# Comparing Panel Data Cointegration Tests with an Application to the "Twin Deficits" Problem

Suzanne McCoskey
United States Naval Academy

Chihwa Kao*
Syracuse University

First Draft: August 1, 1997
This Draft: November 5, 1999

**Abstract**

This paper surveys recent developments and provides Monte Carlo comparison on various tests proposed for cointegration in panel data. In particular, five tests for hetergeneous panel data are tested for empirical size and size-adjusted power. In addition an application is shown where a panel of data for a subset of OECD countries is used to test for a long run relationship between trade deficits and budget deficits: the so-called "twin deficit" problem.

*Key Words and Phrases: Panel Cointegration.*

*JEL classifications: C23, C22.*

## 1 Introduction

Evaluating the statistical properties of data along the time dimension has proven to be very different from analysis in the cross-section dimension. As economists have gained access to better data with more observations across time, understanding these properties has grown increasingly important. An area of particular concern in time series econometrics is the use of non-stationary data. With the desire to study the behavior of a cross-section of observations over time and the increased use of panel data, one new research area is examining the properties of non-stationary time series data in panel form. It is an intriguing question to ask: how exactly does this hybrid style of data combine the statistical elements of traditional cross-sectional analysis and time series analysis? In particular, what is the correct way to analyze non-stationarity, the spurious regression problem, and cointegration in panel data?

Adding the cross-section dimension to the time dynamics offers a real advantage in the testing for non-stationarity and cointegration. The hope of the econometrics of non-stationary panel data is to combine

*We thank participants of the Eighth International Conference on Panel Data, Göteborg, Sweden, 1998, for helpful comments. An electronic version of the paper in postscript format can be retrieved from http://web.syr.edu/~cdkao. Address correspondence to: Chihwa Kao, Center for Policy Research, 426 Eggers, Syracuse University, Syracuse, NY 13244-1020; e-mail: cdkao@maxwell.syr.edu.

the best of both worlds: the method of dealing with non-stationary data from the time series and the increased data and power from the cross-section. The addition of the cross-section dimension, under certain assumptions, can act as repeated draws from the same distribution. Thus as the time and cross-section dimension increase, e.g., using the sequential limit theory or the joint limit theory of Phillips and Moon (1999), panel test statistics can be derived which converge in distribution to normally distributed random variables. Also within the testing framework, the addition of the cross-section dimension seemingly adds power to the tests.

The challenge in taking advantage of these properties is the difficulty in deriving the moments of the complex combinations of Brownian bridges and functionals of Brownian motion which often arise from the asymptotics in the time series literature. Several of the tests discussed in this paper use Monte Carlo simulations, as in the pure time series literature, to pin down these moments. Another difficulty which does not disappear in the panel setting is the difficulty in obtaining good estimates of long-run autocovariances. Finally, the panel setting offers a variety of models: common intercepts, common slopes, common intercepts and common slopes, differing intercepts and differing slopes; which have strong consequences for the estimation. In particular, asymptotics and estimation of common slopes is difficult. Also, the homogeneity or heterogeneity of the deterministic time structure of the cross-sectional observations needs to be considered. In this paper we only consider the heterogeneous model: varying intercepts and varying slopes. McCoskey and Kao (1998b) does consider results for two tests where there are common slopes and varying intercepts. These tests are based on results from Kao (1999) and offer a corrected version of Kao's test which, in general, can improve upon the power of the original test.

Unit root tests in the literature test the stationarity of a given series. These tests can be adapted for residual-based cointegration tests by testing the series of estimated residuals for stationarity. There are unit root tests for panel data already in the literature such as Levin and Lin (1993), Im, Pesaran and Shin (1995) and Maddala and Wu (1999). Once again, as in the time series case, moving from the unit root tests to cointegration tests is complicated by the estimation. The cointegration tests which test the null hypothesis of no cointegration must take into consideration the so-called "spurious regression" problem. Tests based on the null hypothesis of cointegration must take into consideration efficient estimation of a cointegrated relationship. Further, the concept of "pooled" estimation is different from pooling the cross-section testing results. In the case of unit root testing, most tests treat each individual cross-section independently. In the case of cointegration, treating each cross-section independently may translate into allowing for varying slopes and varying intercepts. This has strong implications for the model.

This paper outlines and compares statistics appropriate for testing for cointegration in heterogeneous panel data. A recent paper by Wu and Yin (1999) also looks at this problem as they compare size and power properties for ADF and maximum eigenvalue based tests. This paper, however, compares tests across null hypotheses allowing for both the null of no cointegration and the null of cointegration. Such a comparison can be of use as the debate continues as to the appropriate null hypothesis for different applications.

At least a brief mention should be given to the dynamics of the relationship between time series econometrics and economic theory. Most developments within the time series literature have been criticized as having more to do with a particular data set than economic theory in general. To be sure, there is no economic theory behind the techniques used to estimate lag orders for autoregressive representations, for example. Yet the cointegration literature offers a promising cross-over between economic theory and econometric techniques. The error-correction form, for example, captures short run deviations from the long run relationship between non-stationary variables. While this theory has not been totally resolved, it at least gives a practical motivation for theorists in all fields to familiarize themselves with applied time series techniques. *The Economic Journal* (Jan 1997) includes a discussion of the philosophy of modelling the long run using time series and cointegration results. Included in the journal are articles by Taylor and Dixon, Granger, Pesaran, and Harvey. The major issue mentioned for panel data in this discussion is the issue of how similar are the cross-sections in the panel and the difficulties of pooling heterogeneous cross-sections.

The paper is outlined as follows: Section 2 summarizes tests of the null hypothesis of no cointegration in panel data with varying intercepets and varying slopes. A total of four tests are presented. Section 3 presents a test of the null hypothesis of cointegration in panel data. This test assumes varying intercepts and varying slopes. Section 4 explains the Monte Carlo design for the comparison of the tests. Section 5 summarizes the results of the Monte Carlo experiment. Section 6 presents an application of several of the tests with regards to the "twin deficit" problem: a possible long run relationship between the trade balance and government budget. Section 7 provides some concluding thoughts.

A word on notation used throughout the paper: $\xrightarrow{p}$ is used to denote convergence in probability, $\Rightarrow$ to denote weak convergence, and $I(1)$ to signify a time series that is integrated of order one. All limits in this paper are taken as $T \to \infty$ and followed by $N \to \infty$ sequentially of Phillips and Moon (1999), except where otherwise noted.

# 2 Testing for Cointegration in Panels with the Null Hypothesis of No Cointegration

## 2.1 Average Augmented Dickey-Fuller Test for Varying Slopes

Here we propose a simple ADF test for varying slopes and varying intercepts panel data model. This is the same test used in Wu and Yin (1999).

$$y_{it} \quad = \quad \alpha_i + x_{it}^{'}\beta_i \quad + \quad e_{it}, \; i = 1, ..., N, \; t = 1, ..., T. \tag{1}$$

Each cross-section regression is estimated individually and the pooling from the panel is done in the final step where the panel test statistic is based on some average of the individual cross-section statistics. Each cross-section is allowed its individual cointegrating vector. Each test is constructed such that the cross-sections are assumed independent of each other and heteroskedasticity across the cross-sections is allowed. Im, Pesaran and Shin (1995) present a panel data unit root test based on the average of the ADF statistic of the cross-sections.[1] Using an analogous approach this style of test statistic can be used to test for cointegration. Recall that the ADF test can be constructed as:

$$\hat{e}_{it} = \rho_i \hat{e}_{it-1} + \sum_{j=1}^{p} \vartheta_{ij} \Delta \hat{e}_{it-j} + v_{itp}, \tag{2}$$

where $\hat{e}_{it}$ are OLS residuals from (1). An equivalent way to write equation (2) is given in Phillips and Ouliaris (1990):

$$\Delta \hat{e}_{it} = \rho_i \hat{e}_{it-1} + \sum_{j=1}^{p} \vartheta_{ij} \Delta \hat{e}_{it-j} + v_{itp}.$$

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

$$t_{iADF} = \frac{(u_{-1}' Q_{xp} u_{-1})^{\frac{1}{2}} \hat{\rho}_i}{s_v},$$

where $X_p$ is the matrix of observations on the $p$ regressors $(\Delta \hat{u}_{t-1}, ..., \Delta \hat{u}_{t-p})$, $\hat{u}_{-1}$ is the vector of observations of $\hat{u}_{t-1}$, $Q_{X_p} = I - X_p(X_p' X_p)^{-1} X_p'$ and $s_v^2 = \frac{1}{T} \Sigma_{t=1}^{T} \hat{v}_{tp}^2$.

The average ADF statistic is simple the average of the t-statistic across the cross-sections:

[1] In his paper, Pedroni (1997) also identifies the possibility of extending the logic of Im, Pesaran and Shin from panel unit root tests to panel cointegration tests. Although he does not present the ADF-$t$ statistic in the body of his paper, he does examine some of the properties of the test in his Monte Carlo simulations. Wu and Yin (1999) also present such an average ADF test.

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

It can be shown that:

$$\sqrt{N}(\bar{t}_{ADF} - \mu_{Adf}) \Rightarrow N(0, \sigma^2_{Adf}).$$

As Phillips and Ouliaris note, the limiting distribution of the ADF test statistic is free of nuisance parameters and depends only in the number of regressors. They provide cut-off tail values for the test in the time series case. It is a mere extension of the logic to then simulate the moments $\mu_{Adf}$ and $\sigma^2_{Adf}$ . Using RDNS procedure in Gauss with 50,000 replications the moments were found to be $\mu_{Adf} = -2.026$ and $\sigma_{Adf} = .8200$ in the case of one regressor. Appropriate values for the mean and standard deviation for 1 to 5 regressors is provided in the Appendix B. Tail values are also provided for comparison with the Phillips and Ouliaris results.

## 2.2 Average Phillips $Z_t$ Statistic for Varying Slopes

As Phillips and Ouliaris (1990) show, corrections for autocorrelation and contemporaneous correlation can either be performed through differencing and the ADF-t statistic method or through non-parametric corrections. In accordance with this idea, another test can be considered which is based on the average, across the cross-sections, of the Phillips $Z_t$ statistics. This statistic is by definition, for the varying intercepts and varying slopes model.

Phillips and Ouliaris provide exact details on how to calculate the Phillips $Z_t$ test. The first step, as in the $ADF$ test, is to calculate the estimated residuals from the original regression equation using OLS. Then using the estimated residuals, $\hat{e}_{it}$, perform the following regression:

$$\hat{e}_{it} = \alpha_i \hat{e}_{it-1} + v_{it}.$$

Note that this is similar to the $ADF$ test, although here without the lagged terms, the $v_{it}$ may have some effects from then cross-correlation and autocorrelation.

Define:

$$s^2_{iv} = \frac{1}{T}\sum_{t=1}^{T} \widehat{v}^2_{it}$$

and

$$s^2_{iTl} = \frac{1}{T}\sum_{i=1it}^{T}\widehat{v}^2_{it} + \frac{2}{T}\sum_{s=1}^{l} w_{sl}\sum_{t=s+1}^{T}\widehat{v}_{it}\widehat{v}_{it-s}.$$

These terms are used to calculate the final statistic:

$$\hat{Z}_{it} = \frac{(\hat{\alpha}-1)}{\frac{s_{iTl}}{(\sum_{t=2}^{T}\hat{e}^2_{it-1})^{\frac{1}{2}}}} - \frac{\frac{1}{2}(s^2_{iTl} - s^2_{iv})}{s_{iTl}(\frac{1}{T^2}\sum_{t=2}^{T}\hat{e}^2_{it-1})^{\frac{1}{2}}}. \tag{3}$$

Phillips and Ouliaris show that this $t$-statistic converges in distribution to the same functional of Brownian motion as the $ADF$ $t$-statistic. Thus, for the purposes here, it is convenient to note that a test based on the $Z_t$ test uses the same simulated moments as the $ADF$ test-statistic given above.

The average of the cross-section $\hat{Z}_{it}$ statistics can be defined as $\bar{Z}_t$

$$\bar{Z}_t = \frac{1}{N}\sum_{i=1}^{N}\hat{Z}_{it}$$

and it can be shown that:

$$\sqrt{N}(\bar{Z}_t - \mu_{Adf}) \Rightarrow N(0, \sigma^2_{Adf}). \tag{4}$$

It should be noted that in his paper, Pedroni (1997) also considers a version of the average of the $\hat{Z}_{it}$ statistic. His test is constructed as follows:

$$\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i)}{(\sum_{t=1}^{T}\hat{L}^{-2}_{11ie}\hat{e}^2_{it-1})^{\frac{1}{2}}} \tag{5}$$

where

$$\hat{\lambda} = \frac{1}{2}(s^2_{iTl} - s^2_{iv})$$

and

$$\hat{L}^{-2}_{11ie} = \sigma^2_{0u} - \frac{\sigma^2_{0u\varepsilon}}{\sigma^2_{0\varepsilon}}$$

in the scalar case based on the estimates for $\hat{\Omega}$ outlined in the previous section.

The two tests are conceptually quite similar[2] with the exception that the form from Phillips and Ouliaris only uses the kernal estimate of $\hat{v}_{it}$ for the non-parametric corrections while Pedroni also uses $\hat{\Omega}$. Intuitively,

[2] In fact $Z_t$ can be written as $\frac{\sum \hat{e}_{it-1}v_{it} - T\lambda}{(\sum \hat{e}^2_{it-1})^{\frac{1}{2}} s_{itL}}$. Under the null hypothesis, $v_{it}$ is equivalent to $\Delta\hat{e}_{it}$.

the form by Phillips and Ouliaris is clear, a t-statistic on $\hat{\alpha}$ is adjusted with a non-parametric component. To be consistent with the original article, we follow the format from Phillips and Ouliaris. Although it should be noted that Pedroni's simulated moments for the case with an intercept are 2.03 for the mean and $\sqrt{0.66}$ for the standard deviation-values very close to our own. Pedroni also provides moments for a model without intercept and one including a time trend.

## 2.3 Pedroni (1997)

Pedroni (1997) also proposes several tests for the null hypothesis of cointegration in panel data. His tests for heterogeneous slopes and intercepts fall into two categories. The first set, as discussed above, involves averaging test statistics for cointegration in the time series across cross-sections. The second set groups the statistics such that instead of averaging across statistics, the averaging is done in pieces so that the limiting distributions are based on limits of piecewise numerator and denominator terms.

The first set of statistics as discussed includes a form of the average of the Phillips and Ouliaris $Z_t$ statistic. Pedroni also proposes averaging the $Z_{\hat{\alpha}-1}$ statistic. His form for this statistic is

$$\tilde{Z}_{\hat{\alpha}-1} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} (\hat{e}_{it-1} \Delta \hat{e}_{it} - \hat{\lambda}_i)}{(\sum_{t=1}^{T} \hat{e}_{it-1}^2)} \tag{6}$$

This form directly corresponds to the statistic proposed by Phillips and Ouliaris.[3]

For his second set of statistics[4], Pedroni defines four panel variance ratio statistics. In this study we consider only one of these:

$$Z_{t_{\hat{\rho}_{NT}}} = \frac{\Sigma_{i=1}^{N} \Sigma_{t=2}^{T} \hat{L}_{11i}^{-2} (\hat{e}_{it-1} \Delta \hat{e}_{it} - \hat{\lambda}_i)}{\sqrt{\tilde{\sigma}_{NT}^2 (\Sigma_{i=1}^{N} \Sigma_{t=2}^{T} \hat{L}_{11i}^{-2} \hat{e}_{it-1}^2)}} ,$$

where

$$\tilde{\sigma}_{NT} = \frac{1}{N} \sum_{i=1}^{N} (\frac{\hat{\sigma}_i}{\hat{L}_{11i}})^2.$$

Using consistent estimates of $\Omega$, the long-run variance-covariance matrix given in the previous section, define $\hat{L}_i$ to be the lower triangular Cholesky composition of $\hat{\Omega}$ such that in the scalar case $\hat{L}_{22i} = \hat{\sigma}_\varepsilon$ and $\hat{L}_{11i} = \hat{\sigma}_u^2 - \frac{\hat{\sigma}_{u\varepsilon}^2}{\hat{\sigma}_\varepsilon^2}$, the long-run conditional variance. It should be noted that Pedroni thus bases his test on

[3] The format for this statistic from Phillips and Ouliaris is given as $\hat{Z}_\alpha = T(\hat{\alpha} - 1) - (\frac{1}{2})(s_{Tl}^2 - s_k^2)(\frac{\Sigma_2^T \hat{u}_{t-1}^2}{T^2})^{-1}$, the panel version can be written as $\frac{\sqrt{N}(\bar{Z}_\alpha + 9.05)}{\sqrt{35.98}}$

[4] Again with these statistics, Pedroni uses corrections which do not follow directly from Phillips and Ouliaris. In the simulations, therefore we adapt Pedroni's $Z_{t_{\hat{\rho}_{NT}}}$ by eliminating the $\hat{L}_{11i}^{-2}$ terms. With these corrections the performance of the test improves dramtically.

the average on the numerator and denominator terms respectively, rather than the average for the statistic as a whole.

Using results on convergences of functionals of Brownian motion, Pedroni finds the following result:

$$Z_{t_{\hat{\rho}_{NT}}} + 1.73\sqrt{N} \Rightarrow N(0, 0.93).$$

Note that this distribution applies to the model including an intercept and not including a time trend. Asymptotic results for other model specifications can be found in Pedroni (1997). The intuition on these tests with varying slopes is not straightforward. The convergence in distribution is based on individual convergence of the numerator and denominator terms. What is the intuition of rejection of the null hypothesis? Using the average of the overall test statistic allows more ease in interpretation: rejection of the null hypothesis means that enough of the individual cross-sections have statistics "far away" from the means predicted by theory were they to be generated under the null.

# 3 Testing for Cointegration in Panels with the Null Hypothesis of Cointegration: McCoskey and Kao (1998a)

In this section, a panel test of the null hypothesis of cointegration is presented. Tests of this null hypothesis were first introduced in the times series literature as a response to some critiques of the null hypothesis of no cointegration. For example, testing the null of cointegration rather than the null of no cointegration could be very appealing in applications where cointegration is predicted a priori by economic theory. Also, failure to reject the null of no cointegration could be caused, in many cases, by the low power of the test and not by the true underlying nature of the data.

The residual-based test for null of cointegration in panel data proposed by McCoskey and Kao is an extension of the Lagrange Multiplier (LM) test and locally best invariant (LBI) test for an MA unit root in the time series literature. This test is also discussed in McCoskey and Kao (1998a). Cointegration tests of the null of cointegration in the time series case have been proposed by Harris and Inder (1994) and Shin (1994). Under the null, the asymptotics no longer depend on the asymptotic properties of estimating spurious regression, rather the asymptotics of the estimation of a cointegrated relationship are needed. For models which allow the cointegrating vector to change across the cross-sectional observations, the asymptotics depend merely on the time series results as each cross-section is estimated independently. For models with common slopes, the estimation is done jointly and therefore the asymptotic theory is based on the joint estimation of a cointegrated relationship in panel data.

For the residual based test of the null of cointegration, it is necessary to use an efficient estimation technique of cointegrated variables. In the time series literature a variety of methods have been shown to be efficient asymptotically. These include the fully modified (FM) estimator of Phillips and Hansen (1990) and the dynamic least squares (DOLS) estimator as proposed by Saikkonen (1991) and Stock and Watson (1993). For panel data, Kao and Chiang (1998) show that both the FM and DOLS methods can produce estimators which are asymptotically normally distributed with zero means.

The model presented allows for varying slopes and intercepts:

$$y_{it} \quad = \quad \alpha_i + x_{it}^{'}\beta_i \quad + \quad e_{it},\ i = 1, ..., N,\ t = 1, ..., T, \tag{7}$$

$$x_{it} = x_{it-1} + \varepsilon_{it} \tag{8}$$

$$e_{it} = \gamma_{it} + u_{it}, \tag{9}$$

and

$$\gamma_{it} = \gamma_{it-1} + \theta u_{it}. \tag{10}$$

The null of hypothesis of cointegration is equivalent to $\theta = 0$.

The test statistic proposed by McCoskey and Kao is the following:

$$\overline{LM} = \frac{\frac{1}{N}\Sigma_{i=1}^{N}\frac{1}{T^2}\Sigma_{t=1}^{T}S_{it}^{+2}}{s^{+^2}}, \tag{11}$$

where $S_{it}$ is partial sum process of the residuals,

$$S_{it}^{+} = \sum_{j=1}^{t}\widehat{e}_{ij}$$

with

$$s^{+2} = \frac{1}{NT}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\hat{e}_{it}^{+2}.$$

($\hat{\varpi}_{1.2}^2$ is defined as a consistent estimator of $\sigma_v^2$, the long-run conditional variance under the $H_0$ and is used in place of $s^{+^2}$ if the residuals are estimated using the FM estimator.) The FM estimator non-parametrically corrects for the possible serial correlation and weakly exogenous regressors in a cointegrated regression. The DOLS estimator uses lagged and future differences of $x_{it}$ to correct for these effects.

The asymptotic result for the test is:

$$\sqrt{N}(\overline{LM} - \mu_v) \Rightarrow N(0, \sigma_v^2), \tag{12}$$

where $\mu_v = .1162$ and $\sigma_v^2 = .0109$ and are defined in McCoskey and Kao (1998a). The constants $\mu_v$ and $\sigma_v^2$ are moments of a complex functional of Brownian motion, which depend only on the number of regressors and can be found through Monte Carlo simulation.

The limiting distribution of $\overline{LM}$ is then free of nuisance parameters and robust to heteroskedasticity. For the Monte Carlo results, the fully modified estimation method is used.

# 4 The Monte Carlo Design

The ultimate goal of this Monte Carlo study is to compare the size and power of different residual based tests or cointegration in heterogeneous panel data: varying slopes and varying intercepts. A total of five tests are considered: $ADF^*$, $PO_t^*$, $PO_\alpha^*$, $APG^*$ and $LM^*$.[5]. Wu and Yin (1999) perform a similar comparison for panel tests in which they consider only tests for which the null hypothesis is no cointegration. In their study, the compare ADF statistics with maximum eigenvalue statistics in pooling information on means and p-values respectively. They find that the Average ADF performs better with respect to power and their their maximum eigenvalue based p-value (Fisher type test) performs better with regards to size. We consider only methods which pool "t-statistic" information for the null of no cointegration.

The first four statistics we consider are constructed under the null of no cointegration and the last under the null of cointegration. We first compare the four tests of the null of no cointegration for size and power and then select two of these tests to compare with $LM^*$. To compare these three tests, the study follows Harris and Inder (1994) who suggest testing the ability of the tests to properly identify the underlying nature of the data using two different Data Generating Processes. Thus, each null hypothesis is represented in the final experiment. This entails a two-step procedure outlined below. The simulations were performed in GAUSS using the package COINT 2.0

The random number generator used for this study is URN22 from Karian and Dudewicz (1991)[6].

[5] For a comprehensive comparison of tests on the strict time series case, see Haug (1996). It should be noted that Haug considers tests of both null hypotheses although he performs seperate Monte Carlo simulations for each type.

[6] This generator is a multiplicative congruential method of the following form:

$$x_{i+1} = (x_i * 69069 + 1) \ mod \ 2^{32}$$

The original seed is set to 32007779.

The uniform random numbers are transformed into standard normal random variables using the Box-Muller transformation (Karian and Dudewicz, 1991, p. 160). The standard normal random variables are used to construct a random walk variable for $x_{it}$ in the regression equation and the residuals. The program automatically saves the last number generated to use as the beginning seed in the next replication.

The RNDNS command within GAUSS uses the fast acceptance-rejection algorithm proposed by Kindermand and Ramage

## 4.1 Experimental Design

### 4.1.1 Varying Slopes and Varying Intercepts

$$y_{it} \quad = \quad \alpha_i + x_{it}^{'}\beta_i \quad + \quad e_{it}, \ i = 1, ..., N, \ t = 1, ..., T. \tag{13}$$

As discussed in Section 6.1, after the preliminary comparisons of size and power of all four tests of the null hypothesis of cointegration, three tests- $ADF^*$, $APG^*$, and $LM^*-$ are selected for comparison across both the null of cointegration and null of no cointegration. To compare the tests with varying intercepts, the study considers the two different data generating processes (DGP):

DGP-A, null of no cointegration:

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it},$$

and

$$e_{it} = \rho e_{it-1} + v_{it}.$$

Under the null hypothesis of no cointegration, $\rho = 1$. Under the null, the error term is a random walk. The study includes the following possible values for $\rho :$

$$\rho \in \{1, 0.95, 0.85, 0.75\}.$$

DGP-B, null of cointegration:

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it},$$

and

$$e_{it} = \theta \sum_{k=1}^{t} v_{ik} + v_{it}.$$

Under the null hypothesis of cointegration $\theta = 0$. Under the null the error term does not remember past errors and collapses to a standard normal random variable. The study includes the following values for $\theta :$

$$\theta \in \{0, 0.05, 0.15, 0.25\}$$

(1976) for generating standard normal variables from uniform random numbers. The uniform random numbers are generated according to the multiplicative-congruential method discussed in Kennedy and Gentle (1980). This method can be written in the form $x_i = (ax_{i-1} + c)(\mathrm{mod}\, m)$. In GAUSS the default values are $c = 0$; $m = 2^{31} - 1$; and $a = 397204094$. This particular method performed well in spectral testing as done by Hoaglin (1976) but was outperformed by a number of other generating techniques when tested by Karian and Dudewicz (1991). In particular, URN 22 outperformed this routine.

Notes for both DGP-A and DGP-B we assume that $v_{it}$ is distributed $N(0,1)$ and

$$x_{it} = x_{it-1} + \varepsilon_{it},$$

where $\varepsilon_{it}$ is distributed $N(0, \sigma_i^2)$.

Unless otherwise specified, the study considers the following dimensions for $N$ and $T$: $N \in \{1, 15, 25, 50, 100\}$ and $T \in \{15, 25, 50, 100\}$ and the number of replications for each dimension is $10,000$. The choice of N and T for this experiment underlines an important point: this test is not really appropriate with severely unbalanced data sets, for example extremely large N and small T.

$\alpha_i$, $\beta_i$ and $\sigma_i$ are generated using the default uniform random number generator in GAUSS, i.e., $\alpha_i \sim U[0, 10]$, $\beta_i \sim U[0, 2]$, and $\sigma_i \sim U[0.5, 1.5]$.

### 4.1.2 Two Stages of the Experiment

Because the tests are not all derived under the same null hypothesis, it is difficult to compare their performance directly. A two stage procedure is used here to make sure the results are comparable. The first stage is to compute 5% and 95% critical values under the null of each DGP, A and B. These critical values are used to set the probability of rejecting the null of the particular DGP to 0.05 for all three tests. For the tests of the null hypothesis of no cointegration, with DGP-A, this 5% is simply the value which leads to size equal to 0.05. For the third test of the null of cointegration, this 5% critical value with DGP-A does not, strictly speaking, relate to the size of the test but rather is simply a probability of rejection. With DGP-B the logic is just the reverse. The 5% critical value of the test of the null hypothesis is the value which leads to a size of 0.05 whereas for the other two tests it does not have this exact interpretation. The size is directly related to the null hypothesis of the test not the DGP of the experiment. To summarize:

DGP-A (generated under the null of no cointegration), choose critical values such that

$$\begin{array}{cccc} \text{Test 1} & \text{Test 2} & \text{Test 3} & \\ \Pr(\text{reject null}) = & \Pr(\text{reject null}) = & \Pr(\text{reject null}) = & 0.05 \; . \\ \text{Size} & \text{Size} & & \end{array}$$

DGP-B (generated under the null of cointegration), choose critical values such that

$$\begin{array}{cccc} \text{Test 1} & \text{Test 2} & \text{Test 3} & \\ \Pr(\text{reject null}) = & \Pr(\text{reject null}) = & \Pr(\text{reject null}) = & 0.05 \; . \\ & & \text{Size} & \end{array}$$

The empirical rejection rates, using a one-sided $N(0,1)$ distribution are given in Table 3. These critical values are constructed to insure the tails of all three tests have equivalent density.

The second stage is to calculate rejection rates for the alternative values of the parameters for DGPS A and B. In other words, the second step is used to calculate the ability of the tests to properly reject the null hypothesis of the DGP based on critical values found in the first stage. The intuition behind these two steps is analogous to finding the power of a test after adjusting the critical values for the size of the test. Again, the strict concepts of size and power must be used with caution in this experiment as the null hypothesis used in the DGP is not necessarily the null hypothesis used to construct the test.

Let $c_j^k(N,T)$ be the critical value calculated in Stage 1 for test $j = 1, 2, 3$ and $DGP$, $k = A$ or $B$ for a given $N$ and $T$.

DGP-A given a specific $\rho < 1$ (data generated under the alternative of cointegration):

$$\begin{array}{ccc} \text{Test 1} & \text{Test 2} & \text{Test 3} \\ \Pr\left(\text{test1} < c_1^A(N,T)\right) & \Pr\left(\text{test2} < c_2^A(N,T)\right) & \Pr\left(\text{test3} < c_3^A(N,T)\right) \\ \text{Power (Size=0.05)} & \text{Power (Size=0.05)} & \end{array}.$$

In all cases this is the probability of correctly rejecting the null hypothesis of no cointegration.

DGP-B given a specific $\theta > 0$ (data generated under the alternative of no cointegration):

$$\begin{array}{ccc} \text{Test 1} & \text{Test 2} & \text{Test 3} \\ \Pr\left(\text{test1} > c_1^B(N,T)\right) & \Pr\left(\text{test2} > c_2^B(N,T)\right) & \Pr\left(\text{test3} > c_3^B(N,T)\right) \\ & & \text{Power (Size=0.05)} \end{array}.$$

These probabilities represent the probability of the test correctly rejecting the null hypothesis of cointegration. These comparisons of the tests' ability to properly reject the null hypothesis of DGPs A and B are given in Tables 4 and 5.

## 4.2 Test Statistics

Results from the following forms of the tests are reported. Define the following standardized statistics for varying slopes and varying intercepts:

$$ADF^* = \frac{\sqrt{N}(\bar{t}_{ADF} + 2.026)}{.82},$$

$$PO_t^* = \frac{\sqrt{N}(\bar{Z}_t + 2.026)}{.82}$$

$$PO_\alpha^* = \frac{\sqrt{N}(\bar{Z}_\alpha + 9.05)}{\sqrt{35.98}}$$

$$APG^* = \left(\frac{\Sigma_{i=1}^N \Sigma_{t=2}^T (\hat{e}_{it-1\Delta}\hat{e}_{it} - \hat{\lambda}_i)}{\sqrt{\bar{\sigma}_{NT}^2 (\Sigma_{i=1}^N \Sigma_{t=2}^T \hat{e}_{it-1}^2)}} + 1.73\sqrt{N}\right)/\sqrt{.93}$$

where

$$\bar{\sigma}_{NT} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i)^2,$$

$$LM^* = \frac{\sqrt{N}(\overline{LM} - .1162)}{\sqrt{.0109}}.$$

## 4.3 Interpreting the results

Ultimately the goal of simulations with varying slopes is to see how well these three tests can distinguish between the true character of the DGP and its alternative. For each of these tests under the two different DGPS, the following probability is desired:

$$\Pr_{DGP-H_o}(\text{Rejecting the null } DGP|H_A),$$

i.e., the probability of rejecting the null of the DGP when the alternative is true.

Call this probability $rej_j^k$, the rejection rate of test $j$ under DGP $k$ :

$$rej_j^A = rej_j^A(N, T, \rho)$$

and

$$rej_j^B = rej_j^B(N, T, \theta).$$

Each of these individual experiments can be considered as the sum of Bernoulli random variables where

$$X_i = \begin{cases} 1 & \text{if reject} \\ 0 & \text{otherwise} \end{cases}$$

and the pdf is given:

$$p_X(X) = \begin{cases} p^x(1-p)^{1-x} \;\; x = 0, 1, ... \\ \\ 0 \text{ otherwise} \end{cases}$$

with $E(X) = p$ and $Var(X) = p(1-p)$. Using the results from the Bernoulli distribution (see Appendix B for details) a test is considered "significantly better" if the difference between the two rejection rates is at least as large as two times the standard deviation of comparison, .007071.

## 5 Monte Carlo Results

In Table 1 we show a preliminary comparison of the four tests of the null hypothesis of no cointegration in terms of empirical size. Theory would predict that size should converge to .05 for all tests. Of the four tests, $ADF^*$ seemingly performs the best. (This result on size is consistent with Haug (1996) for the time series case.) In fact $ADF^*$ has a small range across all N and T. The maximum rejections rate reported is .1112 and the minimum .0485. Both of these occur when the T and N are very different in magnitude. $P-O_t^*$ has a strong tendency to over-reject when T $\leq$ 25 and tendency to under-reject for $T \geq 50$. This is especially true for large N. For example, when T=15 and N=100, $P-O_t^*$ will reject the null hypothesis almost 28% of the time. When T=100 and N=100, $P-O_t^*$ will only reject the null .19% of the time. $P-O_a^*$, to the contrary, underrejects the null in all cases expect when T=100 and N=1. This is especially severe for small T. For example, when T=15 and $N \geq 15, P-O_a^*$ never rejects the null hypothesis. $APG^*$, like $P-O_t^*$, has a tendency to overreject for small T and underreject for large T, although in neither cases is the problem as severe. The range of size values for $P-O_t^*$ is [.0019, .2789] while for $APG^*$ the range is [.0035, .1371]. For $APG^*$, under-rejection is especially a problem when $T \geq 25$ and $N \geq 25$. These results underline the intuition of panel data that relative sizes of the T and N dimension can significantly impact the characteristics of the test.

In Table 2 we compare results for the power of the test to correctly reject the null hypothesis of no cointegration. In this preliminary comparison, only results for $\rho = 0.95$ or $\rho = 0.75$ are reported. A '*' is used to indicate when a test performs significantly better than any of the other three when compared pairwise. At first glance it is clear the $APG^*$ performs well with regard to power. In all cases when a most powerful test can be determined, it is $APG^*$. All tests show that decreasing the value of $\rho$ increases the power of the test. It is also clear that increasing the T dimension increases the power of the test more than increasing the cross-section dimension. This result is especially important in applications where researchers may have more ability to increase the cross-sections in their data rather than find more time series data. For example, holding T constant at 25, increasing N from 15 to 100 has the following impact when $\rho = 0.95$, the power of $ADF^*$ increases from .1114 to .3510; the power of $P-O_t^*$ increases from .1572 to .5621; the power of $P-O_a^*$ increases from .1445 to .5036; and the power of $APG^*$ increases from .1709 to .7289. However, when N is held constant at 25 and T increases from 15 to 100 ($\rho = 0.95$), the power of $ADF^*$ increases from .1017 to .8090; the power of $P-O_t^*$ increases from .1291 to .9636; the power of $P-O_a^*$ increases from .1218 to .9314; and the power of $APG^*$ from .1510 to .9826. Another strong result is how using a panel rather than strict time series can increase power dramatically. In the strict time series case ($N = 1$ and $T = 100$), when $\rho = 0.95$ (i.e. data is almost non-stationary) the power of the four tests ranges from .1255 to .1251.

However, increasing N from 1 to 15 shifts the range to a minimum at .5811 and maximum at .8610. Thus, the difficulty in correctly rejecting the null for "near non-stationary data" is alleviated.

Given the results from Tables 1 and 2, we choose $ADF^*$ and $APG^*$ as the tests to compare with $LM^*$ for the second phase of the Monte Carlo experiments. $ADF^*$ performed the best with regard to empirical size and $APG^*$ with regard to power.

As the results on empirical size for $ADF^*$ and $APG^*$ have already been discussed, we now turn our attention to size results for $LM^*$. From Table 3 we observe that, in general, holding T fixed and increasing N decreases the size of the test. In the case of $T < 50$, this causes the test to under-reject the null hypothesis for large N. For $T > 50$, in all cases the test underrejects the null with a range in size of [.0533, .0917], the minimum being reached when $T = 50$ and $N = 100$ and the maximum when $T = 100$ and $N = 15$. Again we see the importance of the relative size of the dimensions of the panel.

Turning our attention to the ability of the three tests to correctly reject the null hypothesis of DGP-A shown in Table 4, we see some surprising results. Intuition seems to suggest that $ADF^*$ and $APG^*$ should outperform $LM^*$ as they are tests derived under this null. However, $LM^*$ clearly outperforms the other two tests for the cases where $\rho = 0.95$ and $\rho = 0.85$. $APG^*$ outperforms the other two when $\rho = 0.75$ for all the cases when $N = 1$, the strict time series case, and when $T \leq 25$ and $N = 15$.

The dominance of $LM^*$ in the cases of $\rho$ close to 1 is significant as these errors which show cointegration are very close to being non-stationary, and it is these "nearly non-stationary" errors which can give researchers the most difficulty. In some cases the power of $LM^*$ is very far away from the other two tests. For example, when $\rho = 0.95$, $T = 50$ and $N = 25$, $LM^*$ correctly rejects the null hypothesis 91.57% of the time, $APG^*$ properly rejects the null only 57.20% of the time and $ADF^*$ only 29.66% of the time. All of the tests show the nicely behaved properties that power increases as $\rho$ decreases, and N and T increase.

Considering Table 5 and the tests' ability to correctly reject the null hypothesis of DGP-B, we see that again $LM^*$ dominates. In fact in all cases when a most powerful test could be determined, $LM^*$ was the most powerful test. However, this results is not quite so surprising as $LM^*$ is the only test of the three derived under the null hypothesis of cointegration. For both Tables 4 and 5 we see the benefits of using panel data rather than the simple time series. Consider the results for LM* in Table 5, when $T = 15$ and $\theta = 0.15$, increasing N from 1 to 15 increases the power from .1066 to .2872; when $T = 25$ an identical increase changes power from .1705 to .6375; when $T = 50$ power increases from .3943 to .9937; and for $T = 100$ power increases from .7224 to .9999

## 6 Testing for "Twin Deficits"

Is there a long run relationship between government budget deficits and trade balances, the current account? Recent developments in the United States' economy, in particular, have caused many economists to study this question. The very nature of the question makes it a natural for cointegration and non-stationary data analysis.

The theory underlying this question comes from Keynesian open macroeconomic models:

$$S = Y - C - G = S^p + S^g$$

where national saving is defined as national income less consumption and government spending, or the sum of private and government saving.

$$S^p + S^g = I + CA$$

In an open economy total savings can also be thought of as equal to domestic investment and the current account, where the current account is the trade balance (exports-imports). Thus

$$S^p = I + CA + (G - T).$$

Krugman and Obstfeld (1997) summarize this identity "a country's private saving can take three forms: investment in domestic capital (I), purchases of wealth from foreigners (CA), and purchases of the domestic government's newly issued debt (G-T)." In the international literature this relationship has been called the Mundell-Fleming idea whereas changes in the government deficit will impact the supply and demand of loans and thus impact the interest rate which will in term impact the balance of trade.

Granger causality testing on the relationship has been done by Miller and Russek (1989) and Tallman and Rosensweig (1991). Both find evidence that the budget deficit (US) leads to the trade deficit while the opposite does not seem to hold. Rosensweig and Tallman (1993) and Mohammadi and Skaggs (1996) (among others) test whether the variables involved are non-stationary and conclude that except for the exchange rate, which might be included in the specification, the pertinent variables are I(1). Cointegration testing has been done by Miller and Russek (1989), Rahmen and Mishra (1992), Bachman (1992) and in general they cannot reject the null of no cointegration in the case of pure time series data. One notable exception is Mohammedi and Skaggs (1996) which uses the Johansen testing (and not the residual based tests) and finds evidence of cointegration.

Several papers go beyond the simple case of just the United States and look at cross-national evidence. Fry (1992) looks at the relationship for developing countries; Karras and Song (1995) look at Australia, Italy, Sweden, the U.K. and the U.S., and Masson, Kremers and Horne (1994) consider the U.S., Japan and Germany. Laney (1984) tests 59 countries for a significant relationship between the deficit and current account. In this paper we will construct a panel with the countries: Austria, Belgium, Canada, Finland, France, Ireland, Italy, The Netherlands, Norway, Spain, Sweden, Great Britain, and the United States; for the years 1975-1994. We will test two different specifications of the twin deficit relationship using the average ADF and LM[7] statistics. Comparing our countries to those included in Laney (1984), Austria, Belgium, Canada, Ireland, Italy, Norway, and Sweden all showed a significant relationship; the Netherlands, Great Britain and the United States did not show a significant relationship. It should be noted that his results did not account for potential non-stationarity and therefore may be susceptible to the spurious regression problem.[8]

The is from the World Bank "1998 World Development Indicators" and we will use the variables: BDGDP-the total budget deficit as % of GDP[9]; CAGDP-the current account balance as % of GDP; and ERI, the real effective exchange rate[10]. For this application we will not perform the tests on stationarity but rather use the previous results to proceed immediately to testing for cointegration, thus we do not consider a possible model with just the current account and exchange rate.

Model 1:

$$CAGDP_{it} = \alpha_i + \beta_i BDGDP_{it} + \varepsilon_{it}$$

[7] It should be noted that the estimation is done using dynamic OLS (the ps procedure in GAUSS) and not Fully Modified estimation. As can be seen in Tables 4 and 5, the fm routine in GAUSS cannot always estimate with a small time dimension. Asymptotically the two methods are identical (see Kao and Chiang (1998)): as would be any appropriate method for estimating a cointegrated system.

[8] Austria, Belgium, Canada, Ireland, Italy, Norway, and Sweden all showed a significant relationship in Laney and the estimated coefficient on the relationship between the trade deficit (negative of the current account) and budget deficit was positive in all cases. The dates used in his study varied somewhat but, in general, covered the years between 1950 and 1980.

[9] Rosenzweig and Tallman (1993) discuss differences in model specification and suggest deflating variables by GNP for real rather than nominal measurements.

[10] The description for real effective exchange from the World Bank CD is as follows: "Real effective exchange rates are derived by deflating a trade-weighted average of the nominal exchange rates that apply between trading partners. For most industrial countries the weights are based on trade in manufactured goods with other industrial countries during 1989-91, and an index of relative, normalized unit labor costs is used as the deflator. (Normalization smooths a time series by removing short-term fluctuations while retaining changes of a large amplitude over the longer economic cycle.) For other countries, prior to 1990, the weights take into account trade in manufactured and primary products during 1980-82; from January 1990 onward weights are based on trade in manufactured and primary products during 1988-90, and an index of relative changes in consumer prices is used as the deflator. An increase in the real effective exchange rate represents an appreciation of the local currency. Because of conceptual and data limitations, changes in real effective exchange rates should be interpreted with caution."

Model 2:

$$CAGDP_{it} = \alpha_i + \beta_{1i} BDGDP_{it} + \beta_{2it} ERI + \varepsilon_{it}.$$

Individual testing results for each model are given in Table 6. The pooled test results are somewhat disappointing: for neither model were the panel tests able to reject the null hypothesis for either no cointegration or cointegration. (ADF* is a one-tailed test which would reject null of no cointegration with a value less than -1.645; LM* is a one tailed test which would reject the null of cointegration with a value greater than 1.645.) In this application, even pooling into the panel seemingly did not allow for enough power.

It is true that the panel dimensions are modest, 13 countries and 20 years. Based our our earlier Monte Carlo results, if DGP A is true then pooling into a panel and increasing the cross-section dimension from one to approximately fifteen (with about fifteen years) could increase the power of the $\text{ADF}^*$ in the most extreme case from .0753 to .3042. For LM* the most extreme improvement would be from .0771 to .6030. But if the autoregressive parameter was closer to one, the improvements would be much more modest. Similar results would be true if DGP B were true. In both cases $\text{LM}^*$ would be more powerful than $\text{ADF}^*$.

Looking at the individual country results, only the Netherlands rejects the null of no cointegration for Model 2.

| | Model 1 | Model 2 |
|---|---|---|
| ADF* | 1.1057 | −0.4248 |
| LM* | −1.4758 | −3.9491 |

One real benefit to the residual based tests is that estimating is done simultaneously to the testing. In Tables 7 and 8 we present the potential cointegration vectors estimated using the OLS and dynamic OLS procedures. OLS estimates and t-statistics should be interpreted with great caution: if the systems are not cointegrated, estimation will be spurious; if the systems are cointegrated then dynamic OLS estimates should be more reliable. Based on Mundell-Fleming idea of twin deficits, we both expect coefficients to be negative. An increase in the budget deficit should cause a decrease in the current account and an increase in the real effective exchange rate (an appreciation of the currency) should also cause a decrease in the current account.

Reviewing the results for the dynamic OLS estimates (we cannot reject cointegration), in Model 1 we see that of the set of thirteen countries, eight have negative estimates for the budget deficit (four which are statistically significant). Three countries have significant and positive estimates. It is interesting to see that one of these countries is the United States. For Model 2, again eight out of the thirteen countries have negative coefficients for the budget deficit (with four statistically significant). The signs are relatively stable across models: Spain is the only country with a positive estimate in Model 1 and negative in Model 2 and

Great Britain the opposite. For the estimates on the real effective exchange rates in Model 2, eight of the thirteen estimates are negative (five are significant.) Considering the special case of the Netherlands, which can reject no cointegration in Model 2, the cointegrating vector seems "nicely behaved" according to the Mundell-Fleming idea with negative estimates for both variables.

In this case, although a panel test is tempting, especially for the goal of increasing power, some issues of appropriate poolability are raised. For the heterogeneous model, obviously less restrictive than the homogeneous coefficients model, there is still the restriction that the elements of the cointegrating vector are assumed constant across the cross-sections. In their paper, Karras and Song (1995) look at the strict time series cases of Australia, Italy, Sweden, the U.K. and the U.S. They raise the question of whether or not all the countries in their sample could truly be classified as Keynesian. Obviously, that would be a concern here as well.

# 7 Conclusion

The development of non-stationary econometrics in the time series literature allowed for a deeper understanding of the statistics of "long-run steady state" relationships. These relationships were identified as cointegrated relationships among non-stationary variables. Extending these results to panel data offers the new challenge of how to combine results on cross-sectional data combined with the time series. This paper evaluates tests for cointegration in panel data. Which test among these is best?

Th choices presented in this paper vary based on the actual null hypothesis. Two of the tests are constructed under the null hypothesis of no cointegration, $ADF^*$ and $APG^*$. These tests are based on the $ADF$ test and Pedroni's pooled. The third test is based on the null hypothesis of cointegration which is based on the LM test from the time series literature, $LM^*$.

The test of the null hypothesis was originally proposed in response to the low power of the tests of the null of no cointegration, especially in the time series case. Further, in those cases where economic theory predicted a long run steady state relationship, it seemed that a test of the null of cointegration rather than the null of no cointegration would be appropriate. The results from the Monte Carlo study here shows that $LM^*$ does outperform the other two tests. In both experiments, $LM^*$ was seen to be more powerful, especially for cases when the parameters generated were very close to values under the null.

Of the two reasons for the introduction of the test of the null hypothesis of cointegration, low power and attractiveness of the null, the introduction of the cross-section dimension of the panel solves one: all of the tests show decent power when used with panel data. For those applications where the null of cointegration is more logical than the null of no cointegration, this study, at a minimum, concludes that using $LM^*$ does not compromise the ability of the researcher of determining the underlying nature of the data.

Finally, we applied the $ADF^*$ and $LM^*$ to the question whether a sample of OECD countries could have a cointegrating relationship between the current account and budget deficit. Neither test was able to reject the null hypothesis and so we can neither conclude cointegration of no cointegration in the system.

# 8 Appendix A

The following are critical values, including mean and standard deviation, for ADF and $Z_t$ which are used for $\bar{t}_{ADF}$ and $\bar{Z}_t$.

| $k$ | *mean* | *std* | 10% | 5% | 1% |
|---|---|---|---|---|---|
| 1 | −2.0261 | .8200 | −3.0383 | −3.3329 | −3.9197 |
| 2 | −2.4687 | .8000 | −3.4695 | −3.7576 | −4.3290 |
| 3 | −2.8535 | .7800 | −3.8319 | −4.1212 | −4.6750 |
| 4 | −3.1758 | .7668 | −4.1500 | −4.4344 | −4.9978 |
| 5 | −3.4816 | .7583 | −4.4584 | −4.7451 | −5.2998 |

The values were calculated in GAUSS using 50,000 replications. Since asymptotic theory tells us that the ADF test should asymptotically be identical in distribution to a Dickey-Fuller test with iid errors, a Dickey-Fuller test on non-stationary residuals was simulated.

Values for the individual 10%, 5%, and 1% levels are provided as a comparison to the values given in Phillips and Ouliaris (1990).

# 9 Appendix B

In this experiment, each $p_j^A$ and $p_j^B$ is estimated by finding the mean of these Bernoulli random variables:

$$rej_j^A = \frac{\Sigma_{i=1}^{10,000} X_{ji}^A}{10,000}$$

Given that each experiment is iid we obtain:

$$E(rej_j^A) = \frac{1}{10,000} * 10,000 * E(X_{ji}^A) = p_j^A$$

and

$$Var(rej_j^A) = \frac{1}{(10,000)^2} * 10,000 * Var(X_{ji}^A) = \frac{p_j^A(1-p_j^A)}{10,000} .$$

Thus the standard error for each rejection rate is equal to

$$\frac{\sqrt{p_j^A(1-p_j^A)}}{\sqrt{10,000}}.$$

The standard error reaches a maximum at the rejection rate of 0.5 with a standard deviation of 0.005. For a rejection rate of 0.99, the standard deviation would be 0.000995.

How about comparing $rej_1^A$ and $rej_2^A$? Comparing the rejection rates across two tests for the same DGP (i.e., to answer the question: which is better at correctly rejecting?) is equivalent to evaluating the significance of the difference of two random variables:

$$Var(rej_1^A - rej_2^A) = Var(rej_1^A) + Var(rej_2^A) - 2Cov(rej_1^A, rej_2^A)$$

where the covariance is given by

$$\sum_{i=1}^{10,000} \frac{(X_{1i}^A - rej_1^A)(Y_{2i}^A - rej_2^A)}{10,000}.$$

The intuition here is that the covariance is measuring whether the tests will reject for the same data or not. The standard error of comparison is given by:

$$\sqrt{Var(rej_1^A) + Var(rej_2^A) - 2Cov(rej_1^A, rej_2^A)}.$$

This reaches a maximum when the tests are assumed independent and each have a rejection rate of 0.5. In that case the standard deviation for comparison would be .007071. In the results, a test is considered "significantly better" if the difference between the two rejection rates is at least as large as two times the standard deviation of comparison.

# References


[1] Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

[2] Bachman, D. D. (1992), "Why is the U.S. Current Account Deficit So Large? Evidence from Vector Autoregressions," *Souther Economic Journal*, 59, 232-240.

[3] Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, John Wiley & Sons, Inc., New York.

[4] Banerjee, A., Dolado, J., Galbraith, J. W. , and Hendry, D. F.. (1994), *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, New York.

[5] Fry, M. J. (1992), "Foreign Debt Accumulation: Financial and Fiscal Effects and Monetary Policy Reactions in Developing Countries," *Journal of International Money and Finance*, 12, 347-367.

[6] Granger, C. W. J. (1997), "On Modelling the Long Run in Applied Economics," *The Economic Journal*, 107, 169-177.

[7] Harris, D., and Inder, B. (1994), "A Test of the Null Hypothesis of Cointegration," in *Nonstationary Time Series Analysis and Cointegration*, Hargreaves, Colin P. Ed., Oxford University Press, New York.

[8] Harvey, A. (1997), "Trends, Cycles and Autoregressions," *The Economic Journal*, 107, 192-201.

[9] Haug, A. A. (1996), "Tests for Cointegration: A Monte Carlo Comparison," *Journal of Econometrics*, 71, 89-115.

[10] Hoaglin, D.C. (1976), "Theoretical Properties of Congruential Random-Number Generators: an Empirical View," *Memorandum NS-340*, Department of Statistics, Harvard University, Cambridge Massachusetts.

[11] Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge.

[12] Im, K., Pesaran, M. H., and Shin Y. (1995), "Testing for Unit Roots in Heterogenous Panels," Manuscript, Department of Applied Economics, University of Cambridge.

[13] Kao, C. (1999), "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data," *Journal of Econometrics*, 90, 1-44.

[14] Kao, C., and Chiang, M-H. (1998), "On the Estimation and Inference of a Cointegrated Regression in Panel Data," Manuscript.

[15] Karian, Z. A., and Dudewicz, E. J. (1991), *Modern Statistical, Systems, and GPSS Simulation: The First Course*, W. H.. Freeman and Company, New York.

[16] Karras, G., and Song, F. (1995), "Government Spending and the Current Account," *International Economic Journal*, 9, 45-56.

[17] Kennedy, W. J., and Gentle, J. E. (1980), *Statistical Computing*, Marcel Dekker, Inc., New York.

[18] Kinderman, A. J., and Ramage, J. G. (1976), "Computer Generation of Normal Random Numbers," *Journal of the American Statistical Association*, Vol. 71, Number 356, 893-896.

[19] Krugman, P.R and Obstfeld, M. (1997), *International Economics: Theory and Policy.* Addison Wesley Longman Inc., Reading MA, 307-315.

[20] Laney, L. O. (1984), "The Strong Dollar, the Current Account, and Federal Deficits: Cause and Effect," *Economic Review (Federal Reserve Bank of Dallas)*, Jan, 1-14.

[21] Levin, A., and Lin, C.-F. (1993), "Unit Root Tests in Panel Data: New Results," Discussion Paper, Department of Economics, UC-San Diego.

[22] Maddala, G. S., and Wu, S. (1999), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test: Evidence From Simulations and the Bootstrap," *Oxford Bulletin of Economics and statistics*, forthcoming.

[23] Masson, P. R., Kremers, J. and Horne, J. (1994), "Net Foreign Assets and International Adjustment: The United States, Japan and Germany," *Journal of International Money and Finance*, 13, 27-40.

[24] McCoskey, S., and Kao, C. (1998a), "A Residual-Based Test of the Null of Cointegration in Panel Data," *Econometric Reviews*, 17, 57-84.

[25] McCoskey, S., and Kao, C. (1998b), "A Monte Carlo Comparison of Tests for Cointegration in Panel Data," mimeo.

[26] Miller, S., and Russek, F. (1989), "Are the Twin Deficits Really Related?" *Contemporary Policy Issues*, 7, 91-115.

[27] Mohammadi, H., and Skaggs, N. T. (1996), "U.S. Fiscal Policy and Trade Deficits: A Broad Perspective," *Southern Economic Journal*, 62, 675-689.

[28] Murthy, N. R. V., and Phillips, J. M. (1996), "The Relationship Between Budget Deficits and Capital Inflows: Further Econometric Evidence," *The Quarterly Review of Economics and Finance*, 36, 485-494.

[29] Pedroni, P. (1997), "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis," Manuscript, Department of Economics, Indiana University.

[30] Pesaran, M. H. (1997), "The Role of Economic Theory in Modelling the Long Run," *The Economic Journal*, 107, 178-191.

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

[32] Phillips, P. C. B., and Loretan, M. (1991), "Estimating Long-Run Economic Equilibria," *Review of Economic Studies*, 58, 407-436.

[33] Phillips, P. C. B., and Moon, H. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057-1111.

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

[35] Phillips, P. C. B., and Perron, P. (1988), "Testing for a Unit Root in Time Series Regressions," *Biometrika*, 75, 335-346.

[36] Rahman, M., and Banamber, M. (1992), "Cointegration of U.S. Budget and Current Account Deficits: Twins or Strangers?" *Journal of Economics and Finance*, 16, 119-127.

[37] Rosenswieg, J. A., and Tallman, E. W. (1993), "Fiscal Policy and Trade Adjustment: Are the Twin Deficits Really Twins?" *Economic Inquiry*, 31, 580-594.

[38] Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1-21.

[39] Shin, Y. (1994), "A Residual Based Test of the Null of Cointegration Against the Alternative of No Cointegration," *Econometric Theory*, 10, 91-115.

[40] Stock, J., and Watson, M. (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783-820.

[41] Tallman, E. W., and Rosensweig, J. A., (1991), "Investigating U.S. Government and Trade Deficits," *Federal Reserve Bank of Atlanta Economic Review*, 76 (3), 1-11.

[42] Taylor, A. M. R., and Dixon, H. D. (1997), "Controversy: On Modelling the Long Run in Applied Economics," *The Economic Journal*, 107, 165-168.

[43] World Bank (1998), *World Development Indicators on CD-ROM.*

[44] Wu, S., and Yong, Y. (1999) "Tests for Cointegration in Hetergeneous Panel: A Monte Carlo Study," mimeo.

Table 1: Preliminary Comparison of Empirical Size-DGP A

| | | ADF* | $PO_t^*$ | $PO_\alpha^*$ | APG* |
|---|---|---|---|---|---|
| T=15 | | | | | |
| | N=1 | .0824 | .0887 | .0063 | .1045 |
| | N=15 | .0919 | .1339 | .0000 | .1014 |
| | N=25 | .0934 | .1557 | .0000 | .1054 |
| | N=50 | .0948 | .1945 | .0000 | .1149 |
| | N=100 | .1112 | .2789 | .0000 | .1371 |
| T=25 | | | | | |
| | N=1 | .0703 | .0690 | .0301 | .0784 |
| | N=15 | .0715 | .0587 | .0039 | .0524 |
| | N=25 | .0701 | .0593 | .0017 | .0467 |
| | N=50 | .0764 | .0563 | .0004 | .0379 |
| | N=100 | .0890 | .0559 | .0003 | .0311 |
| T=50 | | | | | |
| | N=1 | .0536 | .0431 | .0444 | .0487 |
| | N=15 | .0597 | .0257 | .0198 | .0278 |
| | N=25 | .0594 | .0204 | .0171 | .0208 |
| | N=50 | .0611 | .0147 | .0105 | .0150 |
| | N=100 | .0707 | .0083 | .0055 | .0079 |
| T=100 | | | | | |
| | N=1 | .0485 | .0353 | .0593 | .0404 |
| | N=15 | .0567 | .0156 | .0356 | .0222 |
| | N=25 | .0588 | .0106 | .0338 | .0148 |
| | N=50 | .0604 | .0049 | .0266 | .0077 |
| | N=100 | .0663 | .0019 | .0224 | .0035 |

Notes:
(a) Size based on one-sided test with critical value equal to -1.645.

Table 2: Preliminary Comparison of Power-DGP A

| | ADF* | | P-Ot* | | P-Oa* | | APG* | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | 0.95 | 0.75 | 0.95 | 0.75 | 0.95 | 0.75 | 0.95 | 0.75 |
| | | | | T=15 | | | | |
| N=1 | .0534 | .0753 | .0574 | .1025 | .0556 | .0984 | .0573 | .1036 |
| N=15 | .0805 | .3042 | .0993 | .5631 | .1016 | .5678 | .1196* | .6904* |
| N=25 | .1017 | .4664 | .1291 | .7782 | .1218 | .7648 | .1510* | .8896* |
| N=50 | .1307 | .7200 | .1818 | .9682 | .1698 | .9649 | .2379* | .9953* |
| N=100 | .2047 | .9468 | .2910 | .9995 | .2566 | .9992 | .3981* | .9999 |
| | | | | T=25 | | | | |
| N=1 | .0569 | .1082 | .0600 | .1566 | .0597 | .1548 | .0611 | .1620 |
| N=15 | .1114 | .7170 | .1572 | .9549 | .1445 | .9408 | .1709 | .9813* |
| N=25 | .1487 | .8957 | .2086 | .9962 | .1955 | .9939 | .2527* | .9997 |
| N=50 | .2150 | .9951 | .3469 | .9999 | .3068 | .9999 | .4365* | .9999 |
| N=100 | .3510 | .9999 | .5621 | .9999 | .5036 | .9999 | .7289* | .9999 |
| | | | | T=50 | | | | |
| N=1 | .0627 | .2746 | .0780 | .4342 | .0780 | .4292 | .0778 | .4333 |
| N=15 | .2105 | .9998 | .3443 | .9999 | .3118 | .9999 | .3786* | .9999 |
| N=25 | .2966 | .9999 | .5184 | .9999 | .4547 | .9999 | .5720* | .9999 |
| N=50 | .5083 | .9999 | .7849 | .9999 | .7087 | .9999 | .8653* | .9999 |
| N=100 | .7801 | .9999 | .9963 | .9999 | .9282 | .9999 | .9903 | .9999 |
| | | | | T=100 | | | | |
| N=1 | .1255 | .8490 | .1237 | .9398 | .1251 | .9420 | .1231 | .9391 |
| N=15 | .5811 | .9999 | .8287 | .9999 | .7612 | .9999 | .8610* | .9999 |
| N=25 | .8090 | .9999 | .9636 | .9999 | .9314 | .9999 | .9826* | .9999 |
| N=50 | .9801 | .9999 | .9996 | .9999 | .9977 | .9999 | .9999 | .9999 |
| N=100 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 |

Notes:
(a) * indicates the power is "significantly greater" when compared with any of the other three tests.

Table 3: Empirical Rejection Rates

| | $ADF^*$ DGP-A | $APG^*$ DGP-A | $LM^*$ DGP-B |
|---|---|---|---|
| | | T=15 | |
| N=1 | .0824 | .1045 | .0475 |
| N=15 | .0919 | .1014 | .0405 |
| N=25 | .0934 | .1054 | .0266 |
| N=50 | .0948 | .1149 | .0147 |
| N=100 | .1112 | .1371 | (b) |
| | | T=25 | |
| N=1 | .0703 | .0784 | .0638 |
| N=15 | .0715 | .0524 | .0666 |
| N=25 | .0701 | .0467 | .0544 |
| N=50 | .0764 | .0379 | .0371 |
| N=100 | .0890 | .0311 | .0183 |
| | | T=50 | |
| N=1 | .0536 | .0487 | .0735 |
| N=15 | .0597 | .0278 | .0883 |
| N=25 | .0594 | .0208 | .0816 |
| N=50 | .0611 | .0150 | .0655 |
| N=100 | .0707 | .0079 | .0533 |
| | | T=100 | |
| N=1 | .0485 | .0404 | .0736 |
| N=15 | .0567 | .0222 | .0917 |
| N=25 | .0588 | .0148 | .0868 |
| N=50 | .0604 | .0077 | .0880 |
| N=100 | .0663 | .0035 | .0734 |

Notes:
(a) Empirical critical value for DGP-A is -1.645 and for DGP-B, 1.645.
(b) Results for N=100 and T=15 were unobtainable for the fm routine of COINT.

Table 4: Power to Reject: DGP-A

| | $ADF^{*}$ | | | $APG^{*}$ | | | $LM^{*}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | 0.95 | 0.85 | 0.75 | 0.95 | 0.85 | 0.75 | 0.95 | 0.85 | 0.75 |
| | | | | | T=15 | | | | |
| N=1 | .0534 | .0625 | .0753 | .0573 | .0781 | .1036* | .0573 | .0665 | .0771 |
| N=15 | .0805 | .1664 | .3042 | .1196 | .3571 | .6904* | .1403* | .3647 | .6030 |
| N=25 | .1017 | .2551 | .4664 | .1510 | .5303 | .8896* | .1910* | .5454* | .8089 |
| N=50 | .1307 | .3936 | .7200 | .2379 | .8255* | .9953 | .2988* | .7962 | .9961 |
| N=100 | .2047 | .6400 | .9468 | .3981 | .9795 | .9999 | (b) | | |
| | | | | | T=25 | | | | |
| N=1 | .0569 | .0777 | .1082 | .0611 | .0987 | .1620* | .0652 | .0921 | .1357 |
| N=15 | .1114 | .3436 | .7170 | .1709 | .6706 | .9813* | .2397* | .7170* | .9566 |
| N=25 | .1487 | .4987 | .8957 | .2527 | .8769 | .9997 | .3757* | .9062* | .9973 |
| N=50 | .2150 | .7802 | .9951 | .4395 | .9941 | .9999 | .5978* | .9929 | .9999 |
| N=100 | .3510 | .9658 | .9999 | .7280 | .9999 | .9999 | .8605* | .9998 | .9999 |
| | | | | | T=50 | | | | |
| N=1 | .0627 | .1327 | .2746 | .0778 | .1980 | .4333* | .1081* | .2047 | .3923 |
| N=15 | .2105 | .9067 | .9998 | .4084 | .9980 | .9999 | .6652* | .9969 | .9999 |
| N=25 | .2966 | .9905 | .9999 | .5720 | .9999 | .9999 | .9157* | .9999 | .9999 |
| N=50 | .5083 | .9999 | .9999 | .8653 | .9999 | .9999 | .9999* | .9999 | .9999 |
| N=100 | .7801 | .7567 | .9999 | .9903 | .9999 | .9999 | .9999 | .9999 | .9999 |
| | | | | | T=100 | | | | |
| N=1 | .1255 | .4660 | .8490 | .1231 | .5484 | .9391* | .2533* | .5382 | .8817 |
| N=15 | .5811 | .9999 | .9999 | .8610 | .9999 | .9999 | .9999* | .9999 | .9999 |
| N=25 | .8090 | .9999 | .9999 | .9826 | .9999 | .9999 | .9999* | .9999 | .9999 |
| N=50 | .9801 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 |
| N=100 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 |

Notes:
(a) * indicates the rejection rate is "significantly greater" when compared with either of the other two tests.
(b) Results for N=100 and T=15 were unobtainable for the fm routine in COINT.

Table 5: Power to Reject: DGP-B

| | $ADF^*$ | | | $APG^*$ | | | $LM^*$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | 0.05 | 0.15 | 0.25 | 0.05 | 0.15 | 0.25 | 0.05 | 0.15 | 0.25 |
| | | | | | T=15 | | | | |
| N=1 | .0519 | .0643 | .0843 | .0562 | .0646 | .0897 | .0619 | .1066* | .1797* |
| N=15 | .0534 | .0933 | .1822 | .0558 | .1111 | .2504 | .0815* | .2872* | .6460* |
| N=25 | .0559 | .1121 | .2417 | .0587 | .1368 | .3459 | .0951* | .3989* | .8290* |
| N=50 | .0571 | .1411 | .3644 | .0597 | .1789 | .5181 | .1141* | .5798* | .9717* |
| N=100 | (b) | | | | | | | | |
| | | | | | T=25 | | | | |
| N=1 | .0566 | .0965 | .1565 | .0537 | .0886 | .1555 | .0680 | .1705* | .3219* |
| N=15 | .0632 | .2203 | .5502 | .0637 | .2325 | .6364 | .1146* | .6375* | .9676* |
| N=25 | .0718 | .3104 | .7413 | .0660 | .3157 | .7980 | .1423* | .8158* | .9970* |
| N=50 | .0789 | .4638 | .9295 | .0790 | .4993 | .9665 | .1868* | .9659* | .9999* |
| N=100 | .0892 | .6814 | .9965 | .0902 | .7220 | .9996 | .2639* | .9995* | .9999 |
| | | | | | T=50 | | | | |
| N=1 | .0680 | .2164 | .4171 | .0654 | .1947 | .3941 | .1083* | .3943* | .6566* |
| N=15 | .1076 | .7916 | .9960 | .1005 | .7653 | .9964 | .3163* | .9937* | .9999 |
| N=25 | .1313 | .9290 | .9999 | .1201 | .9027 | .9999 | .4190* | .9999* | .9999 |
| N=50 | .1861 | .9949 | .9999 | .1614 | .9941 | .9999 | .6462* | .9999 | .9999 |
| N=100 | .2806 | .9999 | .9999 | .2325 | .9999 | .9999 | .8564* | .9999 | .9999 |
| | | | | | T=100 | | | | |
| N=1 | .1052 | .5113 | .8035 | .0962 | .4692 | .7796 | .2254* | .7224* | .9262* |
| N=15 | .3283 | .9999 | .9999 | .2613 | .9999 | .9999 | .8635* | .9999 | .9999 |
| N=25 | .4855 | .9999 | .9999 | .3814 | .9999 | .9999 | .9674* | .9999 | .9999 |
| N=50 | .7099 | .9999 | .9999 | .5686 | .9999 | .9999 | .9998* | .9999 | .9999 |
| N=100 | .9143 | .9999 | .9999 | .8141 | .9999 | .9999 | .9999 | .9999 | .9999 |

Note:
(a) * indicates the rejection rate is "significantly greater" when compared with either of the other two tests.
(b) Results for T=15 and N=100 were unobtainable for the fm routine in COINT.

Table 6: Individual Testing Results (Models 1 and 2)

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | LM | ADF | LM | ADF |
| Austria | 0.1384 | -1.9151 | 0.0021 | -1.3189 |
| Belgium | 0.0816 | -0.9641 | 0.0036 | -2.2110 |
| Canada | 0.0377 | -2.0519 | 0.0033 | -2.1980 |
| Finland | 0.0981 | -2.4244 | 0.0087 | -1.6158 |
| France | 0.0498 | -2.9220 | 0.0018 | -2.9245 |
| Ireland | 0.0318 | -2.0381 | 0.0044 | -2.0690 |
| Italy | 0.1753 | -3.0456 | 0.0027 | -2.9747 |
| The Netherlands | 0.1289 | -1.9756 | 0.0030 | -4.1550** |
| Norway | 0.0448 | -2.3645 | 0.0088 | -2.1603 |
| Spain | 0.0851 | -2.2966 | 0.0027 | -3.2581 |
| Sweden | 0.0185 | -2.7836 | 0.0026 | -3.4088 |
| Great Britian | 0.0387 | -2.2304 | 0.0028 | -3.1120 |
| United States | 0.0265 | -1.9019 | 0.0024 | -1.8998 |

Notes:
(a) ***, **, and * inidicate significant at 1, 5 and 10 percent.
(b) LM critical values are taken from Harris and Inder (1994).
(c) ADF critical values taken from Phillips and Ouliaris (1990).

Table 7: Estimation Results- Models 1

| | OLS Intercept | OBD | DOLS Intercept | OBD |
|---|---|---|---|---|
| Austria | -7.8242 | -1.4648 | -11.4289 | -2.3793 |
| | (-3.8541) | (-3.2797) | (-2.8524) | (-2.6786) |
| Belgium | 4.6567 | 0.5041 | 6.2360 | 0.7160 |
| | (2.6529) | (2.4148) | (3.7758) | (3.7769) |
| Canada | -6.0685 | -0.7690 | -8.8135 | -1.4960 |
| | (-7.8772) | (-4.2221) | (-6.4871) | (-4.5534) |
| Finland | -2.3456 | -0.0232 | -2.3094 | -0.5233 |
| | (-3.4508) | (-0.1900) | (-1.9860) | (-0.7067) |
| France | -0.4551 | -0.1092 | 0.0790 | 0.1303 |
| | (-1.1939) | (-0.8180) | (0.0735) | (0.2817) |
| Ireland | 2.9936 | 0.8459 | 1.1148 | 0.8167 |
| | (2.1889) | (5.7589) | (0.5013) | (3.8074) |
| Italy | 0.0700 | 0.0575 | -10.3422 | -0.8165 |
| | (0.0345) | (0.3270) | (-1.9388) | (-1.7660) |
| The Netherlands | 2.5838 | -0.0401 | 0.4408 | -0.4046 |
| | (3.0360) | (-0.2110) | (0.3057) | (-1.3092) |
| Norway | 0.1450 | 0.6619 | 0.2353 | -0.4533 |
| | (0.1184) | (1.9638) | (0.1888) | (-0.4533) |
| Spain | -3.4713 | -0.4326 | 2.1773 | 0.6741 |
| | (-3.6093) | (-2.1376) | (0.9406) | (1.3678) |
| Sweden | -1.5077 | -0.0060 | -1.9840 | -0.1126 |
| | (-3.2449) | (-0.0915) | (-1.6273) | (-0.5124) |
| Great Britian | -1.8728 | -0.3880 | -2.8683 | -0.9476 |
| | (-3.2717) | (-2.8717) | (-4.7928) | (-5.1402) |
| United States | 0.1505 | 0.3795 | 6.0026 | 2.0168 |
| | (0.1434) | (1.4002) | (5.5207) | (7.1093) |

Notes:
(a) Estimations done with COINT 2.0.
(b) DOLS estimates obtained with ps command with two lags and two leads.

Table 8: Individual Estimation Results-Model 2

| | OLS Intercept | OBD | ERI | DOLS Intercept | OBD | ERI |
|---|---|---|---|---|---|---|
| Austria | 7.7235 | -1.2603 | -0.1386 | 1.4666 | -1.7974 | -0.1014 |
| | (1.0756) | (-3.0440) | (-2.2396) | (0.3351) | (-8.4879) | (-2.7656) |
| Belgium | 17.0911 | 0.7502 | -0.0913 | 8.5508 | 0.7991 | -0.0135 |
| | (6.5222) | (5.3276) | (-5.2448) | (1.3362) | (1.1057) | (-0.4348) |
| Canada | -2.7981 | -0.7393 | -0.0345 | -22.8792 | -1.7948 | 0.1420 |
| | (-0.8324) | (-4.0067) | (-0.9996) | (-3.5304) | (-7.8644) | (2.3039) |
| Finland | 23.4401 | 0.5318 | -0.2612 | -27.9801 | -1.1097 | 0.2588 |
| | (2.7159) | (2.5161) | (-2.9941) | (-0.3926) | (-1.0095) | (0.3473) |
| France | -0.5423 | -0.1099 | 0.0008 | 128.1765 | 3.7007 | -1.1403 |
| | (0.0898) | (-0.7472) | (0.0145) | (4.6405) | (4.5655) | (-4.6320) |
| Ireland | 5.3855 | 0.7495 | -0.0274 | 64.2915 | 0.7418 | -0.4433 |
| | (0.7777) | (2.4026) | (-0.3527) | (1.2256) | (0.4894) | (-1.1779) |
| Italy | 9.4590 | -0.0731 | -0.1168 | -62.3427 | -0.3007 | 0.6237 |
| | (2.0366) | (-0.4294) | (-2.2018) | (-0.4072) | (-0.1366) | (0.3266) |
| The Netherlands | 11.4159 | 0.0061 | -0.0774 | 16.4141 | -1.1061 | -0.1583 |
| | (4.0107) | (0.0393) | (-3.1989) | (5.5363) | (-3.9893) | (-5.4290) |
| Norway | 45.3439 | 0.5042 | -0.4637 | 957.8015 | -1.6267 | -9.7918 |
| | (1.4091) | (1.4535) | (-1.4055) | (2.0104) | (-0.8912) | (-2.0092) |
| Spain | 3.7399 | -0.4159 | -0.0780 | 90.5182 | -0.3584 | -1.0378 |
| | (0.8777) | (-2.1634) | (-1.7326) | (4.1552) | (-0.4192) | (-3.8255) |
| Sweden | 4.4228 | 0.0853 | -0.0558 | -39.6192 | -2.3946 | 0.2988 |
| | (1.4688) | (1.1257) | (-1.9900) | (-3.6773) | (-3.7934) | (3.4393) |
| Great Britian | -9.2404 | -0.4742 | 0.0714 | -139.8583 | 3.9469 | 1.3520 |
| | (-3.5307) | (-4.0158) | (2.8644) | (-0.5840) | (0.4572) | (0.5695) |
| United States | 0.1705 | 0.3779 | -0.0002 | 7.7021 | 1.5715 | -0.0259 |
| | (0.0865) | (1.2355) | (-0.0121) | (1.7861) | (0.6113) | (-0.5131) |

Notes:
(a) Estimations done with COINT 2.0.
(b) DOLS estimates obtained through ps routine with two leads and two lags.