연세대학교 상경대학
경제연구소
Economic Research Institute
Yonsei Universit

서울시 서대문구 연세로 50
50 Yonsei-ro, Seodaemun-gu, Seoul, Korea
TEL: (+82-2) 2123-4065 FAX: (+82-2) 364-9149
E-mail: yeri4065@yonsei.ac.kr
http://yeri.yonsei.ac.kr/new

UNIT ROOT TESTS IN THE PRESENCE OF MULTIPLE BREAKS IN VARIANCE

SOO-BIN JEONG
Yonsei University
BONG-HWAN KIM
University of California
TAE-HWAN KIM
Yonsei University
HYUNG-HO MOON
University of California

November 2014
2014RWP-70

Economic Research Institute
Yonsei University

UNIT ROOT TESTS IN THE PRESENCE OF MULTIPLE BREAKS IN VARIANCE

SOO-BIN JEONG[a], BONG-HWAN KIM[b], TAE-HWAN KIM[a,*], and HYUNG-HO MOON[b]

[a]School of Economics, Yonsei University, South Korea
[b]Department of Economics, University of California, San Diego, USA

Spurious rejections of the standard Dickey-Fuller (DF) test caused by a single variance break have been reported and some solutions to correct the problem have been proposed in the literature. Kim et al. (2002) put forward a correctly-sized unit root test robust to a single variance break, called the KLN test. However, there can be more than one break in variance in time-series data as documented in Zhou and Perron (2008), so allowing only one break can be too restrictive. In this paper, we show that multiple breaks in variance can generate spurious rejections not only by the standard DF test but also by the KLN test. We then propose a bootstrap-based unit root test that is correctly-sized in the presence of multiple breaks in variance. Simulation experiments demonstrate that the proposed test performs well regardless of the number of breaks and the location of the breaks in innovation variance.

Keywords: Dickey-Fuller test; variance break; wild bootstrap.

JEL Classifications: C12, C15

---

*Corresponding author: Yonsei University, School of Economics,134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea; Tel.: +82-2-2123-5461; fax: +82-2-2123-8638; e-mail address: tae-hwan.kim@yonsei.ac.kr.
We are grateful for the helpful comments from the participants at the 7th Joint Economic Symposium of Five Leading East Asian Universities held at the University of Singapore in January 2013.

1. Introduction

In many empirical studies, the standard Dickey-Fuller (DF) test is routinely employed prior to estimating a main model to judge whether the variables used as inputs to the main model have a unit root. Depending on the test results, either transforming the relevant variables, such as by taking differences, or modifying the main model is required. Ignoring such a step is likely to lead to some spurious regression. For this reason, the DF test has been most frequently used by empirical economists.

However, the DF test can be subject to some irregular behavior when there is a structural break in the innovation variance of a time series. Notably, Kim et al. (2002) demonstrated analytically and by simulation that the DF test can generate severe spurious rejections of the unit root null hypothesis in the presence of such a break. More specifically, they showed that the empirical size of the DF test based on the 5%-nominal size can reach 40%, which implies that the DF test wrongly reject the null hypothesis of unit root 40% of the time when, in fact, the time series under investigation has a unit root. Kim et al. (2002) also proposed a modified DF test (called the KLN test) that is robust to a single variance break.

Despite its potential shortcoming of allowing for only one break in variance, the empirical relevance of the KLN test has been increased further by some applied work on the issue of a structural break in variance. The growing empirical evidence, as fully documented in McConnell and Perez-Quiros (2000), Stock and Watson (2002), Kim et al. (2004), Sensier and van Dijk (2004) and Fukuda (2007), has revealed that there exists a structural break in variance in many US macroeconomic variables. Especially, there has existed a reduction in variance (termed the 'Great Moderation') since the 1980s.

Recently, Perron and Zhou (2008) and Zhou and Perron (2008) proposed a comprehensive procedure for testing jointly for possible multiple structural breaks in both mean and variance, and they employed this new method to re-investigate the 22 macroeconomic variables used by Stock and Watson (2002). Somewhat surprisingly, they uncovered strong evidence indicating that there are two breaks in variance in many important US macroeconomic variables including GDP, inflation and the interest rate. Since all 22 variables have been appropriately transformed to eliminate trends and/or unit roots, these findings have important implications for the unit root test on the corresponding variables. That is, it would be necessary to allow for multiple breaks in variance when testing for a unit root on those macroeconomic variables.

In this paper, we first demonstrate that the KLN test can generate spurious rejections in the presence of multiple breaks in variance. Secondly, we propose a bootstrap-based unit root test that is robust to multiple breaks in variance. The paper is organized as follows. In Section 2, we introduce the model in which the innovation error term possibly exhibits multiple breaks, and we illustrate the irregular behavior of both the standard DF test and the KLN test in such circumstance. We then propose a bootstrap-based unit root test that is correctly-sized in the presence of multiple breaks in variance in Section 3. Monte Carlo simulations are carried out to show the finite performance of the proposed test in terms of size and power in Section 4, and Section 5 contains our conclusions.

2. The Model and Spurious Rejections of the KLN Test

Consider a time series $\{y_t: t = 1,2,\cdots,T\}$ of sample size T generated from the following data generating process (DGP):

$$y_t = \mu + \rho y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} + \varepsilon_t \tag{1}$$

$$\varepsilon_t = \sigma_t \eta_t \tag{2}$$

where η_t is generated from the independent and identically distributed standard normal distribution, denoted by $\mathrm{IIN}(0,1)$. As usual, we assume that all roots of $1 - \sum_{j=1}^{p-1} \varphi_j x^j = 0$ have a modulus greater than unity. We specify the variance of ε_t as follows:

$$\sigma_t^2 = \sum_{i=1}^{k} \sigma_i^2 1[\tau_{i-1} T < t \leq \tau_i T] \tag{3}$$

where τ_i is the break timing of the i^{th} break in variance, satisfying $\tau_0 < \tau_1 < \cdots < \tau_{k-1} < \tau_k$ with $\tau_0 = 0$ and $\tau_k = 1$. For example, the initial variance is σ_1^2 and is changed to σ_2^2 at time $\tau_1 T$. Hence, there can possibly be $k-1$ breaks $(\tau_1, \ldots, \tau_{k-1})$ and k different variances $(\sigma_1^2, \ldots, \sigma_k^2)$. If $k = 2$, then the DGP given in (1) – (3) specializes to the single break model used by Kim et al. (2002).

In this paper, we are interested in testing the unit root hypothesis $\rho = 1$ against the stationary alternative hypothesis $\rho < 1$ in the presence of multiple breaks in variance with $k > 2$. To provide some motivation, we first perform Monte Carlo simulations of the routine application of both the standard DF test (wrongly assuming that there is no break in variance) and the KLN test (wrongly

assuming that there is only a single break in variance) of a time series generated by (1) – (3) with $k = 3$. In this setting, there are two breaks τ_1, τ_2 and three variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$. For notational convenience, we define $\delta_i = \sigma_i/\sigma_1$ which is the i^{th} standard deviation σ_i relative to the initial standard deviation σ_1 with the normalization $\sigma_1 = 1$ so that we can control the degree of variance shift using only δ_i.

Table 1 shows the empirical rejection probabilities of the 5% nominal level by the standard DF test for various values of break locations (τ_1, τ_2), different variances (δ_1, δ_2) and sample size T. All rejection probabilities are calculated through 1,000 replications and we set p = 1. Table 1 shows that the phenomenon of size distortion by the DF test is significant. For example, let us consider the particular case with $(\tau_1, \tau_2) = (0.3, 0.6)$ and $(\delta_1, \delta_2) = (2.5, 0.4)$; that is, the initial standard deviation σ_1 is 1 and it is increased to 2.5 at time $\tau_1 T$. It remains constant until time $\tau_2 T$ when it deceases to 0.4. Table 1 shows that the DF test results in significant spurious rejections. In some extreme cases, the rejection rate can reach around 40%, which means that researchers can wrongly reject the null of unit root when in fact the process does have a unit root 40% of the time.

Next, we employ the KLN test whish is designed to deal with a single variance break to the same situation. The results are shown in Table 2. Although the severity of spurious rejections has been lessened overall, there remain some irregular cases in which the empirical rejections probabilities are around 10%. To investigate what can happen to such irregular cases in large samples, we have increased the sample size T to 500 and 1,000. The results are shown in the bottom panels (c) and (d) in Table 2. The spurious rejection phenomenon is not just a finite sample problem because the size distortion does not vanish as the sample size increases.

3. A New Bootstrap-Based Unit Root Test

Theoretically, it is possible to develop a correctly-sized unit root test by extending the GLS-based approach as in Kim et al. (2002). Such an extension would involve estimating the quasi-maximum likelihood estimator (QMLE) of all break points as follows:

$$(\hat{\tau}_1, \dots, \hat{\tau}_{k-1}) = \underset{\tau_1, \dots, \tau_{k-1}}{argmin}\ \hat{Q}(\tau_1, \dots, \tau_{k-1}) \tag{4}$$

where $\hat{Q}(\tau_1, \dots, \tau_{k-1}) = \sum_{i=1}^{k} \frac{1}{(\tau_i - \tau_{i-1})T} \sum_{t=\tau_{i-1}T+1}^{\tau_i T} \hat{\varepsilon}_t^2$. Note that $\hat{\varepsilon}_t$ is the residual obtained by estimating equation (1) by OLS. Once these break points are obtained, the QML estimators for

variances can be computed by $\hat{\sigma}_i^2(\hat{\tau}_1, \ldots, \hat{\tau}_{k-1}) = \frac{1}{(\hat{\tau}_i - \hat{\tau}_{i-1})T} \sum_{t=\hat{\tau}_{i-1}T+1}^{\hat{\tau}_i T} \hat{\varepsilon}_t^2$. The subsequent GLS (generalized least squares) transformation would be to divide equation (1) by $\hat{\sigma}_i(\hat{\tau}_1, \ldots, \hat{\tau}_{k-1})$. Since the GLS transformation will wipe out all heteroskedasticity in the error term, the usual t-statistic from this GLS regression as in Kim et al. (2002) should not be influenced by multiple breaks in variance. However, obtaining the asymptotic null distribution of the GLS-based statistic can be complicated. In addition, even if the asymptotic distribution is available, its finite sample performance may be questionable because the presence of multiple breaks would divide the sample into smaller sub-samples, which can affect the precision of the computation of the QML estimators. Therefore, the correct finite sample null distribution can be substantially different from those predicted by the asymptotic theory, and relying on the asymptotic distribution can be inaccurate in finite samples.

To solve these two potential problems, we propose a bootstrap-based approach. The basic idea of bootstrapping is to approximate the finite sample distribution of a statistic under the null hypothesis by a computer-intensive method. Specifically, the bootstrap method involves generating a sufficient number of artificial samples (termed 'bootstrap samples') using a DGP artificially constructed by mimicking the true DGP that generated the observed sample. Therefore, if a reasonably-well constructed DGP is used in the bootstrap method, then researchers do not need to rely on the asymptotic theory. However, there is a problem implementing the bootstrap method if there exists heteroskedasticity (either conditional or unconditional) in the DGP and its form is unknown because mimicking the true DGP for generating bootstrap samples can be difficult.

To handle this particular difficulty, Wu (1986) and Liu (1988) proposed the wild bootstrap procedure. They demonstrated the ability of the wild bootstrap to mimic the true DGP with heteroskedasticity of unknown form. A rigorous theoretical justification for the wild bootstrap was also provided by Mammen (1993). Further extensions to the wild bootstrap were studied by several authors such as Härdle and Mammen (1990), Davidson and Flachaire (2001), Gonçalves and Kilian (2004), and Godfrey and Tremayne (2005).

We propose a new test for the unit root hypothesis in the presence of multiple breaks in variance that uses the wild bootstrap. There are several advantages in applying the wild bootstrap method. First, we do not need to know the exact form of heteroskedasticity; that is how many breaks are present, where the break points are located, and what the magnitudes of the different variances are. Therefore, it is not necessary to obtain the QML break estimators explained in (4), which makes our testing procedure applicable to a DGP with multiple breaks, a DGP without any break (the case

typically dealt with the standard DF test), and a DGP with a single break (the case handled by the KLN test). Secondly, the wild bootstrap method can provide a finite sample refinement, especially when the sample size is very small as will be demonstrated in the following simulation section.

We define a Bernoulli random variable i_t defined as taking 1 with probability 1/2 and -1 with probability 1/2. We also let x_t be the vector of all explanatory variables in the OLS regression of equation (1): $x_t = (1, y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p+1})'$ and let X be the matrix consisting of x_t: $X = (x_1', \dots, x_T')'$. With these notations, the proposed wild-bootstrap-based unit root test is implemented as follows:

(a) Estimate equations (1) and (2) by OLS and compute the residuals $\hat{\varepsilon}_t$ given by

$\hat{\varepsilon}_t = y_t - \hat{\mu} - \hat{\rho} y_{t-1} - \sum_{j=1}^{p-1} \hat{\varphi}_j \Delta y_{t-j}$.

(b) Compute the test statistic $T(\hat{\rho} - 1)$.

(c) Construct a bootstrap sample $y^* = (y_1^*, y_2^*, \cdots, y_T^*)'$ by the following recursion with an appropriate correction for initial values:

$y_t^* = \hat{\mu} + y_{t-1}^* + \sum_{j=1}^{p-1} \hat{\varphi}_j \Delta y_{t-j}^* + \hat{\varepsilon}_t^*$,

where $\hat{\varepsilon}_t^*$ is defined as

$\hat{\varepsilon}_t^* = i_t \frac{\hat{\varepsilon}_t}{(1-w_t)}$,

$w_t = x_t'(X'X)^{-1} x_t$.

(d) Using the generated bootstrap sample y^*, we estimate equations (1) and (2) again and compute the bootstrap statistic $T(\hat{\rho}^* - 1)$.

(e) Repeat Steps (c) through (d) B times to obtain B bootstrap statistics.

(f) Approximate the finite sample distribution of the test statistic $T(\hat{\rho} - 1)$ by constructing a (smoothed) histogram using the generated B bootstrap samples of $T(\hat{\rho}^* - 1)$.

(g) Obtain a relevant critical value from the (smoothed) histogram and compare it with the test statistic to make a decision.

Usually, we have better approximation to the finite sample distribution as we increase the value of B. The only cost is computation time. In our simulation study, we set B to be 1,000. We note that there are other variants of the wild bootstrap method which specify the Bernoulli random variable i_t differently, but we use the specification (i.e. taking 1 with probability 1/2 and -1 with probability 1/2)

of Davidson and Flachaire (2001) who demonstrated that this specification outperforms other types of wild bootstrap.

4. Monte Carlo Simulations

For our simulation experiments, the DGP given in (1) and (3) is used throughout. When the DGP is used to generate data, we set $p = 1$ for simplicity, but all simulation results are valid for the general case of $p > 1$. Whenever it is necessary to emphasize this point, we conduct additional simulations with $p > 1$. Once we obtain data $\{y_t : t = 1,2,\cdots,T\}$ using the DGP, we investigate the performance of (i) the standard DF test, (ii) the LKN test, and (iii) the newly proposed WB unit root test in terms of size and power by comparing their empirical rejection probabilities which are calculated through 1000 replications using the 5% nominal level. For the WB unit root test, we follow the wild bootstrap procedure explained in Section 3. In all simulations, the sample size T is set to 50 and 100 to investigate the finite sample properties for each test. In some cases, we extend the sample size to 500 or 1,000 to examine where the empirical rejection probability converges in large samples. As explained in Section 2, we use the relative standard deviation $\delta_i = \sigma_i/\sigma_1$ with the normalization $\sigma_1 = 1$. Hence, all breaks in the variances can be completely characterized by δ_i alone. As mentioned in Section 3, the number of bootstrap samples B is fixed to 1,000 in all cases.

4.1 The Main Multiple Break Case: Since we wish to test the unit root hypothesis $\rho = 1$ against the stationary alternative hypothesis $\rho < 1$ in the presence of multiple breaks in variance with $k > 2$, we first consider the simplest case with two breaks, k = 3. This gives us two break times (τ_1, τ_2) and three variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ which are characterized by (δ_1, δ_2). In this case, the initial standard deviation σ_1 is 1 and is changed to δ_1 at time $\tau_1 T$. It remains constant until time $\tau_2 T$ when it is changed again to δ_2. We choose a variety of values for (τ_1, τ_2) and (δ_1, δ_2) to include early and late breaks as well as increasing and decreasing variance cases. Specifically, we set (τ_1, τ_2) $\in$ {(0.3,0.6), (0.2,08), (0.4,0.8)} and (δ_1, δ_2) $\in$ {(2.5,4.0), (2.5,0.4), (0.4,0.25), (0.4,2.5)}.

It has been shown in Tables 1 and 2 that both the DF test and the KLN test reject the correct unit root hypothesis too often, and that this size distortion can be serious in some cases. Now we turn to the proposed WB unit root test. For comparison, all conditions are identical to those in Tables 1 and 2. The results are displayed in Table 3(a) which demonstrates that size distortion is not present in the results from the new test. There are some cases in which the empirical sizes are slightly larger than the 5% nominal level when the sample size is 50. However, this is a finite sample problem

unlike in the DF and KLN tests, because those rejections tend to vanish as the sample size increases to 100 and 500. The small sample problem is not empirically relevant since, in reality, no researcher wishes to use 50 observations when two breaks are possible. When there are enough observations, the empirical sizes found using the new WB unit root test is reasonably close to the 5% nominal level in all cases. Table 3(b) reports the empirical sizes found using the WB unit root test for the case in which a lagged difference is included in the model, i.e. $p = 2$ in the DGP in (1). We set $\varphi_1 = 0.2$ while all other parameters are identical to those in Tables 3(a). Again the proposed test exhibits good size reliability. We also show the power properties of the new test in Table 4. As expected, the empirical powers of the WB unit root test increases as either the sample size increase or the stationarity parameter ρ decreases.

We now turn to the three break case with $k = 4$. We consider the following values for break times and relative sizes of variances: $(\tau_1, \tau_2, \tau_3) \in \{(0.2, 0.5, 0.8), (0.3, 0.5, 0.7)\}$ and $(\delta_1, \delta_2, \delta_3) \in \{(2.5, 0.8, 0.25), (2.0, 0.25, 1.25), (0.4, 1.25, 4.0), (0.8, 2.5, 0.4)\}$. The empirical sizes found using the DF test and the KLN test are shown in Tables 5 and 6, respectively. Table 5 shows the results for the DF test, which indicates that the size distortion is particularly severe when the innovation variance decreases after the first break point ($\delta_1 < 1$). For instance, the empirical size of the DF test increases to 12.4% when the sample size T is 50. Moreover, this distortion does not vanish as the sample size increases to 1,000. The rejection probability 14.2% when $(\tau_1,\tau_2,\tau_3) = (0.2,0.5,0.8)$, $(\delta_1,\delta_2,\delta_3) = (0.8,2.5,0.4)$ and $T = 1000$. A similar picture emerges for the KNL test as shown in Table 6. In particular, the KNL test tends to over reject the correct null hypothesis when the innovation variance increases after the first break point ($\delta_1 > 1$).[1] On the other hand, as Table 7 shows, the empirical sizes found using the WB unit root test in the presence of three breaks are fairly close to the 5% nominal level in all cases, again demonstrating that the new test is robust to the number of breaks in variance. The empirical powers of the WB unit root test in the three break case are given in Table 8. As before, the empirical powers of the WB unit root test increases when the sample size increase or the stationarity parameter ρ decreases.

---

[1] Although not reported here but available upon request, the KLN test becomes biased when the sample size $T = 50$ and $(\delta_1,\delta_2,\delta_3) = (2.0, 0.25, 1.25)$. For instance, when $(\tau_1,\tau_2,\tau_3) = (0.2, 0.5, 0.8)$ and $\rho = 0.9$, the empirical power is 0.047 whereas the corresponding empirical size is 0.082 as shown in Table 6.

4.2 The Single Break Case: Although the proposed test is designed to deal with the multiple break case, it can be used in the presence of a single break since it must be robust to the number of breaks. In the case of a single break, the KLN test should work well, therefore it is interesting to compare the performance of the KLN test and the WB unit root test in such a single break case with $k = 2$. We consider $\tau_1 \in \{0.2,0.4,0.6,0.8\}$ and $\delta_1 \in \{0.25,0.4,0.6,0.8,1.0,1.25,1.67,2.5,4.0\}$. The empirical sizes of both tests are reported in Tables 9 and 10, respectively for the KLN test and the proposed test and, as expected, they are close to the 5% nominal level in all cases. The empirical power functions of both tests are shown in Tables 11 and 12. It is shown that the proposed test is much more powerful than the KLN test in almost all cases, which indicates that the potential gain in power generated by employing the wild bootstrap procedure is substantial. Therefore, it can be argued that if the possibility of unconditional heteroskedasticity is a concern, then the proposed WB unit root test is preferable to the KNL test because (i) it is robust to multiple breaks and (ii) it is more powerful than the KNL test even in the single break case.

4.3 The Conditional Heteroskedasticity Case: Another case worth investigating is the conditional heteroskedasticity case in which the innovation variance depends on some variables in the information, such as White-type heteroskedasticity or GARCH.[2] We consider the following conditional heteroskedasticity:

$$\sigma_t^2 = (1 + \theta z_t^2)^2 \tag{5}$$

[2] A break in variance is a sudden change in variance that is not related to any variables in the information set, whereas conditional heteroskedasticity implies that the innovation variance is correlated with its pervious terms or covariates in the information set. Hence, the change in variance in conditional heteroskedasticity is smoother and more predictable than a sudden break in variance. One can test the former by employing recent tests such as the White test or the GARCH test while the latter is tested as in Inclán (1993), Bai, et al. (1998) and Bai (2000). However, as simulation results indicate, it is not necessary to distinguish between them as far as the unit root hypothesis is concerned because the proposed test is robust to both sudden breaks in variance and conditional hetroskedasticity.

where z_t is $\mathrm{IIN}(0,1)$. The parameter θ in equation (6) represents the intensity of conditional heteroskedasticity. We choose various values for θ such as 0, 10, 20, and 50. Note that if $\theta = 0$, there is no conditional heteroskedasticity in the innovation variance.

The empirical sizes of the three tests (the DF test, the KLN test, and the proposed test) in the presence of conditional heteroskedasticity specified in (6) are shown in Table 13. It is clear that there is no size distortion in any of the tests. However, the difference in empirical powers of these tests is quite substantial as reported in Table 14. For every value of ρ and θ, the empirical power of the proposed test is the highest. Particularly when the sample size T is as small as 50, the power of the proposed test for some values of (ρ, θ) is more than double that of the other two tests. It can be seen from Tables 13 and 14 that the proposed test not only robust to conditional heteroskedasticity, but is also more powerful than the other two tests in the presence of such heteroskedasticity.

5. Conclusion

In this paper, we have demonstrated that multiple breaks in variance can generate spurious rejections by both the standard DF test and the KLN test. We then proposed a new bootstrap-based unit root test using the wild bootstrap method that is correctly-sized in the presence of multiple breaks in variance. Simulation experiments have shown that the proposed test possesses many advantages: (i) the test is robust to multiple breaks in that researchers do not need know the number or locations of the breaks in variance, (ii) the proposed test is also applicable to a single break in variance and is more powerful than the DF and KLN tests, (iii) the test is also robust to conditional heteroskedasticity and more powerful than the DF and KLN tests in that case as well. Finally, we note that the DGP used in our study is somewhat restricted because it cannot deal with some interesting issues such as stochastic unit root or nonlinear unit root testing problems. Investigating these issues in the presence of a break in variance can be a promising direction for future research.

References

Bai, J (2000). Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices, *Annals of Economics and Finance*, 1, pp. 303-339.

Bai, J, RL Lumsdaine and JH Stock (1998). Testing for and dating breaks in multivariate time series, *Review of Economic Studies*, 65, pp. 395-432.

Davidson, R and E Flachaire (2001). The wild bootstrap, tamed at last. Working Paper IER#1000, Queen's University.

Fukuda, K (2007). A unified approach to detecting unit root and structural break, *Applied Economics*, 39, pp. 279–289.

Godfrey, LG and AR Tremayne (2005). The wild bootstrap and heteroskedaticity-robust tests for serial correlation in dynamic regression models, *Computational Statistics & Data Analysis*, 49, pp. 377–395.

Gonçalves, S and L Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics*, 123, pp. 89–120.

Härdle, W and E Mammen (1990). Bootstrap methods in nonparametric regression, CORE Discussion Paper No. 9049.

Inclán, C (1993). Detection of multiple changes of variance using posterior odds, *Journal of Business and Economic Statistics*, 11, pp. 289–300.

Kim, CJ, CR Nelson and J Piger (2004). The less-volatile U.S. economy: a Bayesian investigation of timing, breadth, and potential explanations, *Journal of Business and Economic Statistics*, 22, pp. 80–93.

Kim, TH, SJ Leybourne and P Newbold (2002). Unit root tests with a break in innovation variance, *Journal of Econometrics*, 109, pp. 365–387.

Liu, R (1988). Bootstrap procedures under some non i.i.d. models, *Annals of Statistics*, 16, pp. 1696–1708.

Mammen, E (1993). Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics*, 21, pp. 255-285.

McConnell M and G Perez-Quiros (2000). Output fluctuations in the United States: what has changed since the early 1980s? *American Economic Review*, 90, pp. 1464–1476.

Perron, P and J Zhou (2008). Testing jointly for structural changes in the error variance and coefficients of a linear regression model, Working Paper, Boston University.

Sensier, M and D van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series, *Review of Economics and Statistics*, 86, pp. 833–839.

Stock, JH and MW Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annuals*, 17, pp. 159–218.

Wu, CFJ (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *Annals of Statistics*, 14, pp. 1261–1295.

Zhou, J and P Perron (2008). Testing for breaks in coefficients and error variance: simulations and applications. Working Paper, Boston University.

Table 1. Empirical sizes of the DF test at the nominal 5% level when there are two breaks in variance.

(δ_1, δ_2)	(2.5, 4.0)	(2.5, 0.4)	(0.4, 0.25)	(0.4, 2.5)
(a) T=50				
(τ_1, τ_2)				
(0.3, 0.6)	0.079	0.080	0.418	0.408
(0.2, 0.8)	0.086	0.073	0.409	0.471
(0.4, 0.8)	0.082	0.102	0.380	0.368
(b) T=100				
(τ_1, τ_2)				
(0.3, 0.6)	0.092	0.110	0.468	0.448
(0.2, 0.8)	0.105	0.088	0.420	0.461
(0.4, 0.8)	0.098	0.094	0.415	0.369

Table 2. Empirical sizes of the KLN test at the nominal 5% level when there are two breaks in variance.

(δ_1, δ_2)	(2.5, 4.0)	(2.5, 0.4)	(0.4, 0.25)	(0.4, 2.5)
(a) T=50				
(τ_1, τ_2)				
(0.3, 0.6)	0.096	0.114	0.037	0.058
(0.2, 0.8)	0.091	0.075	0.039	0.080
(0.4, 0.8)	0.112	0.120	0.044	0.042
(b) T=100				
(τ_1, τ_2)				
(0.3, 0.6)	0.070	0.121	0.041	0.061
(0.2, 0.8)	0.084	0.081	0.047	0.075
(0.4, 0.8)	0.087	0.103	0.035	0.047
(c) T=500				
(τ_1, τ_2)				
(0.3, 0.6)	0.069	0.109	0.048	0.031
(0.2, 0.8)	0.093	0.071	0.036	0.079
(0.4, 0.8)	0.085	0.135	0.024	0.056
(d) T=1000				
(τ_1, τ_2)				
(0.3, 0.6)	0.074	0.116	0.043	0.038
(0.2, 0.8)	0.088	0.069	0.038	0.073
(0.4, 0.8)	0.095	0.127	0.035	0.056

Table 3(a). Empirical sizes of the WB unit root test at the nominal 5% level when there are two breaks in variance.

(δ_1, δ_2)	(2.5, 4.0)	(2.5, 0.4)	(0.4, 0.25)	(0.4, 2.5)
(a) T=50				
(τ_1, τ_2)				
(0.3, 0.6)	0.061	0.057	0.069	0.065
(0.2, 0.8)	0.065	0.053	0.057	0.101
(0.4, 0.8)	0.079	0.067	0.080	0.094
(b) T=100				
(τ_1, τ_2)				
(0.3, 0.6)	0.056	0.060	0.067	0.046
(0.2, 0.8)	0.047	0.047	0.063	0.064
(0.4, 0.8)	0.061	0.049	0.053	0.057
(c) T=500				
(τ_1, τ_2)				
(0.3, 0.6)	0.051	0.053	0.059	0.052
(0.2, 0.8)	0.041	0.039	0.040	0.062
(0.4, 0.8)	0.056	0.054	0.045	0.063

Table 3(b). Empirical sizes of the WB unit root test at the nominal 5% level when there are two breaks in variance ($p = 2$).

(δ_1, δ_2)	(2.5, 4.0)	(2.5, 0.4)	(0.4, 0.25)	(0.4, 2.5)
(a) T=50				
(τ_1, τ_2)				
(0.3, 0.6)	0.064	0.054	0.088	0.079
(0.2, 0.8)	0.071	0.063	0.086	0.106
(0.4, 0.8)	0.075	0.067	0.089	0.092
(b) T=100				
(τ_1, τ_2)				
(0.3, 0.6)	0.070	0.043	0.056	0.065
(0.2, 0.8)	0.050	0.036	0.061	0.078
(0.4, 0.8)	0.066	0.047	0.051	0.061
(c) T=500				
(τ_1, τ_2)				
(0.3, 0.6)	0.056	0.048	0.051	0.039
(0.2, 0.8)	0.049	0.059	0.052	0.051
(0.4, 0.8)	0.042	0.052	0.059	0.049

Table 4. Empirical powers of the WB unit root test at the nominal 5% level when there are two breaks in variance.

	(δ_1, δ_2)	(2.5,4.0)	(2.5,0.4)	(0.4,0.25)	(0.4,2.5)
(a) T=50					
(τ_1, τ_2)	ρ=0.9				
(0.3,0.6)		0.105	0.110	0.087	0.098
(0.2, 0.8)		0.135	0.108	0.089	0.115
(0.4, 0.8)		0.127	0.124	0.094	0.105
(τ_1, τ_2)	ρ=0.8				
(0.3,0.6)		0.283	0.277	0.169	0.261
(0.2, 0.8)		0.285	0.287	0.227	0.232
(0.4, 0.8)		0.314	0.308	0.213	0.216
(b) T=100					
(τ_1, τ_2)	ρ=0.9				
(0.3,0.6)		0.289	0.222	0.161	0.202
(0.2, 0.8)		0.276	0.273	0.196	0.199
(0.4, 0.8)		0.325	0.289	0.186	0.176
(τ_1, τ_2)	ρ=0.8				
(0.3,0.6)		0.749	0.653	0.541	0.667
(0.2, 0.8)		0.743	0.763	0.627	0.575
(0.4, 0.8)		0.816	0.819	0.620	0.456

Table 5. Empirical sizes of the DF test at the nominal 5% level when there are three breaks in variance.

$(\delta_1, \delta_2, \delta_3)$	(2.5,0.8,0.25)	(2.0,0.25,1.25)	(0.4, .25,4.0)	(0.8,2.5,0.4)
(a) T=50				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.034	0.060	0.108	0.124
(0.3, 0.5, 0.7)	0.042	0.076	0.085	0.123
(b) T=100				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.049	0.070	0.116	0.132
(0.3, 0.5, 0.7)	0.044	0.074	0.098	0.129
(c) T=500				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.046	0.088	0.120	0.129
(0.3, 0.5, 0.7)	0.045	0.069	0.078	0.152
(d) T=1000				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.061	0.080	0.114	0.142
(0.3, 0.5, 0.7)	0.051	0.091	0.078	0.135

Table 6. Empirical sizes of the KLN test at the nominal 5% level when there are three breaks in variance.

$(\delta_1, \delta_2, \delta_3)$	(2.5,0.8,0.25)	(2.0,0.25,1.25)	(0.4, .25,4.0)	(0.8,2.5,0.4)
(a) T=50				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.059	0.082	0.073	0.050
(0.3, 0.5, 0.7)	0.091	0.097	0.068	0.055
(b) T=100				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.083	0.065	0.072	0.065
(0.3, 0.5, 0.7)	0.102	0.095	0.060	0.045
(c) T=500				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.077	0.071	0.050	0.048
(0.3, 0.5, 0.7)	0.098	0.091	0.045	0.054
(d) T=1000				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.083	0.066	0.070	0.055
(0.3, 0.5, 0.7)	0.094	0.091	0.039	0.046

Table 7. Empirical sizes of the WB unit root test at the nominal 5% level when there are three breaks in variance.

$(\delta_1, \delta_2, \delta_3)$	(2.5,0.8,0.25)	(2.0,0.25,1.25)	(0.4, .25,4.0)	(0.8,2.5,0.4)
(a) T=50				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.057	0.063	0.056	0.066
(0.3, 0.5, 0.7)	0.056	0.058	0.062	0.068
(b) T=100				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.052	0.059	0.056	0.056
(0.3, 0.5, 0.7)	0.053	0.061	0.056	0.056
(c) T=500				
(τ_1, τ_2, τ_3)				
(0.2, 0.5, 0,8)	0.058	0.057	0.058	0.057
(0.3, 0.5, 0.7)	0.049	0.049	0.048	0.050

Table 8. Empirical powers of the WB unit root test at the nominal 5% level when there are three breaks in variance.

	$(\delta_1, \delta_2, \delta_3)$	(2.5,0.8,0.25)	(2.0,0.25,1.25)	(0.4, .25,4.0)	(0.8,2.5,0.4)
(a) T=50					
(τ_1, τ_2, τ_3)	ρ=0.9				
(0.2, 0.5, 0,8)		0.109	0.080	0.103	0.104
(0.3, 0.5, 0.7)		0.111	0.084	0.108	0.107
(τ_1, τ_2, τ_3)	ρ=0.8				
(0.2, 0.5, 0,8)		0.274	0.184	0.279	0.268
(0.3, 0.5, 0.7)		0.279	0.188	0.282	0.296
(b) T=100					
(τ_1, τ_2, τ_3)	ρ=0.9				
(0.2, 0.5, 0,8)		0.259	0.183	0.247	0.266
(0.3, 0.5, 0.7)		0.269	0.156	0.290	0.266
(τ_1, τ_2, τ_3)	ρ=0.8				
(0.2, 0.5, 0,8)		0.720	0.526	0.768	0.757
(0.3, 0.5, 0.7)		0.721	0.527	0.799	0.733

Table 9. Empirical sizes of the KLN test at the nominal 5% level when there is a single break in variance.

δ_1	0.25	0.4	0.6	0.8	1	1.25	1.67	2.5	4
(a) T=50									
τ_1									
0.2	0.049	0.063	0.051	0.046	0.044	0.056	0.064	0.052	0.050
0.4	0.068	0.046	0.048	0.065	0.047	0.055	0.065	0.055	0.074
0.6	0.062	0.050	0.053	0.054	0.048	0.064	0.068	0.068	0.071
0.8	0.065	0.051	0.043	0.047	0.061	0.060	0.062	0.065	0.057
(b) T=100									
τ_1									
0.2	0.053	0.054	0.039	0.051	0.048	0.059	0.061	0.044	0.057
0.4	0.046	0.044	0.066	0.044	0.044	0.053	0.053	0.053	0.052
0.6	0.050	0.060	0.052	0.039	0.048	0.051	0.052	0.052	0.050
0.8	0.060	0.048	0.047	0.044	0.054	0.049	0.072	0.053	0.061

Table 10. Empirical sizes of the WB unit root test at the nominal 5% level when there is a single break in variance.

	δ_1	0.25	0.4	0.6	0.8	1	1.25	1.67	2.5	4
(a) T=50										
τ_1										
0.2		0.076	0.072	0.062	0.052	0.059	0.048	0.058	0.063	0.060
0.4		0.069	0.068	0.066	0.051	0.057	0.063	0.053	0.062	0.057
0.6		0.060	0.073	0.054	0.061	0.064	0.061	0.057	0.075	0.082
0.8		0.049	0.053	0.062	0.063	0.053	0.064	0.061	0.065	0.075
(b) T=100										
τ_1										
0.2		0.066	0.054	0.052	0.056	0.044	0.057	0.046	0.049	0.044
0.4		0.065	0.051	0.049	0.055	0.046	0.047	0.059	0.046	0.061
0.6		0.042	0.050	0.044	0.071	0.060	0.069	0.062	0.054	0.051
0.8		0.042	0.039	0.055	0.046	0.046	0.057	0.045	0.061	0.067

Table 11. Empirical powers of the KLN test at the nominal 5% level when there is a single break in variance.

	δ_1	0.25	0.4	0.6	0.8	1	1.25	1.67	2.5	4
(a) T=50										
τ_1	ρ=0.9									
0.2		0.057	0.085	0.090	0.084	0.083	0.090	0.088	0.075	0.069
0.4		0.057	0.073	0.081	0.092	0.089	0.098	0.092	0.097	0.112
0.6		0.087	0.090	0.088	0.083	0.067	0.094	0.091	0.117	0.099
0.8		0.072	0.060	0.080	0.086	0.093	0.078	0.095	0.096	0.104
τ_1	ρ=0.8									
0.2		0.171	0.207	0.238	0.236	0.239	0.238	0.242	0.231	0.245
0.4		0.161	0.206	0.217	0.239	0.242	0.229	0.243	0.277	0.247
0.6		0.175	0.216	0.234	0.226	0.245	0.270	0.231	0.230	0.269
0.8		0.174	0.218	0.206	0.231	0.242	0.225	0.229	0.254	0.277
(b) T=100										
τ_1	ρ=0.9									
0.2		0.163	0.190	0.232	0.222	0.238	0.234	0.235	0.233	0.235
0.4		0.160	0.188	0.217	0.235	0.240	0.244	0.229	0.238	0.262
0.6		0.162	0.181	0.189	0.234	0.234	0.220	0.207	0.243	0.237
0.8		0.203	0.213	0.250	0.226	0.240	0.249	0.228	0.242	0.264
τ_1	ρ=0.8									
0.2		0.710	0.753	0.726	0.733	0.762	0.757	0.764	0.751	0.713
0.4		0.675	0.716	0.717	0.736	0.760	0.748	0.729	0.703	0.735
0.6		0.686	0.709	0.726	0.738	0.755	0.715	0.716	0.728	0.748
0.8		0.752	0.747	0.745	0.717	0.756	0.745	0.722	0.755	0.782

Table 12. Empirical powers of the WB unit root test at the nominal 5% level when there is a single break in variance.

	δ_1	0.25	0.4	0.6	0.8	1	1.25	1.67	2.5	4
(a) T=50										
τ_1	ρ=0.9									
0.2		0.115	0.086	0.098	0.091	0.113	0.122	0.115	0.116	0.143
0.4		0.075	0.084	0.093	0.102	0.120	0.118	0.105	0.112	0.139
0.6		0.079	0.092	0.110	0.102	0.107	0.114	0.113	0.102	0.111
0.8		0.098	0.102	0.090	0.120	0.113	0.113	0.109	0.136	0.133
τ_1	ρ=0.8									
0.2		0.175	0.198	0.273	0.313	0.308	0.358	0.345	0.360	0.372
0.4		0.187	0.199	0.283	0.295	0.336	0.304	0.336	0.313	0.272
0.6		0.212	0.257	0.267	0.298	0.331	0.319	0.312	0.295	0.245
0.8		0.303	0.297	0.313	0.300	0.325	0.326	0.298	0.297	0.239
(b) T=100										
τ_1	ρ=0.9									
0.2		0.147	0.195	0.247	0.280	0.301	0.335	0.343	0.333	0.340
0.4		0.149	0.213	0.263	0.266	0.329	0.326	0.307	0.293	0.259
0.6		0.225	0.218	0.290	0.307	0.315	0.301	0.293	0.268	0.207
0.8		0.259	0.293	0.278	0.320	0.318	0.290	0.297	0.287	0.240
τ_1	ρ=0.8									
0.2		0.434	0.641	0.772	0.867	0.849	0.874	0.856	0.866	0.863
0.4		0.540	0.644	0.795	0.826	0.870	0.877	0.832	0.800	0.777
0.6		0.703	0.710	0.822	0.832	0.889	0.866	0.832	0.738	0.594
0.8		0.827	0.822	0.833	0.861	0.865	0.853	0.829	0.692	0.567

Table 13. Empirical sizes of three unit root tests at the nominal 5% level with conditional heteroskedasticity.

	θ	0	10	20	50
(a) T=50					
DF		0.044	0.062	0.044	0.043
KLN		0.051	0.064	0.054	0.056
WB		0.046	0.056	0.054	0.039
(b) T=100					
DF		0.055	0.055	0.047	0.050
KLN		0.051	0.052	0.044	0.058
WB		0.062	0.051	0.052	0.054

Table 14. Empirical powers of three unit root tests at the nominal 5% level with conditional heteroskedasticity.

	θ	0	10	20	50
(a) T=50	ρ=0.9				
DF		0.086	0.066	0.060	0.059
KLN		0.101	0.058	0.064	0.043
WB		0.114	0.135	0.141	0.134
	ρ=0.8				
DF		0.222	0.229	0.209	0.205
KLN		0.233	0.168	0.164	0.166
WB		0.342	0.410	0.414	0.420
(a) T=100	ρ=0.9				
DF		0.228	0.177	0.198	0.213
KLN		0.243	0.183	0.189	0.178
WB		0.312	0.359	0.346	0.388
	ρ=0.8				
DF		0.779	0.803	0.794	0.806
KLN		0.741	0.727	0.733	0.748
WB		0.875	0.884	0.881	0.883