DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 06-09

New Tests for Cointegration in Heterogeneous Panels

Nam T. Hoang

University of Colorado at Boulder

November 2006

Center for Economic Analysis
Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

New Tests for Cointegration in Heterogeneous Panels

Nam T. Hoang*
University of Colorado at Boulder

November 2006

Abstract

This paper makes the following contributions to the existing literature on panel cointegration. First, two new tests based on the principle of weighted symmetric estimation are proposed for panel cointegration testing. Second, the asymptotic distributions of these new tests are examined, and these are shown to be well defined Weiner processes that are free of nuisance parameters. Third, the size and power properties of the proposed tests are studied with a Monte Carlo simulation, and their properties are found to be superior to those of the existing tests across a range of environments.

Keywords: Panel coingtegration test; heterogeneous panel data

JEL classification: C12; C15; C22; C23

*Department of Economics, University of Colorado at Boulder, 256 UCB, Boulder, CO 80309 email: trung.hoang@colorado.edu

1 Introduction

Panel data is commonly used in empirical research today by economists. Following the study of unit root tests in panels, research examining the properties of non-stationary time series in panel form is becoming more and more developed. Kao (1997) and Pedroni (1997) proposed the original tests for cointegration in panels under the null of no cointegration, and these tests are the most commonly used tests in empirical work. The Kao (1997) test is used for homogeneous panels. Pedroni(1997) gives two sets of statistics: the first set is for testing cointegration in homogeneous panels and the second set of statistics is for testing cointegration in heterogeneous panels.

McCoskey and Kao (1998) proposed the use of the average of the Augmented Dickey-Fuller (ADF) statistics over cross-sections based on Im *et al.*(1997) to test the hypothesis of no cointegration in heterogeneous panels.

In this paper, based on the idea of Maddala and Wu (1999) of using the Fisher test and the previously known weighted symmetric (WS) estimation, we proposed three new tests: the average weighted symmetric (AWS) test, the Fisher-ADF (FADF) test and the Fisher weighted symmetric (FWS) test, for testing cointegration in heterogeneous panels with the null hypothesis of no cointegration.

Although weighted symmetric estimation was first introduced by Park and Fuller (1993), this estimation method has not been used by economists doing empirical work in time series. Weighted symmetric estimation usually brings better results compared with the Dickey-Fuller (DF) and ADF estimations, the most commonly used estimation methods in time series. It was shown by Pantula *et al.* (1994) that the test using weighted symmetric estimation is the most powerful test for testing unit roots in a single time series. Hoang and McNown (2006) found that weighted symmetric estimation also dominates the other estimation methods in testing unit roots in panel data in terms of test power. In testing cointegration in heterogeneous panels where the cointegration vectors are allowed different between cross-sections, the Pedroni, McCoskey and Kao tests are the only choices for researchers now. Pedroni uses the DF estimation and adjusts the statistics as the Phillips-Perron test for unit roots, while McCoskey and Kao use the ADF estimation and average the test statistics over cross-sections. The purpose of this study is to give more options for testing for cointegration in heterogeneous panels when the weighted symmetric estimation is used instead of DF or ADF estimation. The asymptotic distribution of weighed symmetric estimator in residual-based testing for cointegration in a set of single series is also derived and it is shown that the limiting distribution of WS estimators in testing cointegration is a function of standard Wiener processes and free of nuisance parameters.

A Monte-Carlo investigation is conducted to examine the performance of the available tests and the proposed tests under different data generating processes (DGPs). Consistent with what is already known about the weighted symmetric estimation, the AWS and FWS tests turn out to be the tests with the best performance, as they consistently dominate the rest of the compared tests in both size and power and in all studied data environments.

This study makes the following contributions to the existing literature on panel cointegration. First, two new tests based on the principle of weighted symmetric estimation are proposed for panel cointegration testing. Second, the asymptotic distributions of these new tests are examined, and these are shown to be well defined Weiner processes that are free of nuisance parameters. Third, the size and power properties of the proposed tests are studied with a Monte Carlo simulation, and their properties are found to be superior to those of the existing tests across a range of environments.

2 Current Tests for Cointegration in Panel Data

2.1 Testing for Cointegration in Homogeneous Panels

Chihwa Kao (1997) considered the following system of cointegrated regressions in the homogeneous panels:

Let

$$x_{it} = x_{it-1} + \epsilon_{it}$$

$$y_{it} = y_{it-1} + v_{it}$$

Consider the regression:

$$y_{it} = \alpha_i + x_{it}\beta + u_{it} \tag{1}$$

$$(i = 1, \ldots, N, t = 1, \ldots, T)$$

where α_i are individual constant terms, β is the slope parameter, ϵ_{it}, v_{it} are stationary distubance terms and so y_{it} and x_{it} are integrated processes of order 1 for all i

The zero mean vector $\xi_{it} = (v_{it}, \epsilon_{it})'$ satisfies

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_{it} \Longrightarrow B_i(\Omega)$$

for all i as $T \longrightarrow \infty$, where $B_i(\Omega)$ is a vector of Brownian motion with asymptotic covariance Ω. Kao derives two types of panel cointegration tests based on residuals from panel least-squares dummy variable (LSDV) estimation.

The first is of a Dickey-Fuller (DF) type, which can be applied to the residuals using:

$$\hat{u}_{it} = \rho \hat{u}_{it-1} + e_{it} \tag{2}$$

The OLS estimate of ρ is:

$$\hat{\rho} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{u}_{it}\hat{u}_{it-1}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{u}_{it-1}^{2}}$$

The null hypothesis that $\rho = 1$ is tested by:

$$\sqrt{N}T(\hat{\rho}-1) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\hat{u}_{it-1}\Delta\hat{u}_{it}}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=2}^{T}\hat{u}_{it-1}^{2}}$$

The second is an Augmented-Dickey-Fuller(ADF) type test which can be calculated from:

$$\hat{u}_{it} = \rho\hat{u}_{it-1} + \sum_{j=1}^{p}\phi_j\Delta\hat{u}_{it-j} + e_{itp} \tag{3}$$

where p is chosen so that the residuals e_{itp} are serially uncorrelated. The ADF test statistic here is the usual t-statistic with $\rho = 1$ in the ADF equation.

The following specification of null and alternative hypotheses is used: $H_0 : \rho = 1$, $H_1 : \rho < 1$.

Kao proposes four DF-type statistics and an ADF statistic. The first two DF statistics are based on assuming strict exogeneity of the regressors with respect to the errors in the equation, while the remaining two DF statistics allow for endogeneity of the regressors. The DF statistic, which allows for endogeneity, and the ADF statistic involve deriving some nuisance parameters from the long-run conditional variances Ω.

Kao showed that the asymptotic distributions of all tests converge to a standard normal distribution as $T \longrightarrow \infty$ and $N \longrightarrow \infty$

Kao is the first author to suggest the test for cointegration in homogeneous panels, The Kao test statistics are calculated by pooling all the residuals of all cross-sections in the panel. It is assumed in Kao's test that all the cointegrating vectors in every cross-section are identical.

2.2 Testing for Cointegration in Heterogeneous Panels

2.2.1 Pedroni (1997)

Pedroni (1997) considers the following model for heterogeneous panel data

$$y_{it} = \alpha_i + x_{it}\beta_i + u_{it} \tag{4}$$

$$(i = 1, \ldots, N, t = 1, \ldots, T)$$

For the processes:

$$x_{it} = x_{it-1} + \epsilon_{it}$$

$$y_{it} = y_{it-1} + v_{it}$$

where α_i are individual constant terms, β_i is the slope parameter for the cross-section i of the panel, ϵ_{it}, v_{it} are stationary disturbance terms and so y_{it} and x_{it} are integrated processes of order 1 for all i

The zero mean vector $\xi_{it} = (v_{it}, \epsilon_{it})'$ is assumed to satisfy

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_{it} \Longrightarrow B_i(\Omega_i)$$

for each cross-section i as $T \longrightarrow \infty$, where $B_i(\Omega_i)$ is a vector of Brownian motion on the interval $r \in [0, 1]$ with asymptotic covariance Ω_i. The asymptotic covariance matrix Ω_i is given by:

$$\Omega_i = \lim_{T \to \infty} E\left[\frac{1}{T}(\sum_{t=1}^{T} \xi_{it})(\sum_{t=1}^{T} \xi_{it}^{'})\right]$$

and can be decomposed as:

$$\Omega_i = \Sigma_i + \Gamma_i + \Gamma_i^{'}$$

where $\Sigma_i = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(\xi_{it}\xi_{it}^{'})$ is the contemporaneous covariance among the components of ξ_{it} for a given cross section i, $\Gamma_i = \lim_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T-1}\sum_{t=k+1}^{T} E(\xi_{it}\xi_{it-k}^{'})$ is the dynamic covariance among the components of ξ_{it}. Ω_i is permitted to vary across individual sections of the panel. Ω_i can be consistently estimated by:

$$\hat{\Omega}_i = \hat{\Sigma}_i + \hat{\Gamma}_i + \hat{\Gamma}_i^{'}$$

or

$$\hat{\Omega}_i = \frac{1}{T}\left[\sum_{t=1}^{T}\hat{\xi}_{it}\hat{\xi}_{it}^{'} + \sum_{s=1}^{k_i}\left(1-\frac{s}{k_i+1}\right)\sum_{t=s+1}^{T}\left(\hat{\xi}_{it-s}\hat{\xi}_{it}^{'} + \hat{\xi}_{it}\hat{\xi}_{it-s}^{'}\right)\right] \tag{5}$$

where $\hat{\xi}_{it} = (\hat{v}_{it}, \hat{\epsilon}_{it})'$ is obtained from autoregressions: $x_{it} = \rho_i x_{it-1} + \epsilon_{it}$ and $y_{it} = \rho_i y_{it-1} + v_{it}$ for each i. Define L_i is the lower triangular decomposition matrix of Ω_i:

$$\Omega_i = L_i^{'}L_i = \begin{bmatrix} L_{11i} & L_{21i} \\ 0 & L_{22i} \end{bmatrix}\begin{bmatrix} L_{11i} & 0 \\ L_{21i} & L_{22i} \end{bmatrix}$$

Then

$$L_{11i} = (\Omega_{11i} - \Omega_{21i}^2/\Omega_{22i})^{1/2}; L_{21i} = \Omega_{21i}/\Omega_{22i}^{1/2}; L_{22i} = \Omega_{22i}^{1/2} \tag{6}$$

Pedroni derived two groups of test statistics. The first group is accounted for in homogeneous panels under the assumption that $\Omega_i = \Omega$ for all i. The second groups of test statistics is for heterogeneous panels which allows the cointegration vector to vary across the cross sections and Ω_i are different in each section of panels. We focus on the second group of test statistics. Pedroni constructed the statistics as follows: In the first step one can estimate the proposed cointegrating regression for each individual member of the panel in the form (4) $y_{it} = \alpha_i + x_{it}\beta_i + u_{it}$. Next, use the original data to estimate $\hat{\xi}_{it}$ by regressing the levels of x_{it} and y_{it} on the lagged levels and use these values of $\hat{\xi}_{it}$ to estimate the appropriate long run covariances $\hat{\Omega}_i$ for each member of the panel as in (5). Then collect the residuals $\hat{u}_{it}$

from the separate regression for each panel member (4) and compute the lower triangular decomposition of the $\hat{\Omega}_i$ given in (5). Finally, run the following regression for each member of the panel:

$$\hat{u}_{it} = \hat{\rho}_i \hat{u}_{it-1} + \hat{e}_{it} \tag{7}$$

and construct the group mean statistics for the null of no cointegration in heterogeneous panels as:

$$\tilde{Z}_{\hat{\rho}_{NT}-1} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T}(\hat{u}_{it-1}\Delta\hat{u}_{it} - \hat{\lambda}_i)}{\sum_{t=1}^{T}\hat{u}_{it-1}^2} \tag{8}$$

$$\tilde{Z}_{t_{NT}} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T}(\hat{u}_{it-1}\Delta\hat{u}_{it} - \hat{\lambda}_i)}{\left[\sum_{t=1}^{T}\frac{1}{\hat{L}_{11i}^2}\hat{u}_{it-1}^2\right]^{1/2}} \tag{9}$$

Where $\hat{L}_{11i} = (\hat{\Omega}_{11i} - \hat{\Omega}_{21i}^2/\hat{\Omega}_{22i})^{1/2}$. $\hat{\Omega}_i$ is estimated as in (5). $\hat{\lambda}_i = \frac{1}{2}(\hat{\sigma}_i^2 - \hat{s}_i^2)$, for which, $\hat{s}_i^2$ is the contemporaneous variance of $\hat{e}_{it}$ and $\hat{\sigma}_i^2$ is the long-run variance of $\hat{e}_{it}$, they are consistently estimated by:

$$\hat{s}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}^2 \tag{10}$$

$$\hat{\sigma}_i^2 = \frac{1}{T}\left[\sum_{t=1}^{T}\hat{e}_{it}^2 + 2\sum_{s=1}^{k_i}\left(1 - \frac{s}{k_i+1}\right)\sum_{t=s+1}^{T}\hat{e}_{it}\hat{e}_{it-s}\right] \tag{11}$$

Pedroni showed that under the null of no cointegration ($\rho_i = 1$ $i = 1, 2, .., N$ in the equation $\hat{u}_{it} = \hat{\rho}_i\hat{u}_{it-1} + \hat{e}_{it}$). $\frac{T}{\sqrt{N}}\tilde{Z}_{\hat{\rho}_{NT}-1}$ and $\frac{1}{\sqrt{N}}\tilde{Z}_{\hat{t}_{NT}}$ converge to the normal distributions with both T and N $\longrightarrow \infty$. With the Monte-Carlo results the asymptotic distributions of these statistics can be written as:

$$\frac{T}{\sqrt{N}}\tilde{Z}_{\hat{\rho}_{NT}-1} + 9.05\sqrt{N} \xrightarrow{L} N(0, 35.98) \tag{12}$$

$$\frac{1}{\sqrt{N}}\tilde{Z}_{\hat{t}_{NT}} + 2.03\sqrt{N} \xrightarrow{L} N(0, 0.66) \tag{13}$$

One can use these results to test the hypothesis of no cointegration in every cross-section of a panel.

2.2.2 McCoskey and Kao (1998)

McCoskey and Kao (1998) propose the average Augmented Dickey-Fuller (ADF) test for varying slopes and varying intercepts across all the members of the panel. They consider the model:

$$y_{it} = \alpha_i + x_{it}\beta_i + u_{it} \tag{14}$$

$$(i = 1, \ldots, N, t = 1, \ldots, T)$$

Individual cointegrating equation are estimated separately for each cross-section and individual ADF statistics are computed for each cross-section. Each test is constructed such that the cross-sections are assumed independent of each other and heteroskedasticity across the cross-section is allowed. Using an analogous approach of Im *et al.* (1995) of average of the ADF statistics of the cross-sections in testing unit roots in panels, McCoskey and Kao computed the panel test statistics in the same way. The individual ADF test can be constructed as:

$$\hat{u}_{it} = \rho_i \hat{u}_{it-1} + \sum_{j=1}^{p} \phi_j \Delta \hat{u}_{it-j} + e_{itp} \tag{15}$$

where $\hat{u}_{it}$ are OLS residuals from (14). We can write equation (15) as:

$$\Delta \hat{u}_{it} = \rho_i \hat{u}_{it-1} + \sum_{j=1}^{p} \phi_j \Delta \hat{u}_{it-j} + e_{itp}$$

The null hypothesis is $H_0 : \rho_i = 0$ and the t-statistic for each i is constructed:

$$t_{iADF} = \frac{\left(\hat{u}'_{-1} Q_{X_p} \hat{u}_{-1}\right)^{1/2} \hat{\rho}_i}{s_e}$$

where $\hat{u}_{-1}$ is the vector of observations of $\hat{u}_{it-1}$, and $Q_{X_p} = I - X_p(X'_p X_p)^{-1} X'_p$ where X_p is the matrix of observations on the p regressors $(\Delta\hat{u}_{it-1}, ..., \Delta\hat{u}_{it-p})$ and $s_e^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{itp}^2$. Phillips and Ouliaris (1990) show that the t_{iADF} converges to a functional of Brownian motion. Finally, McCoskey and Kao construct the panel statistics as:

$$\bar{t}_{ADF} = \frac{1}{N}\sum_{i=1}^{N} t_{iADF}$$

Define $E[t_{iADF}] = \mu_{ADF}$, and $Var[t_{iADF}] = \sigma^2_{ADF}$. Then the central limit theorem can be applied to give:

$$\sqrt{N}(\bar{t}_{ADF} - \mu_{ADF}) \xrightarrow{L} N(0, \sigma^2_{ADF})$$

Phillips and Ouliaris note that the limiting distribution of ADF test statistics in each cross-section is free of nuisance parameters and depends only on the number of regressors in (14). McCoskey and Kao also did a simulation to find the values of μ_{ADF} and σ_{ADF}, where they found that in the case of one regressor, $\mu_{ADF} = -2.206$ and $\sigma_{ADF} = .8200$. In their paper, McCoskey and Kao also suggest the average Phillip Z_t test statistic based on the result that the Z_t statistic in each cross-section converges to the same functional of Brownian motion as that of the t_{iADF}.

3 Weighted Symmetric Estimation Method

Weighted Symmetric (WS) estimation was first proposed by Park and Fuller (1993). The use of WS estimation is not widely used in time series applications, although there is evidence that WS estimation produces test statistic with greater power than those based on simple least squares. In this section, we give the limiting distribution of the WS estimator and the WS pivotal statistics when using WS estimation in testing unit root and cointegration in a single time series.

3.1 Alternative Representations of an Autogregressive Process

The following theorem is taken from Fuller (1996).

Theorem 3.0

Let $\{X_t\}$ *be a time series defined on the integers with* $E\{X_t^2\} < K$ *for all* t*. Suppose* X_t *satisfies*

$$X_t + \sum_{j=1}^{p} \alpha_j X_{t-j} = e_j$$

$$t = 0, \pm 1, \pm 2, ...$$

where $\{e_t\}$*,* $t = 0, \pm 1, \pm 2, ...$ *, is a sequence of uncorrelated* $(0, \sigma^2)$ *random variables. Let* $m_1, m_2, ..., m_p$ *be the roots of the characteristic equation*

$$m^p + \sum_{j=1}^{p} \alpha_j m^{p-j} = 0$$

and assume $|m_i| < 1$*,* $i = 1, 2, ..., p$*. Then* X_t *is covariance stationary. Furthermore,* X_t *is given as a limit in mean square by*

$$X_t = \sum_{j=0}^{\infty} w_j e_{t-j}$$

where $\{w_j\}_{j=0}^{\infty}$ *is the unique solution of the homogeneous difference equation* $w_j + \alpha_1 w_{j-1} + ... + \alpha_p w_{j-p} = 0$ *,* $j = p, p+1, ...,$ *subject to the boundary conditions* $w_0 = 0$ *and* $w_j + \alpha_1 w_{j-1} + ... + \alpha_p w_{j-p} = 0$ *,* $j = 1, 2, ..., p-1$

Proof: See Fuller (1996, p.59).

We now know that a time series satisfying a pth order difference has a representations as an infinite moving average.

Proposition 3.1

If the time series $\{X_t\}$, $t = 0, \pm 1, \pm 2, ...$, *with zero mean satisfies*

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} = e_t$$

where $\{e_t\}$ *is a sequence of uncorrelated* $(0, \sigma^2)$ *random variables, and the roots* $m_1, m_2, ..., m_p$ *of the characteristic equation*

$$m^p + \alpha_1 m^{p-1} + \alpha_2 m^{p-2} + ... + \alpha_p = 0$$

are less than 1 in absolute value. Then we also have

$$X_t + \alpha_1 X_{t+1} + \alpha_2 X_{t+2} + ... + \alpha_p X_{t+p} = v_t$$

where $\{v_t\}$ *is a sequence of uncorrelated* $(0, \sigma^2)$ *random variables.*

Proof: See Appendix

3.2 Weighted Symmetric Estimator

Consider the process:

$$y_t = \rho y_{t-1} + \epsilon_t \tag{16}$$

$$t = 2, 3, \ldots, T; \epsilon_t \sim N(0, \sigma^2)$$

Following proposition 3.1, if a stationary process satisfies this equation, then it also satisfies the equation:

$$y_t = \rho y_{t+1} + u_t \tag{17}$$

$$t = 1, 2, \ldots, (T-1); u_t \sim N(0, \sigma^2)$$

Consider an estimator of ρ that minimizes:

$$Q = \sum_{t=2}^{T} w_t (y_t - \rho y_{t-1})^2 + \sum_{t=1}^{T-1} (1 - w_{t+1})(y_t - \rho y_{t+1})^2 \tag{18}$$

where w_t, t=2,3,...,T, are weights. The estimator obtained by setting $w_t = T^{-1}(t-1)$ is called the weighted symmetric (WS) estimator.

Proposition 3.2

The weighted symmetric estimator of ρ in the (16) process is

$$\hat{\rho} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T-1} y_t^2 + \frac{1}{T} \sum_{t=1}^{T} y_t^2} \tag{19}$$

If an intercept α is added in the regression equation as

$$y_t = \hat{\alpha} + \hat{\rho} y_{t-1} + e_t \tag{20}$$

then the weighted symmetric estimator of ρ would be

$$\hat{\rho} = \frac{\sum_{t=2}^{T} (y_t - \overline{y})(y_{t-1} - \overline{y})}{\sum_{t=2}^{T-1} (y_t - \overline{y})^2 + \frac{1}{T} \sum_{t=1}^{T} (y_t - \overline{y})^2} \tag{21}$$

with $\overline{y} = T^{-1} \sum_{t=1}^{T} y_t$

Proof: See Appendix

It can be extended to find the weighted symmetric estimator for the pth order autoregressive time series

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + e_i \tag{22}$$

$$t = p+1, \dots, T$$

We consider a class of estimators of α, obtained by minimizing:

$$Q = \sum_{t=p+1}^{T} w_t \left[Y_t - \alpha_0 - \sum_{i=1}^{p} \alpha_i Y_{t-i} \right]^2 + \sum_{t=1}^{T-p} (1 - w_{t+1}) \left[Y_t - \alpha_0 - \sum_{i=1}^{p} \alpha_i Y_{t+i} \right]^2 \tag{23}$$

where w_t, $t = 1, 2, ..., T$ are weights and $0 \leq w_t \leq 1$. The value of α that minimizes (23) with:

$$w_t = \begin{cases} 0 & t = 1, 2, ..., p \\ (T - 2p + 2)^{-1}(t - p) & t = p+1, p+2, ..., T-p+1 \\ 1 & t = T-p+2, T-p+3, ..., T \end{cases} \tag{24}$$

is called the weighed symmetric estimator for the pth order autoregressive time series (22).

3.3 Asymptotic Distribution of a Weighted Symmetric Estimator in Testing Unit Roots

Proposition 3.3

For $\hat{\rho}$ as in (19) then under the hypothesis of $\rho = 1$

$$T(\hat{\rho} - 1) \xrightarrow{L} \frac{\frac{1}{2}[K^2 - 1] - G}{G} \tag{25}$$

$$t_{\hat{\rho}} \xrightarrow{L} \frac{\frac{1}{2}[K^2 - 1] - G}{G^{1/2}} \tag{26}$$

where $K = W(1)$, $G = \int_0^1 [W(r)]^2 dr$. W(r) is a standard Wiener process on [0,1].

Proof: See Appendix

Proposition 3.4

For $\hat{\rho}$ as in (21) then under the hypothesis of $\rho = 1$

$$T(\hat{\rho} - 1) \xrightarrow{L} \frac{\left(\frac{1}{2}[K^2 - 1] - G - KH + 2H^2\right)}{[G - H^2]} \tag{27}$$

$$t_{\hat{\rho}} \xrightarrow{L} \frac{\left(\frac{1}{2}[K^2 - 1] - G - KH + 2H^2\right)}{[G - H^2]^{1/2}} \tag{28}$$

where $K = W(1)$, $G = \int_0^1 [W(r)]^2 dr$, $H = \int_0^1 W(r)dr$. $W(r)$ is a standard Wiener process on [0,1].

Proof: See Appendix

Notes: The propositions 3.3 and 3.4 are asserted in Fuller (1996) without proofs. We found the proofs and present them in the Appendix.

3.4 Asymptotic Distribution of Cointegration Test Statistics Using Weighted Symmetric Estimators

Consider two processes $\{x_t\}$ and $\{y_t\}$, suppose that they are both I(1) processes and are totally independent random walks:

$$\begin{cases} x_t = x_{t-1} + \epsilon_t & t = 1, 2, ..., T \\ y_t = y_{t-1} + v_t & t = 1, 2, ..., T \end{cases} \tag{29}$$

where both $\{\epsilon_t\}$ and $\{v_t\}$ are $(0, \sigma^2)$ processes. Consider the regression

$$y_t = \alpha + \beta x_t + u_t, t = 1, 2, ..., T \tag{30}$$

Run the regression (30) and get the residual $\hat{u}_t$. We apply the weighted symmetric unit root test on $\hat{u}_t$. The estimation equation is

$$\hat{u}_t = \rho \hat{u}_{t-1} + e_t, t = 1, 2, ..., T \tag{31}$$

The weighted symmetric estimator of ρ is the one to minimize

$$Q = \sum_{t=2}^{T} w_t (\hat{u}_t - \rho \hat{u}_{t-1})^2 + \sum_{t=1}^{T-1} (1 - w_{t+1})(\hat{u}_t - \rho \hat{u}_{t+1})^2 \tag{32}$$

$$w_t = \frac{t-1}{T}, t = 2, 3, ..., T$$

Following proposition 3.2, the weighted symmetric estimator of ρ is

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T-1} \hat{u}_t^2 + \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2} \tag{33}$$

We find the distribution of $T(\hat{\rho} - 1)$, $t_{\hat{\rho}}$ under the null hypothesis of no cointegration $\rho = 1$, implies $\{x_t\}$ and $\{y_t\}$ are independent random walks as in (29)

Proposition 3.5

Under the null hypothesis of no cointegration $H_0 : \rho = 1$ in (31)

$$T(\hat{\rho} - 1) \xrightarrow{L} \frac{\left[\frac{1}{2}(B_v - 2\zeta A + \zeta^2 B_w) + \frac{1}{2}D - C - \frac{1}{2}(1 + \zeta^2)\right]}{[C]} \tag{34}$$

$$t_{\hat{\rho}} \xrightarrow{L} \frac{\left[\frac{1}{2}(B_v - 2\zeta A + \zeta^2 B_w) + \frac{1}{2}D - C - \frac{1}{2}(1 + \zeta^2)\right]}{[(1 + \zeta^2)C]^{1/2}} \tag{35}$$

where

$$A = V(1)W(1) - W(1) \int_0^1 V(r)dr - V(1) \int_0^1 W(r)dr - \int_0^1 V(r)dr \int_0^1 W(r)dr$$

$$B_v = \left[\int_0^1 V(r)dr\right]^2 - 2V(1)\int_0^1 V(r)dr + [V(1)]^2$$

$$B_w = \left[\int_0^1 W(r)dr\right]^2 - 2W(1)\int_0^1 W(r)dr + [W(1)]^2$$

$$C = \int_0^1 [V(r)]^2 dr - \left(\int_0^1 V(r)dr\right)^2 - \zeta^2\left[\int_0^1 [W(r)]^2 dr - \left(\int_0^1 W(r)dr\right)^2\right]$$

$$D = \left[\int_0^1 V(r)dr\right]^2 + \zeta^2\left[\int_0^1 W(r)dr\right]^2 - 2\zeta\int_0^1 V(r)dr\int_0^1 W(r)dr$$

$$\zeta = \frac{\int_0^1 V(r)W(r)dr - \int_0^1 V(r)dr\int_0^1 W(r)dr}{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)dr\right]^2}$$

$V(r)$ and $W(r)$ are the standard Wiener processes on [0,1].

Proof: See Appendix

The limiting distribution of WS test statistics for cointegration in single set of time series is free of nuisance parameters and depends only on the number of regressors. A simulation can provide the values of moments of the distribution which can be used to test for cointegration in heterogeneous panels.

4 New Tests for Cointegration in Heterogeneous Panels

In this section, new tests are presented that are based on the idea of average test statistics for each member of a panel, used in Im *et al.*(2003) to test for unit roots in panels, and Fisher's approach to combine p-values from individual test in each cross-section which is introduced in Maddala and Wu (1999). We introduce the use of three new tests for testing cointegration in heterogeneous panels.

First we briefly recall the Im *et al.* and Fisher tests. The Im *et al.* statistic is based on the average individual Dickey-Fuller unit root tests as

$$t_{IPS} = \frac{\sqrt{N}(\bar{t} - E[t_i|\rho_i = 1])}{\sqrt{Var[t_i|\rho_i = 1]}} \xrightarrow{L} N(0, 1)$$

where $\bar{t} = N^{-1} \sum_{i=1}^{N} t_i$. The moments of $E[t_i|\rho_i = 1]$ and $Var[t_i|\rho_i = 1]$ are obtained by Monte Carlo simulation. Fisher shows that the statistic is given by $P = -2 \sum_{i=1}^{N} log_e \pi_i$, where π_i is the p-value of a test in cross-section i, is distributed as χ^2 with degrees of freedom $2N$ under the null hypothesis (See Im *et al.* (2003) and Maddala and Wu (1999) for more details).

4.1 Average Weighted Symmetric Test (AWS)

The average weighted symmetric (AWS) procedure is similar to the McCoskey and Kao (1998) test, but instead of using the ADF equation in each cross-section, we use the weighted symmetric estimation for each cross-section. This test is based on the result in proposition 3.5 that the WS t-statistics for testing cointegration in single time series will converge to a function of standard Wiener processes with no nuisance parameter. The WS panel statistics can be constructed as: run the following regression for each member i of the panel

$$y_{it} = \alpha_i + \beta x_{it} t + u_{it}, t = 1, 2, ..., T \tag{36}$$

and get the residual $\hat{u}_{it}$. The estimation equation is

$$\hat{u}_{it} = \rho \hat{u}_{it-1} + e_{it}, t = 1, 2, ..., T \tag{37}$$

Obtain the WS estimator of (37) and then compute the WS t-statistics t_{iWS} for (37) in each cross-section i. Finally, compute the WS panel statistic

$$\bar{t}_{WS} = \frac{1}{N}\sum_{i=1}^{N} t_{iWS}$$

Define $E[t_{iWS}] = \mu_{WS}$, and $Var[t_{iWS}] = \sigma^2_{WS}$. Then the central limit theorem can be applied to give:

$$\sqrt{N}\frac{(\bar{t}_{WS} - \mu_{WS})}{\sigma_{WS}} \xrightarrow{L} N(0,1)$$

.

In the case that $\{e_{it}\}$ are correlated, instead of using (37), we can use the augmented equation to account for the correlation between $\{e_t\}$

$$\hat{u}_{it} = \rho\hat{u}_{it-1} + \sum_{j=1}^{p} \phi_j \Delta\hat{u}_{it-j} + e_{itp} \tag{38}$$

μ_{WS} and σ^2_{WS} values are found through a simulation as shown in table 1.

4.2 Fisher Augmented Dickey-Fuller Test (FADF)

The FADF test procedures are similar to the Maddala-Wu Fisher test for panel unit roots. This test is based on the results from Phillips and Ouliaris (1990) that the ADF t-statistics in testing cointegration converge to a function of standard Wiener processes. Run the following regression for each member i of the panel

$$y_{it} = \alpha_i + \beta x_{it} t + u_{it}, t = 1, 2, ..., T \tag{39}$$

and get the residual $\hat{u}_{it}$. The estimation equation is

$$\hat{u}_{it} = \rho \hat{u}_{it-1} + \sum_{j=1}^{p} \phi_j \Delta \hat{u}_{it-j} + e_{itp} \tag{40}$$

The FADF test requires deriving the distribution of the Dickey-Fuller t-statistic, for which the simulated values were generated for different T and p. Then the p-values π_{iadf} for each ADF t-statistic could be derived. Consequently, the panel FADF statistic P_{FADF} is calculated as: $P_{FADF} = -2\sum_{i=1}^{N} log_e \pi_{iadf} \sim \chi^2_{2N}$. For the critical values of the FADF statistic we use the χ^2 table.

4.3 Fisher Weighted Symmetric Test (FWS)

The FWS test is the same as the FADF test except here we use the WS estimation procedure for each cross section rather than the ADF estimation procedure. The FWS test is based on the proposition 3.5 that the WS t-statistics for testing cointegration in single time series will converge to a function of standard Wiener processes with no nuisance parameters. The FWS test also requires deriving the distribution of the WS t-statistic, for which simulations were generated and the p-values π_{iws} for each WS t-statistic could be computed. The panel FWS statistic P_{FWS} is calculated as: $P_{FWS} = -2\sum_{i=1}^{N} log_e \pi_{iws} \sim \chi^2_{2N}$. For the critical values of FWS statistics P_{FWS} for each N, we use the χ^2 table. The advantage of the Fisher test is that it does not require a balanced panel as the average test does.

5 Monte-Carlo Investigation

5.1 Test statistics

Pedroni-ρ

$$\tilde{Z}_{\hat{\rho}_{NT}-1} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T}(\hat{u}_{it-1}\Delta\hat{u}_{it} - \hat{\lambda}_i)}{\sum_{t=1}^{T}\hat{u}_{it-1}^2}$$

Pedroni-t_ρ

$$\tilde{Z}_{t_{NT}} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T}(\hat{u}_{it-1}\Delta\hat{u}_{it} - \hat{\lambda}_i)}{\left[\sum_{t=1}^{T}\frac{1}{\hat{L}_{11i}^2}\hat{u}_{it-1}^2\right]^{1/2}}$$

McCoskey & Kao

$$\bar{t}_{ADF} = \frac{1}{N}\sum_{i=1}^{N} t_{iADF}$$

$$t_{ADF} = \sqrt{N}\frac{(\bar{t}_{ADF} - \mu_{ADF})}{\sigma_{ADF}}$$

AWS

$$\bar{t}_{WS} = \frac{1}{N}\sum_{i=1}^{N} t_{iWS}$$

$$t_{WS} = \sqrt{N}\frac{(\bar{t}_{WS} - \mu_{WS})}{\sigma_{WS}}$$

FADF

$$P_{FADF} = -2\sum_{i=1}^{N} log_e \pi_{iadf} \sim \chi^2_{2N}$$

FWS

$$P_{FWS} = -2\sum_{i=1}^{N} log_e \pi_{iws} \sim \chi^2_{2N}$$

5.2 Data Generation Processes

The DGP for all six tests based on the null hypothesis of no cointegration is as follows:

$$y_{it} = \alpha_i + \beta_i x_{it} + u_{it} \tag{41}$$

$$(i = 1, \ldots, N, t = 1, \ldots, T)$$

and

$$u_{it} = \rho u_{it-1} + v_{it}$$

It is also assumed that

$$x_{it} = x_{it-1} + \epsilon_{it}$$

where $\epsilon_{it} \sim N(0, \sigma_i^2)$, so we have that y_{it} and x_{it} are random walks, so they are cointegrated series if $|\rho| < 1$ but are not cointegrated, implying (41) is a spurious regression, if $\rho = 1$. The size of the tests are investigated under the null hypothesis $\rho = 1$. For studying the power of tests we set $\rho = 0.9$. The autocorrelation in v_{it} takes the form of moving average component as

$$v_{it} = v_{it}^* + \theta_i v_{it-1}^*$$

where $v_{it}^* \sim N(0, 1)$ and

$$\begin{bmatrix} v_{it}^* \\ \epsilon_{it} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \delta\sigma_i \\ \delta\sigma_i & \sigma_i^2 \end{bmatrix} \right)$$

α_i, β_i and σ_i are generated using the uniform distribution as: $\alpha_i \sim U[0, 10]$, $\beta_i \sim U[0, 2]$ and $\sigma_i \sim U[0.5, 1.5]$. α_i, β_i and σ_i are generated once and fixed in all replications. The choice of N and T for the experiment is : $N \in \{5, 10, 25, 50\}$ and $T \in \{10, 25, 50, 100\}$.

We examine four groups of DGPs by controlling the values of δ and θ_i

(1)-There is no endogeneity between x_{it} and u_{it} and no autocorrelation in v_{it}: $\delta = 0$ and $\theta_i = 0$

(2)-There is endogeneity between x_{it} and u_{it} but no autocorrelation in v_{it}: $\delta = 0.5$ and $\theta_i = 0$

(3)-There is no endogeneity between x_{it} and u_{it} but autocorrelation in v_{it}: $\delta = 0$ and $\theta_i \sim U[-0.4, 0.4]$

(4)-There is both endogeneity between x_{it} and u_{it} and autocorrelation in v_{it}: $\delta = 0.5$ and $\theta_i \sim U[-0.4, 0.4]$

To reduce the effect of initial conditions, *T+50* observations are generated and the first *50* observations are eliminated, using only last T observations. The number of replications is set to 3,000 for computing the empirical size and power of all five tests. The size and power of tests are computed at the 5% nominal level.

The data generating processes here is adopted similarly to the DGP used in the McCoskey and Kao (1998) paper.

5.3 Size and Power of Tests

The results of the simulation experiments are reported in Tables 2-5. In all cases, 3000 trials are used to examine the size and power properties. Table 2[1] presents the first set of experiments when there is no endogeneity between x_{it} and u_{it} and also no autocorrelation in v_{it}. The critical value of the 5% significance level is used to calculate the size and power of the tests. The Pedroni-t test has a strong size distortion, growing large when T or N increases. This size distortion makes the test impractical with large probability of a type I error. The Pedroni-ρ test has a little size distortion also, the size of this test is slightly under the significance level of 5% when T is small. It becomes larger when T increases, so at

[1] All simulations are performed using Matlab 7.0 on a 3.44 GHz, 2GB Ram PC. The programs are available upon requested.

T=100, the size is around 10%. It is noted that the size does not change when N increases. The size of the McCoskey and Kao test is under 5% also, and it is smaller when both N and T are larger, and it decreases faster when N increases compared to when T increases. Although the Pedroni-ρ and McCoskey and Kao tests are undersized, they are still acceptable, as it would lower the chance one could commit a type I error in testing a hypothesis. For all three proposed tests, AWS, FADF and FWS, the size is good, mostly lying between 4% and 6%. In terms of the power of tests, all five tests have good power when T is large. The power increases when N increases, but the speed of the increase is slower compared to the case when T increases. Overall the two tests AWS and FWS have more power which dominates those of other four tests, especially when T is less than 50. For instance, when N=25 and T=25, the powers of AWS and FWS are 50% and 35% compared to the power of other four tests. (14% (Pedroni-t), 10% (Pedroni-ρ), 6% (McCoskey & Kao) and 17% (FADF)). When both N and T are large, AWS and FWS are still the most powerful tests. In the first set of experiments all three tests AWS, FADF, FWS have good size and the tests AWS and FWS have the highest power. The FADF test does not have as good power as the AWS and FWS tests when N is small, but when N is larger than 25 its power is also good. In this case the Pedroni-t test should not be used in practice because of serious size distortion. In both size and power the performances of AWS and FWS dominate the other tests. The AWS test has slightly more power than the FWS, so in this case AWS is the best choice. However, it should be noted that AWS is only for balanced panels, so in the case when we have an unbalanced panel, the FWS would be of first choice.

Table 3 reports the size and power of tests when there is endogeneity between x_{it} and u_{it} but there is no autocorrelation in v_{it}. The Pedroni-t test has size distortion, most of them are close to zero, unlike the environment of table 2. Here the correction for endogeneity of the t-statistics of the Pedroni-t test through the term L_{11i}^{-2} helps to reduce the size if endogeneity really exists. The fact that the size is almost zero is acceptable in practice because it will help to reduce the chance of type I error. The size of the Pedroni-ρ test tends to under 5% when T is small and become a little over 5% when T is larger than 50, and when N=50, T=100, it gets to a maximum of 13%. The McCoskey & Kao test's size is a bit under the nominal size of 5%, and it does not change much when N or T changes. All three new tests, AWS, FADF, FWS, have good size, all falling between 4% to 6%. The power of the Pedroni-t test is low when T and N are small, especially when T is less than 50, and the power decreases even when N increases. When T=100, the power increases as N increase. The cause of the low power of the Pedroni-t test when T is small could be from adding the correction term L_{11i}^{-2}. The power of the Pedroni-ρ test is better: it is powerful when N≥ 25 and T≥ 50 but it is poor when N< 25 or T< 50. Again, the two test AWS and FWS are the most powerful tests. The AWS is a little more powerful than the FWS, but they both dominate other four tests in terms of power. When both N and T are large, the power of those two tests approaches 100% quickly, where the speed of increase is the same with respect to T or to N. The McCoskey and Kao test has good power when T=100, but when T≤ 50 this test is not a powerful one. Compared to table 2, we see that the effect of endogeneity is strong on the Pedroni tests, but it does not affect the AWS and the FWS tests very much, these two tests still perform well under endogeneous data. It also should be mentioned here that the AWS and FWS tests do not have any procedure that correct for the endogeneity but they both have the best size and are the most powerful tests compare with the rest.

Table 4 summarizes the size and power of tests in the case when there is autocorrelation in v_{it} but no endogeneity between x_{it} and u_{it}. We see here again, as in table 2, the Pedroni-t test has serious size distortion especially when T is large. The reason for this, as we observe in table 2, may be the over-correction for endogeneity when it does not exist. This makes the Pedroni-t test unusable in practice. The Pedroni-ρ test has better size, but it is still strongly distorted upward, with the size is getting worse for large T. At N=25, the size seems to be good, but for N=5,10,50, the size is bad. This may be because the difference in magnitude of the correlation coefficient θ_i that is randomly set as $\theta_i \sim [-0.4, 0.4]$. Apparently when the correlation coefficient θ_i in the moving average term is large, the correction for this in the Pedroni-ρ test through the term $\hat{\lambda}_i$ does not work well. The McCoskey and Kao test has a small (under-size) distortion, which is acceptable as it reduces the chance to get type I error in testing a hypothesis. For all three of the proposed tests, AWS, FWS, FADF, the size is good, lying close to the nominal size of 5% for every value of N and T. This simulation evidence confirms that the limiting distributions of all these three tests are approximated well by the standard normal and Chi-squared distributions, based on the fact that they exhibit good and stable size under different panel dimensions. In terms of power the Pedroni-t test perform well but because it has serious size distortion its strong power in this case is not useful. The Pedroni-ρ test is less powerful when T$\leq$ 25 and has good power when T$\geq$ 50, but when T$\geq$ 50 its size is small. In general, the Pedroni-ρ test in this case is unreliable, requiring a trade off between good power and bad size or good size and bad power. The tests with the most power are still the AWS and FWS, while the third best is FADF which is a little more powerful than the McCoskey and Kao test. Compared to table 2 and table 3, they are even more powerful in the presence of moving average correlation in v_{it}. So in this case, both Pedroni tests are unreliable, while the McCoskey and Kao test has an acceptable size but low power when T$\leq$ 50. The AWS and FWS tests turn out to be the best in this experiment, with the third choice being the FADF test.

The last set of experiments is shown in table 5, where in the DGP, both endogeneity and moving average autocorrelation are allowed. We see the size of the Pedroni-ρ test is worse, compared with table 3, and we see that combining both endogeneity and moving average make the Pedroni-ρ test perform badly in size, but if there is only endogeneity then this test performs satisfatory. The size of the Pedroni-t test is much better compared with the table 4, as we discussed before, so when there is endogeneity between x_{it} and u_{it} the Pedroni-t test will have acceptable size. The McCoskey and Kao test is a little under-sized. The three tests, AWS, FWS, FADF, consistently have good size as they do in the three previous tables. For power, the two Pedroni-t and McCoskey and Kao tests have low power when T$\leq$ 50 and the Pedroni-ρ test has good power when T$\geq$ 50, but when T$\geq$ 50 this test has serious size distortion. The AWS and FWS are the most powerful tests in all cases of N and T. The FADF test has less power compared to AWS and FWS tests but is more powerful than the McCoskey and Kao test.

Overall, for all four sets of experiments, the AWS, FWS, FADF tests consistently have good size, close to the 5% nominal size. The AWS and FWS are also consistently the most powerful tests in all data environments. They dominate the four tests in every aspect of performance. The AWS has a bit more power than the FWS, but the AWS test only works with a balanced panel while the FWS can be applied to any kind of panel, balanced or unbalanced.

6 Conclusion

In this paper, we propose three new tests for cointegration in heterogeneous panels, the AWS, FWS and FADF tests. The limiting distributions of weighted symmetric statistics in testing cointegration in a single set of time series are derived, and it is shown that the weighted symmetric limiting distribution is a functional of standard Wiener processes and free of nuisance parameters. The AWS and FWS are based on this result. The Monte-Carlo investigation is also conducted with different data generation environments and we found that, in every panel dimension and data characteristic, the AWS and FWS tests are the most powerful, strongly dominating the other four tests in terms of power . All three new tests also have appropriate size in all DGPs. In the previous literature, the Pedroni tests and the McCoskey and Kao tests are the only tests available for testing cointegration in heterogeneous panel. This Monte-Carlo study shows that the AWS and FWS usually perform much better in terms of size and power compared to the Pedroni tests and the McCoskey and Kao test, in every combination of endogeneity or moving average autocorrelation considered in this study. We propose that in practice, if the data is a balanced panel, the AWS test should be used and if the panel is unbalanced the FWS should be the first choice. The programs for these two tests are available from the author and make it easy to use in empirical work. An important extension for future study would be the effect of cross-section correlation on testing cointegration of these tests in panels.

REFERENCES

Fuller W.A (1996) *"Introduction to Statistical Time Series"*, second edition. John Wiley & Sons, Inc. New York

Im K.S, Pesaran M.H, and Shin Y (2003) *"Testing for Unit Roots in Heterogeneous Panels"*, Journal of Econometrics 115 (revise version of 1997's work), 53-74

Hamilton J.D. (1995) *" Time Series Analysis"*, Priceton University Press.

Kao Chihwa (1999) *"Spurious Regression and Residual-based Tests for Cointegration in Panel Data"*, Cambridge University Press

Maddala G.S and Shaowen Wu (1999)*"A Comparative Study of Unit Root Tests with Panel data and new Simple Test"*, Oxford Bullertin of Economics and Statistics, Speccial issue, 631-652

McCoskey S. and Kao C. (1998) *"A Monte Carlo Comparison of Tests for Cointegration in Panel Data"*, working paper, Syracuse University and United States Naval Academy.

Nam T. Hoang and Robert F. McNown (2006) *"Panel Data Unit Roots Tests Using Various Estimation Methods"*, working paper, Department of Economics, University of Colorado at Boulder.

Pedroni P. (1997) *"Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests, with an Application to the PPP Hypothesis: New results"*, working paper, Indiana University

Pantula S.G, Gonzalez-Farias G and Fuller W.A (1994) *"A comparison of unit-root test criteria"*, Journal of Business and Economic Statistics 12, 449-459

Park H.J and Fuller W.A (1995) *"Alternative Estimators and Unit Root Test for the Autoregressive Process"*, Journal of Time Series Analysis 16, 415-429

P.C.B Phillips (1986) *"Understanding Spurious Regressions in Econometrics"*, Journal of Econometrics 33, 311-340.

P.C.B Phillips and S Ouliaris (1990) *"Asymptotic Property of Residual Based Tests for Cointegration"*, Econometrica 58, 165-193.

TABLE 1: Moments of t-statistics of the Weighted Symmetric Estimator in Testing Cointegration in a Single Equation

p	T	MEAN	VARIANCE
2	7	-1.1343	1.1680
2	8	-1.1953	1.1649
2	9	-1.2412	1.1263
2	10	-1.2802	1.1084
2	15	-1.4039	1.0750
2	20	-1.4700	1.0439
2	25	-1.5117	1.0154
2	30	-1.5494	0.9866
3	40	-1.5513	0.9713
3	50	-1.5784	0.9515
3	60	-1.6083	0.9398
3	70	-1.6202	0.9231
3	80	-1.6392	0.9095
3	90	-1.6454	0.9034
4	100	-1.6137	0.8966
4	150	-1.6510	0.8855
4	200	-1.6657	0.8761

NOTES: means and variances in this table are computed by stochastic simulations with 100,000 replications. The underlying data is generated by:

$$y_t = \alpha + \beta x_t + u_t$$

$$(t = 1, \ldots, T)$$

and

$$x_t = x_{t-1} + \epsilon_t$$

$$u_t = u_{t-1} + v_t$$

where

$$v_t \sim N(0, 1)$$

$$\epsilon_t \sim N(0, \sigma^2)$$

α, β and σ are generated using the uniform distribution as: $\alpha_i \sim U[0, 10]$, $\beta \sim U[0, 2]$ and $\sigma \sim U[0.5, 1.5]$. *T+50* observations are generated and the first *50* observations are eliminated, using only the last T observations. p is decided by $p = Int[c(T/100)^{1/d}]$ with c=4, d=4.

TABLE 2: Size and Power of Tests for Cointegration in the Panels with Heterogeneity of the Intercepts and Slopes: ($\delta = 0$; $\theta_i = 0$)

N	T	PEDRONI-t_ρ	PEDRONI-ρ	MCCOSKEY & KAO-t_{ADF}	AVERAGE-t_{WS}	FISHER-t_{ADF}	FISHER-t_{WS}
		SIZE					
5	10	0.000	0.000	0.032	0.051	0.050	0.047
	25	0.058	0.012	0.028	0.049	0.053	0.054
	50	0.216	0.044	0.032	0.045	0.055	0.049
	100	0.315	0.072	0.025	0.042	0.045	0.045
10	10	0.000	0.000	0.013	0.048	0.040	0.045
	25	0.066	0.010	0.016	0.047	0.042	0.050
	50	0.268	0.044	0.021	0.044	0.043	0.038
	100	0.463	0.093	0.025	0.053	0.052	0.050
25	10	0.000	0.000	0.003	0.052	0.047	0.049
	25	0.039	0.006	0.008	0.055	0.052	0.062
	50	0.445	0.052	0.010	0.057	0.044	0.052
	100	0.709	0.102	0.020	0.059	0.055	0.052
50	10	0.000	0.000	0.000	0.055	0.047	0.052
	25	0.024	0.003	0.002	0.052	0.043	0.054
	50	0.640	0.056	0.005	0.052	0.040	0.047
	100	0.913	0.131	0.007	0.062	0.045	0.053
		POWER					
5	10	0.001	0.000	0.030	0.077	0.052	0.067
	25	0.093	0.042	0.062	0.147	0.092	0.132
	50	0.425	0.316	0.151	0.353	0.173	0.296
	100	0.884	0.913	0.545	0.826	0.534	0.739
10	10	0.000	0.000	0.019	0.098	0.055	0.078
	25	0.118	0.063	0.058	0.246	0.109	0.193
	50	0.655	0.552	0.252	0.630	0.284	0.477
	100	0.992	0.998	0.848	0.988	0.828	0.957
25	10	0.000	0.000	0.008	0.136	0.061	0.108
	25	0.137	0.102	0.064	0.501	0.165	0.346
	50	0.935	0.914	0.487	0.954	0.534	0.832
	100	1.000	1.000	0.997	1.000	0.996	1.000
50	10	0.000	0.000	0.002	0.215	0.071	0.151
	25	0.184	0.173	0.069	0.775	0.273	0.573
	50	0.997	0.998	0.788	1.000	0.822	0.985
	100	1.000	1.000	1.000	1.000	1.000	1.000

NOTES: Data Generating Processes for all six tests based on the null hypothesis of no cointegration as following: $y_{it} = \alpha_i + \beta_i x_{it} + u_{it}$ $(i = 1, \ldots, N, t = 1, \ldots, T)$ and $x_{it} = x_{it-1} + \epsilon_{it}$; $u_{it} = \rho u_{it-1} + v_{it}$; $v_{it} = v^*_{it} + \theta_i v^*_{it-1}$ with:

$$\begin{bmatrix} v^*_{it} \\ \epsilon_{it} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \delta\sigma_i \\ \delta\sigma_i & \sigma_i^2 \end{bmatrix} \right)$$

where $v^*_{it} \sim N(0,1)$, $\epsilon_{it} \sim N(0, \sigma_i^2)$. α_i, β_i and σ_i are generated using the uniform distribution as: $\alpha_i \sim U[0, 10]$, $\beta_i \sim U[0, 2]$ and $\sigma_i \sim U[0.5, 1.5]$. α_i, β_i and σ_i are generated once and fixed in all replications. $\rho = 1$ for size and $\rho = 0.9$ for power. *T+50* observations are generated and the first *50* observations are eliminated, using only last T observations. Number of replications is set to 3,000 for computing the empirical size and power of all five tests. The size and power of tests are computed at the 5% nominal level.

$\delta = 0$ and $\theta_i = 0$ is set in this table, so there is no endogeneity between x_{it} and u_{it} and no serial correlation in v_{it}.

TABLE 3: Size and Power of Tests for Cointegration in the Panels with Heterogeneity of the Intercepts and Slopes: ($\delta = 0.5$; $\theta_i = 0$)

N	T	PEDRONI-t_ρ	PEDRONI-ρ	MCCOSKEY & KAO-t_{ADF}	AVERAGE-t_{WS}	FISHER-t_{ADF}	FISHER-t_{WS}
				SIZE			
5	10	0.000	0.000	0.031	0.054	0.051	0.053
	25	0.002	0.009	0.028	0.046	0.053	0.051
	50	0.009	0.050	0.029	0.042	0.044	0.052
	100	0.021	0.073	0.027	0.043	0.047	0.046
10	10	0.000	0.000	0.020	0.053	0.055	0.048
	25	0.000	0.012	0.019	0.042	0.050	0.038
	50	0.003	0.051	0.024	0.053	0.050	0.058
	100	0.013	0.075	0.023	0.046	0.047	0.045
25	10	0.000	0.000	0.004	0.059	0.048	0.052
	25	0.000	0.006	0.005	0.050	0.049	0.060
	50	0.001	0.055	0.015	0.050	0.052	0.051
	100	0.004	0.095	0.015	0.051	0.044	0.050
50	10	0.000	0.000	0.000	0.056	0.046	0.053
	25	0.000	0.002	0.002	0.047	0.049	0.055
	50	0.000	0.058	0.008	0.057	0.052	0.054
	100	0.001	0.126	0.015	0.054	0.056	0.051
				POWER			
5	10	0.000	0.000	0.036	0.069	0.056	0.063
	25	0.004	0.035	0.049	0.121	0.075	0.109
	50	0.027	0.234	0.110	0.253	0.133	0.202
	100	0.213	0.839	0.414	0.723	0.415	0.619
10	10	0.000	0.000	0.019	0.077	0.051	0.067
	25	0.001	0.035	0.043	0.151	0.087	0.127
	50	0.017	0.390	0.153	0.451	0.190	0.323
	100	0.363	0.983	0.711	0.954	0.688	0.883
25	10	0.000	0.000	0.007	0.117	0.063	0.090
	25	0.000	0.049	0.035	0.310	0.117	0.232
	50	0.008	0.767	0.303	0.831	0.365	0.629
	100	0.702	1.000	0.985	1.000	0.970	0.998
50	10	0.000	0.000	0.001	0.137	0.062	0.098
	25	0.000	0.081	0.032	0.521	0.174	0.358
	50	0.002	0.967	0.510	0.986	0.609	0.892
	100	0.944	1.000	1.000	1.000	0.999	1.000

NOTES: Data Generating Processes for all six tests based on the null hypothesis of no cointegration as following: $y_{it} = \alpha_i + \beta_i x_{it} + u_{it}$ $(i = 1, \ldots, N, t = 1, \ldots, T)$ and $x_{it} = x_{it-1} + \epsilon_{it}$; $u_{it} = \rho u_{it-1} + v_{it}$; $v_{it} = v^*_{it} + \theta_i v^*_{it-1}$ with:

$$\begin{bmatrix} v^*_{it} \\ \epsilon_{it} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \delta\sigma_i \\ \delta\sigma_i & \sigma_i^2 \end{bmatrix} \right)$$

where $v^*_{it} \sim N(0, 1)$, $\epsilon_{it} \sim N(0, \sigma_i^2)$. α_i, β_i and σ_i are generated using the uniform distribution as: $\alpha_i \sim U[0, 10]$, $\beta_i \sim U[0, 2]$ and $\sigma_i \sim U[0.5, 1.5]$. α_i, β_i and σ_i are generated once and fixed in all replications. $\rho = 1$ for size and $\rho = 0.9$ for power. *T+50* observations are generated and the first *50* observations are eliminated, using only last T observations. Number of replications is set to 3,000 for computing the empirical size and power of all five tests. The size and power of tests are computed at the 5% nominal level.
$\delta = 0.5$ and $\theta_i = 0$ is set in this table, so there is endogeneity between x_{it} and u_{it} but no serial correlation in v_{it}.

TABLE 4: Size and Power of Tests for Cointegration in the Panels with Heterogeneity of the Intercepts and Slopes: ($\delta = 0$; $\theta_i \sim U[-0.4, 0.4]$)

N	T	PEDRONI-t_ρ	PEDRONI-ρ	MCCOSKEY & KAO-t_{ADF}	AVERAGE-t_{WS}	FISHER-t_{ADF}	FISHER-t_{WS}
				SIZE			
5	10	0.001	0.000	0.035	0.057	0.054	0.053
	25	0.090	0.077	0.029	0.055	0.057	0.066
	50	0.342	0.053	0.026	0.043	0.047	0.048
	100	0.625	0.176	0.029	0.047	0.054	0.054
10	10	0.000	0.000	0.016	0.054	0.051	0.049
	25	0.154	0.025	0.020	0.059	0.056	0.059
	50	0.435	0.153	0.022	0.052	0.048	0.055
	100	0.802	0.273	0.024	0.049	0.055	0.052
25	10	0.000	0.000	0.004	0.067	0.050	0.056
	25	0.264	0.017	0.011	0.052	0.055	0.055
	50	0.848	0.070	0.013	0.049	0.044	0.048
	100	0.973	0.075	0.020	0.053	0.051	0.049
50	10	0.000	0.000	0.001	0.048	0.040	0.047
	25	0.376	0.030	0.003	0.060	0.056	0.068
	50	0.963	0.369	0.010	0.059	0.059	0.056
	100	0.995	0.662	0.006	0.056	0.049	0.053
				POWER			
5	10	0.002	0.000	0.037	0.104	0.060	0.095
	25	0.238	0.007	0.049	0.139	0.073	0.125
	50	0.785	0.734	0.170	0.382	0.201	0.322
	100	0.982	0.997	0.557	0.852	0.554	0.762
10	10	0.000	0.000	0.025	0.119	0.066	0.105
	25	0.307	0.035	0.053	0.240	0.101	0.187
	50	0.897	0.723	0.253	0.641	0.287	0.492
	100	1.000	1.000	0.861	0.990	0.828	0.952
25	10	0.000	0.000	0.008	0.161	0.071	0.119
	25	0.461	0.035	0.060	0.453	0.155	0.314
	50	0.996	0.993	0.525	0.970	0.567	0.861
	100	1.000	1.000	0.998	1.000	0.996	1.000
50	10	0.000	0.000	0.002	0.218	0.078	0.157
	25	0.623	0.536	0.084	0.810	0.297	0.616
	50	1.000	1.000	0.809	0.999	0.836	0.989
	100	1.000	1.000	1.000	1.000	1.000	1.000

NOTES: Data Generating Processes for all six tests based on the null hypothesis of no cointegration as following: $y_{it} = \alpha_i + \beta_i x_{it} + u_{it}$ ($i = 1, \ldots, N, t = 1, \ldots, T$) and $x_{it} = x_{it-1} + \epsilon_{it}$; $u_{it} = \rho u_{it-1} + v_{it}$; $v_{it} = v_{it}^* + \theta_i v_{it-1}^*$ with:

$$\begin{bmatrix} v_{it}^* \\ \epsilon_{it} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \delta\sigma_i \\ \delta\sigma_i & \sigma_i^2 \end{bmatrix} \right)$$

where $v_{it}^* \sim N(0,1)$, $\epsilon_{it} \sim N(0, \sigma_i^2)$. α_i, β_i and σ_i are generated using the uniform distribution as: $\alpha_i \sim U[0, 10]$, $\beta_i \sim U[0, 2]$ and $\sigma_i \sim U[0.5, 1.5]$. α_i, β_i and σ_i are generated once and fixed in all replications. $\rho = 1$ for size and $\rho = 0.9$ for power. *T+50* observations are generated and the first *50* observations are eliminated, using only last T observations. Number of replications is set to 3,000 for computing the empirical size and power of all five tests. The size and power of tests are computed at the 5% nominal level.
$\delta = 0$ and $\theta_i = 0 \sim U[-0.4, 0.4]$ is set in this table, so there is no endogeneity between x_{it} and u_{it} but there is serial correlation in v_{it}.

TABLE 5: Size and Power of Tests for Cointegration in the Panels with Heterogeneity of the Intercepts and Slopes: ($\delta = 0.5$; $\theta_i \sim U[-0.4, 0.4]$)

N	T	PEDRONI-t_ρ	PEDRONI-ρ	MCCOSKEY & KAO-t_{ADF}	AVERAGE-t_{WS}	FISHER-t_{ADF}	FISHER-t_{WS}
		SIZE					
5	10	0.000	0.000	0.030	0.041	0.051	0.038
	25	0.027	0.082	0.033	0.058	0.057	0.064
	50	0.099	0.053	0.031	0.049	0.049	0.051
	100	0.048	0.354	0.031	0.047	0.050	0.050
10	10	0.000	0.000	0.017	0.050	0.052	0.043
	25	0.000	0.038	0.018	0.049	0.045	0.051
	50	0.039	0.532	0.029	0.055	0.057	0.059
	100	0.081	0.497	0.023	0.048	0.046	0.049
25	10	0.000	0.000	0.004	0.066	0.050	0.057
	25	0.000	0.069	0.010	0.063	0.059	0.062
	50	0.037	0.394	0.015	0.066	0.049	0.062
	100	0.348	0.306	0.017	0.054	0.052	0.052
50	10	0.000	0.000	0.000	0.057	0.052	0.050
	25	0.000	0.030	0.003	0.067	0.060	0.068
	50	0.033	0.368	0.005	0.052	0.043	0.051
	100	0.062	0.630	0.010	0.053	0.047	0.048
		POWER					
5	10	0.000	0.000	0.040	0.064	0.060	0.057
	25	0.040	0.155	0.044	0.121	0.074	0.112
	50	0.143	0.232	0.108	0.252	0.133	0.212
	100	0.695	0.994	0.433	0.742	0.438	0.646
10	10	0.000	0.000	0.023	0.083	0.051	0.068
	25	0.001	0.119	0.038	0.180	0.084	0.143
	50	0.369	0.974	0.193	0.523	0.237	0.390
	100	0.964	1.000	0.752	0.968	0.716	0.901
25	10	0.000	0.000	0.006	0.120	0.063	0.091
	25	0.000	0.312	0.040	0.356	0.135	0.248
	50	0.382	0.990	0.327	0.864	0.402	0.688
	100	0.999	1.000	0.980	1.000	0.968	0.998
50	10	0.000	0.000	0.001	0.153	0.064	0.104
	25	0.000	0.253	0.035	0.568	0.187	0.382
	50	0.535	0.999	0.541	0.984	0.630	0.909
	100	1.000	1.000	1.000	1.000	1.000	1.000

NOTES: Data Generating Processes for all six tests based on the null hypothesis of no cointegration as following: $y_{it} = \alpha_i + \beta_i x_{it} + u_{it}$ $(i = 1, \ldots, N, t = 1, \ldots, T)$ and $x_{it} = x_{it-1} + \epsilon_{it}$; $u_{it} = \rho u_{it-1} + v_{it}$; $v_{it} = v_{it}^* + \theta_i v_{it-1}^*$ with:

$$\begin{bmatrix} v_{it}^* \\ \epsilon_{it} \end{bmatrix} \sim \text{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \delta\sigma_i \\ \delta\sigma_i & \sigma_i^2 \end{bmatrix} \right)$$

where $v_{it}^* \sim N(0,1)$, $\epsilon_{it} \sim N(0, \sigma_i^2)$. α_i, β_i and σ_i are generated using the uniform distribution as: $\alpha_i \sim U[0,10]$, $\beta_i \sim U[0,2]$ and $\sigma_i \sim U[0.5, 1.5]$. α_i, β_i and σ_i are generated once and fixed in all replications. $\rho = 1$ for size and $\rho = 0.9$ for power. *T+50* observations are generated and the first *50* observations are eliminated, using only last T observations. Number of replications is set to 3,000 for computing the empirical size and power of all five tests. The size and power of tests are computed at the 5% nominal level.

$\delta = 0.5$ and $\theta_i = 0 \sim U[-0.4, 0.4]$ is set in this table, so there is endogeneity between x_{it} and u_{it} and serial correlation in v_{it}.

APPENDIX

Proof of Proposition 3.1:

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} = e_t$$

Multiply both side by $X_{t-h} \Longleftrightarrow$

$$X_{t-h}X_t + \alpha_1 X_{t-h}X_{t-1} + ... + \alpha_p X_{t-h}X_{t-p} = X_{t-h}e_t$$

$\Longleftrightarrow$

$$E[X_{t-h}X_t] + \alpha_1 E[X_{t-h}X_{t-1}] + ... + \alpha_p E[X_{t-h}X_{t-p}] = E[X_{t-h}e_t]$$

$\Longleftrightarrow$

$$\gamma(h) + \alpha_1\gamma(h-1) + ... + \alpha_p\gamma(h-p) = E[X_{t-h}e_t]$$

with $\gamma(h-p) = E[X_{t-h}X_{t-p}]$ and $h = 1, 2, ...$

From the theorem 3.0 we have that X_{t-h} can be expressed as a weighted average if e_{t-h} and previous $e's$ then X_{t-h} and e_t are uncorrelated $\longrightarrow Cov(X_{t-h}, e_t) = E[X_{t-h}e_t] = 0$ with $h = 1, 2, ...$ If $h = 0$ then

$$X_t = e_t + w_1 e_{t-1} + w_2 e_{t-2} + ... + ...$$

$\longrightarrow$

$$E[X_t e_t] = E[e_t^2] + w_1 E[e_t e_{t-1}] + w_2 E[e_t e_{t-2}] + ... + ... = E[e_t^2] = \sigma^2$$

So we have

$$\gamma(h) + \alpha_1\gamma(h-1) + ... + \alpha_p\gamma(h-p) = \begin{cases} 0 & h = 1, 2, ... \\ \sigma^2 & h = 0 \end{cases} \tag{42}$$

note that, from the theorem 3.0, $\{X_t\}$ is a covariance stationary series, then:

$$\gamma(h) = E[X_{t-h}X_t] = E[X_{t+h}X_t] = \gamma(-h)$$

and

$$E[X_{t-h}X_t] = E[X_{t+h}X_t]$$
$$E[X_{t-h}X_{t-1}] = E[X_{t+h}X_{t+1}]$$
$$\vdots$$
$$E[X_{t-h}X_{t-p}] = E[X_{t+h}X_{t+p}]$$

adding up and combine to (42)

$$E[X_{t+h}(X_t + \alpha_1 X_{t+1} + \alpha_2 X_{t+2} + ... + \alpha_p X_{t+p})] = \begin{cases} 0 & h = 1, 2, ... \\ \sigma^2 & h = 0 \end{cases}$$

or

$$E[X_{t+h}v_t] = \begin{cases} 0 & h = 1, 2, ... \\ \sigma^2 & h = 0 \end{cases} \tag{43}$$

To prove the proposition, we need to prove: $\{v_t\}$ is an un correlated $(0, \sigma^2)$ random variables, this means: $E(v_t) = 0$; $E[v_t^2] = \sigma^2$ and $E[v_{t+j}v_t] = 0$; $j = 1, 2, ...$

$$E(v_t) = E[X_t + \alpha_1 X_{t+1} + ... + \alpha_p X_{t+p}] = E[X_t] + \alpha_1 E[X_{t+1}] + ... + \alpha_p E[X_{t+p}]) = 0$$

remember that $X_t = e_t + w_1 e_{t-1} + ... + ...$ then $E[X_t] = E[e_t] + w_1 E[e_{t-1}] + ... + ... = 0$

$$E[v_t^2] = E[(X_t + \alpha_1 X_{t+1} + ... + \alpha_p X_{t+p})v_t] = E[X_t v_t + \alpha_1 X_{t+1} v_t + ... + \alpha_p X_{t+p} v_t] =$$

$$= E[X_t v_t] + \alpha_1 E[X_{t+1} v_t] + ... + \alpha_p E[X_{t+p} v_t] = \sigma^2$$

because of (43)

$$E[v_{t+j}v_t] = E[v_t(X_{t+j} + \alpha_1 X_{t+j+1} + ... + \alpha_p X_{t+j+p})] =$$

$$= e[X_{t+j}v_t] + \alpha_1 E[X_{t+j+1}v_t] + ... + \alpha_p E[X_{t+j+p}v_t] = 0$$

with $j = 1, 2, ...$ because of (43)

Proof of Proposition 3.2:

Weighed symmetric estimator of ρ in (16) is a $\hat{\rho}$ which minimizes:

$$Q = \sum_{t=2}^{T} w_t (y_t - \rho y_{t-1})^2 + \sum_{t=1}^{T-1} (1 - w_{t+1})(y_t - \rho y_{t+1})^2$$

$\Longleftrightarrow$

$$Q = \sum_{t=2}^{T} [(\sqrt{w_t})y_t - \rho(\sqrt{w_t})y_{t-1}]^2 + \sum_{t=1}^{T-1} \left[(\sqrt{1 - w_{t+1}})y_t - \rho(\sqrt{1 - w_{t+1}})y_{t+1}\right]^2 \tag{44}$$

if we denote: $X = \begin{bmatrix} (\sqrt{w_2})y_1 \\ \vdots \\ (\sqrt{w_T})y_{T-1} \\ (\sqrt{1 - w_2})y_2 \\ \vdots \\ (\sqrt{1 - w_T})y_T \end{bmatrix}$; $Y = \begin{bmatrix} (\sqrt{w_2})y_2 \\ \vdots \\ (\sqrt{w_T})y_T \\ (\sqrt{1 - w_2})y_1 \\ \vdots \\ (\sqrt{1 - w_T})y_{T-1} \end{bmatrix}$ and estimate (44) by normal OLS method, we get

$$\hat{\rho} = [X'X]^{-1}X'Y = \frac{\sum_{t=2}^{T} w_t(y_t y_{t-1}) + \sum_{t=1}^{T-1} y_t y_{t+1} - \sum_{t=1}^{T-1} w_{t+1}(y_t y_{t+1})}{\sum_{t=2}^{T} w_t y_{t-1}^2 + \sum_{t=1}^{T-1} y_{t+1}^2 - \sum_{t=1}^{T-1} w_{t+1} y_{t+1}^2}$$

put $w_t = T^{-1}(t-1)$, then simplify the expressions, we have:

$$Numerator = T^{-1}\sum_{t=2}^{T}(t-1)y_t y_{t-1} + \sum_{t=1}^{T-1} y_t y_{t+1} - T^{-1}\sum_{t=1}^{T-1} t(y_t y_{t+1}) = \sum_{t=2}^{T} y_t y_{t-1}$$

$$Denominator = T^{-1}\sum_{t=2}^{T}(t-1)y_{t-1}^2 + \sum_{t=1}^{T-1} y_{t+1}^2 - T^{-1}\sum_{t=1}^{T-1} t(y_{t+1}^2) = \sum_{t=2}^{T} y_t^2 + \frac{1}{T}\sum_{t=1}^{T} y_t^2$$

If an intercept α is added in (16), then we can demean (16) as: $(y_t - \overline{y}) = \rho(y_{t-1} - \overline{y}) + \epsilon_t$ with $\overline{y} = T^{-1}\sum_{t=1}^{T} y_t$ and the same result of $\hat{\rho}$ is derived as in (21)

Proof of Proposition 3.3:

$(\hat{\rho} - 1)$ distribution:

$$(\hat{\rho} - 1) = \frac{\sum_{t=2}^{T} y_t y_{t-1} - \sum_{t=2}^{T-1} y_t^2 - \frac{1}{T}\sum_{t=1}^{T} y_t^2}{\sum_{t=2}^{T-1} y_t^2 + \frac{1}{T}\sum_{t=1}^{T} y_t^2}$$

Numerator:

$$\sum_{t=2}^{T} y_{t-1}(y_{t-1} + \epsilon_t) - \sum_{t=2}^{T-1} y_t^2 - \frac{1}{T}\sum_{t=1}^{T} y_t^2 =$$

$$= \sum_{t=2}^{T} y_{t-1}^2 + \sum_{t=2}^{T} y_{t-1}\epsilon_t - \sum_{t=2}^{T-1} y_t^2 - \frac{1}{T}\sum_{t=1}^{T} y_t^2 =$$

$$= \sum_{t=2}^{T} y_{t-1}\epsilon_t + y_1^2 - \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 + \frac{y_T^2}{T}$$

then

$$\frac{1}{T}(numerator) = \frac{1}{T}\sum_{t=2}^{T} y_{t-1}\epsilon_t + \frac{y_1^2}{T} - \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 - \frac{1}{T}\left(\frac{y_T}{\sqrt{T}}\right)^2 \xrightarrow{L} \frac{1}{2}\sigma^2[W^2(1) - 1] - \sigma^2\int_0^1 [W(r)]^2 dr$$

Denominator:

$$\sum_{t=2}^{T-1} y_t^2 + \frac{1}{T}\sum_{t=1}^{T} y_t^2 = \sum_{t=2}^{T} y_t^2 - y_T^2 + \frac{1}{T}\sum_{t=1}^{T} y_t^2 = (1 + \frac{1}{T})\sum_{t=1}^{T} y_t^2 - (y_1^2 + y_T^2) =$$

$$= (1 + \frac{1}{T})\sum_{t=1}^{T} y_{t-1}^2 - (y_1^2 + y_T^2) + (1 + \frac{1}{T})y_T^2 = (1 + \frac{1}{T})\sum_{t=1}^{T} y_{t-1}^2 - y_1^2 + \frac{1}{T}y_T^2$$

then

$$\frac{1}{T^2}(denominator) = (1 + \frac{1}{T})\sum_{t=1}^{T} y_{t-1}^2 - \frac{y_1^2}{T} + \frac{y_T^2}{T^2} \xrightarrow{L} \sigma^2\int_0^1 [W(r)]^2 dr$$

finally:

$$T(\hat{\rho}-1) \xrightarrow{L} \frac{\frac{1}{2}[W^2(1)-1]-\int_0^1[W(r)]^2dr}{\int_0^1[W(r)]^2dr} = \frac{\frac{1}{2}[K^2-1]-G}{G}$$

with $K = W(1)$ and $G = \int_0^1[W(r)]^2dr$

<u>$t_{\hat{\rho}}$ distribution:</u>

$$t_{\hat{\rho}} = \left[\hat{\sigma}^2\left(\sum_{t=2}^{T-1} y_t^2 + \frac{1}{T}\sum_{t=1}^{T} y_t^2\right)^{-1}\right]^{-1/2} (\hat{\rho}-1)$$

where $\hat{\sigma}^2$ is the estimation of σ^2 by: $\hat{\sigma}^2 = \frac{Q(\hat{\rho})}{T-2} \xrightarrow{p} \sigma^2$, then:

$$t_{\hat{\rho}} = \frac{T(\hat{\rho}-1)}{\left[\frac{\hat{\sigma}^2}{\frac{1}{T}\left(\sum_{t=2}^{T-1} y_t^2+\frac{1}{T}\sum_{t=1}^{T} y_t^2\right)}\right]^{-1/2}} \xrightarrow{L} \frac{\frac{1}{2}[W^2(1)-1]-\int_0^1[W(r)]^2dr}{\left(\int_0^1[W(r)]^2dr\right)^{1/2}} = \frac{\frac{1}{2}[K^2-1]-G}{G^{1/2}}$$

Proof of Proposition 3.4:

Before proving the proposition, we prove the following results:

<u>*Lemma 1:*</u> *Under the null hypothesis of* $\rho = 1$

i) $\frac{1}{T}(y_1-\overline{y})^2 \xrightarrow{L} H^2$

ii) $\frac{1}{T}(y_T-\overline{y})^2 \xrightarrow{L} H^2 - 2KH + K^2$

iii) $\overline{y}.\overline{\epsilon} \xrightarrow{L} KH$

iv) $\frac{1}{T}\sum_{t=1}^{T} y_t\epsilon_t \xrightarrow{L} \frac{1}{2}[K^2-1]+1$

v) $\frac{1}{T^2}\sum_{t=1}^{T}(y_t-\overline{y})^2 \xrightarrow{L} G - H^2$

where $\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$; $K = W(1)$; $H = \int_0^1 W(r)dr$; $G = \int_0^1[W(r)]^2dr$; and assume $\sigma = 1$; $y_0 = 0$

<u>*Proof:*</u>

i) $T^{-1}(y_1-\overline{y})^2 = T^{-1}(y_1^2-2\overline{y}y_1+(\overline{y})^2) = T^{-1}\left[y_1^2 - 2y_1(T^{-1}\sum_{t=1}^{T} y_t) + T^{-2}(\sum_{t=1}^{T} y_t)^2\right] =$

$$= T^{-1}\left[y_1^2 - 2y_1T^{-1}(\sum_{t=1}^{T} y_{t-1}+y_T) + T^{-2}(\sum_{t=1}^{T} y_{t-1}+y_T)^2\right] =$$

$$= \frac{y_1^2}{T} - \left(\frac{2y_1}{\sqrt{T}}\right)\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}\right) - \left(\frac{2y_1}{T^{3/2}}\right)\left(\frac{y_T}{\sqrt{T}}\right) + \left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}\right)^2 +$$

$$+2\left(\frac{1}{T}\right)\left(\frac{y_T}{\sqrt{T}}\right)\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{t-1}\right) + \left(\frac{1}{T}\right)\left(\frac{y_T}{\sqrt{T}}\right)^2 \xrightarrow{L} \left[\int_0^1 W(r)dr\right]^2 = H^2$$

ii) $T^{-1}(y_T-\overline{y})^2 = T^{-1}(y_T^2-2\overline{y}y_T+(\overline{y})^2) = T^{-1}\left[y_T^2 - 2y_T(T^{-1}\sum_{t=1}^T y_t) + T^{-2}(\sum_{t=1}^T y_t)^2\right] =$

$$= T^{-1}\left[y_T^2 - 2y_T T^{-1}(\sum_{t=1}^T y_{t-1} + y_T) + T^{-2}(\sum_{t=1}^T y_{t-1} + y_T)^2\right] =$$

$$= \left(\frac{y_T}{\sqrt{T}}\right)^2 - 2\left(\frac{y_T}{\sqrt{T}}\right)\left(\frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1}\right) - 2\left(\frac{1}{T}\right)\left(\frac{y_T}{\sqrt{T}}\right)^2\left(\frac{y_T}{\sqrt{T}}\right) + \left(\frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1}\right)^2 +$$

$$+2\left(\frac{1}{T}\right)\left(\frac{y_T}{\sqrt{T}}\right)\left(\frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1}\right)+\left(\frac{1}{T}\right)\left(\frac{y_T}{\sqrt{T}}\right)^2 \xrightarrow{L} \left[\int_0^1 W(r)dr\right]^2 - 2W(1)\int_0^1 W(r)dr +$$

$+[W(1)]^2 = H^2 - 2KH + K^2$

iii) $\overline{y}.\overline{\epsilon} = T^{-2}\sum_{t=1}^T y_t \sum_{t=1}^T \epsilon_t = \frac{1}{T^2}\left(\sum_{t=1}^T y_{t-1} + y_T\right)\sum_{t=1}^T \epsilon_t =$

$$= \left(\frac{1}{T^{3/2}}\sum_{t=1}^T + \frac{1}{T}\left(\frac{y_T}{\sqrt{T}}\right)\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \epsilon_t\right) \xrightarrow{L} W(1)\int_0^1 W(r)dr = KH$$

iv) $\frac{1}{T}\sum_{t=1}^T y_t\epsilon_t = \frac{1}{T}\sum_{t=1}^T (y_{t-1}+\epsilon_t)\epsilon_t = \frac{1}{T}\sum_{t=1}^T y_{t-1}\epsilon_t + \frac{1}{T}\sum_{t=1}^T \epsilon_t^2 \xrightarrow{L} \frac{1}{2}[K^2-1]+1$

v) $\frac{1}{T^2}\sum_{t=1}^T (y_t-\overline{y})^2 = \frac{1}{T^2}\left(\sum_{t=2}^T y_t^2 - 2\overline{y}\sum_{t=2}^T y_t + (T-2)\overline{y}^2\right) =$

$$= \frac{1}{T^2}\sum_{t=2}^T y_t^2 - \frac{2}{T^3}\sum_{t=1}^T y_t \sum_{t=2}^T y_t + \frac{T-2}{T^4}\left(\sum_{t=1}^T y_t\right)^2 =$$

$$= \frac{1}{T^2}\sum_{t=2}^T y_t^2 - 2\left(\frac{1}{T^{3/2}}\sum_{t=1}^T y_t\right)\left(\frac{1}{T^{3/2}}\sum_{t=2}^T y_t\right) + \left(\frac{T-2}{T}\right)\left(\frac{1}{T^{3/2}}\sum_{t=1}^T y_t\right)^2 \xrightarrow{L}$$

$$\xrightarrow{L} G - 2H.H + H^2 = G - H^2$$

<u>$(\hat{\rho}-1)$ distribution:</u>

$$(\hat{\rho}-1) = \frac{\sum_{t=2}^T (y_t-\overline{y})(y_{t-1}-\overline{y}) - \sum_{t=2}^{T-1}(y_t-\overline{y})^2 - \frac{1}{T}\sum_{t=1}^T (y_t-\overline{y})^2}{\sum_{t=2}^{T-1}(y_t-\overline{y})^2 + \frac{1}{T}\sum_{t=1}^T (y_t-\overline{y})^2}$$

because of: $\sum_{t=2}^T (y_t-\overline{y})(y_{t-1}-\overline{y}) = \sum_{t=2}^T (y_t-\overline{y})(y_t-\overline{y}-y_t+y_{t-1}) =$

$$= \sum_{t=2}^T \left[(y_t-\overline{y})^2 - (y_t-\overline{y})\epsilon_t\right]$$

so,

$$Numerator = (y_T - \overline{y})^2 - \sum_{t=2}^{T} y_t\epsilon_t + \overline{y}\sum_{t=2}^{T}\epsilon_t - \frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2$$

and

$$\frac{1}{T}(numerator) = \frac{1}{T}(y_T - \overline{y})^2 - \frac{1}{T}\sum_{t=2}^{T} y_t\epsilon_t + \overline{y}.\overline{\epsilon} - \frac{1}{T^2}\sum_{t=1}^{T}(y_t - \overline{y})^2 \xrightarrow{L} \text{follow lemma 1}$$

$$\xrightarrow{L} (H^2 - 2KH + K^2) - \left(\frac{1}{2}[K^2 - 1] + 1\right) + KH - (G - H^2) = \frac{1}{2}[K^2 - 1] - G - KH + 2H^2$$

$$\frac{1}{T^2}(denominator) = \frac{1}{T^2}\sum_{t=2}^{T-1}(y_t - \overline{y})^2 + \frac{1}{T}\left(\frac{1}{T^2}\sum_{t=1}^{T}(y_t - \overline{y})^2\right) \xrightarrow{L} \text{follow lemma 1} \xrightarrow{L} (G - H^2)$$

finally:

$$T(\hat{\rho} - 1) \xrightarrow{L} \frac{\left(\frac{1}{2}[K^2 - 1] - G - KH + 2H^2\right)}{[G - H^2]}$$

<u>$t_{\hat{\rho}}$ distribution:</u>

$$t_{\hat{\rho}} = \left[\hat{\sigma}^2\left(\sum_{t=2}^{T-1}(y_t - \overline{y})^2 + \frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2\right)^{-1}\right]^{-1/2}(\hat{\rho} - 1)$$

where $\hat{\sigma}^2$ is the estimation of σ^2 by: $\hat{\sigma}^2 = \frac{Q(\hat{\rho})}{T-2} \xrightarrow{p} \sigma^2$, then:

$$t_{\hat{\rho}} = \frac{T(\hat{\rho} - 1)}{\left[\frac{\hat{\sigma}^2}{\frac{1}{T}\left(\sum_{t=2}^{T-1}(y_t - \overline{y})^2 + \frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2\right)}\right]^{-1/2}} \xrightarrow{L} = \frac{\left(\frac{1}{2}[K^2 - 1] - G - KH + 2H^2\right)}{[G - H^2]^{1/2}}$$

Proof of proposition 3.5:
The following result is taken from Phillips (1986)
<u>*Lemma 2:*</u> *Under the null hypothesis of no cointegration* $H_0 : \rho = 1$ *in (31)*
i)

$$\hat{\beta} \xrightarrow{L} \frac{\int_0^1 V(r)W(r)dr - \int_0^1 V(r)dr\int_0^1 W(r)dr}{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)dr\right]^2} = \zeta$$

ii)

$$\frac{1}{T}\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2 = \frac{1}{T}\sum_{t=2}^{T}(v_t - \hat{\beta}\epsilon_t)^2 \xrightarrow{L} (1 + \zeta^2)\sigma^2$$

<u>*Proof:*</u> see Phillips (1986)

$(\hat{\rho}-1)$ distribution:

$$(\hat{\rho}-1)=\frac{\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}-\sum_{t=2}^{T-1}\hat{u}_t^2-\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2}{\sum_{t=2}^{T-1}\hat{u}_t^2+\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2}$$

we have

$$\sum_{t=2}^{T}(\hat{u}_t-\hat{u}_{t-1})^2=\sum_{t=2}^{T}\hat{u}_t^2-2\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}+\sum_{t=2}^{T}\hat{u}_{t-1}^2$$

$$\longrightarrow\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}=\frac{1}{2}\sum_{t=2}^{T}\hat{u}_t^2+\frac{1}{2}\sum_{t=2}^{T}\hat{u}_{t-1}^2-\frac{1}{2}\sum_{t=2}^{T}(\hat{u}_t-\hat{u}_{t-1})^2$$

then

$$Numerator=\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}-\sum_{t=2}^{T-1}\hat{u}_t^2-\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2=$$

$$=\frac{1}{2}\sum_{t=2}^{T}\hat{u}_t^2+\frac{1}{2}\sum_{t=2}^{T}\hat{u}_{t-1}^2-\frac{1}{2}\sum_{t=2}^{T}(\hat{u}_t-\hat{u}_{t-1})^2-\sum_{t=2}^{T-1}\hat{u}_t^2-\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2=$$

$$=\frac{1}{2}\hat{u}_T^2+\frac{1}{2}\hat{u}_1{}^2-\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2-\frac{1}{2}\sum_{t=2}^{T}(\hat{u}_t-\hat{u}_{t-1})^2$$

so

$$\frac{1}{T}(numerator)=\frac{1}{2}\left(\frac{\hat{u}_T^2}{T}\right)+\frac{1}{2}\left(\frac{\hat{u}_1^2}{T}\right)-\frac{1}{T^2}\sum_{t=1}^{T}\hat{u}_t^2-\frac{1}{2}\left(\frac{1}{T}\sum_{t=2}^{T}(\hat{u}_t-\hat{u}_{t-1})^2\right)$$

(a)

$$\frac{1}{T}\hat{u}_T^2=\frac{1}{T}(y_T-\hat{\alpha}-\hat{\beta}x_T)^2=\frac{1}{T}\left[(y_T-\overline{y})-\hat{\beta}(x_T-\overline{x})\right]^2$$

$$=\frac{1}{T}(y_T-\overline{y})^2-2\frac{1}{T}\hat{\beta}(x_T-\overline{x})(y_T-\overline{y})+\frac{1}{T}\hat{\beta}^2(x_T-\overline{x})^2$$

we have

$$\frac{1}{T}(x_T-\overline{x})(y_T-\overline{y})=\frac{1}{T}\left[x_Ty_T-x_T\frac{1}{T}\sum_{t=1}^{T}y_t-y_T\frac{1}{T}\sum_{t=1}^{T}x_t+\frac{1}{T^2}\sum_{t=1}^{T}x_t\sum_{t=1}^{T}y_t\right]=$$

$$=\frac{x_T}{\sqrt{T}}\frac{y_T}{\sqrt{T}}-\frac{x_T}{\sqrt{T}}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}y_t\right)-\frac{y_T}{\sqrt{T}}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_t\right)+\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_t\right)\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}y_t\right)\xrightarrow{L}$$

$$\xrightarrow{L}\sigma^2\left[V(1)W(1)-W(1)\int_0^1V(r)dr-V(1)\int_0^1W(r)dr-\int_0^1W(r)dr\int_0^1V(r)dr\right]=\sigma^2A$$

so

$$2\frac{1}{T}\hat{\beta}(x_T-\overline{x})(y_T-\overline{y}) \xrightarrow{L} 2\sigma^2\zeta A \tag{45}$$

From lemma 1:

$$\frac{1}{T}(y_T-\overline{y})^2 \xrightarrow{L} \sigma^2\left(\left[\int_0^1 V(r)dr\right]^2 - 2V(1)\int_0^1 V(r)dr + [V(1)]^2\right) = \sigma^2 B_v \tag{46}$$

$$\frac{1}{T}\hat{\beta}^2(x_T-\overline{x})^2 \xrightarrow{L} \sigma^2\zeta^2\left(\left[\int_0^1 W(r)dr\right]^2 - 2W(1)\int_0^1 W(r)dr + [W(1)]^2\right) = \sigma^2\zeta^2 B_w \tag{47}$$

From (45), (46), (47):

$$\frac{1}{T}\hat{u}_T^2 = \frac{1}{T}(y_T-\overline{y})^2 - 2\frac{1}{T}\hat{\beta}(x_T-\overline{x})(y_T-\overline{y}) + \frac{1}{T}\hat{\beta}^2(x_T-\overline{x})^2 \xrightarrow{L}$$

$$\xrightarrow{L} \sigma^2(B_v - 2\zeta A + \zeta^2 B_w) \tag{48}$$

(b)

$$\frac{1}{T^2}\sum_{t=1}^{T}\hat{u}_t^2 = \frac{1}{T^2}\sum_{t=1}^{T}(y_t-\hat{\alpha}-\hat{\beta}x_t)^2 = \frac{1}{T^2}\sum_{t=1}^{T}\left[(y_t-\overline{y})-\hat{\beta}(x_t-\overline{x})\right]^2$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}(y_t-\overline{y})^2 - 2\frac{1}{T^2}\sum_{t=1}^{T}\hat{\beta}(x_t-\overline{x})(y_t-\overline{y}) + \frac{1}{T^2}\sum_{t=1}^{T}\hat{\beta}^2(x_t-\overline{x})^2$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}(y_t-\overline{y})^2 - \hat{\beta}^2\frac{1}{T^2}\sum_{t=1}^{T}(x_t-\overline{x})^2 \xrightarrow{L} \text{follow lemma 1} \xrightarrow{L}$$

$$\xrightarrow{L} \sigma^2\left\{\int_0^1 [V(r)]^2dr - \left(\int_0^1 V(r)dr\right)^2 - \zeta^2\left[\int_0^1 [W(r)]^2dr - \left(\int_0^1 W(r)dr\right)^2\right]\right\} = \sigma^2 C$$

(c)

$$\frac{1}{T}\hat{u}_1^2 = \frac{1}{T}(y_1-\hat{\alpha}-\hat{\beta}x_1)^2 = \frac{1}{T}\left[(y_1-\overline{y})-\hat{\beta}(x_1-\overline{x})\right]^2$$

$$= \frac{1}{T}(y_1-\overline{y})^2 - 2\frac{1}{T}\hat{\beta}(x_1-\overline{x})(y_1-\overline{y}) + \frac{1}{T}\hat{\beta}^2(x_1-\overline{x})^2$$

with

$$\frac{1}{T}(x_1-\overline{x})(y_1-\overline{y}) = \frac{1}{T}\left[x_1y_1 - x_T\frac{1}{T}\sum_{t=1}^{T}y_t - y_1\frac{1}{T}\sum_{t=1}^{T}x_t + \frac{1}{T^2}\sum_{t=1}^{T}x_t\sum_{t=1}^{T}y_t\right] =$$

$$= \frac{x_1}{\sqrt{T}}\frac{y_1}{\sqrt{T}} - \frac{x_1}{\sqrt{T}}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_t\right) - \frac{y_1}{\sqrt{T}}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t\right) + \left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t\right)\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_t\right) \xrightarrow{L}$$

$$\xrightarrow{L} \sigma^2\left[\int_0^1 W(r)dr \int_0^1 V(r)dr\right]$$

From lemma 1:

$$\frac{1}{T}(y_1 - \overline{y})^2 \xrightarrow{L} \sigma^2\left[\int_0^1 V(r)dr\right]^2$$

$$\frac{1}{T}(x_1 - \overline{x})^2 \xrightarrow{L} \sigma^2\left[\int_0^1 W(r)dr\right]^2$$

so

$$\frac{1}{T}\hat{u}_1^2 \xrightarrow{L} \sigma^2\left[\left(\int_0^1 V(r)dr\right)^2 + \zeta^2\left(\int_0^1 W(r)dr\right)^2 - 2\zeta\int_0^1 V(r)dr\int_0^1 W(r)dr\right] = \sigma^2 D$$

combine (a),(b),(c) and lemma 2, we have:

$$\frac{1}{T}(numerator) \xrightarrow{L} \frac{1}{2}\sigma^2(B_v - 2\zeta A + \zeta^2 B_w) + \frac{1}{2}\sigma^2 D - \sigma^2 C - \frac{1}{2}\sigma^2(1+\zeta^2)$$

$$\frac{1}{T^2}(denominator) = \frac{1}{T^2}\left(\sum_{t=2}^{T-1}\hat{u}_t^2 + \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\right) = \frac{1}{T^2}\sum_{t=1}^{T}\hat{u}_t^2 - \frac{\hat{u}_1^2}{T^2} - \frac{\hat{u}_T^2}{T^2} - \frac{1}{T^3}\sum_{t=1}^{T}\hat{u}_t^2 \xrightarrow{L} \sigma^2 C$$

finally:

$$T(\hat{\rho} - 1) \xrightarrow{L} \frac{\left[\frac{1}{2}(B_v - 2\zeta A + \zeta^2 B_w) + \frac{1}{2}D - C - \frac{1}{2}(1+\zeta^2)\right]}{[C]}$$

$\underline{t_{\hat{\rho}} \text{ distribution:}}$

$$t_{\hat{\rho}} = \left[\hat{\sigma}_e^2\left(\sum_{t=2}^{T-1}\hat{u}_t^2 + \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\right)^{-1}\right]^{-1/2}(\hat{\rho} - 1)$$

where $\hat{\sigma}_e^2$ is the estimation of variance of e_t in (31), under the hypothesis of no cointegration, the estimation of variance of e_t in (31) is $\frac{1}{T}\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2$, then

$$\hat{\sigma}_e^2 = \frac{1}{T}\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2 \xrightarrow{L} (1+\zeta^2)^2\sigma^2$$

then:

$$t_{\hat{\rho}} = \frac{T(\hat{\rho}-1)}{\left[\frac{\hat{\sigma}_e^2}{\frac{1}{T}\left(\sum_{t=2}^{T-1}\hat{u}_t^2+\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\right)}\right]^{-1/2}} \xrightarrow{L} \frac{\left[\frac{1}{2}(B_v - 2\zeta A + \zeta^2 B_w) + \frac{1}{2}D - C - \frac{1}{2}(1+\zeta^2)\right]}{[(1+\zeta^2)C]^{1/2}}$$

where

$$A = V(1)W(1) - W(1)\int_0^1 V(r)dr - V(1)\int_0^1 W(r)dr - \int_0^1 V(r)dr \int_0^1 W(r)dr$$

$$B_v = \left[\int_0^1 V(r)dr\right]^2 - 2V(1)\int_0^1 V(r)dr + [V(1)]^2$$

$$B_w = \left[\int_0^1 W(r)dr\right]^2 - 2W(1)\int_0^1 W(r)dr + [W(1)]^2$$

$$C = \int_0^1 [V(r)]^2 dr - \left(\int_0^1 V(r)dr\right)^2 - \zeta^2\left[\int_0^1 [W(r)]^2 dr - \left(\int_0^1 W(r)dr\right)^2\right]$$

$$D = \left[\int_0^1 V(r)dr\right]^2 + \zeta^2\left[\int_0^1 W(r)dr\right]^2 - 2\zeta\int_0^1 V(r)dr\int_0^1 W(r)dr$$

$$\zeta = \frac{\int_0^1 V(r)W(r)dr - \int_0^1 V(r)dr\int_0^1 W(r)dr}{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)dr\right]^2}$$

$V(r)$ and $W(r)$ are the standard Wiener processes on [0,1].