An adaptive truncated product method for combining dependent *p*-values

Xuguang Sheng [a,*], Jingyun Yang [b]

[a] *Department of Economics, American University, United States*
[b] *Methodology Center, Pennsylvania State University, United States*

HIGHLIGHTS

- We propose an adaptive truncated product method (ATPM) to facilitate selecting truncation points.
- We develop a single-layer bootstrap procedure to estimate the distribution of the ATPM.
- The ATPM insures against selecting an inferior truncation point without sacrificing much power.

ARTICLE INFO

ABSTRACT

We propose an adaptive truncated product method that facilitates the selection of the truncation point among a set of candidates. To efficiently estimate the distribution of the proposed method when the *p*-values are correlated, we develop a single-layer bootstrap procedure.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In the multiple testing literature, Zaykin et al. (2002) developed the TPM that combines only the *p*-values less than some pre-specified threshold. As demonstrated by Neuhäuser and Bretz (2005) and Sheng and Yang (forthcoming), the TPM provides more power by removing large *p*-values. However, two major issues remain in the empirical application of the TPM. First, an appropriate truncation point must be specified *a priori*. This arbitrary pre-specification poses the risk of choosing a value too large or too small, which may detrimentally affect power of the test. Second, the null distribution of the TPM is unknown when the *p*-values are correlated. To tackle these two problems, we propose an adaptive truncated product method (ATPM) that purposefully determines the truncation point based on the minimum *p*-value observed at all candidate points. We develop a single-layer bootstrap re-sampling procedure to efficiently estimate the distribution of the ATPM by employing both original and generated data sets as in Ge et al. (2003).

2. An adaptive truncated product method

Suppose that tests have been conducted for N individual hypotheses H_{0i}, $i = 1, 2, \ldots, N$. For each test, the *p*-value, p_i, is calculated. We consider the problem of testing the joint null hypothesis $H_0 = \bigcap_{i=1}^{n} H_{0i}$ at significance level α versus the combined alternative hypothesis H_A: at least one of H_{0i} is false. The test statistic of the TPM is defined as

$$W = \prod_{i=1}^{N} p_i^{I(p_i \leq \tau)}, \tag{1}$$

where $I(\cdot)$ is the indicator function and τ is the truncation point. To use the TPM, an appropriate truncation point must be specified

* Correspondence to: Department of Economics, 4400 Massachusetts Ave., N.W., Washington, DC 20016-8029, United States. Tel.: +1 202 8853782; fax: +1 202 8853790.

E-mail address: sheng@american.edu (X. Sheng).

a priori. The selection of a truncation point can be quite arbitrary as shown by the various values used in the literature. For example, $\tau = 0.05$ is used in Zaykin et al. (2002), $\tau = 0.5$ in Neuhäuser and Bretz (2005) and $\tau = 0.7$ in Neuhäuser (2003).

When the p-values are independent, W has a known distribution under the joint null hypothesis. However, this distribution no longer holds when the p-values are correlated, which poses a problem to the empirical application of the TPM. Zaykin et al. (2002) recommended a Monte Carlo algorithm to estimate the distribution. Unfortunately, this algorithm relies on an estimation of the correlation matrix for the p-value vectors, which is sensitive to the ordering of the tests.

We propose an ATPM that facilitates the selection of the truncation point among a set of candidates. Suppose that there are K candidate truncation points, τ_k, $k = 1, \ldots, K$. Let $W(k)$ be the TPM calculated by combining those p-values that do not exceed τ_k and let $\hat{s}(k)$ be the estimated p-value for $W(k)$. The ATPM is defined as the minimum p-value observed at various candidate truncation points:

$$M = \min_{1 \le k \le K} \hat{s}(k). \tag{2}$$

Many studies such as Dudbridge and Koeleman (2004) and Yu et al. (2009) use the minimum p-value as a test statistic to improve the test power.

We develop a bootstrap re-sampling procedure to efficiently estimate the distribution for the ATPM. First, we calculate the TPM based on the original data set for various truncation points τ_k, denoted by $W_{k,0}$, $1 \le k \le K$. Then we use a bootstrap procedure to generate B data sets under the null hypothesis. The types of bootstrap re-sampling depend on the data generating process and the performed tests. Based on each of the B data sets, we perform statistical tests for N individual hypotheses and calculate the resulting TPM for each τ_k, denoted by $W_{k,b}$, $1 \le k \le K$, $1 \le b \le B$. The overall p-value for the ATPM can be obtained from the original ($b = 0$) and generated ($b \ge 1$) data sets in the following way:

1. Obtain the estimated p-value $\hat{s}_k^b$ for $W_{k,b}$:

$$\hat{s}_k^b = \frac{\sum_{l=0}^{B} I(W_{k,b} \ge W_{k,l})}{B+1}, \quad 1 \le k \le K, 0 \le b \le B.$$

2. Calculate M_b, $0 \le b \le B$, as

$$M_b = \min_{1 \le k \le K} \hat{s}_k^b.$$

3. The overall p-value for M, as defined in Eq. (2), can be estimated by

$$\frac{\sum_{b=0}^{B} I(M \ge M_b)}{B+1}.$$

The proposed approach is based on the "Warp-Speed Method" of Giacomini et al. (forthcoming). The key idea underlying their proposal is that taking just one bootstrap draw for each simulated sample can suffice to provide a useful approximation to the statistic of interest. We reject the joint null hypothesis H_0 if the overall p-value is smaller than the significance level α.

3. Monte Carlo evidence

In this section, we provide simulation evidence on the performance of the ATPM in testing panel cointegration. See Baltagi (2008, chapter 12) for a recent review on nonstationary panels

and p-values. We used the following data generating process (DGP):

$$y_{it} - \alpha_i - \beta_i x_{it} = z_{it},$$

$$z_{it} = \rho_i z_{it-1} + e_{z_{it}}, \qquad x_{it} = x_{it-1} + e_{w_{it}},$$

$$e_{z_{it}} = \lambda_i f_t + \xi_{it}, \qquad e_{w_{it}} = \phi_{it} + \pi \phi_{it-1},$$

where $\begin{bmatrix} \xi_{it} \\ \phi_{it} \\ f_t \end{bmatrix} \sim$ i.i.d. $N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \psi\sigma & 0 \\ \psi\sigma & \sigma^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$.

In conducting the simulation, we choose $\alpha_i \sim \mathcal{U}[0, 5]$, $\beta_i \sim \mathcal{U}[1, 2]$, $\sigma = 1$ and $\psi = 0.5$. We explore all combinations of the following parameter values: $\pi \in \{-0.5, 0\}$, $\lambda_i \sim \mathcal{U}[1, 4]$ or $\mathcal{U}[0, 1]$, $N \in \{20, 40\}$ and $T \in \{50, 100\}$.

When $|\rho_i| < 1$, the error term z_{it} is stationary such that y_{it} and x_{it} are cointegrated with cointegrating vector $(1 \ \alpha_i \ \beta_i)$. We select a common ρ in the simulations to facilitate the interpretation of the results. We examine the size of tests with $\rho = 1$, and the power of tests with $\rho = 0.5$ for $i = 1, \ldots, N_0$ and $\rho = 1$ for $i = N_0 + 1, \ldots, N$, where $N_0 = [\delta * N]$. By varying $\delta \in \{0.1, 0.5, 0.9\}$, we explore the impact of cointegrated relationships in the panel on the power of tests.

The panel is cross-sectionally correlated due to the common factor f_t and the factor loading λ_i, with $\lambda_i \sim \mathcal{U}[1, 4]$ for high cross-section correlation and $\lambda_i \sim \mathcal{U}[0, 1]$ for low correlation. To control for cross-section correlation, we follow the sieve bootstrap procedure advocated by Chang et al. (2006) and Hanck (2009). See the Online Appendix for the implementation details. We use the augmented Dickey–Fuller cointegration test with the number of lags selected in accordance with Ng and Perron (2001). The p-values are obtained using the response surfaces in MacKinnon (1996). We calculate the TPM (denoted by W^*) and the ATPM using data generated from the bootstrap procedure. To limit the effect of multiple comparisons, we use $K = 8$ candidate truncation points ($\tau = 0.05, 0.1, 0.2, \ldots, 0.7$).[1] Due to space limitation, we focus on the results for $\tau = 0.1, 0.3, 0.5$ and 0.7.

Table 1 displays the size of tests. There is no clear pattern for the bootstrap TPM with different truncation points. Despite slightly downward size distortions for $T = 50$ due to weak cross-section correlation, both the ATPM and the bootstrap TPM control the size reasonably well.

As for the power of tests, no truncation point uniformly dominates the others. For example, the TPM with $\tau = 0.3$ may perform quite well in some cases, but has a low power relative to other truncation points in other cases. In Fig. 1, we plot the power of the ATPM against the maximum and minimum power of the TPM obtained at various truncation points. Although it may not always be the most powerful, the ATPM tracks the best truncation point very closely. The Monte Carlo simulations illustrate the strength of the ATPM relative to more risky, arbitrary selection approaches. In conclusion, the ATPM insures against selecting an inferior individual truncation point without sacrificing much power.

Acknowledgments

---

[1] We also allow the truncation points to vary from 0.05 to 0.90 in the increment of 0.05, and explore the influence of changes of the truncation points in a neighborhood of a specified point in the increment of 0.01. In both experiments, we find that the size and power of the ATPM are very similar to those reported in the paper.
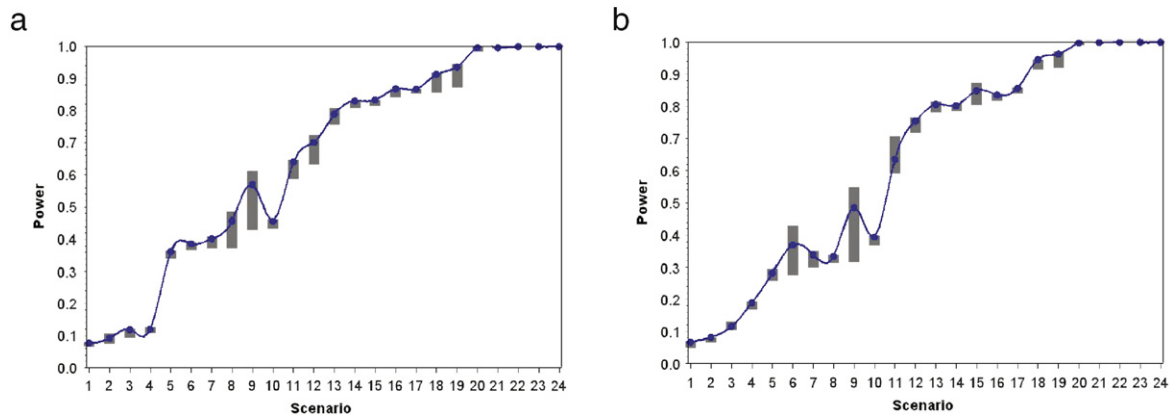
Fig. 1. Power comparison of the TPM and ATPM. The highest and lowest points of the vertical bars represent the maximum and minimum power of the TPM obtained at various truncation points. The vertical bars are ordered by the minimum power of the TPM. The dots represent the power of the ATPM. (a) tests with negative MA errors. (b) tests without MA errors.

Table 1
Size of panel cointegration test.

	N	T	W^*				ATPM
			$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	
			(a) Negative MA error				
	20	50	0.008	0.012	0.011	0.010	0.009
(i)	20	100	0.035	0.046	0.049	0.046	0.041
	40	50	0.009	0.015	0.015	0.015	0.013
	40	100	0.036	0.047	0.052	0.053	0.044
	20	50	0.030	0.050	0.054	0.052	0.045
(ii)	20	100	0.053	0.078	0.083	0.077	0.071
	40	50	0.045	0.070	0.072	0.070	0.062
	40	100	0.072	0.096	0.102	0.100	0.091
			(b) No MA error				
	20	50	0.016	0.017	0.019	0.019	0.013
(i)	20	100	0.032	0.038	0.039	0.037	0.034
	40	50	0.014	0.020	0.019	0.019	0.018
	40	100	0.036	0.046	0.047	0.046	0.043
	20	50	0.032	0.042	0.044	0.043	0.039
(ii)	20	100	0.042	0.051	0.051	0.051	0.048
	40	50	0.031	0.043	0.041	0.039	0.039
	40	100	0.047	0.067	0.067	0.066	0.059

Note: Rejection rates at a nominal level $\alpha = 0.05$, using $\delta = 0$ and 5000 replications. Panels (i) and (ii) correspond to "weak" and "strong" cross-section correlations. Panels (a) and (b) correspond to negative MA error and no MA error. W^* is the bootstrap TPM.

Appendix. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.econlet.2013.02.013.

References

Baltagi, B.H., 2008. Econometric Analysis of Panel Data, fourth ed. John Wiley & Sons Ltd.

Chang, Y., Park, J.Y., Song, K., 2006. Bootstrapping cointegrating regressions. Journal of Econometrics 133, 703–739.

Dudbridge, F., Koeleman, B.P., 2004. Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. The American Journal of Human Genetics 75, 424–435.

Ge, Y., Dudoit, S., Speed, T.P., 2003. Resampling-based multiple testing for microarray data analysis. TEST 12, 1–44.

Giacomini, R., Politis, D.N., White, H., 2012. A Warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Econometric Theory, forthcoming (http://dx.doi.org/10.1017/S0266466612000837).

Hanck, C., 2009. Cross-sectional correlation robust tests for panel cointegration. Journal of Applied Statistics 36, 817–833.

MacKinnon, J.G., 1996. Numerical distribution functions for unit roots and cointegration tests. Journal of Applied Econometrics 11, 601–618.

Neuhäuser, M., 2003. Tests for genetic differentiation. Biometrical Journal 8, 974–984.

Neuhäuser, M., Bretz, F., 2005. Adaptive designs based on the truncated product method. BMC Medical Research Methodology 5, 30.

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

Sheng, X., Yang, J., 2012. Truncated product methods for panel unit root tests. Oxford Bulletin of Economics and Statistics, forthcoming (http://dx.doi.org/10.1111/j.1468-0084.2012.00705.x).

Yu, K., Li, Q., Bergen, A.W., Pfeiffer, R.M., Rosenberg, P.S., Caporaso, N., Kraft, P., Chatterjee, N., 2009. Pathway analysis by adaptive combination of p-values. Genetic Epidemiology 33, 700–709.

Zaykin, D.V., Zhivotovsky, L.A., Westfall, P.H., Weir, B.S., 2002. Truncated product method for combining p-values. Genetic Epidemiology 22, 170–185.