Beyond Panel Unit Root Tests: Using Multiple Testing to Determine the Non Stationarity Properties of Individual Series in a Panel[1]

H.R. Moon[2]
University of Southern California

B. Perron[3]
Université de Montréal, CIREQ, CIRANO

[1]We have benefited from comments from the guest editors, three referees, Olivier Scaillet, and the participants at the conference in honor of Peter Phillips held at Singapore Management University in July 2008, the CIREQ-CIRANO conference "The Econometrics of Interactions" on Oct. 23-24, 2009, and the Conference on Resampling Methods and High Dimensional Data at Texas A&M University on March 25-26, 2010. The usual disclaimer applies.

[2]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.

[3]*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.

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 suggest and demonstrate how to use the false discovery rate (FDR) in evaluating $I(1)/I(0)$ classifications.

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

JEL classification: C32, C33, C44

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 joint null hypothesis. This paper suggests how a researcher could proceed in classifying the individual series into stationary and nonstationary sets.

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. To discipline and 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 expectation of the proportion of rejected hypotheses that are true, or, in other words, the expected fraction of series classified as $I(0)$ that are in fact $I(1)$. 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 as in Romano, Shaikh, and Wolf (2008).

Alternative approaches to classifying the series among $I(0)$ and $I(1)$ components have been proposed. Chortareas and Kapetanios (2008) proposed the Sequential Panel Selection Method (SPSM) which consists of carrying out a sequence of panel unit root tests on panels of decreasing size. After a rejection, a researcher removes from the panel the series with the most evidence in favor of stationarity. One then continues until the joint test of a unit root for the remaining series in the panel is no longer rejected. A different approach was suggested by Ng (2008) who 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 classifying 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. Other economic applications of the FDR concept include Barras, Scaillet, and Wermers (2010) to mutual fund performance, Bajgrowicz and Scaillet (2009) to technical trading rules, and Deckers and Hanck (2009) to growth econometrics.

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. Finally, section 6 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:

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

where d_{it} is the deterministic component, and z_{it}^0 the stochastic component. The component y_{it} is assumed stationary so that non stationarity of the stochastic component follows if $\rho_i = 1$. 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 non-stationary:

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

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 Multiple testing: False discovery rate

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

We have n separate testing problems (one for each series in the panel) 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:

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

The FDR is the expected value of the false discovery proportion. To be more precise, suppose that we denote by FDP to be the false discovery proportion, or the proportion of rejections that are incorrect:

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

The FDR originally proposed by Benjamini and Hochberg (1995) is the expectation of the FDP:

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

It is possible to use FDP as a large n (the number of tests) approximation to FDR_P and establish that

$$\operatorname*{plim}_n FDP = \lim_n FDR_P = \frac{\alpha\pi_0}{\Pr\left(\text{reject the null at level } \alpha\right)}, \tag{2}$$

where $\pi_0 = \frac{n_0}{n}$, the fraction of true null hypotheses. In the context of a mixture model where the number of tests n gets large, Storey (2003) provides an interesting Bayesian interpretation of the FDR, that is the FDR is the posterior probability of the null being true given that we have rejected a particular null hypothesis.

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. We discuss each in turn.

4.1 Approaches to control FDR

Benjamini and Hochberg (1995) (BH hereafter) have suggested to adjust the level of individual tests in the multiple testing procedure to keep the FDR below a level pre-specified by the researcher, say 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 (1979) method which compares increasing p-values to an increasing critical value sequentially. We start with the hypothesis with the smallest p-value. We reject it if $\hat{p}_1 < \gamma\frac{1}{n}$ and move on to the second hypothesis. We compare the second p-value with $\gamma\frac{2}{n}$. If we reject, we move to the third hypothesis and so on. We proceed in this way until the first hypothesis j such that $\hat{p}_j \geq \gamma\frac{j}{n}$. BH prove that this method controls the FDR in the sense that $FDR < \gamma$.

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^*.We will consider estimation of n_0 in the next subsection.

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. The bootstrap is used to approximate the joint distribution of the individual test statistics and calculate an appropriate set of critical values. This requires n computations (from least significant to most significant) using up to n dimensional integrals and is subject to curse of dimensionality.

We need a bootstrap method that allows for serial dependence, cross-sectional dependence and non-stationarity. To accomplish this. 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 (2010) 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 equation (7) in Romano et al. (2008) 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 (2010), we set it equal to choice of bandwidth for long-variance estimation in Andrews (1991) in our simulation below 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 only consider choices of $n \leq 30$.

4.2 Approaches to estimate FDR

Remember FDR in the limit (as the number of tests gets large) is given by (2) :

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

where α is the fixed, user-specified level of the individual tests. We estimate this quantity by replacing π_0 and the denominator by estimators. The denominator is easy to estimate by looking at the fraction of rejections:

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

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:

$$\hat{\pi}_0(\lambda) = \frac{1 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{p}_i \leq \lambda)}{(1-\lambda)} \tag{3}$$

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. Storey et al. (2004) provide a data-dependent choice of the tuning parameter 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 $V_t = \frac{1}{n}\sum_{i=1}^{n}(z_{it} - \bar{z}_t)^2$ is approximately linear in t with coefficient π_0 :

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

for some constant c, which suggests the estimator:

$$\hat{\pi}_0 = \sum_{t=1}^{T} \Delta V_t. \tag{4}$$

With an estimator of π_0 we can get an estimate of FDR as:

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

which is consistent if $\hat{\pi}_0 \xrightarrow{p} \pi_0$.

5 Simulation

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

We consider the basic dynamic panel data model (1) with heterogenous intercepts:

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

where y_{it} exhibits cross-sectional dependence through a factor model introduced in the residuals as in Moon and Perron (2004) and Pesaran (2007) :

$$y_{it} = \lambda_i f_t + u_{it}$$

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 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, the idiosyncratic component u_{it} is ARMA(1,1):

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

and $\varepsilon_{it} \sim i.i.d.N(0,1)$. 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. To preserve space, and as the results change in an obvious way with π_0, we report results only for $\pi_0 = .5$. We also report results only for three pairs of ϕ and θ : when u_{it} is i.i.d. ($\phi = \theta = 0$), when it has a negative MA root ($\phi = 0, \theta = -.5$), and when it has a positive AR root ($\phi = .5, \theta = 0$). We have also looked at the case where the units are cross-sectionally independent (which we can interpret as $\lambda_i = 0$ for all i). Results for other (ϕ, θ) pairs and for independent cross-sections are very similar to those reported here. All other results are available upon request.

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.

In Table 1, we report the average FDP over the replications (which approaches FDR as the number of replications 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 (4), and the last one uses Storey's estimator (3). We report both the mean and standard deviation of the last two estimators.

From this table, we first notice that FDR estimators can be quite conservative. Secondly, there is 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 because the estimator inherits the large size distortions of univariate unit root tests see Schwert, 1989). The negative MA root makes the observed series look like a stationary series, thus biasing the estimator of π_0 downward.

In the last two columns of table 1, we compare with two other methods of classifying series into $I(0)$ and $I(1)$ units. The first method was proposed by Ng (2008). After having estimated the fraction of nonstationary series, one can order the series according to the estimated largest autoregressive root and treat the $\hat{\pi}_0 n$ series with the highest roots as non stationary and the rest as stationary. The second method we consider is the Sequential Panel Selection Method (SPSM) of Chortareas and Kapetanios (2009) which is based on a series of unit root tests on panels of decreasing dimensions. Because our DGP includes cross-sectional dependence, we use the $CIPS$ test of Pesaran (2007) as the panel unit root test in the procedure (Chortareas and Kapetanios used the Im et al. (2003) test which assumes cross-sectional independence).

In the last two columns of table 1, we report the FDR of these two methods. Note that these quantities *cannot* be estimated in practice and that one cannot use some estimated FDR as the basis for comparing classification methods. Since neither method is geared towards control of the FDR, it is not surprising that both methods have a higher FDR than **the** method based on individual tests. Secondly, one can note that Ng's method has much higher FDR than other methods. Finally, the SPSM does well in terms of FDR and is competitive with a sequence of individual tests.

In table 2, 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.

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. 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.

6 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.

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] Bajgrowicz, P., Scaillet, O. 2009. Technical trading revisited: False discoveries, persistence tests, and transaction costs. mimeo.

[3] Barras, L., Scaillet,O., Wermers, R. 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. Journal of Finance LXV, 179-216.

[4] Benjamini, Y. 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.

[5] Benjamini, Y., Yekutieli, Y. 2001. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics 29, 1165-1188.

[6] Breitung, J., Pesaran, M.H. 2008. Unit roots and cointegration in panels, in: Matyas, L., Sevestre, P. (eds.) The econometrics of panel data (third edition), Kluwer Academic Publishers, 279-322.

[7] Chortareas, G., Kapetanios, G. 2008. Getting PPP right: Identifying mean-reverting real exchange rates in panels. Journal of Banking and Finance 33, 390-404.

[8] Deckers, T., Hanck, C. 2009. Multiple testing techniques in growth econometrics. mimeo.

[9] Gonçalves, S. 2011. The moving blocks bootstrap for panel linear regression models with individual fixed effects. Econometric Theory forthcoming.

[10] Hanck, C. 2009. For which countries did PPP hold? A multiple testing approach. Empirical Economics 37, 93-103.

[11] Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandianavian Journal of Statistics 6, 65-70.

[12] Hounkannounon, B. 2011. Bootstrap for panel regression models with random effects. mimeo.

[13] Im, K.S., Pesaran, M.H., Shin, Y. 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115, 53-74.

[14] Lehmann, E.L., Romano, J.P. 2005.Testing Statistical Hypotheses, third edition. Springer.

[15] Moon, H.R., Perron, B. 2004. Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122, 81-126.

[16] Ng, S. 2008. A simple test for non-stationarity in mixed panels. Journal of Business and Economic Statistics 26, 113-127

[17] Ng, S., Perron, P. 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554.

[18] Palm, F.C., Smeekes, S. Urbain, J.-P. 2008. Cross-sectional dependence robust block bootstrap panel unit root tests. METEOR research memorandum 0RM/08/048, mimeo.

[19] Pesaran, M. H. 2007. A simple panel unit root test in the presence of cross section dependence. Journal of Applied Econometrics 22, 265-312.

[20] Romano, J. P., Shaikh, A.M., Wolf, M. 2008. Control of the false discovery rate under dependence using the bootstrap and subsampling. Test 17, 417-442.

[21] Schwert, G. W. 1989. Tests for unit roots: A monte carlo investigation. Journal of Business and Economic Statistics 7, 147-159.

[22] Storey, J.D. 2003. The positive false discovery rate: A bayesian interpretation and the q-value. Annals of Statistics 31, 2013-2035

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

Table 1. FDR and estimates of FDR (%) for different classification schemes

				ADF tests at fixed 5% level				Ng	SPSM (Chortareas and Kapetanios, 2009)
n	T	ϕ	θ	FDR	$\widehat{FDR}$				
					π_0	$\hat{\pi}_0^{Ng}$	$\hat{\pi}_0^{Storey}$		
10	100	0	-.5	1.9	6.9	2.8 (3.4)	8.3 (4.7)	36.4	6.1
		0	0	2.0	6.5	6.4 (6.1)	7.6 (3.8)	18.9	6.9
		.5	0	2.5	8.0	6.8 (7.1)	9.9 (6.9)	22.9	7.6
10	500	0	-.5	3.1	4.9	1.0 (1.0)	5.3 (2.1)	43.7	3.7
		0	0	2.6	4.9	5.0 (4.0)	5.4 (2.1)	19.2	5.8
		.5	0	3.7	4.8	4.1 (3.2)	5.2 (2.1)	21.8	3.9
30	100	0	-.5	1.7	6.5	2.8 (2.3)	8.5 (3.6)	35.3	4.8
		0	0	2.1	6.1	6.1 (3.9)	8.1 (3.3)	15.4	4.8
		.5	0	2.5	7.3	6.4 (3.8)	9.6 (3.8)	18.1	6.5
30	500	0	-.5	3.4	4.8	1.0 (.7)	5.7 (2.2)	44.1	1.4
		0	0	2.3	4.9	4.6 (2.8)	5.9 (2.5)	16.0	2.7
		.5	0	3.4	4.8	4.0 (1.9)	5.5 (2.1)	18.5	1.4

Note: The first column reports the proportion of false rejections. The next three 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. Finally, the last two columns report the false discovery rate associated with different classification schemes, one based on Ng's ordering autoregressive roots and Chortareas amd Kapetanios's scheme based on a sequence of panel unit root tests.

Table 2. FDR control (%)

n	T	ϕ	θ	BH	BH^*	RSW
10	100	0	-.5	1.0	14.3	4.0
		0	0	.9	6.5	3.2
		.5	0	.6	8.5	4.8
10	500	0	-.5	1.8	21.4	6.8
		0	0	1.5	9.5	5.6
		.5	0	2.1	12.1	6.2
30	100	0	-.5	.6	16.2	3.8
		0	0	.9	3.5	3.5
		.5	0	.8	4.4	4.0
30	500	0	-.5	1.8	34.0	7.0
		0	0	1.1	5.0	5.2
		.5	0	1.9	7.4	8.4

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%.