Testing Structural Stability in Heterogeneous Panel Data

[1] **Felix Chan**, [2] **Tommaso Mancini-Griffoli and** [3] **Laurent L. Pauwels**

[1] School of Economics and Finance, Curtin University of Technology
GPO Box U1987 WA 6845, Australia. E-Mail:F.Chan@curtin.edu.au
[2] Paris-Jourdan Sciences Economiques (PSE), CEPREMAP
[3] Hong Kong Monetary Authority and Graduate Institute of International Studies, Geneva

Keywords: *Structural change, cross sectionally dependent errors, heterogeneous panels, Euro effect on trade.*

EXTENDED ABSTRACT

This paper introduces a new test for structural instability among only some individuals at the end of a sample in a panel regression model. Most tests for structural breaks in the literature are appropriate when the break is relatively long lasting and happens in the middle of a sample. The distribution of the corresponding test statistic is suitably found using asymptotics in which the number of observations before and after the break point go to infinity. However, it is often at the end of a sample that researchers and policy-makers alike are interested in testing for instability.

Andrews (2003) proposes a test for structural break which was shown to be particularly useful when the number of post-break observations is small. Unlike the well known Predictive Failure test of Chow (1960), the critical values of Andrews's (1993) test statistic are calculated using parametric sub-sampling methods making the test robust to non-normal, heteroskedastic and serially correlated errors. The extension of the test to panel data, under the assumption of cross sectional independence, is relatively straightforward as shown in Mancini-Griffoli and Pauwels (2006). This extension assumes an alternative hypothesis that all individuals exhibit a break, as in other tests for structural breaks in the panel literature. Yet, these tests do not allow the interesting alternative that only some - and not all - individuals are affected by a break. This paper addresses such question by introducing a standardized Z statistic built from Andrews (2003) statistics averaged across individuals.

Methodologically, the proposed procedure is similar to the approach in Im et al. (2003) which, while focussing on the different question of unit root tests, also considers an average of separate statistics. The test statistic is shown to follow a normal distribution as the number of individuals goes to infinity by using the Lindeberg-Feller Central Limit Theorem (LF-CLT). This greatly simplifies the computation of the critical values with respect to Andrews (2003). As in Andrews (2003), though, the proposed statistic is robust to non-normal, heteroskedastic, serially correlated errors and when the instability occurs at the end of a given sample. Lastly, the test covers the cases of parameter heterogeneity or homogeneity pre- and post-instability. Moreover, it is straightforward to extend the proposed test statistic and the associated asymptotic results to accommodate the presence of cross sectional dependency.

A series of Monte Carlo experiments show that the proposed structural break test performs very well in finite sample. The experiments accommodate serial correlation in the error terms with a mixture of different distributions for the innovations. Monte Carlo results indicate that the test has good size and power with relatively few time series and moderate serial correlation within cross sections. For high levels of serial correlation, the performance of the test improves as the number of time series observations, T, increases. Lastly, the test has good power and size for partial instabilities, when the instabilities are of a small magnitude.

Finally, this paper considers an empirical application of the test to demonstrate its practical usefulness. The question of detecting the effects of Euro on trade has been at the center of lively debates in academic and policy circles alike. However, the papers that have tackled the issue have not provided strong empirical evidence in support of the presumed effect. This is largely due to two empirical issues: the few datapoints available after the Euro's introduction and the heterogeneity of the trade effect over different countries. Given both these characteristics, the test introduced in this paper is particularly well suited. Results show a break at the 10% significance level in Eurozone trade starting in 1998, thereby supporting to the belief commonly expressed in the literature.

1 INTRODUCTION

This paper introduces a new test for structural instability among only some individuals at the end of a sample in a panel regression model. Most tests for structural breaks in the literature, like the celebrated Chow (1960) tests, and those for unknown or multiple break dates in Andrews (1993), Andrews and Ploberger (1994) and Bai and Perron (1998) are appropriate when the break is relatively long lasting and happens in the middle of a sample. The distribution of the corresponding test statistic is suitably found using asymptotics in which the number of observations before and after the break point go to infinity. However, it is often at the end of a sample that researchers and policy-makers alike are interested in testing for instability.

Andrews (2003) proposes a test for structural break which was shown to be particularly useful when the number of post-break observations is small. Unlike the well known Predictive Failure test of Chow (1960), the critical values of Andrews's (1993) test statistic are calculated using parametric sub-sampling methods making the test robust to non-normal, heteroskedastic and serially correlated errors. The extension of the test to panel data, under the assumption of cross sectional independence, is relatively straightforward as shown in Mancini-Griffoli and Pauwels (2006). This extension assumes an alternative hypothesis that all individuals exhibit a break, as in other tests for structural breaks in the panel literature, like in Han and Park (1989), Kao et al. (2005) and De Wachter and Tzavalis (2004). Yet, these tests do not allow the interesting alternative that only some - and not all - individuals are affected by a break. This paper addresses such question by introducing a standardized Z statistic built from Andrews (2003) statistics averaged across individuals.

Methodologically, the proposed procedure is similar to the approach in Im et al. (2003) which, while focussing on the different question of unit root tests, also considers an average of separate statistics. The test statistic is shown to follow a normal distribution as the number of individuals goes to infinity by using the Lindeberg-Feller Central Limit Theorem (LF-CLT). This greatly simplifies the computation of the critical values with respect to Andrews (2003). As in Andrews (2003), though, the proposed statistic is robust to non-normal, heteroskedastic, serially correlated errors and when the instability occurs at the end of a given sample. Lastly, the test covers the cases of parameter heterogeneity or homogeneity pre- and post-instability.

Although the proposed test is initially derived under the assumption of cross sectional independency, it does not impose serious restriction on the practical usefulness of the test. The asymptotic results derived in this paper will still hold in the absence of cross sectional independency as long as the cross sectional dependency can be "filtered out". For example, Chan et al. (2007) modifies the proposed test statistics to accommodate cross sectional dependency using the Common Correlated Effects (CCE) estimator as introduced in Pesaran (2006).

A series of Monte Carlo experiments show that the proposed structural break test performs very well in finite sample. The experiments accommodate serial correlation in the error terms with a mixture of different distributions for the innovations. Monte Carlo results indicate that the test has good size and power with relatively few time series and moderate serial correlation within cross sections. For high levels of serial correlation, the performance of the test improves as the sample size increases. Lastly, the test has good power and size for partial instabilities, when the instabilities are of a small magnitude.

Finally, this paper considers an empirical application of the test, to highlight its properties in a real-world setting: did the introduction of the Euro increase intra-Eurozone trade? The question has been at the center of lively debates in academic and policy circles alike. However, the papers that have tackled the issue have not provided strong empirical evidence in support of the presumed effect. This is largely due to two empirical issues: the few datapoints available after the Euro's introduction and the heterogeneity of the trade effect over different countries. Given both these characteristics, the test introduced in this paper is particularly well suited. Results show a break at the 10% significance level in Eurozone trade starting in 1998, thereby supporting to the belief commonly expressed in the literature.

The paper is organised as follows. Section 2 introduces the panel data stability test for partial break. This is followed by the various asymptotic results in section 3. The finite sample properties will be investigated via Monte Carlo simulation in section 4. Due to the page number constraint, all the proofs are omitted from the current paper. However, interested readers can find the proofs of all the asymptotic results in this paper from Chan et al. (2007). Finally, section 5 illustrates how the test can be put to use to answer the question of the Euro's effect on intra-Eurozone trade. The last section concludes.

2 HETEROGENEOUS PANEL DATA STABILITY TESTS

This section introduces the heterogeneous panel data stability test. Consider the following model for panel data,

$$y_{it} = \begin{cases} \mathbf{\Theta}_{0i}'\mathbf{x}_{it} + u_{it} & t = 1,\dots,T_i \\ \mathbf{\Theta}_{1i}'\mathbf{x}_{it} + u_{it} & t = T_i+1,\dots,T_i+m_i \end{cases} \quad (1)$$

$$u_{it} = \gamma_i'\mathbf{f}_t + \varepsilon_{it} \quad t = 1,\dots,T_i+m_i, \quad (2)$$

for $i = 1,..,N$ with y_{it} being the endogenous variable, $\mathbf{x}_{it} = (x_{it}^{(1)},...,x_{it}^{(d)})'$ is the $d \times 1$ vector of explanatory variables including intercepts and/or seasonal dummies, $\mathbf{\Theta}_{0i}$ and $\mathbf{\Theta}_{1i}$ are the $d \times 1$ vectors of coefficients before and after the breakpoint, respectively. Notice that T_i are the presumed break dates, which can differ for each cross section i and m_i are the number of post-break observations different for each i. Moreover, ε_{it} are the idiosyncratic shocks specific to each individual and assumed to be uncorrelated to $\mathbf{x}_{it}$ and with zero mean, $\mathbf{f}_t$ is the $l \times 1$ vector of unobserved common effects and γ_i are the factor loadings associated with $\mathbf{f}_t$. For the purposes of deriving the test statistics, γ_i is assumed to be 0 for all $i = 1,...,N$. See Chan et al. (2007) for the case where $\gamma_i \neq 0$.

Under equations (1) and (2) with $\gamma_i = 0$, the hypothesis of structural stability is simply

$$H_0 : \mathbf{\Theta}_{1i} = \mathbf{\Theta}_{0i} \quad \forall\, i = 1,\dots,N \text{ and } \forall\, t,$$
$$H_1 : \mathbf{\Theta}_{1i} \neq \mathbf{\Theta}_{0i} \quad \exists\, i \in \{1,\dots,N\} \text{ and for } t > T_i$$

with $t = 1,\dots,T_i,T_i{+}1,\dots,T_i{+}m_i$. Let $N = N_0 + N_1$, where N_0 is the number of individuals for whom $\mathbf{\Theta}_{1i} = \mathbf{\Theta}_{0i}$ and N_1 is the total number of individual that have a break ($\mathbf{\Theta}_{1i} \neq \mathbf{\Theta}_{0i}$). The null hypothesis states that there are no structural breaks across all N individuals, whereas the alternative states that at least one individual experiences a structural break. $\mathbf{\Theta}_{0i}$ can be estimated heterogeneously by simply estimating each time-series $\mathbf{\Theta}_{0i}$ for each i by OLS.

2.1 Panel data stability test

This section introduces a standardised Z statistic to test for stability in panel data models. The Z statistic essentially amounts to comparing two average statistics taken from a pre-break subsample and the post-break sample. The construction of the average statistics for both a pre-break subsample and the post-break sample require to compute the Andrews (2003) test statistic N times for each individual time series. The computation of the individual statistics is identical to the initial time series end-of-sample instability test proposed by Andrews (2003).

Define the test statistics, $S_{i,p}^{\ q}$, for each individual, i, as

$$S_{i,p}^{\ q}(\mathbf{\Theta}_i,\mathbf{\Sigma}_i) = A_{i,p}^{\ q}(\mathbf{\Theta}_i,\mathbf{\Sigma}_i)'[V_{i,p}^{\ q}(\mathbf{\Sigma}_i)]^{-1}A_{i,p}^{\ q}(\mathbf{\Theta}_i,\mathbf{\Sigma}_i), \quad (3)$$

$$A_{i,p}^{\ q}(\mathbf{\Theta}_i,\mathbf{\Sigma}_i) = \mathbf{X}_{i,p}^{'\,q}\,\mathbf{\Sigma}_i^{-1}\,\left(\mathbf{Y}_{i,p}^{\ q} - \mathbf{X}_{i,p}^{\ q}\mathbf{\Theta}_i\right), \quad (4)$$

$$V_{i,p}^{\ q}(\mathbf{\Sigma}_i) = \mathbf{X}_{i,p}^{'\,q}\,\mathbf{\Sigma}_i^{-1}\,\mathbf{X}_{i,p}^{\ q} \quad (5)$$

for all $i = 1,\dots,N$, where $\mathbf{Y}_{i,p}^{\ q}$ and $\mathbf{X}_{i,p}^{\ q}$ are the endogenous variable and the explanatory variables, respectively, for all i starting from the time index p to q. The average statistic $\bar{S} = N^{-1}\sum_{i=1}^{N} S_{i,p}^{\ q}$ amounts to summing each individual $S_{i,p}^{\ q}$ statistic obtained from running the test on the separate time-series. There are two specific variants of $S_{i,p}^{\ q}$ that are used in calculating the standardised Z statistic, namely, $S_i^0 = S_{i,1}^{\ m_i}\left(\widehat{\mathbf{\Theta}}_{i,1}^{\ \bar{T}}, \widehat{\mathbf{\Sigma}}_{i,1}^{\ \bar{T}}\right)$ and $S_i^1 = S_{i,T_i+1}^{\ \bar{T}}\left(\widehat{\mathbf{\Theta}}_{i,1}^{\ \bar{T}}, \widehat{\mathbf{\Sigma}}_{i,1}^{\ \bar{T}}\right)$, where $\bar{T}$ is defined as $\bar{T} = T_i + m_i$ for simplicity. The post-break statistics, S_i^1, are computed for the sample spanning from $p = T_i + 1$ to $q = \bar{T}$, whereas the pre-break sample statistics, S_i^0, are calculated from $p = 1$ to $q = m_i$. Both set of statistics are computed using m observations.

The estimated time-series covariance matrix derived in Andrews (2003) are used as a weight matrix, which estimates the individual i's variances and autocovariances. The covariance matrix is $\widehat{\mathbf{\Sigma}}_{i,1}^{\ \bar{T}} = (T_i+1)^{-1}\sum_{r=1}^{T_i+1}\left(\widehat{\mathbf{U}}_{i,r}^{\ r+m_i-1}\widehat{\mathbf{U}}_{i,r}^{'\ r+m_i-1}\right)$ and $\mathbf{U}_{i,r}^{\ r+m_i-1}$ is individual i's $m_i \times 1$ residual vector resulting from the i^{th} time-series regression, that is $\widehat{\mathbf{U}}_{i,r}^{\ r+m_i-1} = \left(\mathbf{Y}_{i,r}^{\ r+m_i-1} - \mathbf{X}_{i,r}^{\ r+m_i-1}\widehat{\mathbf{\Theta}}_{i,1}^{\ \bar{T}}\right)$. The coefficient vector $\widehat{\mathbf{\Theta}}_{i,1}^{\ \bar{T}}$ is the least square estimates of $\mathbf{\Theta}$ for individual i over the full temporal sample. Under the assumption of cross sectional independency $E\left[\mathbf{U}_{i,t}'\mathbf{U}_{j,s}\right] = 0$, for $i \neq j$ and $\forall\ s,t = 1,\dots,\bar{T}$. Hence, the covariance matrix can be computed for each individual and used respectively in each individual test.

If $m_i \leq d$, the projection matrix collapses to a $m_i \times m_i$ identity matrix and the $S_i(\mathbf{\Theta}_i,\mathbf{\Sigma}_i)$ statistic becomes

$$P_i(\mathbf{\Theta}_i,\mathbf{\Sigma}) = \left(\mathbf{Y}_{i,p}^{\ q} - \mathbf{X}_{i,p}^{\ q}\mathbf{\Theta}_i\right)'\ \mathbf{\Sigma}_i^{-1}\ \left(\mathbf{Y}_{i,p}^{\ q} - \mathbf{X}_{i,p}^{\ q}\mathbf{\Theta}_i\right) \quad (6)$$

2.2 The Z Statistic

The standardised test Z statistic to test for stability in panel data models is built by taking the difference of the post- and pre-break average statistics derived in the previous subsection. It can be written as

$$Z = \frac{\left(\bar{S}^1 - \bar{S}^0\right)}{\sqrt{Var\left(\bar{S}^1 - \bar{S}^0\right)}} \quad (7)$$

where $\bar{S}^1$ and $\bar{S}^0$ are the average statistics for the post- and pre-break sample respectively. Intuitively, if the null hypothesis is true than Z will be centered around 0. However, under the alternative, the Z will centered further away from 0 and hence, the further away from 0 is the Z statistics, the more evidence against the null hypothesis in favor of the null.

The sample size used for $\bar{S}^0$ is the same as $\bar{S}^1$. It is recommended to use the first m_i observations to estimate $\bar{S}^0$ in order to minimise the potential impact of serial correlation in the errors by maximising the distance between the two subsamples. The size of m_i used to calculate the $\bar{S}^0$ can be increased if needed, at the cost of the potentially increasing problem of serial correlation. This is essentially an empirical issue, and any subsample selection problem affecting $\bar{S}^0$ should diminish as N increases, as discussed in the following section 3.

3 ASYMPTOTIC RESULTS

This section provides the asymptotic properties of the proposed test. Define the data set as a sequence of random variables $\{\mathbf{W}_{0,it}\}$ where $\{\mathbf{Y}_i, \mathbf{X}_i\} \subset \{\mathbf{W}_{0,it}\}$. Under H_0, the data are $\mathbf{W}_{0,it}$ for $i = 1, \ldots, T_i + m_i$, while under H_1 the data are $\mathbf{W}_{it} = \mathbf{W}_{0,it}$ for $i = 1, \ldots, T_i$ and $\mathbf{W}_{it} = \mathbf{W}_{T,it}$ for $i = T_i, \ldots, T_i + m_i$. $\{\mathbf{W}_{T,it} : i = T_i, \ldots, T_i + m_i\}$ are some random variables with a joint distribution different from $\{\mathbf{W}_{0,it} : i = T_i, \ldots, T_i + m_i\}$. Note that under H_1 the data are from a triangular array since the breakpoint is changing with $T \to \infty$.

Let $B(\mathbf{\Theta}_{0i}, \epsilon_T)$ be a ball centered around $\mathbf{\Theta}_{0i}$ with radius $\epsilon_T > 0$ as in Andrews (2003). For $m_i > d$, the following assumptions for the S_i^{ν}, $\nu = 0, 1$ are:

A1: $\{\mathbf{W}_{0,it} : t \geq 1\}$ $\forall$ i, is stationary and ergodic.

A2: (a) $\|\widehat{\mathbf{\Theta}}_{i,1}^{T_i+m_i} - \mathbf{\Theta}_{0i}\| \xrightarrow{p} 0$, with $(T_i, N) \to \infty$ with m_i fixed under H_0 and H_1. (b) $\sup_{\mathbf{\Theta} \in B(\mathbf{\Theta}_{0i}, \epsilon_T)} \|\widehat{\mathbf{\Sigma}}_{i,1}^{T_i+m_i} - \mathbf{\Sigma}_{0i}\| \xrightarrow{p} 0$ with $(T_i, N) \to \infty$ with m_i fixed under H_0 and H_1, for some nonsingular matrix $\mathbf{\Sigma}_{0i}$, for all sequences of constant $\{\epsilon_{T,N} : T \geq 1, N \geq 1\}$ and $\epsilon_{T,N} \to 0$ as $(T, N) \to \infty$.

A3: (a) $S_i(\mathbf{\Theta}_i, \mathbf{\Sigma}_i)$ is continuously differentiable in a neigbourhood of $(\mathbf{\Theta}_{0i}, \mathbf{\Sigma}_{0i})$ with probability one under H_0 and H_1, where $\mathbf{\Sigma}_{0i}$ is as in assumption 2(b). (b) Let $(\partial/\partial(\mathbf{\Theta}_i, \mathbf{\Sigma}_i^{-1}))$ denote the partial differentiation with respect to $\mathbf{\Theta}_i$ and the non redundant elements of $\mathbf{\Sigma}_i^{-1}$. S_i is bounded as

$$E \sup_{\substack{\mathbf{\Theta} \in B(\mathbf{\Theta}_{0i}, \epsilon_T), \\ \mathbf{\Sigma}_i \in N(\mathbf{\Sigma}_{0i})}} \|(\partial/\partial(\mathbf{\Theta}_i, \mathbf{\Sigma}_i^{-1})) S_i^0(\widehat{\mathbf{\Theta}}_i, \widehat{\mathbf{\Sigma}}_i)\| < \infty,$$

for some $\epsilon_T > 0$, where $\mathbf{\Sigma}_{0i}$ is as in assumption 2(b). $N(\mathbf{\Sigma}_{0i})$ denotes some neighbourhood of $\mathbf{\Sigma}_{0i}$. (c) The distribution function of $S_i^0(\widehat{\mathbf{\Theta}}_{0i}, \widehat{\mathbf{\Sigma}}_{0i})$ is continuous and increasing at its $1 - \alpha$ quantile, when $m > d$.

A4: (a) $E\left[U_{it}\mathbf{X}_{it}\right] = 0$, $\forall$ i and $t \geq 1$. (b) $E\left[U_{it}^2\right] < \infty$ and $E\|\mathbf{X}_{it}\|^{2+\delta} < \infty$ for some $\delta > 0$ and $\forall$ i and $t \geq 1$. (c) $E\left[\mathbf{X}_{it}\mathbf{X}_{it}^{'}\right]$ and $\mathbf{\Sigma}_0 = E\left[\mathbf{U}_{i,1}^{m_i}\mathbf{U}_{i,1}^{'\,m_i}\right]$ are positive definite, $\forall$ i and $t \geq 1$.

Theorem 1 *Under Assumptions 1 - 4, the Z statistic as described in equation (7)*

$$Z = \frac{\left(\bar{S}^1 - \bar{S}^0\right)}{\sqrt{\widehat{Var}\left(\bar{S}^1 - \bar{S}^0\right)}}$$

has an asymptotic distribution

$$\sqrt{N}\, Z \overset{A}{\sim} N\left(0, 1\right)$$

4 SIMULATIONS

4.1 Monte Carlo design

The experiment uses the linear regression model as follows:

$$\begin{aligned} y_{it} = \mathbf{\Theta}_i^{'}\mathbf{x}_{it} + u_{it}, \quad u_{it} = \rho u_{it-1} + \epsilon_{it} \\ \mathbf{x}_{it} = \tan(\mathbf{w}_{it}), \quad \mathbf{w}_{it} \sim NIID(\mu, \sigma^2) \end{aligned} \qquad (8)$$

The number of regressors is set to $d = 5$, which includes a constant, but does not include a lagged dependent variable. To start with, some benchmark results are generated in order to investigate the normality, the size and power of the test. Firstly, the set of Monte Carlo experiments simulate the null in order to analyse the size of the test. Moreover, a discussion of the properties of the distributions of the test under the LF-CLT is provided. The null hypothesis is simulated over the full sample $\bar{T}$ using the coefficient vector $\mathbf{\Theta}_{1i} = \mathbf{\Theta}_{0i} = 0$, $\forall$ i. Secondly, the power properties of the test are examined. The alternative hypothesis of partial instability is simulated, allowing for some individuals to experience a structural break while some do not. The ratio $\frac{N_1}{N}$ is gradually changed from .10, .50, .65, .80 and 1, in order to allow for a larger proportion of the individuals to experience a structural break. Furthermore, the alternative hypothesis featuring a partial structural break is simulated, using $\mathbf{\Theta}_{0i} = 0$, $\mathbf{\Theta}_{1i} = \frac{1}{10} \times (1, 1, 1, 1, 1)^{'}$, for some i and $\mathbf{\Theta}_{1j} = \mathbf{\Theta}_{0i} = 0$, for some $j \neq i$. The magnitude of the break is very small, equal to .1. Note that when $\frac{N_1}{N} = 1$, the coefficient vector is homogeneous across i's, implying that all individuals experience a structural break and

$\boldsymbol{\Theta}_0 = \boldsymbol{\Theta}_{0i} = 0$ prior to the instability, $\forall\ i$ and $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_{1i} = \frac{1}{10} \times (1,1,1,1,1)^{'}$ after the instability, $\forall\ i$.

To analyze the distribution, size and power of the test, the Monte Carlo experiments are conducted with the following settings: $m = m_i = \frac{1}{10} \times \bar{T}$, $\bar{T} = 30, 50$ and 100, $N = 20, 40, 60, 80$ and 100, where $\bar{T} = T + m = T_i + m_i$. For simplicity, the break dates T_i and post-break observations m_i are identical for all individuals. The regression's error term is generated with an AR(1) process with the following autoregressive parameters: $\rho = .4$ and $.95$ which is common to all individuals, or in other words, all individuals' errors have the same ρ. Four different types of iid distributions for the innovation of the error term are considered: standard $N(0,1)$, a recentered and rescaled χ^2_d and t_5, and an uniform distribution with support $[0,1]$. Different individuals have different innovation processes, such that the four distributions are intermixed evenly in the panel. The number of replications equals 2000. Note that the alternative hypothesis is fixed for all individuals, implying that we test for a break occurring at a known fixed date. All simulations were carried out using Ox 4.02.[1]

4.2 Monte Carlo results

The first results look at the probabilities of a type I error with normal significance level of $.05$. The main results can be summarised as follows: (i) Overall the Monte Carlo experiments reveal that the test statistic is close to Normal with 2000 replications showing that the LF-CLT hold, under moderate serial correlation and relatively small time dimension; (ii) The size of the test is relatively close to the desired value of $.05$ for $N > 20$ when $\bar{T} = 100$ and $m = 10$ when serial correlation is moderate, $\rho = .4$. Moreover, the test has reasonable size when the time horizon is decreased to $\bar{T} = 50$ and $\bar{T} = 30$ given the DGP, and it is relatively unaffected if the number of post break observations are increased to 20 % of $\bar{T}$ and (iii) The normality of the distribution worsen in the presence of extreme serial correlation. The mean grows substantially, the distribution is skewed and the variance shrinks below 1, as expected. The size, on the other hand, deteriorates as the number of individuals increase such that its largest value of $.199$ is attained for $N = 100$ and $\bar{T} = 100$. This result is expected as all individuals are required to exhibit the same high degree of serial correlation, and the impact of serial correlation is compounding as N increased with fixed $\bar{T}$. Moreover, increasing $\bar{T}$ from 100 to 250 observations or more improves the size, as implied by ergodicity.

[1]The programming code is available upon request.

In sum, the test has reasonable size in small temporal and individual sample with moderate serial correlation. However, under extreme serial correlation the size of the test deteriorates substantially.

Overall the test has good power. The power of the test is analysed for a normal significance level of $.05$. The most important results of the Monte Carlo experiments are as follows: (i) The power of the test varies little with $\bar{T}$ except when it is very small ($\bar{T} \leq 50$). This underlines the advantage gained by working with a panel structure. The test remains powerful even if $\bar{T}$ is decreased to 50: when $c = .80$ and $N = 80$, the power is $.65$. The power of the test will also improve if the number of post break observation increases (see Chan et al. (2007)); (ii) The test gains power as either N or c increases. The power of the test is above $.90$ when N and c are high. The power is still good when both c and N are of medium size and $\bar{T} = 100$; for example when $c = .65$ and $N = 60$ the power is $.85$. Moreover, the power of the test is good when N is high (100) and c is low (.50) with $\bar{T} = 100$. The reverse is also true: when $N = 40$ and $c = .80$, the power is $.91$ and (iii) The power of the test is quite robust to serial correlation, especially when N and c are large. Even in extreme cases when serial correlation is $.95$, the test has power of $.71$ when $N = 100$ and $c = .80$.

Overall, the test seems quite powerful given the DGP when serial correlation is moderate. Lastly, the power of test increases as the magnitude of the break increases. The results of the Monte Carlo experiments are available upon request.

5 EMPIRICAL EXAMPLE

This section provides an empirical application to demonstrate the usefulness of the test. Since the publication of Rose (2000), the question on a common currency's effect on trade has been a focus of the trade literature. See Micco et al. (2003) and Flam and Nordström (2003) and Mancini-Griffoli and Pauwels (2006). Overall, the literature mostly found a positive effect starting between 1998 and 1999 of the order of 10 to 20%. Only Mancini-Griffoli and Pauwels (2006), by applying a modification of the Andrews (2003) test to homogeneous panel data (considering the alternative of a common effect of the Euro across all trading partners) shows that the effect is significant, and refines the conclusion by showing that the break first occurred in the growth rate of trade around 1998, and only had noticeable repercussions on the level of trade around 2002.

This section extends the findings of Mancini-Griffoli and Pauwels (2006) to test for a break in the trade between only some of the Eurozone countries. Indeed, while Mancini-Griffoli and Pauwels (2006) can be seen as providing a methodological advancement with respect to the existing literature, the paper is limited by the assumption that all individuals (unilateral trading partners) exhibit the same break in trade due to the Euro. But intuition tells us the contrary. For instance, while it was clear that Germany was going to play a central role in the Euro project from its inception, it was uncertain whether Italy would meet the strict accession requirements almost until the Euro's introduction. It would therefore seem natural that each country's trade pattern would have responded differently, if at all, to the new currency. Thus, the particular test developed in this paper appears especially fitting.

Consider the typical model of trade between countries:

$$\begin{aligned} V_{i,jt} &= \alpha_{i,j} + \lambda_t + \varphi_j + \gamma_1 Y_{it} + \gamma_2 Y_{jt} + \\ &\quad \gamma_3 \xi_{i,jt} + \epsilon_{it} \qquad (9) \\ \Delta V_{i,jt} &= \rho \widehat{\epsilon}_{it-1} + \delta_1 \Delta Y_{it} + \delta_2 \Delta Y_{jt} + \delta_3 \Delta \xi_{i,jt} \\ &\quad + \gamma_1 \Delta V_{\bullet t} + \gamma_2 \Delta Y_{\bullet t} + \nu_{it} \qquad (10) \end{aligned}$$

where $V_{i,jt}$ is the value of imports from country j to country i, Y_{jt} and Y_{it} are nominal GDP, $\xi_{i,jt}$ is the real exchange rate between the two countries engaged in trade, ϵ_{it} is a regression error, $\alpha_{i,j}$ is a pair-specific fixed effect to control for variables of type common border, language, history, legal system, distance and others traditionally shown to matter in gravity equations, λ_t allows for the country pair intercept to be time dependent, and φ_j is a country of origin dummy as used in Rose and van Wincoop (2001).

The model is estimated using difference from sample mean fixed effects and the second using pooled OLS and where the $\bullet$ notation indicates sample average over the subscript it replaces. These additional variables are included to remove cross sectional dependence in the errors, as described in Pesaran (2006) as the common correlated effect (CCE) estimator. The goal is to control for the common factor causing cross correlations, but because it is unknown, Pesaran suggests it can be proxied by a linear combination of the sample averages of the regressors and regressand. Mancini-Griffoli and Pauwels (2006) show that using the additional CCE terms do indeed get rid of cross sectional dependence and prove that this condition allows for the inversion of the covariance matrix in the Andrews (2003) test statistic.

For the regressions, the quarterly data were obtained from Eurostat, IMF DOTS and IFS, as in most other relevant empirical papers. The unilateral import values are used as trade data, obtained from IMF DOTS. Finally, all the data are seasonally adjusted using the standard X.12 smoothing algorithm.

Recall that the Z test statistic in this paper is standardised with the $\bar{S}^0$ average statistic calculated within the pre-instability sample. This paper suggested to estimate the latter from the very beginning of the sample. The Euro data set, however, is not very reliable over the first 4 years (16 quarters) of data, as several series are extrapolated using moving averages from yearly data to fill in some missing observations. Hence, to overcome some of these shortcomings, the benchmark $\bar{S}^0$ is found using data starting from 1985 Q1. This still ensures a large period between the pre-break subsampling and the post-break observations thereby minimizing disturbances due to serial correlation. The particular choice of date to anchor the $\bar{S}^0$ statistic will be tested for robustness.

The empirical results can be summarised as follows: First, recall that Mancini-Griffoli and Pauwels (2006) finds a break in the growth of trade (in the ECM) in 1998Q1 at the 10% significance level and that this break appears to last 6 quarters. Note that the break date is defined as the first quarter for which the null of stability is rejected for 6 quarters with at least 10% significance. Thus, it is reassuring to note that the break date seems to be robust to the alternative hypothesis of a heterogeneous break. Indeed, the null of stability is rejected, this time at the 1% level, for a break in 1998Q1 and lasting 6 quarters. Second, the break appears to last longer with this paper's test - up to 12 quarters (or 3 years). Note that the length of a break is found by fixing the break date and repeating the test while adding on quarter to the post-break sample period with each iteration. These results are also as expected. As different Eurozone countries reacted differently to the Euro (some exhibiting a break and some not, or perhaps a much shorter one), the alternative hypothesis of a common break across the board is restrictive and probably only fitting for a few quarters. On the contrary, the more accommodating and realistic alternative of a heterogeneous break is accepted for a longer time period. Fourth, it is encouraging to note that the test results are barely sensitive to the choice of pre-break sampling date. This is as argued in the paper and stands given the stability and ergodicity of the pre-break data. In summary, it appears that indeed, the introduction of the Euro did have a noticeable impact on intra Eurozone trade, as anticipated by the original Rose (2000) hypothesis. What, exactly, in the new currency caused this rise in trade, is another question well worth considering in other research. But at least, end of sample instability tests, like the one presented here, lay solid and precise foundations for such research to continue its course.

6 CONCLUDING REMARKS

This paper built a stability test for heterogeneous panel data in the light of the IPS test for unit roots, to test the null of stability for all cross sections versus the alternative that some cross sections experience the instability and some do not. The test statistic is constructed as a standardised average of independent test statistics computed for each cross section. Asymptotic results show that the test is Normally distributed as per the Lindeberg-Feller central limit theorem.

Monte Carlo results show that the test performs well in terms of power and size, even when the time and individual dimensions are small. The results show that the test performs relatively well in the presence of serial correlation in the errors and that the results can be improved by increasing the time dimension. These results allow the test to be used widely in finance and economics applications. This paper explored one particular example, showing the existence of a trade effect of the Euro's introduction.

7 ACKNOWLEDGEMENT

The authors would like to thank Jushan Bai, In Choi, Hans Genberg, Michael McAleer, Hashem Pesaran and Michael Salemi for insightful discussions, constructive comments and suggestions. The authors would also like to thank the participants of the Third Symposium on Econometric Theory and Applications (SETA), 2007 Far Eastern Meeting of Econometric Society and The 14^{th} International Conference on Panel Data as well as the seminar participants at HKUST and HKIMR. This paper was written while the third author was a visiting fellow at the School of Economics and Finance at Curtin University of Technology, Perth, Australia. The hospitality of the school is gracefully acknowledged. The views expressed in this paper are those of the authors, and do not necessarily reflect those of the Hong Kong Monetary Authority.

8 References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Andrews, D. W. K. (2003). End-of-sample instability tests. *Econometrica*, 71(6):1661–1694.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62:1383–1414.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78.

Chan, F. Mancini-Griffol, T. and Pauwels, L. (2007). Stability Tests for Heterogeneous Panel Data. Curtin University Working Paper.

Chow, G. C. (1960). Tests of equality between sets of coefficient in two linear regressions. *Econometrica*, 28:591–605.

De Wachter, S. and Tzavalis, E. (2004). Detection of structural breaks in linear dynamic panel data models. Working Paper No. 505, Queen Mary, University of London, Department of Economics.

Flam, H. and Nordström, H. (2003). Trade volume effects of the euro: Aggregate and sector estimates. Manuscript, Institute for International Economic Studies.

Han, A. K. and Park, D. (1989). Testing for structural change in panel data: Application to a study of u.s. foreign trade in manufacturing goods. *Review of Economics and Statistics*, 71:135–142.

Im, K. S., Pesaran, H., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115:53–74.

Kao, C., Trapani, L., and Urga, G. (2005). Modelling and testing for structural brekas in panels with common and idiosyncratic stochastic trends. Syracuse University, Department of Economics.

Mancini-Griffoli, T. and Pauwels, L. L. (2006). Did the euro affect trade? answers from end-of-sample instability tests. HEI Working paper, Graduate Institute of International Studies, Geneva, Economics Section.

Micco, A., Ordoñez, G., and Stein, E. (2003). The currency union effect on trade: Early evidence from emu. *Economic Policy*, 18(37):316–356.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Rose, A. (2000). One market one money: estimating the effect of common currencies on trade. *Economic Policy*, 15(30):7–45.

Rose, A. K. and van Wincoop, E. (2001). National money as a barrier to international trade: The real case for currency union. *American Economic Review, Papers and Proceedings*, 91(2):386–390.