The presence of \( u_i \) also makes (6) different from the specification that does not allow for individual effects. Here, the group heterogeneity due to individual effects is characterized by the random variables \( u_i \) and absorbed into the error term. Thus, (6) is known as the random-effects model.

If the OLS method is applied to (6), the OLS estimators of \( b \) and \( a \) are, respectively,

\[
\hat{b}_p = \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (z_{it} - \bar{z})(z_{it} - \bar{z})' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (z_{it} - \bar{z})(y_{it} - \bar{y}) \right), \quad (7)
\]

and \( \hat{a} = \bar{y} - \bar{z}' \hat{b}_p \). Comparing with the within-groups estimator (6), \( \hat{b}_p \) is based on the deviations of \( y_{it} \) and \( z_{it} \) from their respective averages \( \bar{y} \) and \( \bar{z} \), whereas \( \hat{b}_w \) is based on the deviations from groups averages \( \bar{y}_i \) and \( \bar{z}_i \). Alternatively, pre-multiplying \( \ell_T / T \) through (6) yields

\[
\bar{y}_i = a + \bar{z}_i b + (u_i + \bar{e}_i), \quad i = 1, \ldots, N,
\]

(8)

where \( \bar{e}_i = \sum_{t=1}^{T} e_{it} / T \). By noting that the sample averages of \( \bar{y}_i \) and \( \bar{z}_i \) are just \( \bar{y} \) and \( \bar{z} \), the OLS estimators for (8) are

\[
\hat{b}_b = \left( \sum_{i=1}^{N} (\bar{z}_i - \bar{z})(\bar{z}_i - \bar{z})' \right)^{-1} \left( \sum_{i=1}^{N} (\bar{z}_i - \bar{z})(\bar{y}_i - \bar{y}) \right), \quad (9)
\]
and $\hat{a}_b = \bar{y} - \bar{z}'\bar{b}_b$. The estimator $\bar{b}_b$ is known as the *between-groups estimator* because it is based on the deviations of group averages from the overall averages. It can also be shown that the estimator (7) is a weighted sum of the within-groups estimator (6) and the between-group estimator (9). Thus, $\hat{b}_p$ is known as the *pooled estimator*.

Suppose that the classical conditions: $Z$ are non-stochastic and for every equation $i$ such that

$$E(y_i) = \ell_T a_0 + Z_i b_0, i = 1, \ldots, N.$$ 

Then,

$$E(y) = \ell_N a_0 + Z b_0,$$

and

$$E(y_i) = a_0 + \bar{z}_i b_0, i = 1, \ldots, N.$$ 

It follows that $\hat{a}_b$ and $\bar{b}_b$, as well as $\hat{a}_p$ and $\hat{b}_p$, are unbiased for $a_0$ and $b_0$ in the random-effects model. Moreover, write

$$y_i = \ell_T a_0 + Z_i b_0 + e_i^*,$$

where $e_i^*$ is the sum of two components: the random effects $\ell_T u_i$ and the disturbance $e_i$ for equation $i$. Then

$$\text{var}(y_i) = \sigma^2_u \ell_T \ell_T' + \text{var}(e_i) + 2 \text{cov}(\ell_T u_i, e_i),$$

where $\sigma^2_u$ is var($u_i$). As the first term on the right-hand side above is a full matrix, var($y_i$) is not a scalar variance-covariance matrix in general. Consequently, $\hat{a}_p$ and $\hat{b}_p$ are not the BLUEs. For the specification (8), $\hat{a}_b$ and $\hat{b}_b$ are not the BLUEs unless more stringent conditions are imposed.

**Remark:** If the fixed-effects model (3) is correct, the random-effects model (6) can be viewed as a specification that omits $N - 1$ relevant dummy variables. This implies the pooled estimator $\hat{b}_p$ and the between-groups estimator $\hat{b}_b$ are biased for $b_0$ in the fixed-effects model.

To perform FGLS estimation for the random-effect model, more conditions on var($y_i$) are needed. If var($e_i$) = $\sigma_0^2 I_T$ and $E(u_i e_i) = 0$, var($y_i$) becomes:

$$S_0 := \text{var}(y_i) = \sigma^2_u \ell_T \ell_T' + \sigma_0^2 I_T.$$
Suppose, alternatively, \( E(u_iu_j) = 0 \), \( E(u_ie_j) = 0 \) and \( E(e_ie_j) = 0 \) for all \( i \neq j \), then \( \text{cov}(y_i, y_j) = 0 \). Hence,

\[
\Sigma_0 := \text{var}(y) = I_N \otimes S_0
\]

which is a block diagonal matrix and is not a scalar variance-covariance matrix unless \( \sigma_u^2 = 0 \).

It can be verified that the desired transformation matrix for GLS estimation is \( \Sigma_0^{-1/2} = I_N \otimes S_0^{-1/2} \), where

\[
S_0^{-1/2} = I_T - \frac{c}{T} \ell_T \ell_T',
\]

and \( c = 1 - \sigma_u^2/(T \sigma_u^2 + \sigma_0^2)^{1/2} \). The transformed data are then \( S_0^{-1/2} y_i \) and \( S_0^{-1/2} z_i \), \( i = 1, \ldots, N \), and their \( t \)th elements are, respectively, \( y_{it} - c\bar{y}_i \) and \( z_{it} - c\bar{z}_i \). Regressing \( y_{it} - c\bar{y}_i \) on \( z_{it} - c\bar{z}_i \) gives the desired GLS estimator.

It can be shown that the GLS estimator is also a weighted average of the within-groups and between-groups estimators. For the special case that \( \sigma_0^2 = 0 \), we have \( c = 1 \) and

\[
\Sigma_0^{-1/2} = I_N \otimes (I_T - \ell_T \ell_T'/T) = I_{TN} - P_D,
\]

as in the fixed-effects model. In this case, the GLS estimator of \( b \) is nothing but the within-groups estimator \( \hat{b}_w \). When \( c = 0 \), the GLS estimator of \( b \) reduces to the pooled estimator \( \hat{b}_p \).

To compute the FGLS estimator, the parameters \( \sigma_u^2 \) and \( \sigma_0^2 \) in \( S_0 \) have to be estimated first. As the random effects \( u_i \) can be eliminated by taking the difference of \( y_i \) and \( \ell_T \bar{y}_i \):

\[
y_i - \ell_T \bar{y}_i = (Z_i - \ell_T \bar{z}_i)b_0 + (e_i - \ell_T \bar{e}_i),
\]

the OLS estimator of \( b_0 \) is just the within-groups estimator \( \hat{b}_w \). Since \( u_i \) has been eliminated, \( \sigma_0^2 \), the variance of \( e_{it} \), can be estimated by

\[
\hat{\sigma}_e^2 = \frac{1}{T(N - N - k + 1)} \sum_{i=1}^{N} \sum_{t=1}^{T} [(y_{it} - \bar{y}_i) - (z_{it} - \bar{z}_i)' \hat{b}_w]^2,
\]

which is also the variance estimator \( \hat{\sigma}_w^2 \) in the fixed-effects model. To estimate \( \sigma_u^2 \), as

\[
\bar{y}_i = a_0 + \bar{z}_i b_0 + (u_i + \bar{e}_i), \quad i = 1, 2, \ldots, N,
\]

which corresponds to the specification (8) for computing the between-groups estimator. When \( \text{var}(y_i) = \sigma_0^2 I_T \) also holds form every \( i \),

\[
\text{var}(u_i + \bar{e}_i) = \sigma_u^2 + \sigma_0^2 / T.
\]
This variance may be estimated by \( \sum_{i=1}^{N} \hat{e}_{b,i}^2 / (N - k) \), where
\[
\hat{e}_{b,i} = (y_i - \bar{y}) - (z_i - \bar{z})' \hat{b}, \quad i = 1, \ldots, N.
\]
Consequently, the estimator for \( \sigma_u^2 \) is
\[
\hat{\sigma}_u^2 = \frac{1}{N-k} \sum_{i=1}^{N} \hat{e}_{b,i}^2 - \hat{\sigma}_e^2 T.
\]
The estimators \( \hat{\sigma}_u^2 \) and \( \hat{\sigma}_e^2 \) now can be used to construct the estimated transformation matrix \( \hat{S}^{-1/2} \). It is clear that the FGLS estimator is a very complex function of \( y \).

3 Panel Unit Roots

Testing for unit roots has become a standard procedure in time series analyzes. For panel data, panel unit root tests have been proposed by Levin and Lin (1992), Im, Pesaran and Shin (1997), Harris and Tzavalis (1999), Madala and Wu (1999), Choi (1999), Hadri (1999), and Levin, Lin and Chu (2002). Originally, Bharagava et al. (1982) proposed modified Durbin-Watson statistic based on fixed effect residuals and two other test statistics on differenced OLS residuals for random walk residuals in a dynamic model with fixed effects. Boumahdi and Thomas (1991) generalized the Dickey-Fuller (DF) test for unit roots in panel data. Various modified DF test statistics have been applied by Breitung and Meyer (1994). Quah (1994) suggested a test for unit root in a panel data without fixed effects when both \( N \) and \( T \) go to infinity at the same rate such that \( N/T \) is constant. Levin and Lin (1992) extended the model to allow for fixed effects, individual deterministic trends and heterogeneous serially correlated errors. Both \( T \) and \( N \) are assumed to tend to infinity and \( N/T \to 0 \). Phillips and Moon (1999b) pointed out the asymptotic properties of estimators and tests proposed for nonstationary panels are determined by the way of \( N \) and \( T \) tending to infinity. First, take one index, say \( N \) is fixed and \( T \) is allowed to increase to infinity, an intermediate limit is obtained. Then by letting \( N \) tend to infinity subsequently, a sequential limit theory is established. Phillips and Moon (1999b) argued that these sequential limits are easy to derive and are helpful in extracting quick asymptotics. However, Phillips and Moon (1999b) gave a simple example to illustrate how sequential limits could mislead asymptotic results. Second, Quah (1994) and Levin and Lin (1992) allowed \( N \) and \( T \) to pass to infinity along a specific diagonal path in the two dimensional array. The path can be determined by a monotonically
increasing functional relation of $T = T(N)$. Phillips and Moon (1999b) demonstrated that the limit theory established by this assumption is dependent on the functional relation $T = T(N)$ and the assumed expansion path may not provide an appropriate approximation for a given $(T, N)$ situation. Third, a joint limit theory is established by allowing both $N$ and $T$ tend to infinity simultaneously without placing specific diagonal path restriction on the divergence. Phillips and Moon (1999b) argued that, in general, joint limit theory is more robust than either sequential limit or diagonal path limit. The multi-index asymptotic theory in Phillips and Moon (1999a, b) is applied to joint limits in which $T, N \to \infty$ and $(T/N) \to \infty$, i.e., to situations where the time series sample is large relative to the cross-section sample. However, the general approach of Phillips and Moon (1999b) is also applicable to situations $(T/N) \to 0$ although different limit results will usually be obtained.

3.1 Levin and Lin (1992) Tests

Consider the model

$$y_{it} = \rho_i y_{i(t-1)} + z_{it}' \gamma + u_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T,$$

(10)

where $z_{it}$ is the deterministic component and $u_{it}$ is a stationary process. $z_{it}$ could be zero, one, the fixed effects, $\mu_i$, or fixed effects as well as a time trend, $t$. The Levin and Lin (LL) tests assume that $\rho_i = \rho$ for all $i$ and are interesting in testing the null hypothesis

$$H_0 : \rho = 1$$

against the alternative hypothesis

$$H_a : \rho < 1.$$

The model in (10) can be written as, given $\rho_i = \rho$,

$$\Delta y_{it} = \delta y_{i(t-1)} + z_{it}' \gamma + u_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T,$$

(11)

where $\delta = \rho - 1$. Then the null of $\rho = 1$ is equivalent to $\delta = 0$.

Let $\hat{\rho}$ be the OLS estimator of $\rho$ in (10) and define

$$z_t = \begin{bmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{Nt} \end{bmatrix} = \begin{bmatrix} 1 & \mu_1 & t \\ 1 & \mu_2 & t \\ \vdots & \vdots & \vdots \\ 1 & \mu_N & t \end{bmatrix}.$$
\[ h(t, s) = z_i' \left( \sum_{t=1}^{T} z_t z_t' \right) z_s, \]

\[ \tilde{u}_{it} = u_{it} - \sum_{s=1}^{T} h(t, s) u_{is}, \]

\[ \tilde{y}_{it} = y_{it} - \sum_{s=1}^{T} h(t, s) y_{is}, \]

then we have

\[ \sqrt{NT}(\hat{\rho} - 1) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \tilde{y}_{i,t-1} \tilde{u}_{it} \]

\[ \sqrt{NT}(\hat{\rho} - 1) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \tilde{y}_{i,t-1} \tilde{u}_{it} \]

and the corresponding t-statistic

\[ t_{\hat{\rho}} = \frac{(\hat{\rho} - 1) \sqrt{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{i,t}^2}}{s_e}, \]

where

\[ s_e^2 = \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{u}_{it}^2. \]

Assume that there existing a scaling matrix \( D_T \) and piecewise continuous function \( Z(r) \) such that

\[ D_T^{-1/2} \sum_{t=1}^{[Tr]} z_t \Rightarrow Z(r) \]

uniformly for \( r \in [0,1] \). For a fixed \( N \),

\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \tilde{y}_{i,t-1} \tilde{u}_{it} \Rightarrow \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \int W_i Z dW_i Z \]

and

\[ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{y}_{i,t}^2 \Rightarrow \frac{1}{N} \sum_{i=1}^{N} \int W_i^2 Z, \]

as \( T \to \infty \), where \( W_i Z(r) = W(r) - [\int W Z'][\int Z Z']^{-1} Z(r) \) is an \( L_2 \) projection residual of \( W(r) \) on \( Z(r) \). Assume that \( \int W_i Z dW_i Z \) and \( \int W_i^2 Z \) are independent across \( i \) and have finite moments. Then it follows that

\[ \frac{1}{N} \sum_{i=1}^{N} \int W_i^2 Z \to_p E \left[ \int W_i^2 Z \right] \]
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \int W_i Z dW_i - E \left[ \int W_i Z dW_i \right] \right) \Rightarrow N \left( 0, \text{var} \left( \int W_i Z dW_i \right) \right)
\]

as \( N \to \infty \) by a law of large number and the Lindeberg-Levy central limit theorem. By Levin and Lin (1992), we have

\[
z_{it} E[\int W_i Z dW_i] \quad \text{var}[\int W_i Z dW_i] \quad E[\int W_i^2 Z] \quad \text{var}[\int W_i^2 Z]
\]

0 0 \( \frac{1}{2} \) \( \frac{1}{2} \) \( \frac{1}{3} \)
1 0 \( \frac{1}{3} \) \( \frac{1}{2} \) ?
\( \mu_i \) \( -\frac{1}{2} \) \( \frac{1}{12} \) \( \frac{1}{6} \) \( \frac{1}{45} \)
\((\mu_i, t)\) \( -\frac{1}{2} \) \( \frac{1}{60} \) \( \frac{1}{15} \) \( \frac{11}{6300} \)

Using above results, Levin and Lin (1992) obtain the following limiting distribution of \( \sqrt{NT} (\hat{\rho} - 1) \) and \( t_{\hat{\rho}} \):

\[
z_{it} \quad \hat{\rho} \quad t_{\hat{\rho}}
\]

0 \( \sqrt{NT} (\hat{\rho} - 1) \Rightarrow N(0, 2) \quad t_{\hat{\rho}} \Rightarrow N(0, 1) \)
1 \( \sqrt{NT} (\hat{\rho} - 1) \Rightarrow N(0, 2) \quad t_{\hat{\rho}} \Rightarrow N(0, 1) \)
\( \mu_i \) \( \sqrt{NT} (\hat{\rho} - 1) + 3\sqrt{N} \Rightarrow N(0, 51/5) \quad \sqrt{1.25 t_{\hat{\rho}} + \sqrt{1.875N}} \Rightarrow N(0, 1) \)
\((\mu_i, t)\) \( \sqrt{N(T(\hat{\rho} -1) + 7.5) \Rightarrow N(0, 2895/112) \quad \sqrt{448/277(t_{\hat{\rho}} + \sqrt{3.75N})} \Rightarrow N(0, 1) \)

Sequential limit theory, i.e., \( T \to \infty \) followed by \( N \to \infty \), is used to derive the limiting distributions. In case \( u_{it} \) is stationary, the asymptotic distributions of \( \hat{\rho} \) and \( t_{\hat{\rho}} \) need to be modified due to the presence of serial correlation.

Harris and Tzavalis (1999) also derived unit root tests for (10) with \( z_{it} = \{0\}; \{\mu_i\} \), or \( \{\mu_i, t\} \) when the time dimension of the panel, \( T \), is fixed. This is typical case for micro panel studies. The main results are

\[
z_{it} \quad \hat{\rho} \quad t_{\hat{\rho}}
\]

\( \mu_i \) \( \sqrt{N} (\hat{\rho} - 1 + \frac{3}{T+1}) \Rightarrow N \left( 0, \frac{3(17T^2 - 20T + 17)}{(T-1)(T+1)^4} \right) \)
\((\mu_i, t)\) \( \sqrt{N} (\hat{\rho} - 1 + \frac{15}{2(T+2)}) \Rightarrow N \left( 0, \frac{15(193T^2 - 278T + 1147)}{112(T+1)^4(T-2)} \right) \)

Harris and Tzavalis (1999) also showed that the assumption that \( T \) tends to infinity at a faster rate than \( N \) as in tests of Levin and Lin (1992) rather than \( T \) fixed as in the case in
micro panels, yields tests which are substantially undersized and have low power especially when $T$ is small.

Recently, tests of Levin and Lin (1992) have been implemented too test the PPP hypothesis using panel data. O’Connel (1998), however, showed that tests of Levin and Lin (1992) suffered from significant size distortion in the presence of correlation among contemporaneous cross-sectional error terms. O’Connel highlighted the importance of controlling for cross-sectional dependence when testing for a unit root in panels of real exchange rates.

Virtually all the existing nonstationary panel literature assume cross-sectional independence. It is true that the assumption of independence across $i$ is rather strong, but it is needed in order to satisfy the requirement of the Lindeberg-Levy central limit theorem. Moreover, as pointed out by Quah (1994), modeling cross-sectional dependence is involved because individual observations in a cross-section have no natural ordering. Driscoll and Kraay (1998) presented a simple extension of common nonparametric covariance matrix estimation techniques which are robust to very general forms of spatial and temporal dependence as the time dimension becomes large. Conley (1999) presented a spatial model of dependence among agents using a metric of economic distance that provides cross-sectional data with a structure similar to time-series data. Conley proposed a generalized method of moment (GMM) using such dependent data and a class of nonparametric covariance matrix estimators that allow for a general form of dependence characterized by economic distance.

It is worthy to mention that a method called PANIC (Panel Analysis of Non-stationary in the Idiosyncratic and Common components) has been suggested by Bai and Ng (2003) to overcome the problem of cross-sectional dependence across $i$. This method will be introduced later.

3.2 Im, Pesaran and Shin (1997) Tests

The tests of Levin and Lin (1992) are restrictive in the sense that it requires $\rho$ to be homogeneous across $i$. As Maddala (1999) pointed out, the null may be fine for testing convergence in growth among countries, but the alternative restricts every country to converge at the same rate. Im, Pesaran and Shin (1997) allow for heterogeneous coefficient of $y_{it-1}$ and proposed an alternative testing procedure based on the augmented DF tests when $u_{it}$ is serially correlated with different serial correlation properties across cross-sectional units, i.e.,
\[ u_{it} = \sum_{j=1}^{p_i} \psi_{ij} u_{it-j} + \epsilon_{it}. \] Substituting this \( u_{it} \) in (10), we get
\[ y_{it} = \rho_i y_{it-1} + \sum_{j=1}^{p_i} \psi_{ij} \Delta y_{it-j} + z_{it}' \gamma + \epsilon_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T, \quad (13) \]
The null hypothesis is
\[ H_0 : \rho_i = 1 \]
for all \( i \) against the alternative hypothesis
\[ H_a : \rho_i < 1 \]
for at least one \( i \). The \( t \)-statistic suggested by Im, Pesaran and Shin (1997) is defined as
\[ \bar{t} = \frac{1}{N} \sum_{i=1}^{N} t_{\hat{\rho}_i}, \quad (14) \]
where \( t_{\hat{\rho}_i} \) is the individual \( t \)-statistic of testing \( H_0 : \rho_i = 1 \) in (13). It is known that for a fixed \( N \),
\[ t_{\hat{\rho}_i} \Rightarrow \frac{1}{\sqrt{W_i}} \frac{dW_i}{dZ} \left[ \frac{1}{\sqrt{W_i}} \frac{1}{2} \right] = t_{iT} \quad (15) \]
as \( T \to \infty \). Im, Pesaran and Shin (1997) assume that \( t_{iT} \) are i.i.d. and have finite means and variances. Then
\[ \sqrt{N} \left( \frac{1}{N} \sum_{i=1}^{N} t_{iT} - E[t_{iT} | \rho_i = 1] \right) \Rightarrow N(0,1) \]
as \( N \to \infty \) by the Lindeberg-Levy central limit theorem. Hence, the test statistic of Im, Pesaran and Shin (1997) has the limiting distribution as
\[ t_{IPS} = \frac{\sqrt{N} (\bar{t} - E[t_{iT} | \rho_i = 1])}{\sqrt{\text{var}[t_{iT} | \rho_i = 1]}} \Rightarrow N(0,1) \]
as \( T \to \infty \) followed by \( N \to \infty \) sequentially. The values of \( E[t_{iT} | \rho_i = 1] \) and \( \text{var}[t_{iT} | \rho_i = 1] \) have been computed by Im, Pesaran and Shin (1997) via simulations for different values of \( T \) and \( p_i \)'s.

The local power of tests of Levin and Lin (1992) and Im, Pesaran and Shin (1997) has been investigated by Breitung (2000). Breitung (2000) finds that those tests suffer from a dramatic loss of power if individual specific trends are included. Besides, the power of these tests is very sensitive to the specification of the deterministic terms. McCoskey and
Selden (1998) applied the test of Im, Pesaran and Shin (1997) for testing unit root for per capita national health care expenditures and gross domestic product for a panel of OECD countries and the null is rejected for both panels. Gerdtham and Löthgren (1998) obtained different results by adding a time trend into the regression model (13).

3.3 Combining p-value Tests

Let $G_{i,T}$ be a unit root test statistic for the $i$-th group in (10) and assume that as $T_i \to \infty$, $G_{i,T_i} \Rightarrow G_i$. Let $p_i$ be the $p$-value of a unit root test for cross-section $i$, i.e., $p_i = 1 - F(G_{i,T_i})$, where $F(\cdot)$ is the distribution function of the random variable $G_i$. Note that the null of unit root is not rejected when the $p$-value is larger than 5% if the significance level is set at 95%. In contrast, the null is rejected when $p$-value is smaller than 5%. Maddala and Wu (1999) and Choi (1999) proposed a Fisher type test:

$$P = -2 \sum_{i=1}^{N} \ln p_i$$

(16)

which combines the $p$-value from unit root tests for each cross-section $i$ to test for unit root in panel data. $P$ is distributed as $\chi^2$ with $2N$ degrees of freedom as $T_i \to \infty$ for all $N$. When $p_i$ closes to 0 (null hypothesis is rejected), $\ln p_i$ closes to $-\infty$ so that large value $P$ will be found and then the null hypothesis of existing panel unit root will be rejected. In contrast, when $p_i$ closes to 1 (null hypothesis is not rejected), $\ln p_i$ closes to 0 so that small value $P$ will be found and then the null hypothesis of existing panel unit root will not rejected.

Choi (1999) pointed out the advantages of the Fisher test: (1) the cross-sectional dimension, $N$, can be either finite or infinite, (2) each group can have different types of nonstochastic and stochastic components, (3) the time series dimension, $T$ can be different for each $i$ (imbalance panel data), and (4) the alternative hypothesis would allow some groups to have unit roots while others may not. A main disadvantage involved is that the $p$-value have to be derived by Monte Carlo simulations.

When $N$ is large, Choi (1999) also proposed a $Z$ test,

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\frac{-2 \ln p_i - 2}{2})$$

(17)

since $E[-2 \ln p_i] = 2$ and var$[-2 \ln p_i] = 4$. Assume $p_i$’s are i.i.d. and use the Lindeberg-Levy
central limit theorem to get

\[ Z \Rightarrow N(0,1) \]

as \( T_i \to \infty \) followed by \( N \to \infty \).

3.4 Residual Based LM Test

Hadri (1999) proposed a residual based Lagrange Multiplier (LM) test for the null that the time series for each \( i \) are stationary around a deterministic trend against the alternative of a unit root in panel data. Consider the following model

\[ y_{it} = z_{it}' \gamma + r_{it} + \epsilon_{it} \quad (18) \]

where \( z_{it} \) is the deterministic component, \( r_{it} \) is a random walk

\[ r_{it} = r_{it-1} + u_{it} \]

\( u_{it} \sim i.i.d.(0, \sigma_u^2) \) and \( \epsilon_{it} \) is a stationary process. (18) can be written as

\[ y_{it} = z_{it}' \gamma + e_{it} \quad (19) \]

where

\[ e_{it} = \sum_{j=1}^{t} u_{ij} + \epsilon_{it}. \]

Let \( \hat{e}_{it} \) be the residuals from the regression in (19) and \( \hat{\sigma}_e^2 \) be the estimate of the error variance. Also, let \( S_{it} \) be the partial sum process of the residuals,

\[ S_{it} = \sum_{j=1}^{t} \hat{e}_{ij}. \]

Then the LM statistic is

\[ LM = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \frac{S_{it}^2}{\hat{\sigma}_e^2}. \]

It can be shown that

\[ LM \overset{p}{\to} E \left[ \int W_i Z \right] \]

as \( T \to \infty \) followed by \( N \to \infty \) provided \( E[\int W_i^2 Z] < \infty \). Also,

\[ \frac{\sqrt{N}(LM - E[\int W_i^2 Z])}{\sqrt{\text{var}[\int W_i^2 Z]}} \Rightarrow N(0,1) \]

as \( T \to \infty \) followed by \( N \to \infty \).
4 Panel Cointegration Tests

Entorf (1997) studied spurious fixed effects regressions when the true model involves independent random walks with and without drifts. For \( T \to \infty \) and \( N \) finite, the nonsense regression phenomenon holds for spurious fixed effects models and inference based on \( t \)-value can be highly misleading. Kao (1999) and Phillips and Moon (1999) derived the asymptotic distributions of the least squares dummy variable estimator and various conventional statistics from the spurious regression in panel data.

Consider a spurious regression model for all \( i \) using panel data:

\[
y_{it} = x_{it}' \beta + z_{it} - 1 \gamma + e_{it},
\]

where

\[
x_{it} = x_{it-1} + \epsilon_{it},
\]

and \( e_{it} \) is I(1). The OLS estimator of \( \beta \) is

\[
\hat{\beta} = \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it} \right),
\]

(20)

where \( \tilde{y}_{it} \) is defined in (12) and

\[
\tilde{x}_{it} = x_{it} - \sum_{t=1}^{T} h(t, s)x_{is}.
\]

It is known that if a time-series regression for a given \( i \) is performed in model (20), the OLS estimator of \( \beta \) is spurious. It is easy to see that

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \rightarrow^p \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it} \rightarrow^p E \left[ \int W_{it}W_{it}' \right] \Omega_{\epsilon}
\]

and

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it} \rightarrow^p E \left[ \int W_{it}W_{it}' \right] \Omega_{ut}
\]

where

\[
\begin{align*}
z_{it} & = E[\int W_{it}W_{it}'] \\
0 & = \frac{1}{2} \\
1 & = \frac{1}{2} \\
\mu_i & = \frac{1}{6} I_k \\
\mu_i(t) & = \frac{1}{15} I_k
\end{align*}
\]
Then we have
\[ \hat{\beta} \xrightarrow{p} \Omega^{-1}_{\epsilon} \Omega_{ue}. \] (21)

(21) shows that the OLS estimator of \( \beta, \hat{\beta} \), is consistent for its true value, \( \Omega^{-1}_{\epsilon} \Omega_{ue} \). This is due to the fact that the noise, \( e_{it} \), is as strong as the signal, \( x_{it} \), since both \( e_{it} \) and \( x_{it} \) are I(1). In the panel regression (20) with a large number of cross-sections, the strong noise of \( e_{it} \) is attenuated by pooling the data and a consistent estimate of \( \beta \) can be extracted. The asymptotics of the OLS estimator are very different from those of the spurious regression in pure time series. This has an important consequence for residual-based cointegration tests in panel data, because the null distribution of residual-based cointegration tests depends on the asymptotics of the OLS estimator. This point is explained in the next section.

4.1 Kao Tests

Kao (1999) presented two types of cointegration tests in panel data, the DF and ADF types tests. The DF type tests from Kao can be calculated from the estimated residuals in (20) as:
\[ \hat{e}_{it} = \rho \hat{e}_{it-1} + v_{it}, \] (22)

where
\[ \hat{e}_{it} = \tilde{y}_{it} - \tilde{x}_{it}' \hat{\beta}. \]

The null of no cointegration is represented as \( H_0 : \rho = 1 \). The OLS estimate of \( \rho \) and the \( t \)-statistic are given as
\[ \hat{\rho} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{it} \hat{e}_{it-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{it}^2}, \]

and
\[ t_{\hat{\rho}} = \frac{(\hat{\rho} - 1) \sqrt{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{it}^2}}{s_\epsilon}, \]

where
\[ s_\epsilon^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{e}_{it} - \hat{\rho} \hat{e}_{it-1})^2}{TN}. \]
Kao (1999) proposed the following four DF type tests by assuming $z_{it} = \{\mu_i\}$:

$$DF_\rho = \frac{\sqrt{NT(\hat{\rho} - 1) + 3\sqrt{N}}}{\sqrt{10.2}},$$

$$DF_t = \sqrt{1.25t_{\hat{\rho}} + \sqrt{1.875N}},$$

$$DF^{*}_\rho = \frac{\sqrt{NT(\hat{\rho} - 1) + 3\sqrt{N}\hat{\sigma}^2_v}}{\sqrt{3 + \frac{3\hat{\sigma}^2_v}{5\hat{\sigma}^2_{0v}}}},$$

$$DF^{*}_t = \frac{t_{\hat{\rho}} + \sqrt{6N\hat{\sigma}^2_v}}{\sqrt{\frac{\hat{\sigma}^2_v}{2\hat{\sigma}^2_{0v}} + \frac{3\hat{\sigma}^2_v}{10\hat{\sigma}^2_{0v}}}}.$$

where $\hat{\sigma}^2_v = \hat{\Sigma}_u - \hat{\Sigma}_{ue}\hat{\Sigma}^{-1}_e$ and $\hat{\sigma}^2_{0v} = \hat{\Omega}_u - \hat{\Omega}_{ue}\hat{\Omega}^{-1}_e$. While $DF_\rho$ and $DF_t$ are based on the strong exogeneity of the regressors and errors, $DF^{*}_\rho$ and $DF^{*}_t$ are for the cointegration with endogenous relationship between regressors and errors. For the ADF tests, the following regression is considered

$$\hat{e}_{it} = \rho\hat{e}_{it-1} + \sum_{j=1}^{p} \psi_j \Delta \hat{e}_{it-j} + v_{it}. \quad (23)$$

With the null hypothesis of no cointegration, the ADF test statistics can be constructed as:

$$ADF = \frac{t_{ADF}}{\sqrt{\frac{\hat{\sigma}^2_v}{2\hat{\sigma}^2_{0v}} + \frac{3\hat{\sigma}^2_v}{10\hat{\sigma}^2_{0v}}}} + \sqrt{6N\hat{\sigma}^2_v}$$

where $t_{ADF}$ is the $t$-statistic of $\rho$ in (23). The asymptotic distributions of $DF_\rho$, $DF_t$, $DF^{*}_\rho$, $DF^{*}_t$ and $ADF$ converge to a standard normal distribution $N(0,1)$.

### 4.2 Residual Based LM Test

McCoskey and Kao (1998) derived a residual-based test for the null of cointegration in panels. For the residual based test, it is necessary to use an efficient estimation technique of cointegrated variables. There are several methods that have been shown to be efficient asymptotically in literature. These include the fully modified estimator of Phillips and Hansen (1990) and the dynamic least squares estimator of Saikonen (1991) and Stock and Watson (1993).

The model considered in McCoskey and Kao (1998) allows for varying slopes and intercepts:

$$y_{it} = \alpha_i + x'_{it}\beta + e_{it}, \quad (24)$$
\[ x_{it} = x_{it-1} + \epsilon_{it} \]  
\[ \epsilon_{it} = r_{it} + u_{it} \]  
\[ r_{it} = r_{it-1} + \theta u_{it} \]  

where \( u_{it} \) are i.i.d. \((0, \sigma_u^2)\). The null hypothesis of cointegration is equivalent to \( \theta = 0 \). The test statistic proposed by McCoskey and Kao (1998) is defined as:

\[ LM = \frac{1}{N} \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_{it}^2}{\hat{\sigma}_e^2} \]  

where \( S_{it} \) is the partial sum process of the residuals,

\[ S_{it} = \sum_{j=1}^{t} \hat{e}_{ij} \]

and \( \hat{\sigma}_e^2 \) is defined in McCoskey and Kao (1998). The asymptotic result for the test is:

\[ \sqrt{N}(LM - \mu_v) \Rightarrow N(0, \sigma_v^2). \]

The moments, \( \mu_v \) and \( \sigma_v^2 \), can be found through Monte Carlo simulation. The limiting distribution of LM is then free of nuisance parameters and robust to heteroskedasticity.

4.3 Pedroni Tests of Pedroni (1997)

4.4 Likelihood-Based Cointegration Test of Larsson, Lyhagen, and Löthgren (1998)
APPENDIX 1.

A. Some important properties of normalized Brownian motion:

1. \( W(0) = 0; \)
2. \( E[W(r)] = 0, 0 \leq r \leq 1; \)
3. \( \text{var}[W(r)] = r, 0 \leq r \leq 1; \)
4. \( W(r) \sim N(0, r), 0 \leq r \leq 1; \)
5. \( dW(r)/dt = \epsilon_r; \)
6. \( \text{cov}[W(s), W(t)] = E[W(s)W(t)] = \min(s, t); \)
7. \( E[W^2(r)] = \text{var}[W(r)] = r; \)
8. \( E[W^4(r)] = 3(\text{var}[W(r)])^2 = 3r^2; \)
9. \( \text{var}[W^2(r)] = E[W^4(r)] - (E[W^2(r)])^2 = 3r^2 - r^2 = 2r^2; \)
10. \( E[W^2(s)W^2(t)] = st + 2\min(s^2, t^2); \)

Proof: For \( s \leq t, \)

\[
E[W^2(s)W^2(t)] = \frac{1}{2} E\left\{ (W(t) - W(s))^2 \right\} \cdot \frac{1}{2} E\left\{ (W(t) - W(s))^2 \right\} = st + 2\min(s^2, t^2).
\]

11. \( \text{cov}[W^2(s), W^2(t)] = 2\min(s^2, t^2). \)

Proof: For \( s \leq t, \)

\[
\text{cov}(W^2(s), W^2(t)) = E\left\{ (W(s)^2 - EW^2(s))(W(t)^2 - EW^2(t)) \right\}
\]
\[ E[W^2(s)W^2(t)] - st - ts + st \]
\[ = st + 2s^2 - st - st + st = 2s^2 = 2 \min(s^2, t^2). \]

B. Some Brownian Motion Algebra:

1. \( E \int_0^1 W(r) dW(r) = 0 \) and \( \text{var} \int_0^1 W(r) dW(r) = 1/2; \)

   \text{Proof:} By Ito’s Lemma,
\[ \int_0^1 W(r) dW(r) = \frac{1}{2} [W^2(1) - 1], \]
and \( E[W^2(1)] = 1, \text{var}[W^2(1)] = 2, \) then
\[ \text{var} \int_0^1 W(r) dW(r) = \text{var} \left( \frac{1}{2} [W^2(1) - 1] \right) \]
\[ = \frac{1}{4} \text{var}[W^2(1)] = \frac{1}{4} \times 2 = \frac{1}{2}. \]

2. \( E \int_0^1 W(r) dr = 0 \) and \( \text{var} \int_0^1 W(r) dr = 1/3. \)

   \text{Proof:} For \( r \) is a number, then we can interchange \( E \) and \( \int \) so that
\[ E \int_0^1 W(r) dr = \int_0^1 E[W(r)] dr = 0, \]
and
\[ \text{var} \int_0^1 W(r) dr = E\{ \int_0^1 W(r) dr - E[\int_0^1 W(r) dr] \}^2 \]
\[ = E[\int_0^1 W(r) dr]^2 \]
\[ = E[\int_0^1 \int_0^1 W(r) W(s) dr ds] \]
\[ = \int_0^1 \int_0^1 E[W(r) W(s)] dr ds \]
\[ = \int_0^1 \int_0^1 \min(s, r) dr ds \]
\[ = \int_0^1 \left( \int_0^r s ds + \int_r^1 r ds \right) dr \]
\[ = \int_0^1 \left( \frac{s^2}{2} \big|_0^r + rs \big|_r^1 \right) dr \]
\[ = \int_0^1 \left( \frac{r^2}{2} + (r - r^2) \right) dr \]
\[ = \frac{r^2}{2} \big|_0^1 - \frac{r^3}{6} \big|_0^1 = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}. \]
3. $E \int_0^1 W^2(r) dr = 1/2$ and $\text{var} \int_0^1 W^2(r) dr = 1/3$.

Proof:

$$E \int_0^1 W^2(r) dr = \int_0^1 E[W^2(r)] dr = \int_0^1 r^2 |_{r=0} = \frac{1}{2},$$

and

$$\text{var} \int_0^1 W^2(r) dr = E \left\{ \int_0^1 W^2(r) dr - E \left[ \int_0^1 W^2(r) dr \right] \right\}^2$$

$$= E \left\{ \int_0^1 \int_0^1 \left[ W^2(r) - EW^2(r) \right] \left[ W^2(s) - EW^2(s) \right] dr ds \right\}$$

$$= \int_0^1 \int_0^1 \text{cov}[W^2(r), W^2(s)] dr ds$$

$$= \int_0^1 \int_0^1 2 \min(r^2, s^2) dr ds$$

$$= \int_0^1 \left[ \int_0^r 2s^2 ds + \int_r^1 2r^2 ds \right] dr$$

$$= 2 \int_0^1 \left( \frac{r^3}{3} \bigg|_0^r + \frac{r^2 s}{s=0} \right) dr$$

$$= 2 \int_0^1 \left( \frac{r^3}{3} + r^2 - r^3 \right) dr$$

$$= 2 \left( \frac{1}{3} - \frac{2r^4}{12} \right) |_{r=1}$$

$$= 2 \left( \frac{1}{3} - \frac{2}{12} \right) = \frac{1}{3}.$$!

4. $\Psi(r) = W(r) - \int_0^1 W(r) dr$ is a demeaned Brownian motion, then

(a) $E \int_0^1 \Psi(r) dW(r) = -\frac{1}{2}$ and $\text{var} \int_0^1 \Psi(r) dW(r) = \frac{1}{12}$;

(b) $E \int_0^1 \Psi^2(r) dr = \frac{1}{6}$ and $\text{var} \int_0^1 \Psi(r) dr = \frac{1}{45}$

Proof:

$$E \int_0^1 \Psi(r) dW(r) = E \int_0^1 \left( W(r) - \int_0^1 W(r) dr \right) dW(r)$$

$$= E \left( \int_0^1 W(r) dW(r) - \int_0^1 \int_0^1 W(r) W(s) dr ds \right)$$

$$= 0 - E \int_0^1 \int_0^1 W(r) W(s) dr ds$$

$$= -\frac{1}{3}. \text{(check!!!)}$$
5. $E \int_0^1 W(r)W(1)dr = \frac{1}{2}$;
6. $E \int_0^1 rW(r)W(s)dr = \frac{5}{24}$;
7. $E \int_0^1 rW(r)W(1)dr = \frac{1}{3}$;
8. $E \int_0^1 rsW(r)W(s)dr = \frac{2}{15}$;
9. $E \int_0^1 W^3(1)rW(r)dr = 1$;
10. $E \int_0^1 W^2(r)W^2(1)dr = \frac{7}{6}$;
11. $E \int_0^1 \int_0^1 W^2(r)W^2(s)drds = \frac{7}{12}$;
12. $E \int_0^1 \int_0^1 W^2(1)rW(r)W(s)drds = \frac{18}{24}$;
13. $E \int_0^1 \int_0^1 W^2(1)rW(r)sW(s)drds = \frac{16}{35}$;
14. $E \int_0^1 \int_0^1 W(1)W(r)W^2(s)drds = \frac{8}{12}$;
15. $E \int_0^1 \int_0^1 \int_0^1 W(1)rW(r)W(s)W(t)drdsdt = \frac{1}{2}$;
16. $E \int_0^1 \int_0^1 \int_0^1 W^2(r)W(s)W(t)drdsdt = \frac{13}{36}$;
17. $E \int_0^1 \int_0^1 \int_0^1 W^2(r)W(s)tW(t)drdsdt = \frac{197}{120}$;
18. $E \int_0^1 \int_0^1 \int_0^1 W^2(r)sW(s)tW(t)drdsdt = \frac{11}{63}$;
19. $E \int_0^1 \int_0^1 \int_0^1 rW(r)W(s)W(t)W(1)drdsdt = \frac{23}{72}$;
20. $E \int_0^1 \int_0^1 \int_0^1 rW(r)sW(s)W(t)W(1)drdsdt = \frac{37}{180}$;
21. $E \int_0^1 \int_0^1 \int_0^1 \int_0^1 W(r)W(s)W(t)W(u)drdsdtdu = \frac{1}{3}$;
22. $E \int_0^1 \int_0^1 \int_0^1 \int_0^1 rW(r)W(s)W(t)W(u)drdsdtdu = \frac{5}{24}$;
23. $E \int_0^1 \int_0^1 \int_0^1 \int_0^1 rW(r)sW(s)W(t)W(u)drdsdtdu = \frac{189}{1440}$;
24. $E \int_0^1 \int_0^1 \int_0^1 \int_0^1 W(r)sW(s)tW(t)uW(u)drdsdtdu = \frac{1}{12}$;
25. $E \int_0^1 \int_0^1 \int_0^1 \int_0^1 rW(r)sW(s)tW(t)uW(u)drdsdtdu = \frac{4}{75}$;
References
