DEPARTMENT OF ECONOMICS
WORKING PAPERS

economics.ceu.edu

Testing for Unit Roots in Panel Data with Boundary Crossing Counts

by

Peter Farkas[1]
Central European University
and
Laszlo Matyas
Central European University

November 3, 2015 8.30 am

2015/5

[1] Corresponding author,Central European University, Department of Economics, peter.farkas@xlberg.com.

Abstract

This paper introduces a nonparametric, non-asymptotic method for statistical testing based on boundary crossing events. The method is presented by showing it's use for unit root testing. Two versions of the test are discussed. The first is designed for time series data as well as for cross sectionally independent panel data. The second is taking into account cross-sectional dependence as well. Through Monte Carlo studies we show that the proposed tests are more powerful than existing unit root tests when the error term has *t*-distribution and the sample size is small. The paper also discusses two empirical applications. The first one analyzes the possibility of mean reversion in the excess returns for the S&P500. Here, the unobserved mean is identified using Shiller's CAPE ratio. Our test supports mean reversion, which can be interpreted as evidence against strong efficient market hypothesis. The second application cannot confirm the PPP hypothesis in exchange-rate data of OECD countries.

JEL: C12, C23, C52, F14

Keywords: Nonparametric statistical testing, Panel data, Unit root, Mean reversion in financial markets, PPP hypothesis

1 Introduction

Technological innovation and the IT revolution have brought us into a new era of data abundance. This previously unseen richness of data creates an opportunity for nonparametric methods, especially in fields where there is a genuine need for flexible stochastic modeling.

In this paper, we introduce a new nonparametric method for hypothesis testing. We present it using the specific example of panel unit root testing, instead of discussing it under an abstract setting,

Depending on the underlying data generating process (DGP), the literature considers two basic model structures for unit root testing. The first one is suited for data without a deterministic trend, such as, for example, real exchange rates, inflation rates or interest rates, etc. The second one can be used to formalize DGP with deterministic trends, such as the GDP, etc. Let us take, as a starting point, the first structure without deterministic trend, which is also discussed in the review of Breitung and Pesaran (2008, p. 295):

$$X_{it} = \mu_i + \alpha_i X_{it-1} + u_{it} \tag{1}$$

where X_{it} are the data series to be analysed, μ_i are the individual-specific fixed effect and u_{it} are the composite error terms, with the number of cross sections being $i = 1, \dots, N$ and the number of time dimensions $t = 1, \dots, T_i$, that is we cater for unbalanced panels as well.

In line with the literature, as in Levin and Lin (1992), Maddala and Wu (1999) and Im et al. (2003), our null hypothesis assumes that all α_i are 1:

$$H_0 : \alpha_1 = \cdots = \alpha_N = 1 \text{ for all } i = 1, \dots, N \tag{2}$$

We consider the heterogeneous alternative where some, but not necessarily all, cross sections are stationary:

$$H_1 : \exists\ N_0\colon \alpha_{N_0} < 1,\ \ 0 < N_0 \leq N \tag{3}$$

As for the individual effects, they may or may not be zero under the null hypothesis, we come back to this issue later on.

As for the composite error terms, Hurlin et al. (2007) highlight two main approaches. First, authors may use various factor structures, such as in Choi (2006) or in Pesaran (2007). Others, for example Chang (2002), propose to work with the residuals covariance matrix and to rely on instrumental variables. Here, we apply a covariance matrix-based technique, but instead of relying on the residual covariance matrix, we work with the covariance matrix of variables describing boundary crossing events. The panel unit root testing literature, in general, is structured as follows (see Banerjee (1999), Baltagi and Kao (2001), Hurlin et al. (2007) and Breitung and Pesaran (2008)). First, there are first generation tests (see Levin and Lin (1992)), where the error terms of the model are assumed to be independent across i, and second generation tests (see Pesaran (2007)), where the errors terms are allowed to be contemporaneously correlated.[1] The independence assumption can be quite problematic, as cross-sectional dependence may arise in many applications, for example in the case of output growth equations, as in Pesaran (2004), or due to spatial dependence as in Baltagi et al. (2007).

Second, the alternative hypothesis may be homogeneous or heterogeneous. The former, used for example by Levin et al. (2002), assumes that $\alpha_1 = ... = \alpha_i = ...\alpha_n$. This homogeneous alternative is somewhat restrictive, for example in the case of convergence hypothesis for different countries in a macro model, it would imply that all countries or regions converge at

[1] A more elaborate classification of first and second generation tests is provided in Hurlin et al. (2007, p. 3, Table 1.).

the same rate if indeed they converge at all. Consequently, the less restrictive heterogeneous alternative which allow for cross-sectional differences has been introduced, for example, by Im et al. (2003).

Finally, tests differ in how they aggregate across different cross sections. There is in fact quite a variety of different aggregational techniques to combine individual cross sections. Maddala and Wu (1999), for example, suggest making use of the early results of meta analysis described in Tippett et al. (1931) and Fisher (1932), or more recently by Wolf (1986) who combines individual significance levels. Alternatively, Im et al. (2003) propose to merge individual t-statistics. What we propose in this paper is to aggregate by counting the number of boundary crossing events. To summarize: The test we propose in this paper can be classified as a second generation unit root test, with heterogeneous alternatives, which aggregates across individuals using the number of boundary crossing events (to be introduced below).

The intuition behind the test is very simple. Let us assume that a series X_{it} is enclosed by an upper and a lower boundary. If the process is stationary, a boundary crossing event is less likely than if it is unit root, as the demeaned process, unlike the unit root one, has the tendency to return to zero. By counting the number of boundary crossings therefore we can distinguish between the two processes.

Formally, let us introduce a new class of discrete stochastic process called boundary crossing counting process or BCC process, Y_{it}, which counts the number of boundary crossing events. Let U_i be some upper boundary and L_i be some lower boundary (the decision on the boundaries will be discussed in the next section). Also, upon each boundary crossing, the underlying stochastic process is restarted at some restarting value, X^0_{it}. (the choice of this restarting value will also be detailed in the next section). Let the restarted process be denoted by X^*_{it}. Note that the process may be restarted several times.[2] Let us differentiate between the following counting processes.

1. $Y^U_{it}(X^*_{it})$ counts the number of upper crossing events, that is how many times X^*_{it} needs to be restarted after an upper-crossing event.

2. $Y^L_{it}(X^*_{it})$ counts the number of lower crossing events, that is how many times X^*_{it} needs to be restarted after a lower-crossing event.

3. $Y^A_{it}(X^*_{it}) = Y^U_{it}(X^*_{it}) + Y^L_{it}(X^*_{it})$ counts all crossing events.

4. $Y^D_{it}(X^*_{it}) = Y^U_{it}(X^*_{it}) - Y^L_{it}(X^*_{it})$ is the difference between the number of upper and lower crossing events.

5. Finally, sometimes there is a need to refer to all of these processes at once, in this case we use the notation $Y_{it}(X^*_{it})$.

Thus, the counting process is a function of the restarted process, $Y_{it}(X^*_{it})$. Also, the restarted process is a function of the underlying data, the two boundaries and finally the restarting value, $Y_{it}(X^*_{it}(X_{it}, U_i, L_i, X^0_{it}))$. These dependencies are suppressed in the rest of the paper for ease of notation.

Our approach has several desirable properties. Besides the usual favorable properties of nonparametric tests, our method is non-asymptotic (although we briefly discuss the large sample properties as well). Moreover, the technique can also be used in the case of unbalanced panels, or panels with

[2] Boundary classification may be found in Karlin and Taylor (1981, p. 234), where they differentiate between "regular", "absorbing", "natural" and "entrance" types. The type of boundary applied in our paper does not have a one to one correspondence to any of these cases: They could be called "restarting boundaries". If one must classify, restarting boundaries are attainable and regular boundaries, where the process is restarted upon boundary crossing events.

missing values, or when the data generating process is sampled with an uneven frequency. Also, the test is relatively powerful when the error term is not normal, for example, when it follows a t-distribution.

Naturally, the BCC test suffers from certain drawbacks. Similar to Fisher's exact test, the distribution of the test statistics is a discrete one. Consequently, selecting the usual 1%, 5% and 10% as critical value is somewhat problematic, and we have to make use of the closest available discrete value.

Our paper is structured as follows. Section 2 sets out the model and discusses some estimation issues. Section 3 compares the finite sample properties of this newly introduced test to other frequently used unit root tests using Monte Carlo simulations. Section 4 is dedicated to two applications. Section 5 discusses additional technical details while the last section concludes.

2 Testing for Unit Roots Using Boundary Crossing Events

In this section, we show how to use the number of boundary crossing events for testing for unit roots in panel data. We proceed with the derivation in two steps. First, we discuss how to construct the test statistics in an ideal case when errors are independent. Then we continue by extending the derivation for cases when this is not true.

We assume that the DGP is characterized by Equation 1. Moreover, let us assume that X_{it} starts from minus infinity. Also, for the time being, we assume that the individual effects are zero, the boundaries are chosen exogenously and they are symmetric, $L_i = -U_i$.[3] Moreover, the restarting

[3]These assumptions are discussed in detail in Section 5.

value is zero for all i and t, that is $X^0_{it} = 0$. Finally, let the restarted process be defined over $\widetilde{X}_{it} = X_{it} - X_{i0}$ for ease of notation.

2.1 Test Statistics in the Case of Independent Errors

Under these condition $Y^D_{ik-1} > 0$ implies that $\widetilde{X}_{it}$ moves in a positive range in between the two boundary crossing events, that is in between T^*_{ik-1} and T^*_{ik}. Also, if $Y^D_{ik-1} < 0$, then $\widetilde{X}_{it}$ moves in a negative range. Finally, if $Y^D_{ik-1} = 0$, then $\widetilde{X}_{it}$ fluctuates around zero.

Furthermore, let Z_{ik} describe the k^{th} boundary crossing event in a way that $Z_{ik} = 1$ in case of an upper crossing and $Z_{ik} = -1$ is case of lower crossing. We aim to exploit the relationship between Y^D_{ik-1} and Z_{ik}.

Under the null hypothesis, $\Delta\widetilde{X}_{it} = \mu_i + u_{it}$. Consequently, the following upper-crossing probabilities are equal:

$$\underbrace{p(Z_{ik} = 1|Y^D_{ik-1} > 0)}_{p_{11}} = \underbrace{p(Z_{ik} = 1|Y^D_{ik-1} < 0)}_{p_{12}} \tag{4}$$

Likewise, the lower crossing probabilities below are also equal:

$$\underbrace{p(Z_{ik} = -1|Y^D_{ik-1} > 0)}_{p_{21}} = \underbrace{p(Z_{ik} = -1|Y^D_{ik-1} < 0)}_{p_{22}} \tag{5}$$

By combining Equation (4) and Equation (5), we obtain the following equality under the null:

$$H_0 : p_{11} + p_{22} = p_{12} + p_{21} \tag{6}$$

Under the stationary alternative hypothesis, $\Delta\widetilde{X}_{it} = \mu_i + (\alpha_i - 1)\widetilde{X}_{it-1} + u_{it}$. Since $(\alpha - 1) < 0$, a lower crossing event is more likely in case $Y^D_{k-1} > 0$, than in case $Y^D_{k-1} < 0$.

$$H_1 : \underbrace{p(Z_{ik} = -1|Y^D_{ik-1} < 0)}_{p_{11}} < \underbrace{p(Z_{ik} = -1|Y^D_{ik-1} > 0)}_{p_{12}} \tag{7}$$

Also, an upper crossing event is more likely in case $Y^D_{ik-1} < 0$, than in case $Y^D_{ik-1} > 0$.

$$H_1 : \underbrace{p(Z_{ik} = 1 | Y^D_{ik-1} < 0)}_{p_{21}} > \underbrace{p(Z_{ik} = 1 | Y^D_{ik-1} > 0)}_{p_{22}} \tag{8}$$

Consequently, the alternative hypothesis can be described as follows:

$$H_1 : p_{11} + p_{22} < p_{12} + p_{21} \tag{9}$$

In the data, we can observe five kinds of events, which are summarized in Table 1. We essentially differentiate between three cases. First, in the case of

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="3">Cumulative Upper minus Lower Crossing</td></tr>
<tr><td>$Y^D_{ik-1} < 0$</td><td>$Y^D_{ik-1} = 0$</td><td>$0 < Y^D_{ik-1}$</td></tr>
<tr><td rowspan="2">Next BC Event</td><td>$Z_{ik} = -1$</td><td>$E_{11} + 0.25$ (Divergence)</td><td rowspan="2">E_{00} (Non Informative)</td><td>$E_{12} + 0.25$ (Convergence)</td></tr>
<tr><td>$Z_{ik} = 1$</td><td>$E_{21} + 0.25$ (Convergence)</td><td>$E_{22} + 0.25$ (Divergence)</td></tr>
</table>

Table 1: *Contingency table based on boundary crossing events. E_{jk} indicates the number of events observed in the data. Note that 0.25 is added to each cells for technical reasons in order to avoid any division with zero.*

events E_{11} or E_{22}, $\widetilde{X}_{it}$ drifts further away from the origin, in other words, it diverges. Also, in the case of events E_{12} and E_{21}, it converges back towards the origin. Finally, in the case of events E_{00}, $\widetilde{X}_{it}$ is close to the origin, thus these boundary crossing events are considered to be noninformative in this regard and hence, they are not taken into account.

From what has been noted above, the right hand side of Equation (6) expresses the convergence probabilities, $p_c = p_{12}+p_{21}$, which can be estimated as follows:

$$p_c^{\#} = \frac{E_{12} + E_{21} + \frac{1}{2}}{E_{11} + E_{12} + E_{21} + E_{22} + 1} = \frac{B_c}{B_T}, \tag{10}$$

where $p_c^{\#}$ is the counting estimator, B_c is the number of convergence events and $B_T > 0$ is the total number of informative boundary crossing events.

To conclude, under the null hypothesis, the convergence probability is 0.5. Under the stationary alternative, the convergence probability is greater than 0.5. Note that we could also analyze the explosive alternative hypothesis, which would imply that the convergence probability is less than 0.5 as well as the joint stationary or explosive alternatives, which would imply that $p_c^{\#} \neq 0.5$, but these cases are not discussed due to space constraints.

We continue by discussing how to test the $p_c^{\#} = 0.5$ hypothesis. Essentially, the test statistics can be obtained by a quasi-binomial distribution as each boundary crossing event can be interpreted as a Bernoulli trial which takes the value of one upon convergence and the value of zero upon divergence.

The number of Bernoulli trials are $B_T = E_{11} + E_{12} + E_{21} + E_{22} + 1$, the number of successful trials are $B_c = E_{12}+E_{21}+0.5$ and the success probability is 0.5. The difference between our case and the pure binomial distribution is that here, the number of trials is stochastic. This can, however, be easily accounted for as the test distribution can be conditioned on the realized number of trials: the resulting conditional distribution is a binomial one,

$$(B_c|B_T) \overset{H_0}{\sim} Bin(B_c, 0.5|B_T), \tag{11}$$

where $Bin(.)$ denotes the binomial distribution. Since the stationary alternative states that $p_c^{\#} > 0.5$, the test is one sided.

2.2 Test Statistics in the Case of Dependent Errors

When the errors are dependent, the boundary crossing events can no longer be described by independent Bernoulli trials and hence, the binomial distribution cannot be used anymore. Yet, under cross-sectional dependence, the null hypothesis described in Equation 6 is still valid, only the variance of the test distribution is affected. Let us begin by dealing with cross-sectional dependence. We return to the question of autocorrelation later.

Potentially, there are three methods to adjust for cross-dependence. The first one is to rely on the law of dependent large numbers. The second one, which is somewhat theoretical in nature, aims to restore the independence of the Bernoulli trials by modifying the counting procedure. Finally, the last one is built on the fact that under the null hypothesis, the variance of the sum of the variables describing the individual trials can be estimated from the individual boundary crossing events. We continue our analysis with this last solution while the first two approaches are discussed in Section 5.

Next, we show how to capture the cross-dependence with the covariance matrix of the individual trials. First, for some boundary crossing k, let us define C_{ik} in the following way:

$$C_{ik} = -1 \quad \text{if} \qquad \begin{cases} Z_{ik} = 1 \text{ and } Y^D_{ik-1} > 0 \text{ or} \\ Z_{ik} = -1 \text{ and } Y^D_{ik-1} < 0. \end{cases} \tag{12}$$

Also, in case the boundary crossing event points toward convergence:

$$C_{ik} = 1 \quad \text{if} \qquad \begin{cases} Z_{ik} = 1 \text{ and } Y^D_{ik-1} < 0 \text{ or} \\ Z_{ik} = -1 \text{ and } Y^D_{ik-1} > 0. \end{cases} \tag{13}$$

Using this notation, the null hypothesis can be restated as

$$H_0 : \sum_{i=1}^{N} \sum_{k=1}^{B_i} C_{ik} = S_c = 0 \tag{14}$$

where B_i is the total number of quasi-Bernoulli[4] trials for cross-section i, that is for those boundary crossing events where $Y^D_{ik-1} \neq 0$. From now on, we refer to C_{ik} as convergence dummies and S_c as convergence sum. For the ease of notation, we suppress the indexes of the summations. Under the stationary alternative, the convergence sum is greater than zero:

$$H_1 : \sum \sum C_{ik} > 0 \tag{15}$$

Next, let us define C as follows:

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{T1} & c_{T2} & \cdots & c_{Tn} \end{pmatrix} \tag{16}$$

where the elements of C may be one, zero or minus one:

$$c_{it} = \begin{cases} 0 \text{ if } Y^A_{it} = Y^A_{it-1} \text{ or } Y^D_{it-1} = 0 \\ C_{ik(t)} \text{ otherwise.} \end{cases} \tag{17}$$

where the subscript in $C_{ik(t)}$ indicates that the k^{th} boundary crossing event occurs in time t. Also, let $\Sigma = C'C./(1_{n\times1}B)$, where $B = [B_1, B_2, ...B_N]$ describes the number of quasi-Bernoulli trials for each cross sections and ./ indicates element by element division. Note that if no boundary crossing events are observed for a particular cross section, then it needs to be removed from the sample. The variance of the test statistics can be expressed using Σ:

$$var(S_c) = E({S_c}^2) = E(B \times \Sigma \times 1_{n\times1})) \tag{18}$$

[4] In case of Bernoulli trials, the outcome is either +1 or zero. Here, the outcome is either +1 or -1.

where the first equality is due to the fact that $E(S_c) = 0$ under the null hypothesis, while the second equality is due to the fact that adding zeros to a sum does not modify its value.

We construct the empirical distribution by simulation. The idea is to make use of the fact that under the null, if we simulate $\sum(B_i)$ random numbers having zero mean and Σ as covariance matrix, then the sum of these simulated random numbers will have the same mean and the same variance as S_c. Hence, under the null, we can approximate the confidence interval for S_c using the sum of these simulated random numbers. In theory, we could draw the elements for the summation from an arbitrarily distribution, yet in practice, the convenient choice for this simulated distribution would be the normal one, or any other distribution with a closed form inverse.

Potentially, depending on the given application, there may be many different methods to carry out this simulation. Here, we show a solution based on the Cholesky decomposition. The algorithm to obtain the test distribution and it's critical value is the following:

1. Count the number of boundary crossing events. Multiple boundary crossings between two observations shall be recorded as two consecutive boundary crossing events.[5]

2. Estimate Σ from the sample.

3. Simulate 1000 correlated random numbers using normal distribution with mean zero and covariance matrix Σ and a sample size of B_T. In practice, especially when N is large and T is small, it may happen that Σ is not positive definite. Since we are dealing with matrixes containing

[5]This is in fact a sampling issue, the sampling frequency is not fine enough to record what happens between two observations.

only 1,0,-1, we are more and more likely to observe perfect dependence by chance as the number of cross sections increases even if the underlying DGP does not involves dependent cross sections. In practice, this can be corrected by finding the nearest symmetric positive semi-definite matrix as described by Higham (1988).[6]

4. The test statistics can be obtained by summing up each simulated sample, the critical value being the 95% percentile of these sums.

5. The null hypothesis is rejected if the convergence sum observed in the sample is larger than this critical value.

To sum up: the algorithm simulates random numbers having the same mean and variance as the sample convergence sum. Hence, although they may differ in higher moments, they are likely to be distributed similarly, and so the simulated sums can be used to approximate the true test distribution.

Regarding the BCC-test, there are a couple of additional issues of interest. First of all, the inference is based on constructing the empirical distribution of the test statistic by simulation. Hence, the covariance matrix is not necessarily generated under the null hypothesis because it is computed from the panel. Note that our test is very similar to many of the tests analyzed above as they all suggest capturing dependence based on the data.

Moreover, the dependence may also be captured using principle component analysis. In fact, the ideal method to capture dependence varies from application to application. If the method to capture the dependencies is inappropriate, then the BCC-test also suffers. Here, we only have room to describe one potential solution for illustrative purposes.

[6]The procedure was implemented based on John D'Errico's Matlab code. We would like to take this opportunity to acknowledge his contribution.

In addition, we use the normal distribution to construct the empirical distribution. Although we present a technical device to correct for possible singularity, if the matrix is near-singular then the normality assumption may not be appropriate.

3 Comparative Monte Carlo Analysis

This section compares the performance of the BCC test with other, commonly used unit root tests. It is important to mention that, strictly speaking, direct power comparisons between the different tests are not valid, since they have different null and/or alternative hypotheses. These differences are discussed in more detail in Section 5. Yet, we still present the actual size and the power of the different tests in one table, but these results need to be interpreted with some caution.

The design of the Monte Carlo experiments aims to simulate data for which nonparametric, in a distribution-free sense, methods may be reasonably applied. In particular, throughout the experiments, we assume that error terms have t-distribution with three degrees of freedom, which is approximately equal to the estimated degree of freedom of the daily log returns of the S&P500.

The DGP may be unit root or stationary with $\alpha_i = 0.9$, which is a typical choice also taken by Maddala and Wu (1999). For the stationary case, we assume that the initial observation is close to the long-term mean.

The Monte Carlo trials are repeated 2000 times, and for each repetition we carry out the tests at 5% significance level, the probability of rejecting the null hypothesis is obtained by dividing the number of cases when the null is rejected by the total number of Monte Carlo trials.

As for the BCC-test, we select boundaries according to the rule described in Equation (23). As explained in Section 5, this is a heuristic rule based on Monte Carlo experiments.

We begin by analyzing the case of time-series data, continue by first generation unit root test and conclude by the second generation tests. The simulations were implemented in Matlab, the time series tests used the build-in functions while the panel data tests were inspired by Hurlin's Matlab codes[7] for which we are very grateful.

3.1 Unit Root Tests for Time-Series Data

Time-series unit root tests are important for two reasons. First of all, the BCC test can also be used for time-series data. Also, panel unit root tests typically combine individual time-series tests. Hence, the performance of the time-series unit root tests provide some indication regarding the performance of those panel data tests, that are derived from them.

The BCC test is compared to several parametric unit-root tests. In particular, we compare the BCC test to the Augmented Dickey-Fuller test (Dickey and Fuller (1981)), further referred to as ADF test, to the Phillips-Perron test (Phillips and Perron (1988)), further referred to as PP test and finally to the variance ratio test (Lo and MacKinlay (1988)) further referred to as VR test. We implement these tests using the corresponding build-in Matlab functions.

The design of the experiment is as follows. We simulate 2000 sample paths, each consisting of either 50, 100 or 200 observations. As for the error term, u_t is assumed to follow a t-distribution whose parameters match the log-returns of the S&P500. Table 2 shows the rejection frequencies. The first

[7]The libraries were downloaded from the website of Orlean's University.

part of the table shows the actual size of the test, while the second part the power.

DGP			Unit Root Tests			
α_i	N	T	ADF	PP	VR	BCC
1	1	50	0.0535	0.0535	0.0510	0.0445
1	1	100	0.0455	0.0455	0.0505	0.0705
1	1	200	0.0605	0.0605	0.0480	0.0610
0.9	1	50	0.1085	0.1085	0.0565	0.1415
0.9	1	100	0.3210	0.3210	0.0500	0.3450
0.9	1	200	0.8865	0.8865	0.0915	0.4975

Table 2: *Monte Carlo results for time series unit root tests. The table shows the rejection frequencies for the BCC test as well as other, commonly used, time series unit root tests. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

Table 2 reveals that the BCC test performs better than the standard ADF and PP test in case the sample size is small, 50 in our case, while the ADF and the PP test performs better for larger sample sizes. Although the differentiating power in a small sample is modest, by combining multiple cross-sections in the case of panel-data, even this small difference may result in sizable gain of statistical power for the panel data case. This possibility will be explored in the next section. Also, the BCC test performs better than the Variance Ratio tests. Note that the Variance Ratio test has been primarily designed to identify heteroscedasticity and not to differentiate between unit root and near unit root processes.

Furthermore, the statistical power of the BCC test in larger samples may be improved further by perfecting the counting mechanism. The idea is as follows. Right now, we discard those boundary crossing events for which the Y_t^D is zero. Yet, one of the characteristics of stationary processes is that they are more likely to cross the long-term mean than a unit root process. Hence, the number of such events, E_{00} in Table 1, also contains information. Incorporating this information into the testing procedure may improve performance further, especially for larger samples.

To conclude, the BCC test is relatively powerful in case the small sample. For larger ones, tests based on the ADF regression dominate the presented version of the BCC test.

3.2 First Generation Panel Unit Root Tests

We begin by comparing the BCC test with three different first generation tests under cross-sectional independence. First, we consider several versions of Im et al. (2003)'s IPS tests. More specifically, we calculate the w-bar test which is based on the t-values, the t-bar test which is based on the moments of the DF distribution and finally the z-bar test which is based on the assumption of no autocorrelation of the residuals. Since the error term in the DGP is not autocorrelated, the results of w-bar, t-bar and z-bar tests should not be substantially different.

In addition, we consider two versions of Maddala and Wu (1999)'s test, further referred to as MW test, and Choi (2001)'s test, further referred to as CH test. The two tests differ in how they combine the individual p-values. In Maddala and Wu (1999), the p-values are calculated based on the critical values of Fisher's statistics, while in Choi (2001), they are based on the individual ADF statistics. Since these tests rely on meta-analysis-based techniques to combine p-values, from now on, we will refer to them as meta-analysis based tests. For both tests, we consider two versions. In the first one, the autocorrelation is estimated from the simulated data, while in the second one, the lag parameter is set to zero.

Finally, as for the BCC test, since the Bernoulli trials are independent, we rely on the binomial distribution-based version. The test statistics are obtained by pooling the number of boundary crossing events over the cross sections and the p-values are obtained from the right hand side of the corresponding binomial distribution.

As for the DGP, besides the baseline assumptions detailed at the beginning of this section, we assume that the cross sections are independent

DGP				First Generation Unit Root Tests							
f_i	α_i	N	T	IPS (w-bar)	IPS (t-bar)	IPS (z-bar)	MW (lag = 0)	MW (DF-lag)	CH (lag = 0	CH (DF-lag)	BCC
0	1	12	25	0.1030	0.1505	0.0605	0.0985	0.0325	0.1190	0.0460	0.0630
0	1	20	25	0.1015	0.1720	0.0590	0.1155	0.0430	0.1350	0.0515	0.0575
0	1	12	50	0.0550	0.0545	0.0575	0.0380	0.0405	0.0480	0.0505	0.0855
0	1	20	50	0.0595	0.0570	0.0580	0.0400	0.0420	0.0505	0.0505	0.0750
0	0.9	12	25	0.3455	0.4205	0.2830	0.2995	0.1635	0.3335	0.1960	0.5010
0	0.9	20	25	0.4300	0.5615	0.4175	0.3945	0.2320	0.4315	0.2665	0.6855
0	0.9	12	50	0.8440	0.8395	0.8230	0.6710	0.6230	0.7105	0.6735	0.9150
0	0.9	20	50	0.9745	0.9720	0.9660	0.8755	0.8365	0.8995	0.8615	0.9880

Table 3: *Monte Carlo results for first generation panel data unit root tests. The table shows the probability of rejecting the null hypothesis for the BCC test and other, commonly used, first generation panel data unit root tests in case of balanced panels and cross-sectional independence. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

Table 3 reveals that the BCC test has the highest power. As the sample size grows, the difference in the power between the existing tests and the BCC test diminishes. These results are in line with the findings of the time series analysis detailed in the previous subsection. Moreover, pooling the number of boundary crossing events over all cross-sections seems to be an effective aggregational technique. Combining relatively powerful individual tests in an effective way results in a panel data test which has high differentiating power.

As for the other non-BBC tests, in this particular Monte Carlo setup, the t-bar version of the IPS-tests has higher power than the other one, in small samples, even when T is small, and it also suffers from minor size-distortion. Also, the IPS tests typically show a somewhat higher differentiating power than the tests based on meta-analysis. As for the meta-analysis based tests, the test of Choi (2001) has somewhat higher differentiating power than the test of Maddala and Wu (1999). Finally, by providing additional information on the lag structure, the power of the meta-analysis based tests, especially in small sample, can be improved.

In an additional Monte Carlo experiments (results not presented here due to space constraints) we can show that missing data causes size-distortion in parametric first generation tests. In case of BCC test, some size-distortion is also present but to a much lesser degree. Under time series settings, the ADF test does not exhibit significant size distortion, hence the problem is probably caused by aggregational techniques.

Overall, the binomial BCC test has favourable properties, when the cross-sections are independent and the panel has missing observations.

3.3 Second Generation Panel Unit Root Tests

We continue by studying second generation unit root tests. Since Maddala and Wu (1999) already conducted a set of experiments in which the simulated data is spatially dependent, here, we focus on factor models. Table 1. in Hurlin et al. (2007)'s review differentiates between two main approaches to deal with cross-sectional dependencies. First, authors may use various factor models, such as in Choi (2006) or in Bai and Ng (2002). Others, for example Chang (2002), propose to work with the residuals' covariance matrix.

Factor models assume that the dependence is captured by one or more factors. In the case of a single common factor, Pesaran (2007) proposes to deal with cross-sectional dependence by further augmenting the Augmented Dickey-Fuller regressions by both the cross section average of the lagged levels and of the lagged first differences. These cross-sectionally augmented Augmented Dickey-Fuller equations, CADFs, are estimated by OLS, and the individual t–ratios of the OLS estimates are combined to obtain the test statistics. The advantage of this approach is its simplicity, while it may not be able to fully capture those cross-dependence structures which consist of several factors.

Multiple common factors are typically quantified using principle components. Bai and Ng (2004), for example, propose to first separate the common factors and the idiosyncratic terms and then to test them separately. The advantage of their method is that the properties of the common components may also be of economic interest, not just the those of the original data. Moon and Perron (2004), on the other hand, promote testing for unit roots on the de-factored series, which allows for a rather general specification of the common components. Since both multi-factor approaches described above rely on principle component analysis, the results may depend on the scale on which the variables are measured. Also, the differentiating power of these tests in finite samples when N is large and T is small may in some cases be limited.

Alternatively, cross-dependence may be captured via the covariance matrix. Fundamentally, the difficulty arises from the fact that the limiting distribution of the OLS or GLS estimators is dependent on certain nuisance parameters and hence, the usual Wald type of test cannot directly be applied. There are, however, some methods to overcome this problem. First, bootstrap-based estimators starting perhaps from Maddala and Wu (1999) may be used. Also, Chang (2002) proposes using a special, non-linear instrumental variable estimator and make use of the fact that the proposed individual IV estimates for the t–ratio statistics are asymptotically independent even for dependent cross-sectional units. Finally, Demetrescu et al. (2006) explain how to combine individual p–values in the cases where there is constant correlation among the p–values of the individual estimates.

In the next Monte Carlo exercise, we are interested in how tests perform under general conditions. Hence, we simulate data using the following multiple common factor model:

$$u_{it} = f_i^1 \times \Theta_t^1 + f_i^2 \times \Theta_t^2 + f_i^3 \times \Theta_t^3 + \epsilon_{it}, \tag{19}$$

where Θ_t^1, Θ_t^2, and Θ_t^3 are i.i.d. random unobserved common components and f_i^1, f_i^2, and f_i^3 are the factors. The random components are assumed to be drawn from a t-distribution with three degrees of freedom. As for the value of f_i^1, f_i^2, and f_i^3, we assume that they are randomly chosen from the uniform distribution centered around some predefined constants, detailed in Table 4. Hence, the loadings are different for each cross section.

	f_i^1		f_i^2		f_i^3	
	neg. dep.	pos. dep.	neg. dep.	pos. dep	neg. dep.	pos. dep.
w.f.d.	-0.2	0.4	-0.1	0.3	-0.3	0.2
s.f.d.	-0.6	0.8	-0.5	0.7	-0.75	0.7

Table 4: *Assumption for factor loadings on the individual cross sections in case cross-dependence arise out of multiple common factors.*

The first row, $w.f.d$, abbreviates weaker factor dependence while the second row, $s.f.d.$ abbreviates stronger factor dependence. For example, in the first row, for the first factor loading, half of the cross-sections are assumed to have f_i^1 around -0.2 while for the remaining cross sections, f_i^1 is assumed to be around 0.4. The first row models a case when more than half of the variation is driven by the idiosyncratic term. For the second row, most of the variation is driven by the unobserved common factors.

We compare the BCC test to the PANIC unit root test of Bai and Ng (2004), further referred to as BNG test, and the Cross-sectionally Augmented Dickey-Fuller test, or CADF test, of Pesaran (2007), further referred to as PS test. As for the former unit root test, once the factor structure has been removed, the p-values of the different cross sections are aggregated either by Choi (2001)'s method, shown in the first column, or by Maddala and Wu (1999)'s method, shown in the second column. As for the latter unit root test, the first column, titled $CIPS$, shows the cross-sectionally augmented version of the IPS test, which is based on t-bar statistics while the second column, titled $CIPS^*$, shows the suitably truncated version of the cross-sectionally augmented DF-statistics. As for the BCC test, we rely on the simulation-based version. The test statistics are obtained by pooling the convergence dummies from the cross sections. The p-values are obtained from the distribution of the simulated convergence sums. Table 5 shows the rejection frequencies. The first part of the table shows the actual size of the tests and the second part of the table shows the power. The nominal size is set at 5.0% for all tests.

DGP				Second Generation Panel Unit Root Tests				
f_i	α_i	N	T	BNG (Choi)	BNG (MW)	PS (CIPS)	PS (CIPS*)	BCC (sim)
w.f.d	1	12	25	0.066	0.057	0.045	0.041	0.096
w.f.d	1	20	25	0.046	0.039	0.054	0.050	0.077
w.f.d	1	12	50	0.043	0.036	0.042	0.040	0.088
w.f.d	1	20	50	0.032	0.025	0.046	0.045	0.105
w.f.d	0.9	12	25	0.251	0.217	0.094	0.094	0.520
w.f.d	0.9	20	25	0.416	0.381	0.129	0.129	0.692
w.f.d	0.9	12	50	0.827	0.784	0.331	0.331	0.906
w.f.d	0.9	20	50	0.974	0.967	0.570	0.570	0.971
s.f.d	1	12	25	0.072	0.061	0.084	0.084	0.097
s.f.d	1	20	25	0.047	0.040	0.100	0.097	0.105
s.f.d	1	12	50	0.060	0.046	0.081	0.080	0.109
s.f.d	1	20	50	0.044	0.037	0.093	0.093	0.134
s.f.d	0.9	12	25	0.276	0.241	0.149	0.149	0.482
s.f.d	0.9	20	25	0.388	0.353	0.225	0.225	0.557
s.f.d	0.9	12	50	0.801	0.768	0.405	0.405	0.820
s.f.d	0.9	20	50	0.963	0.949	0.571	0.571	0.884

Table 5: *Monte Carlo results for second generation panel unit root tests. The table shows the probability of rejecting the null hypothesis for the BCC test and other, commonly used, second generation panel data unit root tests for balanced panels in case dependence arises out of multiple common factors. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

The BCC test continues to be the most powerful when the sample size is small. For larger sample sizes, the BNG test dominates the BCC test in a sense that it has comparable power while the size of the BNG test is closer to the nominal size than the size of the BCC test.

As for the BNG test, Choi's aggregational technique is slightly more powerful than the alternative method. The test of Bai and Ng (2004) is more powerful than the test of Pesaran (2007) which is in line with the expectation as the former is a test specifically designed for multiple factor models.

As for the BCC-test, Table 5 suggest that its empirical size increases with the amount of information. This is related to how the dependence is captured. Hence for larger panels, we may need to develop additional techniques for capturing dependence. The use of principle component analysis for example could improve the properties further.

To conclude, the BCC test can be applied in cases when the cross-sectional dependence arises out of a common unobserved component. It dominates existing tests when the sample size is small. However, its performance weakens as the sample size increases. Hence, it may be reasonable to combine the BCC-technique with factor-analysis based procedures.

4 Empirical Applications

4.1 A Financial Application

Based on the first equation of Balvers et al. (2000), we analyze the possibility of mean reversion in financial markets. The starting point is as follows:

$$\log(R_{t+1}) - \log(R_t) = \mu + \beta(P_t - P_t^*) + u_t \tag{20}$$

where R_t is the market return, P_t^* is the long term mean or the equilibrium value of the market, which is unobserved and P_t is the price level. Moreover, we assume that u_t is independent but not necessarily identically or normally distributed. The null hypothesis is:

$$H_0 : \beta = 0 \tag{21}$$

The alternative hypothesis is:

$$H_1 : \beta < 0 \tag{22}$$

The fundamental problem is that P_t^* is unobserved. Hence, parametric estimation may be problematic. Our nonparametric method, however, seems to be useful in overcoming this problem. In our approach, we assume that we can infer $(P_t - P_t^*)$, that is the difference between the equilibrium value and the current market value, based on some fundamental measure, F_t. As for this fundamental measure, we use the price earnings ratio as suggested by Shiller (2005, p. 186).

Shiller states that there is probably a weak relationship between the price earnings ratio and the long-run stock returns. However, quantifying this relationship is problematic because the observations are overlapping.

Our approach for measuring this relationship differs from Shiller's as we do not work with overlapping observations for the returns. Instead, we measure returns using boundary crossing events. Consequently, we do not work with constant sampling frequency but rather with random sampling frequency.

Also, in our approach, we do not need to specify the exact relationship between the equilibrium value and this fundamental measure, it is sufficient to assume the following:

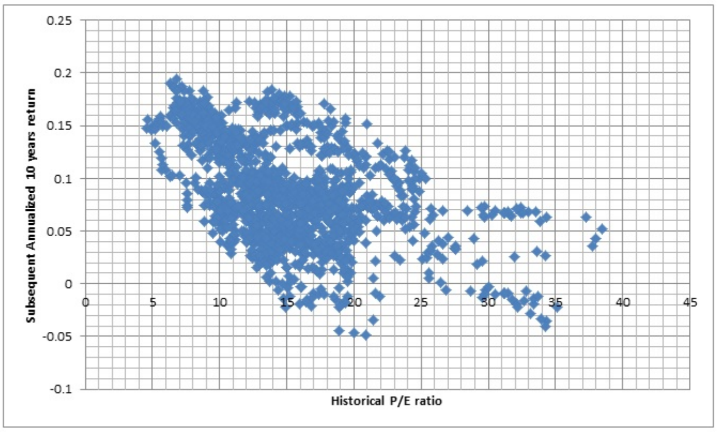

Figure 1: *P/E ratio and annualized returns. This figure is based on Shiller (2005, p. 186). The data is downloaded from Shiller's website. Note that observations are overlapping.*

Assumption 1 ***Market is overvalued,*** *that is* $P_t - P_t^* > 0$ *if the price earnings ratio,* F_t*, is sufficiently above its long-term average, that is* $F_t > \bar{F} + C$*, where* $\bar{F}$ *is the median P/E ratio and* C *is an exogenously chosen constant.*

If the price-earnings ratio is above its long-term average, then it is a sign that the market is overvalued, that is P_t is above its equilibrium value, P^*.

Assumption 2 ***Market is undervalued****, that is* $P_t - P_t^* < 0$ *if the price earnings ratio is sufficiently below its long-term average, that is* $F_t < \bar{F} - C$*.*

Likewise, if the price-earnings ratio is below its long-term average, the market is undervalued, hence P_t is below its equilibrium value, P^*.

Under the null hypothesis, the upper crossing probability is not influenced by the market fundamentals. Under the alternative hypothesis, an upper crossing event is more likely when the market is undervalued than when it is overvalued. Similarly, a lower crossing event is more likely when the market is overvalued than when it is undervalued. We test the null hypothesis using the convergence probability described in Equations (10).

The data is the monthly S&P500 obtained from Shiller's website[8]. We measure the returns as excess log returns. For the risk-free rate, we use the 10-year rate as provided by Shiller. As for the boundaries, $U = 0.2070 = 5 \times \sigma^{telr}$, where $telr$ indicates total excess log returns and $L = -0.2070$. For F_t, that is for the fundamental value, we use Shiller's cyclically adjusted price earnings ratio, or CAPE ratio, as calculated on the spreadsheet provided by Shiller. This is the 10 year moving average of the real price-earnings ratios. For $\bar{F}$, we use the median value of the CAPE ratios which is 16.0. We chose $C = 1.5$, that is we assume that the market is undervalued if the $F_t < 14.5$. Likewise, we assume that the market is overvalued if $F_t > 17.5$. The baseline results are shown in the following contingency table.

		Market valuation		
		undervaluation $(F_t < \bar{F} - C)$	neutral $(\bar{F} - C \leq F_t \leq \bar{F} + C)$	overvaluation $\bar{F} + C < F_t$
Next BC	$Z_k = -1$	8	7	10
Event	$Z_k = 1$	28	8	14

Table 6: *Stock market valuation and boundary crossing events. In total, we observe 75 boundary crossing events over the periods of 135 years. Thus, the average holding period is 1.80 years. The convergence probability is 0.63 which is significantly above 0.5 at 5% significance level, the p-value is 0.0259.*

The data rejects the null hypothesis. The convergence probability as

[8] `http://www.econ.yale.edu/~shiller/data.htm`

defined in Equation (10) is 0.63 which is significantly above 0.5. The p-value is 0.0259. In total, we observe 75 counting events over 135 years. Hence, the average holding period is 1.80 years which is much lower than the holding period of 10 years suggested by Shiller. We could increase the average holding period by applying wider boundaries. However, such an increase would reduce the number of boundary crossing events which would make inference more difficult.

The results can be interpreted as follows. In the past 135 years, under the current boundary setting, there were 25 occasions on which the excess return of the S&P500 over some random investment horizon were -20.7%. One could have avoided 17 occasions, that is approximately 68% of the cases by exiting the market when the CAPE ratio rises above 14.5. Of course, the price to pay for such market timing strategy is to avoid 22 occasions in which the market increased by 20.7%.

Let us conclude this application by analyzing to what extent the choice of parameters influences our results. We consider two factors: boundary selections and the choice of C for determining over and undervaluation. As for the choice on boundaries, we also consider upper boundaries placed at 4 and 6 standard deviation distance, that is for example, in the case of log returns to $U = 0.1657 = 4 \times \sigma^{telr}$ and $U = 0.2485 = 6 \times \sigma^{telr}$, $L = -U$. Finally, as for the choice on C, we also consider $C = 1$ and $C = 2$. The results are shown in Table 7.

Table 7 further confirms what Shiller proposes in his book: the relationship between the seasonally adjusted price-earnings ratio and the excess returns is probably significant. None of the settings accept the null hypothesis at 10% significance level. This finding may be interpreted as supportive evidence against strong efficient market hypothesis and in favour for funda-

p–values of the BCC-test			
Boundaries in	C		
standard deviation	1.0	1.5	2.0
-4,4	0.0272	0.0140	0.0147
-5,5	0.0178	0.0259	0.0240
-6,6	0.0871	0.0586	0.0266

Table 7: *Robustness exercise for the BCC-test on mean-reversion. The null hypothesis is rejected at 10% significance level regardless of how the parameters are chosen for the BCC test.*

mental analysis and market timing strategies.

4.2 PPP hypothesis

This section applies our new test to evaluate whether the PPP hypothesis holds. This is a common application for the above-reviewed panel unit tests, see in Chang (2002) or Pesaran (2007). We use the data of Pesaran (2007) as downloaded from the data archive of the Journal of Applied Econometrics for comparability. This panel covers the period of 1974 -1998 for 17 OECD countries. The test is applied to log real exchange rates which are computed as $x_{it} = s_{it} + p_{ust} - p_{it}$, where s_{it} is the log of the nominal exchange rate of the currency of country i in terms of US dollars, and p_{ust} and p_{it} are logarithms of consumer price indices for the United States and country i respectively.

We define the counting process over $\widetilde{x}_{it} = x_{it} - x_{i0}$. We chose the boundaries as described in Equation (23). We construct the empirical distribution by simulation as described earlier in the case of dependent errors. The results are shown in Table 8.

Based on the data, the unit root hypothesis cannot be rejected. There-

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="3">Cumulative Upper minus Lower Crossing</td></tr>
<tr><td>$Y_{ik-1}^D < 0$</td><td>$Y_{ik-1}^D = 0$</td><td>$0 < Y_{ik-1}^D$</td></tr>
<tr><td rowspan="2">Next BC Event</td><td>$Z_{ik} = -1$</td><td>161</td><td rowspan="2">93</td><td>102</td></tr>
<tr><td>$Z_{ik} = 1$</td><td>206</td><td>91</td></tr>
</table>

Table 8: *Testing the PPP hypothesis using boundary crossing events. The convergence probability is* 0.5500 *which is not significant at the usual significance levels. The p-value is 0.1722.*

fore our test does not support the PPP hypothesis. This result is in-line with some of the findings in the literature, while it contradicts others. For example, Pesaran (2007) found that the CIPS test does not reject the unit root hypothesis for the same dataset. On the other hand, using a similar dataset, the test of Chang (2002) strongly rejects the unit root hypothesis.

Our findings are robust to the parameter settings. In particular, as shown in Table 9, applying narrower or wider boundaries results in a similar conclusion.

L_i	U_i	convergence probability	p-value
-1	1	0.5419	0.1796
-1.18	1.18	0.5500	0.1722
-2	2	0.5749	0.1116

Table 9: *Sensitivity analysis for the BCC test on the PPP hypothesis. The null hypothesis of unit root cannot be rejected even if applying narrower or wider boundaries.*

Note that some caution may be needed when interpreting this result. In particular, the presence of autocorrelation, which is discussed in the next section, may influence this finding. Thus, further development of the test may be needed to fully confirm this result.

5 Discussion

Next, we consider some additional issues. First, we discuss how to set the boundaries. Then, we analyze the role of the individual effects. We continue by briefly discussing some methodological issues. Next, we analyze how to deal with autocorrelation and finally, we conclude with a brief discussion on the large sample properties.

5.1 Boundary Selection

Boundaries may be set exogenously or endogenously. The latter is well beyond the scope of our paper. Hence, we restrict our analysis to an illustrative Monte Carlo experiment.

The design of this Monte Carlo study is as follows. We simulate 2000 sample paths, each consisting of either 100, 500 or 1000 observations. We consider six different set of boundaries, measured in the standard deviations of Δx_t, ranging from $+/-1$ to $+/-6$. After counting the number of boundary crossing events, we obtain the p–values from the corresponding binomial distribution. We accept the null hypothesis if the p–value is larger than 5%. Rejection frequencies are calculated as the ratio of the number of cases when the null hypothesis is rejected and the total number of simulated sample paths. Hence, the first part of the table shows the actual size of the test, while the second part shows the power.

Table 10 reveals that as sample size increases, the ideal boundaries widen. For small sample sizes, wider boundaries are not practical since such setup does not generate enough boundary crossing events for inference.

DGP			Probability of Rejecting the Null Hypothesis					
			Lower boundaries at the standard deviation of Δx_{1t}					
α	N	T	[-1,1]	[-2,2]	[-3,3]	[-4,4]	[-5,5]	[-6,6]
1	1	100	0.0780	0.0490	0.0280	0.0060	0.0000	0.0000
1	1	500	0.1040	0.0600	0.0550	0.0500	0.0330	0.0310
1	1	1000	0.0920	0.0480	0.0410	0.0410	0.0430	0.0460
0.9	1	100	0.2600	0.1730	0.0280	0.0000	0.0000	0.0000
0.9	1	500	0.6530	0.6270	0.7170	0.7980	0.5620	0.1420
0.9	1	1000	0.6750	0.6570	0.7450	0.8390	0.8860	0.6520

Table 10: *Boundary selection for univariate unit root test. The table shows the rejection frequencies. The nominal significance level is 5%. Rejection occurs when the p-value is less than 0.05.*

Hence, neither the null, nor the alternative hypothesis can be rejected. For larger sample sizes, having enough boundary crossing event is less of an issue and hence, we can focus on having more informative events.

Characterizing the optimal boundaries analytically is beyond the scope of this paper, but clearly, the goal, as far as possible, is to minimize size distortion while maximize the power of the test. Here, we settle for the following heuristic rule.

$$U_i = \begin{cases} \sigma_i \text{ if } T < 100 \\ (1 + min(1, \frac{N}{100}) \times \frac{(T_i - 100)}{225}) \times \hat{\sigma_i} \text{ otherwise,} \end{cases} \tag{23}$$

where $\hat{\sigma_i}$ is the sample standard deviation for cross section i. The test provides reasonable differentiating power while, at the same time, it's empirical size is close to the nominal one. Hence we use this heuristic rule for setting up the boundaries.

5.2 The Role of Individual Effects

Let us begin by analyzing what happens if the individual effects are not zero under the null hypothesis of unit root. Table 11 summarizes the necessary additional notations.

	Cumulative Upper minus Lower Crossing	
	$Y^D_{ik-1} < 0$	$0 < Y^D_{ik-1}$
$Z_{ik-1} = -1$	$E_{11} = (\frac{1}{2} - \delta) \times B_T \times (1 - p_{21})$	$E_{12} = (\frac{1}{2} + \delta) \times B_T \times (1 - p_{22})$
$Z_{ik-1} = 1$	$E_{21} = (\frac{1}{2} - \delta) \times B_T \times p_{21}$	$E_{21} = (\frac{1}{2} + \delta) \times B_T \times p_{22}$
Total	$(\frac{1}{2} - \delta) \times B_T$	$(\frac{1}{2} + \delta) \times B_T$

Table 11: *The role of individual effect for the panel BCC test.*

Thus, δ captures the effect of μ_i. Let us substitute $E_{12} = (0.5 - \delta) \times B_T \times (1 - p_{22})$ and $E_{21} = (0.5 + \delta) \times B_T \times p_{21}$ in Equation(10).

$$p^{\#}_c = \frac{(0.5 - \delta) \times B_T \times (1 - p_{22}) + (0.5 + \delta) \times B_T \times p_{21}}{B_T} \tag{24}$$

Simplifying yields

$$p^{\#}_c = \frac{1}{2} + \frac{1}{2}(p_{22} - p_{21}) + \delta(1 - (p_{21} + p_{22})). \tag{25}$$

Under the unit root hypothesis, assuming away from autocorrelation for the time being, $p_{21} = p_{22}$. As for the individual effect of Equation (1), there are three cases.

1. If $\mu_i = 0$, then $\delta = 0$ and $p^{\#}_c = 0.5$.
2. If $\mu_i > 0$, then $\delta > 0$ and $(p_{21} + p_{22}) > 1$. Hence, $p^{\#}_c < 0.5$.
3. Finally, if $\mu_i < 0$, then $\delta < 0$ and $(p_{21} + p_{22}) < 1$. Hence, $p^{\#}_c < 0.5$.

To conclude, if the null hypothesis assume that $\mu_i = 0$ and in the true process, $\mu_i <> 0$, then the actual size of the BCC test will be less than the nominal size. Note that this is similar to other unit root tests which are based on the Dickey-Fuller asymptotics, since in this case, they use the DF-statistics when in reality, they should be using the standard OLS t-statistics. In both cases, the actual inference is unlikely to be negatively affected by the potential misspecification of the individual effect in the null hypothesis because economics theory, most of the time, postulates that a stochastic process is either stationary or unit root. If the true process is unit root with a drift then it is easier to identify the lack of stationarity.

Let us continue by analyzing the stationary case. Here, the critical assumption is that X_{it} starts from minus infinity which, for simplicity, is quantified as $X_{i0} = \mu_i/(1-\alpha_i)$. What happens if the process does not start from minus infinity? Substituting $X_{i0} = X_{i0} + \sum_{j=1}^{t-1} \Delta X_{ij}$ to Equation 1 results

$$X_{it} = \mu_i + \alpha_i(X_{i0} + \sum_{j=1}^{t-1} \Delta X_{ij}) + u_{it} \tag{26}$$

Assuming that $X_{i0} = (\mu_i + \gamma_i)/(1-\alpha_i)$, where γ_i captures the difference between the process initial value and its long-term mean, Equation (26) can be reformulated as follows:

$$X_{it} = \mu_i + \alpha_i(\frac{\mu_i + \gamma_i}{1-\alpha_i} + \sum_{j=1}^{t-1} \Delta X_{ij}) + u_{it} \tag{27}$$

Substituting $\mu_i = \mu_i \times (1-\alpha_i)/(1-\alpha_i)$ and $\widetilde{X}_{it} = X_{it} - X_{i0}$ results

$$\widetilde{X}_{it} = -\gamma_i + \alpha_i \sum_{j=1}^{t-1} \Delta X_{ij} + u_{it}. \tag{28}$$

Thus, if the process is assumed to start at minus infinity, that is $\gamma_i = 0$, then it is free of individual effects in the sample period. Consequently, $\delta = 0$ in Equation (25) and the convergence probability captures only the difference in p_{11} and p_{12}. On the other hand, if the process starts far away from the long-term mean, that is $\gamma_i <> 0$, then the individual effect is not zero, thus $\delta <> 0$ and this effects the convergence probability.

Note that depending on the initial value, X_{i0}, the stochastic process may exhibit fairly different behavior in finite samples, which is illustrated in Figure 2. In the left diagram of Figure 2, the stochastic process does not

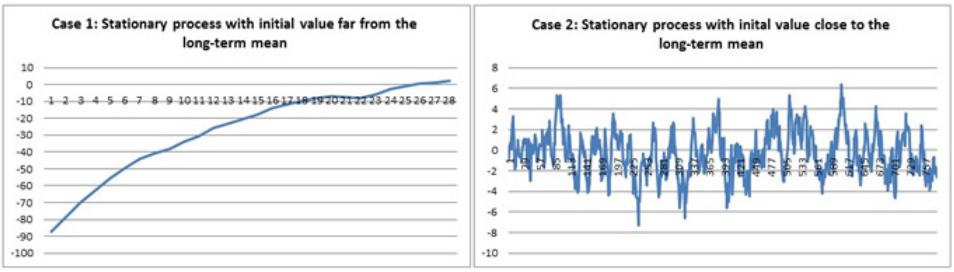

Figure 2: *Stationary process with an off-equilibrium and near-equilibrium initial value. The DGP for both cases is as in Equation* (1) *with* $\alpha_i = 0.9$. *In the first case, the stationary process is started from an off-equilibrium position while in the second case, the first observation is close to its long-term mean.*

appear to exhibit mean-reverting behavior. Since the BCC test is designed to measure the tendency to return to the mean, it would not reject the null hypothesis in this case. Note that economic theory often postulates the case shown in the right diagram of Figure 2: It not only predicts that a stochastic process is stationary, but the theory often also implicitly infers that it is close to its long-term mean. Hence, the BCC test incorporates this implicit assumption as well.

To conclude, the BCC test essentially quantifies the tendency to return to the mean. If the null hypothesis is accepted, then there is no such tendency, which implies that the DGP is either unit root process, or a stationary process for which the initial value is far away from the process's long term mean. In other words, the stochastic process has been initiated from an out of equilibrium position and it has not been measured for a sufficiently long time period for the process to return to its long-term mean. The BCC test is less suitable for differentiating between these two cases. If the null hypothesis is rejected, then there is a tendency to return to the mean and hence, the DGP is likely to be stationary near to its long-term mean.

5.3 Methodological Issues

In this paper we mirror closely the typical assumptions made by the panel unit root tests, such as in Im et al. (2003), Chang (2002) or in Breitung and Pesaran (2008), including their strengths as well as their weaknesses. Potentially, there are several meaningful research questions, for example:

1. Is the DGP unit root without deterministic trend or stationary process without deterministic trend?

2. Is the DGP unit root with drift or trend-stationary?

Although the second question is fully relevant, in this paper, we only deal with the first question due to space constraints. In the literature, it is common to begin the discussion with the first case as seen in Im et al. (2003), Chang (2002) or in Breitung and Pesaran (2008). Additional cases may be discussed in a separate paper.

When dealing with the first question, we make two implicit assumptions. First, we assume that the deterministic trend is excluded based on economic theory as well as based on the nature of the problem being modeled. Moreover, we assume that economic theory predicts stationarity. These implicit assumptions are claimed to be true for many commonly tested hypothesis such as the PPP hypothesis, for example.

Using Carl Popper's terminology, in order to test any theory, one needs to try to falsify it. Thus, in order to test for stationarity, we need to assume a non-stationary null hypothesis. There are many options for choosing this null hypothesis. We considered the one without individual effect because among all the non-stationary models described by Equation (1), it has the highest convergence probability. If $\mu_i \neq 0$ then as explained in Equation (25), the convergence probability is less than half. Consequently, if we reject the null hypothesis under the model with no individual effect then we would reject it in cases of positive or negative individual effect as well.

On the other hand, accepting the null hypothesis implies that there is at least one realistic non-stationary model which is supported by the data. Hence, economic theory suggesting stationarity cannot be confirmed. (It cannot necessarily be falsified either due to potential issues related to statistical power but that is a separate case.) Of course, there may be infinite many non-stationary models which could be rejected. However, using Popper's terminology, in order to falsify economic theory, it is sufficient to show one realistic counter-example.

5.4 Boundary Crossing Counts and Autocorrelation

Let us discuss the case when there is autocorrelation in the error term in Equation (1). Typically, such as in Bai and Ng (2002), Choi (2006), Im et al. (2003), Chang (2002), Levin et al. (2002), Maddala and Wu (1999) and Pesaran (2007), panel unit root tests are based on individual ADF tests and the number of lags are estimated from the data. This can safely be done as (Said and Dickey (1984)) the variable of interest in the ADF test and in the DF test have the same limiting distribution even in the case when the number of lags, m, is unknown, if $m^3/T \to \infty$ as $T \to \infty$. This result holds under more general conditions as well, as discussed in Chang and Park (2002). Alternatively, as in Moon and Perron (2004), instead of using the ADF test, one may incorporate the lag structure into the factor model as well.

In case of BCC test, based on Equation (25), autocorrelation effects the convergence probability in finite samples. In particular, in the case of positive autocorrelation, the convergence probability is less than 0.5 while in the case of negative autocorrelation, the convergence probability is greater than 0.5 under the null hypothesis in finite samples. Thus, autocorrelation may induce size distortion. Similarly, in the case of stationary process, autocorrelation may induce loss of power.

Let us outline two potential solutions to the problem of autocorrelation in finite samples. Note that we do not provide detailed solutions due to space constrains. The first option is to filter out the autocorrelation of the error term in Equation (1) and continue by carrying out the BCC test on the autocorrelation-adjusted data.

In order to implement this option, the first step is to assume that the autocorrelation structure can be captured by the following model:

$$\Delta X_{it} = \mu_i + (\alpha_i - 1)X_{it-1} + \rho_{i1}\Delta X_{it-1} + \rho_{i2}\Delta X_{i,t-2} + ... \rho_{im_i}\Delta X_{i,t-m_i} + \epsilon_{it} \quad (29)$$

where ϵ_{it} is free of autocorrelation and m_i is the number of lags. The convergence probability can be restored to 0.5 by first estimating the autocorrelation structure and second by defining the counting process over the autocorrelation-adjusted differences as shown in Equation (30).

$$\widetilde{X'}_{it} = \begin{cases} 0 \text{ if } t \leq m_i \\ \sum_{j=m_i+1}^{t} \Delta X_{ij} - (\rho_{i1}^E \Delta X_{it-1} + ... + \rho_{im_i}^E \Delta x_{i,t-m_i}) \text{ for } t > m_i. \end{cases}$$

(30)

where $\rho^E_{i1}, ...\rho^E_{im_i}$ are the estimated autocorrelation coefficients. This option may be problematic because, besides the usual issues such as selection of lags as discussed in Harris (1992), it requires a parametric estimation, which is somewhat alien to the original nonparametric philosophy of the test.

Alternatively, we can account for potential autocorrelation via the contingency table. The idea is similar to what has been discussed in Section 2, but instead of analyzing the relationship between the full history of the process, Y^D_{ik-1}, and the next boundary crossing event,Z_{ik}, as in Table 1, here, we analyze the relationship between the "immediate history", represented by the last, Y^U_{ik-1}, or last few, boundary crossing events, and the next boundary crossing event.

More precisely, the condition $Y^U_{ik-1} = 1$ implies that the DGP's immediate history was characterized by positive shocks. If there is no autocorrelation, this information should not affect the next boundary crossing event.

$$p(Z_{ik} = 1|Y^U_{ik-1} = 0) = p(Z_{ik} = 1|Y^U_{ik-1} = 1); \tag{31}$$

Likewise, the lower crossing probabilities below should also not depend on the previous boundary crossing events:

$$p(Z_{ik} = -1|Y^U_{ik-1} = 0) = p(Z_{ik} = -1|Y^U_{ik-1} = 1); \tag{32}$$

By combining these two equations, we obtain the following equality under the null:

$$\begin{aligned} p(Z_{ik} = 1|Y^U_{ik-1} = 0) + p(Z_{ik} = -1|Y^U_{ik-1} = 1) = \\ = p(Z_{ik} = -1|Y^U_{ik-1} = 0) + p(Z_{ik} = 1|Y^U_{ik-1} = 1) \end{aligned} \tag{33}$$

For ease of notation, let us introduce an additional variable which describes the effect of the process's immediate history.

$$A_{ik} = -1 \text{ if } \qquad \begin{cases} Z_{kt} = 1 \text{ and } Y^U_{ik-1} = 0 \text{ or} \\ Z_{kt} = -1 \text{ and } Y^U_{ik-1} = 1. \end{cases} \tag{34}$$

Also, the events of the right hand side of Equation (33) are denoted as follows:

$$A_{ik} = 1 \text{ if } \qquad \begin{cases} Z_{kt} = 1 \text{ and } Y^U_{ik-1} = 1 \text{ or} \\ Z_{kt} = -1 \text{ and } Y^U_{ik-1} = 0. \end{cases} \tag{35}$$

Using this notation, the null hypothesis, which is somewhat analogue with the null hypothesis of no autocorrelation in the parametric case, stating that the stochastic process's immediate history does not affect its next realization, can be described as follows:

$$H_0 : p(A_{ik} = 1) = p(A_{ik} = -1) = \frac{1}{2} \tag{36}$$

From now on, we refer to $p(A_{ik} = 1)$, as autocorrelation probability. Under the alternative hypothesis of autocorrelation, these probabilities are no longer equal, the stochastic process's immediate history has an effect on the next event.

$$H_1 : p(A_{ik} = 1) \neq p(A_{ik} = 0) \tag{37}$$

More precisely, in case of positive correlation, $p(A_{ik} = 1) > p(A_{ik} = 0)$, while in case of negative autocorrelation, $p(A_{ik} = 1) < p(A_{ik} = 0)$. All this is summarized in Table 12.

In case of the parametric approach, one can include additional lags into the autocorrelation structure. In this nonparametric, state-based approach, it is also possible to include additional states. Assuming we find significant autocorrelation probability in the above-described first step, we can check

		Last Boundary Crossing Event	
		$Y^U_{ik-1} = 0$	$Y^U_{ik-1} = 1$
Next	$Z_{ik} = -1$	$A_{ik} = 1$	$A_{ik} = -1$
BC Event	$Z_{ik} = 1$	$A_{ik} = -1$	$A_{ik} = 1$

Table 12: *Contingency table describing the effect of autocorrelation.*

for additional autocorrelation by adding an additional states. For example, we may ask if

$$p(Z_{ik} = 1 | Y^U_{ik-1} = 0, Y^U_{ik-2} = 0) \stackrel{?}{=} p(Z_{ik} = 1 | Y^U_{ik-1} = 0, Y^U_{ik-2} = 1)? \quad (38)$$

The null hypothesis in this case would state that once we have controlled for the immediate history of the process by controlling for Y^U_{ik-1}, the additional history represented by $Y^U_{ik-2} = 0$ does not affect the next boundary crossing probabilities significantly, while the alternative hypothesis would state that the additional history is of importance.

Finally, let us discuss how to combine the effect of the process's immediate and full history. More precisely, let us assume that we have found significant autocorrelation probability in the first state but did not find a significant relationship in the second state. Table 13 summarizes the potential states for this case.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="3">Cumulative Upper minus Lower Crossing</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>$Y^D_{ik-1} < 0$</td><td>$Y^D_{ik-1} = 0$</td><td>$0 < Y^D_{ik-1}$</td></tr>
<tr><td rowspan="4">Next Boundary Crossing Event</td><td>$Z_{ik} = -1$</td><td>$C^1_{ik} = -1$</td><td rowspan="4">Non Informative</td><td>$C^1_{ik} = 1$</td><td rowspan="2">$Y^U_{ik-1} = 0$</td><td rowspan="4">Previous Boundary Crossing Event</td></tr>
<tr><td>$Z_{ik} = 1$</td><td>$C^1_{ik} = 1$</td><td>$C^1_{ik} = -1$</td></tr>
<tr><td>$Z_{ik} = -1$</td><td>$C^2_{ik} = -1$</td><td>$C^2_{ik} = 1$</td><td rowspan="2">$Y^U_{ik-1} = 1$</td></tr>
<tr><td>$Z_{ik} = 1$</td><td>$C^2_{ik} = 1$</td><td>$C^2_{ik} = 1$</td></tr>
</table>

Table 13: *Contingency table describing the logic of the BCC unit root test in case of autocorrelation.*

The remaining steps are almost identical to what has been described above for the case of no autocorrelation. The null hypothesis states that:

$$H_0 : p(C^1_{ik} = 1) + p(C^2_{ik} = 1) = p(C^1_{ik} = -1) + p(C^2_{ik} = -1), \tag{39}$$

while alternative hypothesis states that

$$H_1 : p(C^1_{ik} = 1) + p(C^2_{ik} = 1) > p(C^1_{ik} = -1) + p(C^2_{ik} = -1). \tag{40}$$

In case of cross-sectionally independent error terms, the distribution of the test statistics can be calculated by using the fact that the sum of two binomial distributions is also binomial. In case of cross-sectional dependence, the simulation-based methods can be used. Further elaborating on the method outlined above is well beyond the scope of our paper. At this stage, we can conclude that potentially, we can adjust for autocorrelation in a nonparametric manner as well.

5.5 Large Sample Properties

Here, we briefly discuss the large sample properties of the BCC test under the assumptions made in Section 2. The structure of the problem in our case differs, to a certain extent, from a typical asymptotic analysis. Here, the properties of test statistics depend basically on the number of restarts, that is on the number of Bernoulli trials, which only indirectly depends on the sample size of the original data.

Fundamentally, we analyze the data in two steps. The first step is to characterize the original data using boundary crossing events. This step converts the original data (having some unknown distribution) to random variables having Bernoulli distribution. The second step is to estimate the convergence probabilities based on the re-sampled data.

The properties of this second-step estimator depend on the properties of the boundary crossing events, namely, on its dependence structure and on its sample size. These properties, in turn, depend both on the original data as well as on the counting procedure, specifically on how the boundaries are selected and how the counting is carried out. The structure of the problem is summarized in Table 14.

	Original DGP		Counting		BC events		Estimator of $p(C_{ik}=1)$
	Sample size	Dep. of u_{it}	Restart	Boundaries	Sample size	Dep. of C_{ik}	
1	$T \to \infty$ or $N \to \infty$	Ind.	Full	Constant	$B_T \to \infty$	Ind.	Consistent
2	$T \to \infty$ and N is finite	Strong crosss-dep.	Full	Constant	$B_T \to \infty$	Dep.	Consistent, as dep. LLN applies
3	$T \to \infty$ and N is finite	Strong cross-dep.	Rand.	Constant	$B_T \to \infty$	Ind.	Consistent
4	$N \to \infty$ and T is finite	Strong cross-dep.	Full	Constant	$B_T \to \infty$	Dep.	Not consistent, dep. LLN does not apply
5	$N \to \infty$ and T is finite	Strong cross-dep.	Rand.	Constant	B_T is finite	Ind.	Not consistent, sample size is finite
6	$N \to \infty$ and T is finite	Weak cross-dep.	Full	Constant	$B_T \to \infty$	Dep.	Consistent, as dep. LLN applies
7	T, N finite	Ind.	Full	$b \to \infty$	$B_T \to \infty$	Ind.	Probably inconsistent

Table 14: *What drives the large sample properties of the BCC test?*

Let us begin by the first case, when the error terms in Equation (1) are independent. In this case, the necessary condition for consistency requires that the re-sampled data's sample size goes in probability to infinity, when the underlying data goes to infinity. This is ensured if the probability of observing a boundary crossing event is positive and either $N \to +\infty$ and T is finite, or $T \to +\infty$ and N is finite, finally when $N \to +\infty$ and $T \to +\infty$ regardless of how N/T behaves.

There are basically two methods to deal with the problem of cross-dependence. The first method (cases 2, 4 and 6 in Table 14) is to carry out the counting the usual way, that is to restart the counting process immediately after the boundary crossing event has been observed. As a result, the re-sampled data will consist of dependent Bernoulli trials. The consistency of the convergence probability estimator in this case depends on whether the law of dependent large number applies, or not.

While, in a way, independence is unique, dependence comes in many different forms. Consequently, there are many different dependent LLNs, as found in Andrews (1988), Hansen (1991) or De Jong (1995). The fundamental idea behind these theorems is essentially very similar: if the dependence between the observations decrease sufficiently quickly as the distance between them increases then the LLN applies.

Such a decrease in dependency occurs in the second case due to the increasing time-distance and in case 6, when the error terms are by definition weak-dependent. This latter case may be used to model spatial dependence when the dependence between the cross-sections decreases as physical distance increases. On the other hand, the law of dependent LLN would probably not apply in case 4 when there is strong cross-dependence. Such strong cross-dependence may arise as a result of common unobserved factors.

So far, we have used general dependent LLN. Alternatively, it may be possible to make use of the specific law of large numbers for dependent Bernoulli trials. To our knowledge, such law does not exists under general specification, but special models have been analyzed. For example, James et al. (2008) examines a special case when the success probability of the trials is conditioned on the total number of successes achieved up to that point. Unfortunately, this model cannot be directly adapted to the case of BCC test as the counting procedure induces effects which are not captured by this model. Yet, a more general model along these lines may provide further insights on the large sample properties in the future.

The second method to deal with cross-dependence, shown as case 3 and 5 of Table 14, is to alter the counting procedure in a way that the resulting counting events consist of independent Bernoulli trials. The idea is as follows:

- Start the counting procedure at the first cross section.
- Assuming that we observe a boundary crossing event in time T_1^*, we do not restart the procedure for this cross-section but continue the counting on the next cross section. More precisely, we continue the counting on $\widetilde{X}_{2t} = \sum_{j=T_1^*+1}^{t} \Delta X_{2j}$.
- Likewise, each time we observe a boundary crossing event, the counting continues in the next cross section. Naturally, if the number of cross sections are finite and the counting on the last cross section is finished, then the counting continues in the first cross section again.
- Finally, if we do not observe boundary crossing for some cross-section, the counting also continues in the next cross section. Simply, such cross sections are ignored.

This procedure re-establishes the independence of the Bernoulli trials as the re-sampled variables are calculated based on observations which come from different time-periods and hence which are independent. Therefore, the only remaining necessary condition for consistency is to make sure that the counting does not stop so that the sample size of the re-sampled data goes to infinity as the sample size of the original data goes to infinity. This condition holds for example in case 3 but does not hold in case 5, when the time dimension is not sufficiently large.

Finally, let us examine case 7, which is interesting from a theoretical point of view. In this case, the sample size in the original data is finite, but $b \to 0$ where $U_i = b \times \sigma_i$ and $L_i = -U_i$. First of all, in order to be able to carry out the counting, we would need to assume that the underlying DGP is continuous from which we can obtain an infinitely fine sample. The dilemma is as follows. If $b \to 0$, then the conjecture[9] is that the sample size of the re-sampled data goes to infinity. Thus, we would be able to estimate the convergence probability with arbitrary precision. Such precision would contradict statistical intuition, as it is unlikely that the limitation posed by the lengths of the observation period could be overcome by re-sampling.

6 Conclusion

In this paper, we introduce a nonparametric method for statistical testing based on boundary crossing events. We present this method by showing how it can be used for unit root testing. We detail two versions of the test. The first one is designed for time series data as well as for cross sectionally independent panel data. The second one is able to take into account cross-sectional dependence as well.

[9] In order to verify this hypothesis, the first step would be to examine the limit of the first exit time distribution as the boundary goes to zero. The limiting function is likely to be similar to the Dirac delta function. The second step would be to show that the sample size of the re-sampled data converge in probability to infinity for any finite T, which would follow from the fact that boundary crossing counting distribution essentially involves repeated convolutions.

We use Monte Carlo studies to show that the proposed tests are more powerful than existing unit root tests in the case when the error term has a t-distribution and the sample size is small.

Two empirical applications are also considered. The first one analyzes the possibility of mean reversion in the excess returns for the S&P500. We identify the unobserved mean using Shiller's CAPE ratio. Our test supports mean reversion, which can be interpreted as evidence against strong efficient market hypothesis. The second application could not confirm the PPP hypothesis.

We have also identified several ways for further improvements. First of all, the counting procedure as well as the choice of boundaries can be improved. Moreover, our method may be combined with principle component analysis. Finally, besides unit root testing, our method can be used to solve other statistical problems as well.

References

Andrews, Donald WK, 1988, Laws of large numbers for dependent non-identically distributed random variables, *Econometric theory* 4, 458–467.

Bai, Jushan, and Serena Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191–221.

Bai, Jushan, and Serena Ng, 2004, A panic attack on unit roots and cointegration, *Econometrica* 72, 1127–1177.

Baltagi, Badi H, Georges Bresson, and Alain Pirotte, 2007, Panel unit root tests and spatial dependence, *Journal of Applied Econometrics* 22, 339–360.

Baltagi, Badi H, and Chihwa Kao, 2001, *Nonstationary panels, cointegration in panels and dynamic panels: A survey*, volume 15 (Emerald Group Publishing Limited).

Balvers, Ronald, Yangru Wu, and Erik Gilliland, 2000, Mean reversion across national stock markets and parametric contrarian investment strategies, *The Journal of Finance* 55, 745–772.

Banerjee, Anindya, 1999, Panel data unit roots and cointegration: an overview, *Oxford Bulletin of economics and Statistics* 61, 607–629.

Breitung, Jörg, and M Hashem Pesaran, 2008, Unit roots and cointegration in panels, in László Mátyás, and Patrick Sevestre, eds., *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (Springer).

Chang, Yoosoon, 2002, Nonlinear IV unit root tests in panels with cross-sectional dependency, *Journal of Econometrics* 110, 261–292.

Chang, Yoosoon, and Joon Y Park, 2002, On the asymptotics of ADF tests for unit roots, *Econometric Reviews* 21, 431–447.

Choi, In, 2001, Unit root tests for panel data, *Journal of international money and Finance* 20, 249–272.

Choi, In, 2006, *Combination unit root tests for cross-sectionally correlated panels* (Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter CB Phillips. Cambridge University Press).

De Jong, RM, 1995, Laws of large numbers for dependent heterogeneous processes, *Econometric Theory* 11, 347–358.

Demetrescu, Matei, Uwe Hassler, and Adina-Ioana Tarcolea, 2006, Combining significance of correlated statistics with application to panel data, *Oxford Bulletin of Economics and Statistics* 68, 647–663.

Dickey, David A, and Wayne A Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 1057–1072.

Fisher, R.A., 1932, *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 4th Edition).

Hansen, Bruce E, 1991, Strong laws for dependent heterogeneous processes, *Econometric Theory* 7, 213–221.

Harris, Richard ID, 1992, Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test, *Economics letters* 38, 381–386.

Higham, Nicholas J, 1988, Computing a nearest symmetric positive semidefinite matrix, *Linear algebra and its applications* 103, 103–118.

Hurlin, Christophe, Valérie Mignon, et al., 2007, Second generation panel unit root tests, *Working paper, https://halshs.archives-ouvertes.fr/halshs-00159842* .

Im, Kyung So, M Hashem Pesaran, and Yongcheol Shin, 2003, Testing for unit roots in heterogeneous panels, *Journal of econometrics* 115, 53–74.

James, Barry, Kang James, and Yongcheng Qi, 2008, Limit theorems for correlated Bernoulli random variables, *Statistics & Probability Letters* 78, 2339–2345.

Karlin, Samuel, and Howard M. Taylor, 1981, *A second course in stochastic processes*, volume 2 (Gulf Professional Publishing).

Levin, Andrew, and Chien-Fu Lin, 1992, Unit root test in panel data: asymptotic and finite-sample properties, *Discussion Paper No. 96-56, University of California at San Diego* .

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu, 2002, Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of econometrics* 108, 1–24.

Lo, Andrew W, and Archie Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of financial studies* 1, 41–66.

Maddala, Gangadharrao S, and Shaowen Wu, 1999, A comparative study of unit root tests with panel data and a new simple test, *Oxford Bulletin of Economics and statistics* 61, 631–652.

Moon, Hyungsik Roger, and Benoit Perron, 2004, Testing for a unit root in panels with dynamic factors, *Journal of econometrics* 122, 81–126.

Pesaran, M Hashem, 2004, General diagnostic tests for cross section dependence in panels, *Cambridge Working Papers in Economics, https://www.repository.cam.ac.uk/handle/1810/446* .

Pesaran, M Hashem, 2007, A simple panel unit root test in the presence of cross-section dependence, *Journal of Applied Econometrics* 22, 265–312.

Phillips, Peter CB, and Pierre Perron, 1988, Testing for a unit root in time series regression, *Biometrika* 75, 335–346.

Said, Said E, and David A Dickey, 1984, Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* 71, 599–607.

Shiller, Robert J., 2005, *Irrational exuberance* (Random House LLC).

Tippett, Leonard Henry Caleb, et al., 1931, *The methods of statistics.* (London: Williams & Norgate Ltd.).

Wolf, Fredric M, 1986, *Meta-analysis: Quantitative methods for research synthesis*, volume 59 (Sage).