Beyond Panel Unit Root Tests: Using Multiple Testing to Determine the Non Stationarity Properties of Individual Series in a Panel*

H.R. Moon†

University of Southern California

B. Perron‡

Université de Montréal, CIREQ, CIRANO

October 2009

Abstract

Most panel unit root tests are designed to test the joint null hypothesis of a unit root for each individual series in a panel. After a rejection, it will often be of interest to identify which series can be deemed to be stationary and which series can be deemed nonstationary. Researchers will sometimes carry out this classification on the basis of n individual (univariate) unit root tests based on some ad hoc significance level. In this paper, we demonstrate how to use the false discovery rate (FDR) in evaluating $I(1)/I(0)$ classifications based on individual unit root tests when the size of the cross section (n) and time series (T) dimensions are large. We report results from a simulation experiment and illustrate the methods on two data sets.

Keywords: False discovery rate, Multiple testing, unit root tests, panel data.

JEL classification: C32, C33, C44

*We have benefited from comments from the guest editors, three referees, and the participants at the conference in honor of Peter Phillips held at Singapore Management University in July 2008 and the CIREQ-CIRANO conference "The Econometrics of Interactions" on Oct. 23-24, 2009.

†Department of Economics, University of Southern California, Los Angeles, CA 90089, U.S.A. (213) 740-2108. E-mail: moonr@usc.edu. Financial support from the faculty development award of USC and the National Science Foundation is gratefully acknowledged.

‡*Dépt. de sciences économiques, Université de Montréal, CIREQ and CIRANO, C.P. 6128, Succ. centre-ville, Montréal, Québec, H3C 3J7, Canada. Tel. (514) 343-2126. E-mail: benoit.perron@umontreal.ca.* Financial support from FQRSC, SSHRC, and MITACS is gratefully acknowledged.

1 Introduction

Most panel unit root tests are designed to test the joint null hypothesis of a unit root for each individual series in a panel (see, for example, Breitung and Pesaran (2008) for a recent survey). This raises the issue of how to interpret a rejection of this null hypothesis. It it obviously unwarranted to treat all series in the panel as stationary in this case since this rejection only implies that a significant proportion of the series can be described as stationary. This paper examines how a researcher should proceed in identifying the individual series that can be deemed to be nonstationary and those that can be deemed stationary.

Often, researchers will carry out this classification in empirical work on the basis of n individual (univariate) unit root tests based on some ad hoc significance level. No statistical evaluation of the aggregated decision based on these n individual decisions is provided. To evaluate the aggregation of individual tests, this paper suggests the use of some concepts from the statistical literature on multiple testing. In particular, we will argue that the use of the false discovery rate (FDR) proposed by Benjamini and Hochberg (1995) provides a useful diagnostic on the aggregate decision. The FDR is the expected ratio of the number of falsely rejected null hypotheses over the total number of rejections. The FDR is interpreted as a posterior of the true null given a rejection of the null hypothesis, (see Storey (2003)).

The main contribution of this paper is to demonstrate how to use the false discovery rate in evaluating $I(1)/I(0)$ classifications based on individual unit root tests when he size of the cross section (n) and time series (T) dimensions are large. We suggest two approaches: the first one adjusts the critical value of the individual unit root tests to achieve a targeted FDR level, while the second approach estimates the FDR based on a fixed choice of level for the individual tests (for example, 5%).

Application of FDR as a controlling mechanism for our classification is faced with two difficulties. The first one is that FDR depends on the (obviously unknown) number of true null hypotheses. Thus FDR is not by itself an identified concept. We solve this problem in our context by the use of the Ng (2008) estimator of the fraction of nonstationary series. The second problem is the presence of cross-sectional dependence among the units in the panel. We solve this problem by applying a bootstrap procedure to estimate the distribution of p-values in the panel and thus control the FDR.

Alternative approaches to classifying the series among $I(0)$ and $I(1)$ components have been proposed. Kapetanios (2003) proposed to carry out a sequence of panel unit root tests on panels of decreasing size. One removes from the panel the series with the most evidence in favor of stationarity. One continues until the joint test of a unit root for the remaining series in the panel is no longer rejected. On the other hand, Ng (2008) estimates the fraction of nonstationary series. She conjectures that one can then identify the $I(1)$ and $I(0)$ series by ordering them according to the magnitude of their autoregressive parameter.

In independent work, Hanck (2009) uses multiple testing in the context of a mixed panel, but he focuses on the family-wise error rate (FWE), a concept that is less desirable when the number of tests performed (equal to the cross-sectional dimension in this case) is large.

The reminder of this paper is organized as follows: the next section describes the standard panel unit root testing problem, while section 3 presents the multiple testing methodology. Section 4 describes how one can control or estimate the false discovery rate. Section 5 presents simulation evidence that our proposal gives useful information. Section 6 reports results from two empirical applications. Finally, section 7 concludes.

2 Panel unit root testing problem

This section introduces briefly the panel unit root testing problem. A more exhaustive review can be found in Breitung and Pesaran (2008).

We suppose that we have panel data z_{it} of individual i that is observed at time t for $i = 1, ..., n$ and $t = 1, ..., T$. Hence, n and T denote the size of the cross section and time series dimensions, respectively. We model our panel using a decomposition among deterministic and stochastic components as:

$$\begin{aligned} z_{it} &= d_{it} + z_{it}^0, \quad (1) \\ z_{it}^0 &= \rho_i z_{it-1}^0 + y_{it}, \end{aligned}$$

where d_{it} is the deterministic component, and z_{it}^0 the stochastic component. Three basic models of the deterministic components are typically of interest: $d_{it} = 0 \; \forall i, t$, $d_{it} = \alpha_i$ (individual intercepts only), and $d_{it} = \alpha_i + \beta_i t$ (individual trends).

The null hypothesis of interest is that all stochastic components are nonstationary:

$$H_0 : \rho_i = 1 \text{ for all } i = 1, \ldots, n,$$

whereas the alternative hypothesis takes the form:

$$H_A : \rho_i < 1 \text{ for some } i,$$

where ρ_i is the largest autoregressive root in the time series of individual i.

Since a panel unit root test is a joint test, one cannot readily interpret a rejection. In particular, it does not provide any information on the properties of individual time series in the panel. Our goal is to identify the stationary series in the panel and provide a certain statistical evaluation of the identification based on the individual unit root tests in the panel.

3 Introduction to multiple testing: False discovery rate approach

In this section, we present briefly the multiple testing methodology; one can see Lehmann and Romano (2005) for further details.

Suppose that there are n separate testing problems that are either true null or true alternative hypotheses. The number of true null hypotheses will be denoted by n_0 and the number of false null hypotheses will be denoted by n_1. The outcome of each test is either to reject or not to reject the null hypotheses. The testing result can be summarized by the 2×2 table:

	# of nulls not rejected	# of rejected nulls	total
the null is true	$M_{0\|0}$	$M_{1\|0}$	n_0
the null is false	$M_{0\|1}$	$M_{1\|1}$	n_1
total	$n - R$	R	n

Thus, R nulls out of n are rejected, and among these R rejections, there are $M_{1|0}$ false rejections and $M_{1|1}$ correct rejections.

The familywise error rate is the probability that we incorrectly reject at least one true null hypothesis:

$$FWE = \Pr\left[M_{1|0} \geq 1\right].$$

When looking at a large number of tests, controlling the FWE becomes difficult and requires a decreasing level of individual tests as we increase the number of tests. In such cases, one is often willing to tolerate a few incorrect rejections. This leads to the $k-FWE$ which is the probability that we reject at most k true null hypotheses:

$$k - FWE = \Pr\left[M_{1|0} \geq k\right]$$

($e.g.$, see Lehmann and Romano (2005)).

In panel unit roots, we often look at a panel of increasing size, $n \to \infty$. Thus, for any fixed k, control of the $k - FWE$ will encounter similar difficulties as control of the simple FWE. It seems natural in this context to let the number of false rejections we are willing to tolerate tend to infinity and use as our control measure the false discovery proportion, i.e. the proportion of rejections that are false or, using the above notation,

$$\begin{aligned} FDP &= \frac{M_{1|0}}{R} \text{ if } R > 0 \\ &= 0 \text{ if } R = 0. \end{aligned}$$

Unfortunately, it is impossible to control this quantity. Instead, Benjamini and Hochberg (1995)'s proposal is to control the expectation of the FDP, which they call the false discovery rate (FDR), and which is defined as

$$FDR_P = E_P\left(\frac{M_{1|0}}{R} 1\left\{R > 0\right\}\right).$$

Although we will not consider this possibility here, one could also try to control the false non-discovery rate (FNR):

$$FNR = E_P\left(\frac{M_{0|1}}{n-R} 1\left\{n - R > 0\right\}\right)$$

which is the proportion of non-rejections that are coming from false null hypotheses or even a weighted average of these two quantities as in Storey (2003).

Storey (2003) provides an interesting Bayesian interpretation of the FDR in the context of a mixture model. Suppose that $H_i = 0$ (=1) if the i^{th} null hypothesis is true (or false) and let $H^m = (H_1, ..., H_m)'$. We denote by $\hat{p}_i$ the p-value associated with i^{th} individual unit root test. We know that if the i^{th} null hypothesis is true, then $\hat{p}_i$ has a uniform distribution on the $[0, 1]$ interval.

We suppose the random mixture model $(\hat{p}_i, H_i) \sim iid$ such that

$$H_i \sim B\left(1 - \pi_0\right)$$

$$\Pr\{\hat{p}_i \leq t\} = \pi_0 U(t) + (1-\pi_0) F(t) = G(t),$$

where $U(t) = t$ is the c.d.f. of a uniform distribution and $F(t)$ is the c.d.f. of p-values under the alternative. The variable π_0 can be interpreted as the probability that the null hypothesis is true, in which case the p-values are $i.i.d. U[0,1]$.

This result is exact if one uses the exact distribution of the test statistics under both the null and alternative hypotheses. In the case where asymptotic $(T \to \infty)$ approximations are used, this result is asymptotic and $F(t) \to 1$ for any consistent test. In this case,

$$G(t) = \pi_0 t + (1-\pi_0).$$

For a common size t for all n tests, the number of rejected null hypotheses is:

$$\begin{aligned} R &= \sum_{i=1}^{n} 1\{\hat{p}_i \leq t\} \\ M_{1|0} &= \sum_{i=1}^{n} 1\{\hat{p}_i \leq t\}(1-H_i) \end{aligned}$$

and one can express the false discovery proportion as

$$FDP\left(t\right)=\frac{\sum_{i=1}^{n}1\left\{\hat{p}_i\leq t\right\}\left(1-H_i\right)}{\sum_{i=1}^{n}1\left\{\hat{p}_i\leq t\right\}+1\left\{\hat{p}_i>t\text{ for all }i\right\}}.$$

where the second term in the denominator avoids division by 0.

When the number of tests n is large,

$$FDP\left(t\right)\overset{p}{\rightarrow}\frac{\pi_0 t}{G\left(t\right)}=E\left(FDP\left(t\right)\right) \tag{2}$$

This limit can be re-expressed as:

$$\begin{aligned}\frac{\pi_0 t}{G\left(t\right)} &= \frac{\pi_0 U\left(t\right)}{\pi_0 U\left(t\right)+\left(1-\pi_0\right)F\left(t\right)}\\ &= \frac{\Pr\left\{\text{Reject the null }|H_i=0\right\}P\left\{H_i=0\right\}}{\Pr\left\{\text{Reject the null}\right\}}\\ &= \Pr\left\{H_i=0|\text{ Reject the null}\right\}.\end{aligned}$$

So, the FDR is the posterior probability of the null being true given that we have rejected a particular null hypothesis as the number of tests $n\rightarrow\infty$.

4 Control and estimation of the FDR

There are two approaches to using FDR in practice. The first one is to adjust the level of individual tests so as to control the resulting FDR. The second approach fixes a level for individual tests and estimates the resulting FDR of this procedure.

4.1 Approaches to control FDR

Benjamini and Hochberg (1995) have suggested to adjust the level of individual tests in the multiple testing procedure to keep the FDR below a level prespecified by the researcher, γ. Suppose that the p-values of the n tests have been ordered in ascending order without loss of generality: $\hat{p}_1 < \hat{p}_2 < ... < \hat{p}_n$. They recommend the sequential Hohm method which compares p-values to an increasing critical value. Hypothesis i is rejected if its p-value is sufficiently small, $\hat{p}_i \leq \gamma \frac{i}{n}$. They prove that with this method controls the FDR in the sense that $FDR < \gamma$ with probability 1 when this method is used.

The BH method of controlling FDR is conservative. It uses the total number of tests in the denominator of the critical values. One can show (Storey et al., 2004) that replacing n by n_0, the number of true null hypotheses, would also control FDR. Since $n_0 < n$, the critical value will be higher for any i, and more hypotheses will be rejected. We will call the FDR-controlling method which rejects null hypotheses when $\hat{p}_i \leq \frac{i}{n_0}\gamma$ the modified BH procedure and denote it BH^*.

A difficulty with the application of FDR in a panel context is the fact that cross-sectional units display cross-sectional dependence. The above rules have been shown to be valid under independence,although some form of dependence can be allowed, see for example. Benjamini and Yuketeli (2001).

As shown by Romano, Shaikh and Wolf (2008) , the bootstrap or subsampling can be used to control for general dependence structures. Their insight is that, for a given set of critical values $\{c_1, ..., c_n\}$, we can decompose FDR as:

$$\begin{aligned} FDR_P &= EE_P\left[\frac{F}{\max\{R,1\}}|R\right] \\ &= \sum_{r=1}^{s}\frac{1}{r}E_P\left[F|R=r\right]\Pr\{R=r\} \\ &= \sum_{r=(n-n_0)+1}^{n}\frac{r-(n-n_0)}{r}\Pr\{T_1 \leq c_1, ..., T_r \leq c_r, T_{r+1} > c_{r+1}\} \quad (3) \end{aligned}$$

We determine critical values to ensure that the above quantity is bounded by the desired FDR level γ for any probability distribution P. This requires n computations (from least significant to most significant) using up to n-

dimensional integrals and is subject to curse of dimensionality.

The bootstrap is used to approximate the joint distribution of the test statistics and calculate the appropriate set of critical values. We need a bootstrap method that allows for serial dependence, cross-sectional dependence and non-stationarity. We bootstrap vectors of first differences of the data using the moving block bootstrap. Similar methods have been used by Palm, Smeekes, and Urbain (2008) for panel unit root tests and Gonçalves (2009) for a panel regression model. However, Palm et al. (2008) bootstrap residuals from a sequence of individual autoregressions. Hanck (2009) uses a sieve bootstrap on the residuals. One could also use the double resampling of Hounkannounon (2009) which is robust to general forms of cross-sectional and serial correlation.

Our algorithm is as follows:

1. Calculate the first difference $\Delta z_{it} = z_{it} - z_{i,t-1}$ and collect these as n-vectors for each time period $\Delta Z_t = (\Delta z_{1,t}, ..., \Delta z_{n,t})'$.

2. For a given block size b,draw $[T/b]$ blocks of b consecutive observations of ΔZ_t with replacement. Then draw a last block of length $T - [T/b]\, b$. Call this bootstrap sample ΔZ^*.

3. Generate the bootstrap sample of level variables by cumulating:

$$Z_t^* = \sum_{j=1}^{t} \Delta Z_j^*.$$

4. Compute an ADF test for each of the n series in the bootstrap sample.

5. Repeat steps 2-4 B times.

6. Compute the n critical values recursively by solving (3) for $n_0 = 1, ..., n$.

7. Having determined the set of critical values, $\{\hat{c}_1, ..., \hat{c}_n\}$, test null hypotheses sequentially. Reject the most significant null hypothesis (the one with the smallest statistic) if the ADF statistic for that series is less than c_1. If it is, reject the second null hypothesis if $T_2 < \hat{c}_2$ and so on until a null hypothesis is no longer rejected, call it j^*. The resulting set of $I(1)$ series are those from j^* to n, and the $I(0)$ series are $1, ..., j^* - 1$.

There are three practical difficulties with this approach: firstly, it requires the choice of block size b. As in Gonçalves (2009), we set it equal to choice of bandwidth for long-variance estimation in Andrews (1991). Secondly, as opposed to the other methods described here which are based on individual p-values, the bootstrap method can only be applied to balanced panels. If the number of cross-sectional units varies over time, the above algorithm would create "holes" in our bootstrap sample. Finally, the method requires the computation of the joint distribution of the n ADF statistics. It is therefore subject to the curse of dimensionality in two ways. Firstly, the accuracy of any estimate of a high-dimensional distribution is likely dubious, even with a large number of bootstrap replications. Second, because we have to compute n critical values, the difficulty of computations increases with n. In the simulation experiments below, we do not consider choices of n larger than 30 for that reason.

4.2 Approaches to estimate FDR

Suppose that we fix the level of the individual tests to some quantity α. Remember FDR in the limit (as the number of tests gets large) is given by (2) :

$$FDR = \frac{\alpha \pi_0}{\Pr\left(reject\ H_{0i}\right)}.$$

The natural estimator of this quantity involves replacing π_0 and the denominator by some estimators. The denominator is easy to estimate by looking at the fraction of rejections:

$$\Pr\left(\widehat{reject}\ H_{0i}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(\hat{p}_{i,T} \leq \alpha\right) = \frac{R}{n}.$$

We now derive its limit under sequential asymptotics as $T \to \infty$ followed by $n \to \infty$.

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(\hat{p}_{i,T} \leq \alpha\right) &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(\hat{p}_{i,T} \leq \alpha\right)H_i + \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(\hat{p}_{i,T} \leq \alpha\right)\left(1-H_i\right) \\ &\overset{T\to\infty}{\to} \frac{1}{n}\sum_{i=1}^{n}H_i + \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(U_i \leq \alpha\right)\left(1-H_i\right) \\ &\overset{n\to\infty}{\to} \left(1-\pi_0\right) + \alpha\pi_0. \qquad (4)\end{aligned}$$

This sequential limit is also joint if the individual unit root tests's weak limit is uniform in i under both the null and the alternative hypotheses.

Finding an estimator of π_0 is more problematic. The fraction of true null hypotheses is partly the problem we are trying to solve.

In the existing literature, Storey et al. (2004) have proposed the following general estimator:

$$\hat{\pi}_0(\lambda) = \frac{1 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{p}_i \leq \lambda)}{(1-\lambda)}$$

for some $\lambda \in (0,1)$. This comes from the fact that large p-values are likely to come from true null hypotheses. Thus, we should expect $\pi_0(1-\lambda)$ p-values above λ. Asymptotically, this estimator is consistent. To see this, the above estimator is:

$$\begin{aligned} \hat{\pi}_0(\lambda) &= \frac{1 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{p}_i \leq \lambda)}{(1-\lambda)} \\ &\rightarrow \frac{1 - ((1-\pi_0) + \lambda\pi_0)}{(1-\lambda)} = \pi_0. \end{aligned}$$

where the second line follows from (4). However, Storey et al. proved that this estimator is conservative in finite samples.

The above estimator depends on a tuning parameter λ. Storey et al. (2004) provide a data-dependent choice of λ that minimize mean square error (MSE).

Instead of relying on the above generic estimator, one can, in the context of panel unit root tests, estimate the proportion of true null hypotheses by using the results in Ng (2008). She estimates the fraction of units in a panel that have a unit root by looking at the behavior of the cross-sectional variance as a function of time. Her key insight is that the cross-sectional variance grows linearly over time with a slope equal to the fraction of the units that are non-stationary.

Ng showed that the cross-sectional variance $V_t = \frac{1}{n}\sum_{i=1}^{n}\left(z_{it} - \bar{z}_t\right)^2$ is approximately linear in t with coefficient π_0 :

$$V_t \approx c + \pi_0 t$$

for some constant c, which suggests the estimator:

$$\hat{\pi}_0 = \sum_{t=1}^{T} \Delta V_t.$$

Ng shows that this estimator converges at rate $\sqrt{n}$ and is asymptotically normal. The estimator is robust to some forms of cross-sectional dependence and one can control for serial correlation by first correcting the scale by estimating an AR process.

With estimates of π_0 and R, we can get an estimate of FDR as:

$$\widehat{FDR} = \frac{\hat{\pi}_0 \alpha}{\hat{R}/n} = \frac{\hat{\pi}_0 \alpha}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\hat{p}_i \leq \alpha\right)},$$

which, following the above discussion, is consistent if $\hat{\pi}_0 \xrightarrow{p} \pi_0$ and the denominator converges to $(1-\pi_0)+\alpha\pi_0$.

5 Simulation

In this section, we report results from a small simulation experiment. We want to analyze the effects on the FDR of the fraction of series with a unit root, the size of n and T, and the extent of cross-sectional dependence.

We first ignore the issue of cross–sectional dependence and consider the basic dynamic panel data model (1) with heterogenous intercepts:

$$
\begin{aligned}
z_{it} &= \alpha_i + z_{it}^0, \\
z_{it}^0 &= \rho_i z_{it-1}^0 + u_{it},
\end{aligned}
$$

where u_{it} is ARMA(1,1):

$$(1-\phi L)\,u_{it} = (1+\theta L)\,\varepsilon_{it}$$

and $\varepsilon_{it} \sim i.i.d.N\,(0,1)$. The autoregressive parameter ρ_i is 1 for the first π_0 fraction of the series and for the remaining $(1-\pi_0)$ fraction, ρ_i is $U\,[0,.9]$.

We consider 3 values for π_0 : .1, .5 and .9. The individual effects α_i are $N(0,1)$. Finally, we consider three values for each of ϕ and θ, -.5, 0, and .5 but do not consider cases where the roots cancel each other out. This means that we have a total of 7 pairs of ϕ and θ.

We consider the n null hypotheses that each series has a unit root. We use an ADF test for this purpose. We choose the degree of augmentation in the regression with the MAIC or Ng and Perron (2001) with a maximum of 4 lags. We consider two choices of n and T, $n = 10, 30$ and $T = 100, 500$. We do not consider larger choices of n because of the heavy computational burden imposed by the bootstrap procedure of Romano et al. (2008). We run each experiment 1000 times.

To begin with, we report 2 estimates of π_0. Because FDR depends on this (unknown) parameter, many properties of the FDR estimators and FDR control methods are directly related to these. The first estimator is the Ng (2008) estimator A, while the other is Storey's estimator with data-dependent choice of λ. The means and standard deviations over the replications are reported in table 1.

*** Insert table 1 here ***

Two main conclusions arise from these results. The first one is that Ng's estimator is much less biased than Storey's which can be quite conservative. However, it has much higher variance which increases with π_0. Second, the Ng estimator is severely biased downward with a negative MA component. This is expected as the $I(1)$ series approach stationary behavior. Unit root tests have been widely documented to suffer from severe size distortions in this case for the same reason (see Schwert, 1989). There is also a downward bias for the positive MA case for the smaller choice of T (100), but this disappears when $T = 500$. A larger T also makes the Storey estimator less conservative. The size of n does not make any difference on the centering of the Ng estimator, but it reduces its variances (since its rate of convergence is $\sqrt{n}$). Increasing n is detrimental to the Storey estimator for $\pi_0 = 10\%$, but beneficial for the other values.

In Table 2, we report the average FDP over the replications (which approaches FDR as the number of replication increases) for a fixed test size of 5% and three (conservative) estimates that differ according to the choice of $\hat{\pi}_0$. The first one uses the true π_0 (and is therefore infeasible), the second uses Ng's estimator, and the last one uses Storey's estimator. We report both the mean and standard deviation of the last two estimators.

*** Insert table 2 here ***

From this table, we notice that FDR increases with π_0.That is, if most series are nonstationary, then findings of stationarity are more likely to be false. It also increases with both n and T. Second, the FDR estimators can be quite conservative, particularly for the larger choice of π_0. There is also not much effect of either n or T on the estimators. Finally, the relative performance of these estimators follows that of the estimators of π_0. Because Ng's estimator of π_0 is less biased but more volatile, the estimator of FDR based on it is less biased but more variable in general. However, it behaves quite poorly in the large MA cases.

In table 3, we change our approach and report results when we try to control the FDR at 5%. We consider three methods described above. The first one is the original Benjamini and Hochberg (BH) method that compares the p-values to an increasing sequence of critical values. This method implicitly assumes that all null hypotheses are correct ($\pi_0 = 1$). The second method is the modified BH method (denoted BH^*) which uses the Ng estimator of π_0 when calculating the increasing critical values. Finally, we report the bootstrap-based method of Romano et al. (2008) implemented as described above. If the methods controlled the FDR perfectly, we would expect 5%

in all cells in the table. Numbers below 5% indicate that the method controls the FDR since the proportion of false rejections is less than the desired level of 5%. However, it lacks power since we could have rejected other null hypotheses without violating the FDR constraint.

*** Insert table 3 here ***

The first thing to note from the table is that the original BH method is very conservative. Despite a desired level of 5%, we reject much less often than that. One thing to note however is that this conservativeness is especially present for the small values of π_0. For $\pi_0 = .9$, the procedure is not that much conservative. This is due to the fact that BH assumes that $\pi_0 = 1$ when constructing the critical values. On the other hand, using the Ng estimator of π_0 alleviates these problems as expected. However, in the cases with large MA components, the FDR is not controlled at all and the method performs quite poorly. Finally, the bootstrap method of Romano et al. performs really well in obtaining an FDR of approximately 5% even in the large MA cases were the modified BH procedure performs poorly.

5.1 Cross-sectional dependence

Our second set of experiments adds cross-sectional dependence through a factor model. The common factor f_t is introduced in the residuals as in Moon and Perron (2004) and Pesaran (2007) :

$$\begin{aligned} z_{it} &= \alpha_i + z_{it}^0, \\ z_{it}^0 &= \rho_i z_{it-1}^0 + y_{it}, \\ y_{it} &= \lambda_i f_t + u_{it} \end{aligned}$$

where the factor loadings are $U[0,1]$ and the factor is an AR(1):

$$f_t = .5 f_{t-1} + v_t$$

where $v_t \sim i.i.d.N(0,1)$. The rest of the design is as above (in particular, u_{it} is an ARMA(1,1) process with parameters ϕ and θ).

Table 4 reports the results of the estimation of π_0 as in table 1 above.

*** Insert table 4 here ***

The results are similar to those of table 1 except that the Ng estimator now overestimates π_0 with a large negative AR process. The negative MA case again leads to a severe downward bias.

Table 5 presents the average proportion of rejections that are false and the same three estimators of the FDR as in table 2 above. First, note that the false discovery rate is lower in this case than in the independent case. This is the usual finding when there is dependence among the tests under consideration. On the other hand, the estimators of FDR are roughly the same as before, thus leading to a larger bias than before.

*** Insert table 5 here ***

In table 6, we look at the performance of the BH, BH* and RSW procedures in controlling the FDR. As expected, the presence of dependence increases the degree of conservativeness of the BH procedure. The BH* procedure works quite well except in the large MA cases where the Ng estimator of π_0 is severely biased. The bootstrap-based procedure of RSW on the other hand provides very good FDR control for all parameter configurations.

*** Insert table 6 here ***

6 Empirical examples

In this section, we employ our proposed approach to classify series in two panels into $I(0)$ and ((1) series.

Our first example uses real income data for households from the PSID. We follow Meghir and Pistaferi (2004) and remove households with female heads, with missing education data, and with outliers. We are left with $n = 154$ households for $T = 26$ years (1968-1993). As in Ng (2008), data is first regressed on individual effect, age, age^2 and education.

Our second empirical illustration uses exchange rate data. We use the long annual data on real exchange rates relative to the US dollar from Taylor (2002). [1] Because we require a balanced panel in the application of the bootstrap to control the false discovery rate, we restrict the sample to the 19 countries for which data is available over the period 1892-1996. Our panel dimensions are thus $n = 19$ and $T = 105$. We only allow for a constant term in the deterministics, but our results are similar with the inclusion of a linear trend. Hanck (2009) uses similar data. He mentions that the differences between his results and those of Taylor (2002) are due to different sample periods, different intrapolation methods for missing wartime data,

[1] We thank Alan M. Taylor for sharing his data.

and different lag length selection. Overall, our individual results are closer to those of Taylor because we only differ by small changes in sample period in order to balance the panel and lag selection rules. As reported by Lopez et al. (2005), the results using this data set are very sensitive to the choice of lag length. Using the same lag lengths as those reported in Taylor gives very close results.

The results of the application of our suggested procedures are presented in table 7. In order to increase the power of the unit root tests, we also report results with the application of the DF-GLS test of Elliott el al. (1996) , and these results are presented in the second column of the table next to those based on ADF tests.

6.1 PSID data

Our estimate of the fraction of households with nonstationary income is about 20% and does not depend on which test is used.

On the other hand, since Storey's estimator of the fraction of true null hypotheses depends on the p-values of the test, it depends on the choice of test. For the ADF test the estimate is very high, 87%. With the use of the DF-GLS test, the estimate is 27% which is close to the one obtained using

Ng's estimator.

Turning now to the results of individual ADF tests, with a fixed level of $\alpha = 5\%$, we reject the unit root for 24 out of 154 households (or about 16%). The two estimates of FDR reflect the large difference in the estimate of π_0. The estimator using the Ng estimator is 6.5%, and the one using the Storey estimator is about 28%. These can be interpreted as the posterior probability that each of these 24 rejections is for a null hypothesis that is false.

Control of FDR at the level of 5% leaves only 9 rejections with the Hohm criterion (the BH method). Use of the bootstrap to allow for dependence leaves a very small number of rejections (2). This result is robust to the choice of block size. It is probably due to the time dimension of the data not being sufficient for the application of the block bootstrap.

Results based on the more powerful DF-GLS test are very similar. With a fixed 5% level, we reject the unit root for 25 out of 154 series (instead of 24). The Storey estimator is however much lower than before and close to the Ng estimate. Thus, the FDR estimates are close to one another and quite small, and we can be fairly confident that those series that have been classified as $I(0)$ are indeed stationary.

*** Insert table 7 here ***

6.2 Real exchange rates

For the real exchange rate panel, Ng's (2008) estimate of π_0 is negative. This is the same result that she reports in her paper. Storey's estimator is 21% with the ADF test and 10.5% with the DF-GLS test. Both estimates suggest a large proportion of stationary series which is rather unusual in tests of PPP. Table 8 provides detailed results for this application.

If we fix the level of tests at 5%, we reject the unit root for 6 countries using the ADF test and 11 for the DF-GLS test. The identity of these countries can be found by looking at table 8. Countries for which we reject the unit root are identified with an asterisk in that table under the heading "5%". The estimate of FDR using Ng's estimator is negative given the negative estimate of π_0, but Storey's estimator is small. Again, we can have reasonable confidence that the rejections are from false null hypotheses.

Controlling for multiplicity using either the Hohm criterion or the bootstrap (the results are identical) leaves a single significant country (Finland) using the ADF test and 10 out 11 using the DF-GLS test (only Denmark drops out).

7 Conclusion

In this paper, we demonstrate how to use the FDR in evaluating $I(1)/I(0)$ classifications based on individual unit root tests. In the literature, most of the analysis of the FDR have been done under independence. Yet, in many interesting applications, cross-sectional data are not independent, and sometimes this dependence is quite strong. We illustrate the methods on two panel data sets and use FDR to measure the probability our confidence in the findings of stationarity.

As developed here, the methods used to control or dependence require the use of the joint distribution of the test statistics. To obtain an estimate of this distribution, we rely on the bootstrap, and this method is subject to the curse of dimensionality. Application to panels with a large number of cross-sections would probably require the use of a parametric model of dependence such as a factor or spatial model.

References

[1] Andrews, D. (1991): Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59, 817–858.

[2] Bai, J. and S. Ng (2004): A *PANIC* Attack on Unit Roots and Cointegration, *Econometrica*, 72, 1127-1177.

[3] Benjamini, Y. and Hochberg, Y. (1995):. Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B*, 57, 289-300.

[4] Benjamini, Y. and Yekutieli, (2001): The Control of the False Discovery Rate in Multiple Testing under Dependency, *The Annals of Statistics*, 29, 1165-1188.

[5] Breitung, J. and M. H, Pesaran (2008): Unit Roots and Cointegration in Panels, in Matyas, L. and P. Sevestre (eds.), *The Econometrics of Panel Data (Third Edition)*, Kluwer Academic Publishers, 279-322.

[6] Elliott, G. T. Rothenberg, and J. Stock (1996): Efficient Tests for an Autoregressive Unit Root, *Econometrica*, 64, 813–836.

[7] Gonçalves, S. (2009): The Moving Blocks Bootstrap for Panel Linear Regression Models with Individual Fixed Effects, mimeo, Université de Montréal.

[8] Hanck, C. (2009): For Which Countries did PPP hold? A Multiple Testing Approach, *Empirical Economics* 37, 93-103.

[9] Hounkannounon, B. (2009): Bootstrap for Panel Regression Models with Random Effects, Mimeo, Université de Montréal.

[10] Kapetanios, G. (2003): Determining the Stationarity Properties of Individual Series in Panel Datasets, Working paper 495, Department of Economics, Queen Mary, University of London.

[11] Lehmann, E.L. and J.P. Romano (2005): *Testing Statistical Hypotheses,* 3rd Ed. Springer.

[12] Lopez, C., C.J. Murray, and D. H. Papell (2005): State of the Art Unit Root Tests and Purchasing Power Parity, *Journal of Money, Credit, and Banking*, 37, 361-369.

[13] Moon, H.R. and B. Perron (2004): Testing for a Unit Root in Panels with Dynamic Factors, *Journal of Econometrics*∫, 122, 81-126.

[14] Moon, H. R. and B. Perron (2007): An Empirical Analysis of Nonstationarity in a Panel of Interest Rates with Factors, *Journal of Applied Econometrics*, 22, 383-400.

[15] Ng, S. (2008): A Simple Test for Non-Stationarity in Mixed Panels, *Journal of Business and Economic Statistics* , 26, 113-127

[16] Ng, S. and P. Perron (2001): Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, *Econometrica*, 69, 1519-1554.

[17] Palm, F.C., S. Smeekes, and J.-P. Urbain (2008): Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests, METEOR Research Memorandum 0RM/08/048, Mimeo.

[18] Schwert, G. W. (1989): Tests for Unit Roots: A Monte Carlo Investigation, *Journal of Business and Economic Statistics*, 7, 147-159.

[19] Storey, J.D. (2003): "The Positive False Discovery Rate: A Bayesian Interpretation and the q-value", *Annals of Statistics, 31,* 2013-2035

[20] Storey, J. D., J. Taylor, and D. Siegmund (2004), "Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach, *Journal of the Royal Statistical Society B*, 66, 187-205.

[21] Taylor, A. M. (2002): A Century of Purchasing-Power Parity, *Review of Economics and Statistics*, 84, 139-150.

Table 1. Estimators of $\pi_0(\%)$ - Independent case

				$\pi_0 = 10\%$		50%		90%	
n	T	ϕ	θ	Ng	Storey	Ng	Storey	Ng	Storey
10	100	-.5	-.5	10.0 (15.2)	31.2 (31.9)	51.6 (38.0)	56.2 (19.2)	96.6 (54.9)	85.3 (21.2)
		-.5	0	11.3 (20.2)	29.7 (30.7)	57.2 (45.7)	56.2 (19.3)	103.5 (63.4)	85.2 (20.9)
		0	-.5	3.5 (7.9)	34.3 (31.9)	16.5 (18.7)	58.3 (20.2)	30.2 (23.4)	87.1 (19.9)
		0	0	10.4 (15.6)	30.9 (31.7)	52.9 (40.9)	57.4 (20.0)	96.0 (55.8)	85.5 (20.1)
		0	.5	6.4 (13.9)	33.1 (35.0)	35.3 (33.9)	52.3 (21.1)	61.6 (47.3)	77.4 (23.1)
		.5	0	10.8 (20.0)	36.3 (31.0)	45.7 (42.8)	59.3 (21.3)	85.7 (56.2)	86.7 (20.4)
		.5	.5	9.8 (14.0)	33.2 (33.2)	51.0 (38.5)	59.5 (19.5)	95.5 (55.0)	84.9 (21.0)
10	500	-.5	-.5	10.2 (14.4)	28.7 (34.1)	51.3 (33.1)	53.2 (20.0)	91.9 (45.4)	82.1 (21.7)
		-.5	0	10.4 (15.0)	29.2 (34.6)	55.0 (37.1)	53.4 (20.0)	99.8 (52.7)	82.6 (21.3)
		0	-.5	1.8 (3.4)	27.6 (33.2)	8.9 (7.5)	53.5 (20.3)	15.9 (10.1)	82.1 (21.6)
		0	0	11.1 (17.3)	30.0 (35.1)	49.7 (32.2)	53.0 (20.5)	90.7 (45.1)	82.7 (21.7)
		0	.5	10.3 (18.2)	32.2 (36.7)	48.9 (36.8)	50.3 (21.0)	92.0 (51.4)	77.9 (22.7)
		.5	0	9.1 (14.0)	29.7 (35.0)	43.4 (31.6)	53.4 (20.0)	79.9 (45.7)	82.7 (21.5)
		.5	.5	9.9 (14.5)	29.4 (34.7)	51.4 (35.0)	52.4 (19.6)	91.8 (45.6)	82.6 (22.0)
30	100	-.5	-.5	10.9 (10.3)	16.5 (11.1)	53.5 (23.1)	62.2 (19.4)	93.6 (28.7)	95.0 (10.6)
		-.5	0	12.2 (12.3)	17.0 (14.3)	57.8 (26.4)	62.4 (18.6)	102.8 (34.9)	94.7 (10.8)
		0	-.5	3.4 (4.2)	18.7 (10.8)	16.2 (10.3)	65.6 (18.4)	30.0 (14.4)	96.1 (10.1)
		0	0	10.4 (9.1)	16.9 (13.0)	52.0 (22.5)	62.5 (19.3)	95.3 (29.7)	95.6 (10.0)
		0	.5	7.6 (9.8)	14.4 (12.6)	35.5 (20.6)	52.5 (18.1)	62.6 (26.1)	87.4 (16.2)
		.5	0	9.8 (10.6)	22.0 (14.4)	47.8 (25.2)	65.8 (18.8)	86.1 (32.9)	95.7 (10.7)
		.5	.5	10.7 (9.6)	17.0 (13.0)	52.9 (22.2)	62.4 (19.4)	93.3 (29.3)	95.2 (10.3)
30	500	-.5	-.5	10.3 (8.6)	13.4 (12.9)	49.8 (18.7)	57.1 (18.8)	91.4 (25.9)	91.1 (14.0)
		-.5	0	11.2 (9.5)	13.8 (14.4)	56.2 (22.1)	57.9 (18.7)	99.9 (29.6)	91.8 (13.7)
		0	-.5	1.7 (1.8)	13.8 (14.1)	8.7 (4.3)	56.8 (19.6)	15.3 (5.6)	92.0 (13.3)
		0	0	10.1 (8.5)	13.8 (13.9)	50.7 (19.1)	56.2 (18.6)	91.4 (25.2)	92.0 (13.2)
		0	.5	10.2 (9.7)	13.6 (15.2)	50.2 (21.0)	52.3 (18.6)	92.1 (28.7)	85.8 (16.0)
		.5	0	9.1 (8.6)	13.5 (13.5)	44.5 (18.3)	56.9 (18.2)	78.2 (26.0)	91.3 (13.4)
		.5	.5	10.5 (8.8)	13.3 (11.2)	51.1 (19.9)	56.5 (18.2)	90.8 (25.8)	91.6 (13.3)

Note: The table reports the estimates of the fraction of series with a unit root (averaged over the replications) and standard deviation undeneath. The number in the first column is Ng′s (2008) estimator A, while the second column is Storey et al.'s (2004) estimator with data-based choice of λ.

Table 2. FDR and estimates of $FDR\,(\%)$ - Independent case

n	T	ϕ	θ	$\pi_0 = 10\%$				50%				90%			
				FDR	$\widehat{FDR}$			FDR	$\widehat{FDR}$			FDR	$\widehat{FDR}$		
					π_0	$\hat{\pi}_0^{Ng}$	$\hat{\pi}_0^{Storey}$		π_0	$\hat{\pi}_0^{Ng}$	$\hat{\pi}_0^{Storey}$		π_0	$\hat{\pi}_0^{Ng}$	$\hat{\pi}_0^{Storey}$
10	100	-.5	-.5	.3	.7	.7 (1.0)	2.0 (2.1)	2.6	5.7	6.0 (4.6)	6.6 (3.0)	12.9	40.4	43.1 (28.0)	38.1 (14.0)
		-.5	0	.5	.6	.7 (1.3)	1.9 (1.9)	3.2	5.7	6.5 (5.6)	6.4 (2.7)	14.8	39.6	46.1 (31.5)	38.1 (13.9)
		0	-.5	.2	.7	.3 (.6)	2.5 (2.4)	1.7	6.7	2.2 (3.0)	8.0 (4.8)	9.3	42.1	14.3 (11.8)	41.1 (12.3)
		0	0	.4	.6	.7 (1.1)	2.0 (2.0)	2.4	5.8	6.1 (5.9)	6.7 (3.3)	13.7	40.5	43.3 (28.3)	38.5 (13.5)
		0	.5	.7	.6	.4 (.8)	2.0 (2.1)	4.8	5.3	3.7 (3.7)	5.6 (3.0)	22.4	35.8	24.8 (21.9)	31.4 (15.0)
		.5	0	.3	.9	.9 (1.8)	3.1 (2.8)	2.5	8.3	7.7 (8.9)	10.0 (6.9)	11.1	42.2	40.4 (27.7)	41.1 (12.4)
		.5	.5	.3	.6	.6 (.9)	2.1 (2.1)	3.0	5.8	6.0 (4.9)	6.7 (3.1)	13.9	40.1	42.3 (26.7)	38.1 (14.2)
10	500	-.5	-.5	.3	.6	.6 (.8)	1.6 (1.9)	2.6	4.9	5.0 (3.3)	5.2 (2.0)	15.4	38.0	39.1 (23.1)	35.3 (14.6)
		-.5	0	.4	.6	.6 (.8)	1.6 (1.9)	3.6	4.8	5.4 (3.7)	5.2 (2.0)	15.6	38.0	42.6 (26.5)	35.6 (14.5)
		0	-.5	.3	.6	.1 (.2)	1.5 (1.8)	2.7	4.9	.9 (.7)	5.2 (2.1)	16.3	37.7	6.9 (4.9)	35.7 (14.5)
		0	0	.5	.6	.6 (1.0)	1.7 (1.9)	2.8	4.9	4.8 (3.2)	5.2 (2.1)	13.9	38.8	38.9 (23.5)	35.6 (14.4)
		0	.5	.5	.6	.6 (1.0)	1.8 (2.0)	4.7	4.8	4.7 (3.6)	4.8 (2.1)	24.6	33.9	36.0 (25.5)	30.3 (15.1)
		.5	0	.4	.6	.5 (.8)	1.6 (1.9)	2.9	4.9	4.2 (3.1)	5.2 (2.0)	15.1	38.2	34.2 (22.9)	35.5 (14.6)
		.5	.5	.3	.6	.6 (.8)	1.6 (1.9)	3.1	4.8	5.0 (3.5)	5.1 (2.0)	15.3	38.1	39.8 (23.6)	35.9 (14.5)
30	100	-.5	-.5	.3	.6	.7 (.7)	1.1 (.7)	2.8	5.6	6.0 (2.7)	7.0 (2.4)	20.3	44.0	46.5 (25.5)	47.0 (10.9)
		-.5	0	.4	.6	.8 (.8)	1.1 (.9)	3.3	5.5	6.3 (3.0)	6.9 (2.2)	20.3	42.6	48.9 (25.3)	45.3 (19.0)
		0	-.5	.2	.7	.2 (.3)	1.3 (.8)	2.0	6.2	2.0 (1.3)	8.2 (2.6)	13.2	53.2	18.3 (12.8)	58.7 (28.0)
		0	0	.4	.6	.7 (.6)	1.1 (.8)	3.2	5.6	5.9 (2.7)	7.1 (2.4)	19.6	44.4	47.5 (25.0)	47.8 (21.2)
		0	.5	.6	.6	.5 (.6)	.9 (.8)	5.6	5.1	3.6 (2.2)	5.4 (2.0)	30.5	34.7	23.9 (13.4)	33.4 (14.1)
		.5	0	.3	.8	.8 (.9)	1.8 (1.3)	2.6	7.3	7.0 (4.0)	9.6 (3.4)	17.4	60.8	60.4 (40.7)	67.0 (35.9)
		.5	.5	.3	.6	.7 (.6)	1.1 (.9)	3.0	5.6	5.9 (2.6)	7.0 (2.5)	19.2	43.8	45.4 (23.6)	46.4 (19.3)
30	500	-.5	-.5	.4	.6	.6 (.5)	.7 (.7)	3.4	4.8	4.8 (1.8)	5.5 (1.9)	21.5	35.3	36.4 (13.5)	36.3 (10.3)
		-.5	0	.5	.6	.6 (.5)	.8 (.8)	3.6	4.8	5.4 (2.2)	5.6 (1.9)	22.3	34.9	39.1 (15.1)	36.0 (10.2)
		0	-.5	.4	.6	.1 (.1)	.8 (.8)	3.3	4.8	.8 (.4)	5.5 (2.0)	20.2	35.9	6.2 (2.7)	37.1 (10.3)
		0	0	.4	.6	.6 (.5)	.8 (.8)	3.0	4.9	4.9 (1.9)	5.5 (1.8)	20.1	36.0	36.4 (13.5)	36.6 (10.1)
		0	.5	.7	.6	.6 (.5)	.8 (.8)	5.4	4.7	4.8 (2.0)	5.0 (1.8)	30.1	31.4	32.6 (14.1)	30.3 (10.4)
		.5	0	.5	.6	.5 (.5)	.8 (.8)	3.5	4.8	4.3 (1.8)	5.5 (1.8)	21.8	35.2	30.8 (12.8)	36.0 (10.3)
		.5	.5	.4	.6	.6 (.5)	.7 (.6)	3.2	4.8	5.0 (2.0)	5.5 (1.8)	21.2	35.5	36.0 (13.2)	36.4 (10.2)

Note: The first column reports the proportion of false rejections. The remaining columns report estimates of the false discovery rate using π_0, Ng's estimator of π_0, and Storey's estimator of π_0 with data-dependent choice of λ.

Table 3. FDR control (%) - Independent case

				$\pi_0 = 10\%$			50%			90%		
n	T	ϕ	θ	BH	BH^*	RSW	BH	BH^*	RSW	BH	BH^*	RSW
10	100	-.5	-.5	.2	4.1	4.8	1.4	7.1	5.2	1.8	4.1	4,6
		-.5	0	.4	4.4	4.5	1.5	8.0	5.0	2.7	6.5	4.8
		0	-.5	.2	4.3	4.6	.6	14.4	4.0	1.0	11.2	3.1
		0	0	.3	4.2	4.9	1.2	6.5	5.6	2.8	4.7	5.1
		0	.5	.6	4.9	5.4	2.8	16.3	8.2	5.9	16.0	9.7
		.5	0	.1	4.2	4.1	.9	8.3	4.1	.9	3.9	3.6
		.5	.5	.3	4.0	5.1	1.4	8.2	4.8	2.0	3.6	4.6
10	500	-.5	-.5	.3	4.5	5.5	1.5	9.1	5.6	3.1	5.9	6.2
		-.5	0	.4	4.2	5.6	2.1	9.3	5.5	2.6	4.6	5.7
		0	-.5	.3	4.7	5.8	1.5	21.9	6.4	3.1	28.4	6.2
		0	0	.5	4.6	5.5	1.5	8.3	5.5	3.4	4.9	6.7
		0	.5	.5	5.0	5.8	2.9	12.0	8.3	5.8	9.7	7.4
		.5	0	.4	4.8	5.9	1.7	11.4	7.2	2.9	7.1	6.1
		.5	.5	.3	4.6	5.4	1.9	8.8	6.0	2.9	5.2	5.8
30	100	-.5	-.5	.3	6.4	5.4	1.2	3.7	5.2	2.7	2.6	5.8
		-.5	0	.4	6.1	4.8	1.6	3.9	5.1	2.6	3.0	3.7
		0	-.5	.2	9.2	5.2	.9	18.7	3.5	.8	7.0	4.5
		0	0	.4	6.4	5.4	1.5	3.9	5.5	2.2	3.4	5.1
		0	.5	.5	8.0	5.9	3.3	12.2	9.1	7.1	12.7	8.5
		.5	0	.2	6.8	4.6	.7	4.2	4.1	1.0	2.4	3.9
		.5	.5	.3	6.3	5.3	1.3	3.6	5.1	2.7	3.1	5.8
30	500	-.5	-.5	.4	6.7	5.8	1.8	5.0	6.0	2.6	3.3	6.3
		-.5	0	.4	6.5	5.8	1.8	4.5	6.0	2.7	3.8	6.3
		0	-.5	.3	9.9	5.9	1.6	35.2	6.1	2.6	28.2	6.5
		0	0	.4	6.8	5.7	1.7	4.7	6.1	2.3	3.4	5.4
		0	.5	.6	7.1	5.9	3.1	7.6	7.2	5.5	6.3	8.5
		.5	0	.4	7.1	5.9	1.9	6.2	5.9	3.0	1.2	6.5
		.5	.5	.4	6.8	5.9	1.6	4.7	6.0	2.6	5.2	6.1

Note: The table reports the proportion of false rejections using the Benjamini-Hochberg method and the bootstrap method of Romano et al. (2008) with a desired FDR level of 5%.

Table 4. Estimators of $\pi_0(\%)$ - Factor model

				$\pi_0 = 10\%$		50%		90%	
n	T	ϕ	θ	Ng	Storey	Ng	Storey	Ng	Storey
10	100	-.5	-.5	12.0 (18.6)	32.1 (30.8)	51.5 (45.0)	61.3 (20.6)	77.5 (50.9)	86.2 (22.4)
		-.5	0	18.7 (33.4)	32.3 (30.8)	64.9 (64.1)	61.7 (22.3)	83.8 (59.5)	85.0 (23.8)
		0	-.5	4.6 (9.8)	31.8 (29.5)	20.5 (21.7)	60.1 (20.7)	33.8 (29.2)	85.4 (22.0)
		0	0	13.7 (22.1)	32.0 (30.8)	51.1 (42.5)	59.6 (19.5)	75.8 (48.8)	84.8 (22.7)
		0	.5	11.8 (19.0)	33.0 (32.6)	45.6 (48.0)	62.9 (23.4)	47.5 (41.1)	82.6 (26.0)
		.5	0	10.3 (21.1)	36.0 (30.8)	43.8 (44.5)	59.7 (21.8)	72.7 (49.0)	85.5 (20.9)
		.5	.5	12.5 (19.7)	31.6 (30.6)	50.9 (45.5)	60.6 (20.3)	76.4 (52.2)	86.6 (21.3)
10	500	-.5	-.5	11.7 (18.2)	30.5 (35.6)	48.3 (41.7)	56.1 (21.4)	68.4 (39.8)	83.3 (24.0)
		-.5	0	19.6 (32.9)	29.4 (34.8)	70.2 (67.3)	58.7 (24.3)	85.8 (55.3)	81.4 (25.6)
		0	-.5	2.1 (4.0)	29.6 (34.9)	10.3 (10.5)	54.1 (21.2)	16.9 (12.9)	81.2 (22.9)
		0	0	11.5 (16.9)	27.5 (33.2)	51.7 (40.8)	55.5 (21.3)	67.2 (39.5)	81.7 (24.0)
		0	.5	12.5 (19.1)	27.1 (32.9)	45.5 (46.7)	60.6 (23.6)	51.3 (39.4)	82.5 (25.5)
		.5	0	9.2 (14.2)	27.5 (33.1)	41.7 (31.8)	53.9 (20.9)	69.6 (40.5)	81.2 (22.0)
		.5	.5	11.5 (18.2)	28.5 (33.9)	44.7 (37.6)	56.0 (22.3)	69.1 (40.8)	81.5 (24.6)
30	100	-.5	-.5	12.2 (13.3)	19.3 (13.7)	50.8 (31.5)	66.3 (23.4)	74.5 (34.7)	93.3 (15.9)
		-.5	0	18.2 (22.4)	20.0 (25.5)	67.1 (63.2)	66.3 (25.4)	83.3 (48.2)	89.7 (21.3)
		0	-.5	4.5 (5.8)	20.7 (15.6)	21.1 (16.6)	65.1 (22.1)	33.4 (18.3)	93.7 (13.4)
		0	0	12.0 (12.9)	19.8 (15.3)	51.2 (32.2)	67.4 (23.3)	74.4 (31.2)	94.0 (14.4)
		0	.5	11.5 (15.8)	19.6 (16.8)	42.9 (43.7)	64.6 (28.4)	50.7 (39.1)	84.1 (27.4)
		.5	0	9.7 (10.9)	21.9 (13.8)	43.6 (23.4)	65.2 (20.0)	75.5 (32.3)	94.9 (11.4)
		.5	.5	12.4 (13.7)	19.2 (13.2)	50.9 (35.9)	67.8 (22.5)	74.9 (34.2)	93.4 (15.4)
30	500	-.5	-.5	11.7 (12.1)	14.7 (15.2)	48.6 (30.0)	58.9 (25.0)	69.5 (27.9)	89.1 (18.5)
		-.5	0	16.1 (18.5)	15.8 (17.0)	66.0 (52.4)	57.1 (27.0)	82.9 (42.5)	82.8 (25.7)
		0	-.5	2.3 (3.5)	14.3 (14.9)	10.4 (7.0)	58.1 (22.0)	16.7 (8.9)	89.7 (17.4)
		0	0	11.5 (11.3)	15.1 (15.4)	47.5 (29.2)	59.8 (24.7)	67.8 (27.0)	88.1 (19.4)
		0	.5	11.6 (13.7)	17.0 (18.9)	43.6 (41.4)	58.3 (29.2)	49.3 (30.7)	77.9 (30.5)
		.5	0	9.0 (8.7)	14.2 (15.0)	41.2 (19.4)	56.4 (20.6)	69.0 (24.6)	90.0 (15.1)
		.5	.5	11.3 (10.9)	14.6 (14.8)	48.6 (29.2)	59.3 (24.5)	68.9 (27.6)	88.2 (19.5)

Note: The table reports the estimates of the fraction of series with a unit root (averaged over the replications). The number in the first column is Ng's (2008) estimator A, while the second column is Storey et al.'s (2004) estimator with data-based choice of λ.

Table 5. FDR and estimates of FDR (%) - Factor model

				$\pi_0 = 10\%$				50%				90%			
n	T	ϕ	θ	FDR	$\widehat{FDR}$			FDR	$\widehat{FDR}$			FDR	$\widehat{FDR}$		
					π_0	$\hat{\pi}_0^{Ng}$	$\hat{\pi}_0^{Storey}$		π_0	$\hat{\pi}_0^{Ng}$	$\hat{\pi}_0^{Storey}$		π_0	$\hat{\pi}_0^{Ng}$	$\hat{\pi}_0^{Storey}$
10	100	-.5	-.5	.3	.7	.8 (1.3)	2.3 (2.3)	1.6	6.4	6.5 (6.6)	8.0 (4.8)	9.1	41.7	36.3 (25.3)	40.6 (13.5)
		-.5	0	.2	.7	1.3 (2.2)	2.3 (2.3)	2.1	6.5	8.7 (10.5)	8.3 (5.7)	9.1	41.9	39.2 (29.4)	40.1 (14.0)
		0	-.5	.2	.7	.4 (.8)	2.4 (2.3)	1.9	6.9	2.8 (3.4)	8.3 (4.7)	8.6	42.6	15.9 (14.2)	40.6 (12.9)
		0	0	.2	.7	.9 (1.5)	2.3 (2.3)	2.0	6.5	6.4 (6.1)	7.6 (3.8)	8.6	41.8	34.9 (23.5)	39.6 (13.9)
		0	.5	.4	.7	.8 (1.3)	2.3 (2.4)	2.4	6.4	5.5 (6.0)	7.9 (5.1)	12.5	40.5	21.9 (20.1)	38.2 (15.2)
		.5	0	.3	.9	.9 (1.7)	3.1 (2.8)	2.5	8.0	6.8 (7.1)	9.9 (6.9)	11.1	41.9	34.4 (24.5)	40.5 (12.5)
		.5	.5	.3	.7	.9 (1.4)	2.2 (2.2)	2.2	6.3	6.4 (6.0)	7.8 (4.4)	10.3	41.8	35.3 (25.7)	40.5 (13.3)
10	500	-.5	-.5	.2	.6	.7 (1.0)	1.7 (2.0)	2.0	4.9	4.7 (4.1)	5.5 (2.1)	9.5	40.7	30.4 (19.4)	37.6 (14.7)
		-.5	0	.3	.6	1.1 (1.8)	1.6 (1.9)	2.5	4.9	6.9 (6.7)	5.7 (2.4)	9.8	40.5	38.7 (27.8)	36.8 (15.4)
		0	-.5	.4	.6	.1 (.2)	1.6 (1.9)	3.1	4.9	1.0 (1.0)	5.3 (2.1)	14.9	38.2	7.4 (6.4)	35.3 (14.8)
		0	0	.3	.6	.6 (.9)	1.5 (1.8)	2.6	4.9	5.0 (4.0)	5.4 (2.1)	10.9	40.1	30.2 (20.0)	36.9 (14.7)
		0	.5	.3	.6	.7 (1.1)	1.5 (1.8)	2.3	4.9	4.5 (4.6)	5.9 (2.3)	11.9	39.7	22.9 (19.7)	36.8 (15.6)
		.5	0	.4	.6	.5 (.8)	1.5 (1.8)	3.7	4.8	4.1 (3.2)	5.2 (2.1)	16.6	37.5	29.6 (20.5)	34.7 (14.8)
		.5	.5	.3	.6	.6 (1.0)	1.6 (1.9)	2.2	4.9	4.4 (3.7)	5.5 (2.2)	11.7	39.8	31.1 (20.4)	37.0 (15.0)
30	100	-.5	-.5	.3	.7	.8 (.9)	1.4 (1.0)	2.2	6.1	6.2 (4.0)	8.1 (3.5)	13.2	51.4	43.4 (31.7)	54.4 (28.6)
		-.5	0	.3	.7	1.3 (1.6)	1.4 (1.2)	2.4	6.1	8.3 (7.8)	8.3 (3.9)	12.5	53.2	49.9 (39.5)	54.8 (31.6)
		0	-.5	.3	.7	.3 (.4)	1.5 (1.2)	1.7	6.5	2.8 (2.3)	8.5 (3.6)	13.4	55.9	21.0 (16.7)	59.3 (31.7)
		0	0	.2	.7	.8 (.9)	1.4 (1.1)	2.1	6.1	6.1 (3.9)	8.1 (3.3)	12.6	53.4	42.8 (28.0)	54.3 (26.8)
		0	.5	.3	.7	.8 (1.1)	1.3 (1.2)	2.6	5.9	5.1 (5.6)	7.8 (4.2)	15.6	49.3	28.0 (28.6)	46.8 (29.0)
		.5	0	.2	.8	.8 (.9)	1.8 (1.3)	2.5	7.3	6.4 (3.8)	9.6 (3.8)	16.5	62.7	52.6 (37.9)	65.7 (35.6)
		.5	.5	.3	.7	.8 (.9)	1.4 (1.0)	2.2	6.1	6.2 (4.4)	8.3 (3.3)	14.2	52.8	41.5 (25.6)	53.3 (26.6)
30	500	-.5	-.5	.2	.6	.7 (.7)	.8 (.8)	2.2	4.9	4.7 (3.0)	5.8 (2.5)	15.5	38.0	29.5 (13.9)	38.1 (12.2)
		-.5	0	.3	.6	.9 (1.0)	.9 (.9)	2.6	4.9	6.4 (5.1)	5.6 (2.7)	15.8	37.9	35.8 (21.2)	36.4 (14.9)
		0	-.5	.4	.6	.1 (.2)	.8 (.8)	3.4	4.8	1.0 (.7)	5.7 (2.2)	18.8	36.6	6.9 (4.1)	36.9 (11.8)
		0	0	.3	.6	.6 (.6)	.8 (.9)	2.3	4.9	4.6 (2.8)	5.9 (2.5)	15.1	38.2	28.9 (13.8)	38.0 (12.4)
		0	.5	.4	.6	.6 (.8)	.9 (1.0)	2.7	4.9	4.3 (4.0)	5.7 (2.9)	15.3	38.1	21.0 (14.9)	33.9 (16.3)
		.5	0	.4	.6	.5 (.5)	.8 (.8)	3.4	4.8	4.0 (1.9)	5.5 (2.1)	21.0	35.5	27.3 (12.5)	35.7 (11.0)
		.5	.5	.3	.6	.6 (.6)	.8 (.8)	2.7	4.9	4.7 (2.9)	5.8 (2.5)	14.5	38.5	29.3 (13.8)	38.0 (12.4)

Note: The first column reports the proportion of false rejections. The remaining columns report estimates of the false discovery rate using π_0, Ng's estimator of π_0, and Storey's estimator of π_0 with data-dependent choice of λ.

Table 6. FDR control (%) - Factor model

				$\pi_0 = 10\%$			50%			90%		
n	T	ϕ	θ	BH	BH^*	RSW	BH	BH^*	RSW	BH	BH^*	RSW
10	100	-.5	-.5	.2	3.6	4.4	.7	6.3	3.5	1.5	4.1	3.8
		-.5	0	.2	3.4	3.9	1.1	6.5	3.2	1.3	3.5	2.9
		0	-.5	.2	4.1	4.5	1.0	14.3	4.0	1.4	8.8	3.8
		0	0	.2	3.7	4.4	.9	6.5	3.2	1.9	4.3	3.3
		0	.5	.3	4.1	4.3	1.1	10.0	5.1	2.4	11.0	4.5
		.5	0	.2	3.9	4.0	.6	8.5	4.8	1.0	4.0	4.8
		.5	.5	.2	3.6	4.4	1.2	6.5	3.4	1.7	4.5	3.6
10	500	-.5	-.5	.2	4.3	5.5	1.2	9.1	5.4	1.2	4.5	4.7
		-.5	0	.3	4.0	5.3	1.6	7.9	5.3	1.4	4.3	4.2
		0	-.5	.4	4.7	5.7	1.8	21.4	6.8	2.6	27.4	5.4
		0	0	.3	4.3	5.4	1.5	9.5	5.6	1.1	5.7	5.3
		0	.5	.3	4.1	5.5	1.5	11.3	5.5	2.1	10.7	4.2
		.5	0	.4	4.8	5.9	2.1	12.1	6.2	3.1	8.5	6.1
		.5	.5	.3	4.2	5.4	1.4	9.6	5.0	1.8	6.6	4.9
30	100	-.5	-.5	.3	6.1	4.7	.9	3.7	3.8	1.3	2.0	3.9
		-.5	0	.2	5.3	3.7	1.0	3.9	3.8	1.3	2.4	2.5
		0	-.5	.2	8.7	5.1	.6	16.2	3.8	1.1	6.0	4.4
		0	0	.1	5.9	4.6	.9	3.5	3.5	1.7	2.3	4.1
		0	.5	.3	6.6	4.5	1.2	9.3	5.5	2.8	8.3	4.1
		.5	0	.2	6.9	4.4	.8	4.4	4.0	1.1	1.9	5.2
		.5	.5	.2	6.0	4.5	.9	3.6	4.1	2.4	3.5	3.7
30	500	-.5	-.5	.2	6.3	5.4	1.1	4.9	5.1	1.7	3.5	4.9
		-.5	0	.3	5.7	5.4	1.3	4.7	5.1	2.0	3.4	3.9
		0	-.5	.4	9.6	5.9	1.8	34.0	7.0	2.3	28.6	5.0
		0	0	.3	6.4	5.7	1.1	5.0	5.2	1.9	3.7	5.5
		0	.5	.4	6.7	5.0	1.4	9.8	5.6	2.1	7.6	4.7
		.5	0	.4	7.1	6.1	1.9	7.4	8.4	2.6	5.2	5.8
		.5	.5	.3	6.3	5.8	1.4	5.5	5.4	1.7	3.2	4.5

ote: The table reports the proportion of false rejections using the Benjamini-Hochberg method and the bootstrap method of Romano et al. (2008) with a desired FDR level of 5%.

Table 7. Empirical results

	PSID data		**Real exchange rates**	
	ADF	DF-GLS	ADF	DF-GLS
# rejections (%)	24 (15.8)	25 (16.2)	6 (31.6)	11 (57.9)
$\hat{\pi}_0^{Ng}$	20.3	20.3	-30.3	-30.3
$\hat{\pi}_0^{Storey}$	86.6	27.3	21.1	10.5
$\widehat{FDR}_{Ng}$	6.5	6.2	-4.8	-2.6
$\widehat{FDR}_{Storey}$	27.8	8.4	3.3	.9
BH	9 (5.8)	8 (5.2)	1 (5.3)	10 (52.6)
RSW	2 (1.3)	2 (1.3)	1 (5.3)	1 (5.3)

Table 8. Detailed empirical results
Real exchange rates

	ADF				DF-GLS			
	τ_i	5%	BH	RSW	τ_i	5%	BH	RSW
Argentina	-2.67				-2.68	*	*	*
Australia	-2.61				-1.83			
Belgium	-3.12	*			-2.80	*	*	*
Brazil	-2.13				-2.44	*	*	*
Canada	-1.79				-1.69			
Denmark	-2.07				-2.01	*		
Finland	-4.45	*	*	*	-4.46	*	*	*
France	-2.93	*			-1.92			
Germany	-1.72				-2.27	*	*	*
Italy	-3.11	*			-3.08	*	*	*
Japan	-0.51				-0.11			
Mexico	-2.16				-1.74			
Netherlands	-1.77				-1.60			
Norway	-2.15				-3.00	*	*	*
Portugal	-1.82				-1.55			
Spain	-2.18				-2.31	*	*	*
Sweden	-2.90	*			-2.31	*	*	*
Switzerland	-0.95				-0.66			
United Kingdom	-2.90	*			-2.86	*	*	*
Total		6	1	1		11	10	10

Note: The table reports the rejections using a fixed 5% critical value, the the Benjamini-Hochberg method and the boostrap method of Romano et al. (2008) with a desired FDR level of 5%.