Local Structural Trend Break in Stationarity Testing*

Anton Skrobotov†
Gaidar Institute for Economic Policy,
The Russian Presidential Academy
of National Economy and Public Administration

July 23, 2013

Abstract

In a recently publicized study, Harvey *et al.* (2012) investigated procedures for unit root testing employing break detection methods under local break in trend. We apply this methodology to analyze asymptotic and finite sample behavior of procedures under local break to test the stationarity null hypothesis local to unit root, against alternative hypothesis about the presence of a unit root. We extend the GLS-based stationarity test proposed by Harris *et al.* (2007) to the case of structural break and obtain asymptotic properties under local trend break. Two procedures are considered. The first procedure uses a with-break stationarity test, but with adaptive critical values. The second procedure utilizes the intersection of rejection testing strategy containing tests with and without a break. Application of these approaches help to prevent serious size distortions for small break magnitude that are otherwise undetectable. Additionally, in a similar approach as Harvey *et al.* (2013) and Busetti and Harvey (2001), we propose a test based on minimizing the sequence of GLS-based stationarity test statistics over all possible break dates. This infimum-test in contrast to Busetti and Harvey (2001) does not require an additional assumption about a faster rate of convergence of break magnitude. Asymptotic and finite sample simulations show that under local to zero behavior of the trend break the asymptotic analysis provides a good approximation of the finite sample behavior of the proposed procedures. Proposed procedures can be used for confirmatory analysis together with tests of Harvey *et al.* (2012) and Harvey *et al.* (2013).
Key words: Stationarity tests, KPSS tests, local break in trend, size distortions, intersection of rejection decision rule.
JEL: C12, C22

*We thank Andrey Polbin for his help in the implementation of simulations.
†E-mail: antonskrobotov@gmail.com

1 Introduction

Presence of a unit root in a time series is a major question in empirical data analysis and unit root testing plays an important role in applications. However, macroeconomic data are often characterized by a break in trend and starting with the work of Perron (1989) considerable research attention has focused on the impact of this break on unit root testing. Similar approaches have been developed to test the null hypothesis of stationarity of time series against a unit root alternative, starting with the work of Kwiatkowski *et al.* (1992) (hereafter KPSS). In a subsequent study, Busetti and Harvey (2001) generalized the KPSS test to the case of structural change and obtained the limiting distributions of the test statistics (see also Harvey and Mills (2003). For unknown break date authors proposed the infimum of the sequence of stationarity statistics for each possible break date. This approach, however, is based on an assumption that the magnitude of trend break converges to zero at a faster rate than $T^{-3/2}$ (for break in level the rate of convergence should be a faster than $T^{-1/2}$). Without this assumption, the limiting distribution of a test statistic will depend not only on the break fraction, but also on its magnitude. Simulations of Busetti and Harvey (2003) showed that the infimum-test has serious size distortions even for small break. Additionally, Busetti and Harvey (2003) proposed another approach based on the superconsistent break date estimator (see also Kurozumi (2002)). In this approach the first step consists of an estimation of a superconsistent estimate of break fraction under the null hypothesis of stationarity. Subsequently, this estimate is substituted into the KPSS statistic as true. Obtained limiting distribution of test statistic coincides with the limit distribution of known break date. However, for small breaks this procedure has significantly lower power than the infimum-test.

In a related study, Müller (2005) investigated the properties of conventional KPSS test under near integration. Results of this study revealed that the use of a bandwidth parameter in the long-run variance estimator, increasing at a slower rate than the length of the sample, leads to an asymptotic size equal to unity under the null hypothesis about near integration. This result helps to address the increasing size with highly autocorrelated stationary data generating process. It is clear that conventional KPSS tests that take into account the structural shift have a similar disadvantage. However, use of the bandwidth parameter in the long-run variance estimator increasing at the same rate as the length of the sample the KPSS test is dominated by point optimal test proposed by author. Harris *et al.* (2007) (hereafter HLM) proposed a modification of KPSS test based on GLS-detrending which was not affected by asymptotic size distortions and was comparable with the point optimal test of Müller[1].

Recent studies Harvey *et al.* (2012) (hereafter HLT12) and Harvey *et al.* (2013) (hereafter HLT13) addressed the problem of uncertainty concerning the presence and dating of structural break in the context of unit root testing. An intuitive approach is to use a pre-test to detect the break and then calculate the test statistic with or without this break. However, these methods are effective only in the case of a fixed or zero trend break, which in finite samples produces “valleys” in the power functions of tests; the power is high for a very small break, but declines rapidly with increasing magnitude of the break until it increases again. HLT12 proposed two strategies to address this issue. The first strategy prescribes to always to use a test with break, but with adaptive critical values. The second approach proposed to use the union of rejection of two tests, taking into account scenarios with an without break tests. Additionally, these authors developed local

[1]It should be noted that GLS-based KPSS test clearly dominates point optimal test in cases of a large initial value. The problem of initial conditions, however, is beyond the scope of this study

asymptotic theory for existing and new procedures by using local to zero behavior of the trend break. HLT13 proposed an alternative approach in which the test statistic is computed similarly to (Zivot and Andrews, 1992), minimizing the sequence of test statistics for all possible break dates using the GLS-detrended data.

In this study we extend GLS-based KPSS test proposed by HLM to the case of structural break through several approaches. The first approach uses the break fraction estimator obtained by minimizing the sum of squared residuals for the model in first differences. This estimator provides a superconsistent estimator of break fraction. The second approach uses modifications proposed in HLT12 in the context of stationarity testing and applies some pre-tests for testing for break presence. The third approach uses an infimum of sequence of GLS-detrended test statistics for all possible break dates. Whereas conventional infimum-based KPSS test requires additional assumption of the break magnitude it is not necessary for GLS-based test. We obtain the limiting distributions of all statistics under local to zero trend break and investigate the asymptotic behavior of these tests. Additionally, we assign asymptotic critical values for the above procedures. Asymptotic size shows similar properties obtained by HLT12. If we use specific tests (with or without break) which are based on some pre-tests, then the size of this testing procedure is small for very small breaks, and then increases as the break magnitude increases before decreasing again. We refer to this effect as size “hills” similar to the power “valleys” in the context of unit root testing. Finite sample simulations confirm a good approximation of the limiting distributions under local behavior of the trend break. Our proposed modifications offers improvement to test robustness for small trend break magnitude and smoothes the described size “hills”. The infimum test shows the most attractive properties alongside with simplicity of calculation.

The paper is organized in six sections and contains an Appendix. Following the introductory section we describe the model with (local to zero) trend break and GLS-based KPSS test statistic in Section 2. In Section 3 we obtain the limiting distributions of test statistics under local break behavior. Alternative procedures in the stationarity testing context are described in Section 4. Asymptotic and finite sample behavior of all considered tests is investigated in Section 5. Section 6 presents our conclusions. All proofs are collected in the Appendix section.

2 Model

Consider the data generating process (DGP) in the case of break in trend as

$$y_t = \mu + \beta t + \gamma_T DT_t(\lambda_0) + u_t,\ t = 1, \dots, T \tag{1}$$

$$u_t = \rho u_{t-1} + \varepsilon_t,\ t = 2, \dots, T \tag{2}$$

where $DT_t(\lambda_0) = (t - \lfloor \lambda_0 T \rfloor)\mathbb{I}(t > \lfloor \lambda_0 T \rfloor)$, $\mathbb{I}(\cdot)$ is the indicator function and the trend break occurs at time $\lfloor \lambda_0 T \rfloor$ (λ_0 is the corresponding break fraction), if break magnitude $\gamma_T \neq 0$. It is assumed that the true break fraction λ_0 is unknown, but belongs to the range $\Lambda = [\lambda_L, \lambda_U]$, $0 < \lambda_L < \lambda_U < 1$, λ_L and λ_U are trimming parameters.

The assumption about initial value in (2) is such that $u_1 = O_p(T^{1/2})$ and linear process ε_t satisfies the standard assumptions (see Phillips and Solo (1992)):

Assumption 1 *Let*

$$\varepsilon_t = \gamma(L)e_t = \sum_{i=0}^{\infty} \gamma_i e_{t-i},$$

with $\gamma(z) \neq 0$ *for all* $|z| \leq 1$ *and* $\sum_{i=0}^{\infty} i|\gamma_i| < \infty$*, where* e_t *is the martingale difference sequence with conditional variance* σ_e^2 *and* $\sup_t \mathbb{E}(e_t^4) < \infty$*. The short-run and long-run variances of* ε_t *are defined as* $\sigma_\varepsilon^2 = E(\varepsilon_t^2)$ *and* $\omega_\varepsilon^2 = \lim_{T\to\infty} T^{-1}\mathbb{E}\left(\sum_{t=1}^{T} \varepsilon_t\right)^2 = \sigma_e^2\gamma(1)^2$*, respectively.*

The autoregressive parameter in (2) is $\rho = \rho_T = 1 - c/T$, where $c \geq 0$. Our purpose is testing the null hypothesis of stationarity against alternative about a unit root regardless of whether the break in trend is present or not. We test the hypothesis $H_0 : c \geq \bar{c} > 0$ against the alternative $H_1 : c = 0$, where $\bar{c}$ is the minimal amount of mean reversion under the stationary null hypothesis. However, in contrast to Kurozumi (2002) and Busetti and Harvey (2001), who considered the break magnitude γ_T as fixed (independent of sample size, T), we consider the break magnitude as local to zero, that is $\gamma_T = \kappa\omega_\varepsilon T^{-1/2}$ as in HLT12 and HLT13.

Because conventional KPSS tests have the asymptotic size equal to unity under local to unit root behavior of ρ_T (see (Müller, 2005)), for stationarity testing we can use a modification of the KPSS test for (quasi) GLS-detrended data proposed by HLM[2] (see also Skrobotov (2013) where uncertainty about the presence of linear trend is investigated). Specifically, the test statistic in the absence of break (only with linear trend) is constructed as

$$S^t(\bar{c}) = \frac{T^{-2}\sum_{t=2}^{T}\left(\sum_{j=2}^{t}\tilde{u}_j^t\right)^2}{\hat{\omega}_\varepsilon^2}, \tag{3}$$

where $\tilde{u}_t$ are OLS residuals from regression of $\mathbf{y}_{\bar{c}} = y_t - \bar{\rho}_T y_{t-1}$ on $\mathbf{Z}_{\bar{c}} = z_t - \bar{\rho}_T z_{t-1}$, $t = 2, \ldots, T$, where $z_t = (1, t)'$ (that is, when $\gamma_T = 0$).

We propose the extension of this test statistic to the case of a single structural break in trend:

$$S^{tb}(\bar{c}, \lambda) = \frac{T^{-2}\sum_{t=2}^{T}\left(\sum_{j=2}^{t}\tilde{u}_j^{tb}\right)^2}{\hat{\omega}_\varepsilon^2}, \tag{4}$$

where $\tilde{u}_t^{tb}$ are OLS residuals from regression of $\mathbf{y}_{\bar{c}} = y_t - \bar{\rho}_T y_{t-1}$ on $\mathbf{Z}_{\bar{c}} = z_t - \bar{\rho}_T z_{t-1}$, $t = 2, \ldots, T$, где $z_t = (1, t, DT_t(\lambda))'$. For all statistics (3) and (4) the $\hat{\omega}_\varepsilon^2$ is any consistent long-run variance estimator of ε_t.[3].

If the break date is unknown, Harris *et al.* (2009) proposed to use break date estimator $\tilde{\lambda}$, based on first differenced regression (1):

$$\tilde{\lambda} = \arg\min_{\lambda\in\Lambda} S(1, \lambda), \tag{5}$$

where $S(1, \lambda)$ is the sum of squared residuals from regression $\mathbf{y}_0 = y_t - y_{t-1}$ on $\mathbf{Z}_0 = z_t - z_{t-1}$, where $z_t = (1, t, DT_t(\lambda))'$. This estimate is superconsistent for all $0 \leq c$ (see Harris *et al.* (2009)).

[2]In HLM authors considered only constant and trend cases and analyzed only the constant case.

[3]E.g., HLM used nonparametric spectral density estimator with quadratic spectral (QS) kernel and the automatic bandwidth selection of (Newey and West, 1994). For calculation corresponding GLS-detrended residuals, $\tilde{u}_t$ or $\tilde{u}_t^{tb}$, are used.

3 Asymptotic behavior of stationarity tests

We investigate the asymptotic size of the S^t ans $S^{tb}(\tilde{\lambda})$ tests under local to unit root behavior and local to zero behavior of break in trend magnitude $\gamma_T = \kappa\omega_\varepsilon T^{-1/2}$, where κ is some constant[4]. Asymptotic distribution of S^t statistic is given in the following theorem.

Theorem 1 *Let $\{y_t\}$ is generated as (1) and (2) and Assumption 1 holds. Then under $\rho_T = 1 - c/T$, $0 \leq c < \infty$*

$$S^t \;\Rightarrow\; \int_0^1 \left(H_{c,\bar{c}}(r, \lambda_0, \kappa) - 6r(1-r)\int_0^1 H_{c,\bar{c}}(s, \lambda_0, \kappa)ds \right)^2 dr, \tag{6}$$

where

$$\begin{aligned} H_{c,\bar{c},\kappa}(r, \lambda_0, \kappa) &= H_{c,\bar{c}}(r) + \kappa l(r, \lambda_0), \\ H_{c,\bar{c}}(r) &= W_c(r) + \bar{c}\int_0^r W_c(s)ds - r\left[W_c(1) + \bar{c}\int_0^1 W_c(s)ds\right], \\ l(r, \lambda_0) &= \left[(r-\lambda_0) + \frac{\bar{c}}{2}(r-\lambda_0)^2\right]\mathbb{I}(r > \lambda_0) - r\left((1-\lambda_0) + \frac{\bar{c}}{2}(1-\lambda_0)^2\right). \end{aligned}$$

Remark 1 Proof of Theorem 1 is given in the Appendix. It should be noted that for $\kappa = 0$ (in the absence of a break) the limiting distribution coincides with the distribution obtained in HLM. However for $\kappa \neq 0$ the limiting distribution of the statistic depends on dating and magnitude of a local trend break.

Additionally, we obtain the limiting distribution of the test with break, $S^{tb}(\lambda)$, implemented for some generic break fraction λ, which may be different from the true break fraction, λ_0. The results is given in Theorem 2.

Theorem 2 *Let $\{y_t\}$ is generated as (1) and (2) and Assumption 1 holds. Then under $\rho_T = 1 - c/T$, $0 \leq c < \infty$*

$$S^{tb}(\lambda) \Rightarrow \int_0^1 H^{tb}_{c,\bar{c}}(r, \lambda_0, \lambda, \kappa)^2 dr, \tag{7}$$

where

$$\begin{aligned} H^{tb}_{c,\bar{c}}(r, \lambda_0, \lambda, \kappa) &= W_c(r) + \bar{c}_\lambda \int_0^r W_c(r)dr + \kappa\left[(r-\lambda_0) + \frac{\bar{c}_\lambda}{2}(r-\lambda_0)^2\right]\mathbb{I}(r > \lambda_0) \\ &- \begin{bmatrix} \bar{c}_\lambda r \\ \bar{c}_\lambda \frac{r^2}{2} + r \\ (\bar{c}_\lambda \frac{r^2}{2} + r + \bar{c}_\lambda \frac{\lambda^2}{2} - \lambda - \bar{c}_\lambda \lambda r)\mathbb{I}(r > \lambda_0) \end{bmatrix}' \begin{bmatrix} \bar{c}_\lambda^2 & \bar{c}_\lambda + \bar{c}_\lambda^2/2 & k_{\bar{c}_\lambda}(\lambda) \\ \bar{c}_\lambda + \bar{c}_\lambda^2/2 & 1 + \bar{c}_\lambda + \bar{c}_\lambda^2/3 & m_{\bar{c}_\lambda}(\lambda) \\ k_{\bar{c}_\lambda}(\lambda) & m_{\bar{c}_\lambda}(\lambda) & d_{\bar{c}_\lambda}(\lambda) \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} a_{c,\bar{c}_\lambda} + \kappa q_{c,\bar{c}_\lambda}(\lambda_0) \\ b_{c,\bar{c}_\lambda} + \kappa f_{c,\bar{c}_\lambda}(\lambda_0) \\ b_{c,\bar{c}_\lambda}(\lambda) + \kappa f_{c,\bar{c}_\lambda}(\lambda_0, \lambda) \end{bmatrix} \end{aligned}$$

[4] Here and below, we omit the dependence of the tests on the $\bar{c}$ parameter for brevity.

with

$$
\begin{aligned}
k_{\bar{c}_\lambda}(\lambda) &= \bar{c}_\lambda + \bar{c}_\lambda^2/2 - \lambda(\bar{c}_\lambda + \bar{c}_\lambda^2 - \bar{c}_\lambda^2\lambda/2), \\
m_{\bar{c}_\lambda}(\lambda) &= 1 + \bar{c}_\lambda + \bar{c}_\lambda^2/3 - \lambda(1 + \bar{c}_\lambda + \bar{c}_\lambda^2/2 - \bar{c}_\lambda^2\lambda^2/6), \\
d_{\bar{c}_\lambda}(\lambda) &= 1 + \bar{c}_\lambda + \bar{c}_\lambda^2/3 - \lambda(1 + 2\bar{c}_\lambda - \bar{c}_\lambda\lambda + \bar{c}_\lambda^2 - \bar{c}_\lambda^2\lambda + \bar{c}_\lambda^2\lambda^2/3), \\
a_{c,\bar{c}_\lambda} &= \bar{c}_\lambda W_c(1) + \bar{c}_\lambda^2 \int_0^1 W_c(s)ds, \\
q_{c,\bar{c}_\lambda}(\lambda_0) &= \bar{c}_\lambda(1-\lambda_0) + \bar{c}_\lambda^2(1-\lambda_0)^2/2, \\
b_{c,\bar{c}_\lambda} &= (1+\bar{c}_\lambda)W_c(1) + \bar{c}_\lambda^2 \int_0^1 sW_c(s)ds, \\
f_{c,\bar{c}_\lambda}(\lambda_0) &= (1-\lambda_0)[1 + \bar{c}_\lambda + \bar{c}_\lambda^2/3 - \bar{c}_\lambda^2\lambda_0(1+\lambda_0)/6], \\
b_{c,\bar{c}_\lambda}(\lambda) &= (1 + \bar{c}_\lambda - \bar{c}_\lambda\lambda)W_c(1) - W_c(\lambda) + \bar{c}_\lambda^2 \int_\lambda^1 (s-\lambda)W_c(s)ds, \\
f_{c,\bar{c}_\lambda}(\lambda_0,\lambda) &= (1-\lambda_0)\{1 + \bar{c}_\lambda + \bar{c}_\lambda^2/3 - \bar{c}_\lambda\lambda - \bar{c}_\lambda^2\lambda(1-\lambda_0)/2 - \bar{c}_\lambda^2\lambda_0(1+\lambda_0)/6\} \\
&\quad -(\lambda-\lambda_0)\{1 - \bar{c}_\lambda^2(\lambda-\lambda_0)^2/6\}\mathbb{I}(\lambda > \lambda_0).
\end{aligned}
$$

Remark 2 Proof of Theorem 2 is given in the Appendix. For $\lambda = \lambda_0$ it can be shown that the limiting distribution of the $S^{tb}(\lambda)$ statistic does not depend on break magnitude κ. At the same time, if we use the break fraction estimate $\tilde{\lambda}$ then under local behavior of the trend break this estimate can not be consistently estimated and will be different from the true break fraction λ_0 in a general case. More specifically, if we use the break date estimate as (5), this estimate under local behavior of both autoregressive root and the break parameter has the following asymptotic distribution (see HLT12):

$$
\tilde{\lambda} \Rightarrow \arg\sup_{\lambda\in\Lambda} \begin{bmatrix} W_c(1) + \kappa(1-\lambda_0) \\ W_c(1) - W_c(\lambda) + \kappa(1-\lambda_0) - \kappa(\lambda-\lambda_0)\mathbb{I}(\lambda > \lambda_0) \end{bmatrix}' \begin{bmatrix} 1 & (1-\lambda) \\ (1-\lambda) & (1-\lambda) \end{bmatrix}^{-1}
\times \begin{bmatrix} W_c(1) + \kappa(1-\lambda_0) \\ W_c(1) - W_c(\lambda) + \kappa(1-\lambda_0) - \kappa(\lambda-\lambda_0)\mathbb{I}(\lambda > \lambda_0) \end{bmatrix}
$$

Then, the limiting distribution of $S^{tb}(\tilde{\lambda})$ test follows directly from the continuous mapping theorem (CMT). I.e., for $\tilde{\lambda} \neq \lambda_0$ the limiting distribution of $S^{tb}(\tilde{\lambda})$ will depend on dating and magnitude of the local break in trend and also on the estimated break fraction $\tilde{\lambda}$.

Remark 3 For $\kappa = 0$ the $S^{tb}(\tilde{\lambda})$ test will have the asymptotic size that is higher than the nominal one for a specified critical value at $c = \bar{c}_\lambda$, for small κ the size of test will increase slightly and only for a moderate κ the size will be close to a nominal one. We set the $\bar{c}$ and $c = \bar{c}_\lambda$ parameters as in corresponding unit root tests, equal to 13.5 (as in Elliott *et al.* (1996)) and 17.6 (as in HLT13).

Remark 4 For $c = \bar{c}$ the S^t and $S^{tb}(\tilde{\lambda})$ tests have the same limiting distribution as in KPSS and Busetti and Harvey (2001), respectively. This allows the use of known critical values for specified $c = \bar{c}$ or $c = \bar{c}_\lambda$ as in HLM.

4 Alternative procedures to mitigate the effect of the size hills

If there is knowledge about presence of the break in data then it is necessary to use the $S^{tb}(\tilde{\lambda})$ test allowing this break as otherwise liberal size distortions will take place. At the same time, if the trend break does not occur ($\gamma_T = 0$), then the $S^{tb}(\tilde{\lambda})$ test allowing the break can be used, but the S^t test that does not allow the break will be effective (as it has smaller size and higher power).

In Harris *et al.* (2009) and Carrion-i-Silvestre *et al.* (2009) testing strategies were proposed in the context of unit root testing, in which under uncertainty over the presence break the test statistics with or without break is selected on the basis of some pre-tests. In these cases, if the break is detected by some artificial statistics then the unit root test (in our case the stationarity test) with break should be used. In this paper let B be some statistics for testing $\gamma_T = 0$ with a priori unknown whether the series is trend-stationary or contains an autoregressive unit root. If the null hypothesis is not rejected, i.e., $B < cv_B$, where cv_B is corresponding critical value for B statistic, then it is recommended to use the stationarity test without break (S^t in our case). If, however, the null hypothesis will be rejected, i.e., $B \geq cv_B$, then the S^t test will have serious size distortions due to a neglected break, and it is necessary to use the $S^{tb}(\tilde{\lambda})$ test. This strategy can be written as follows:

$$S(B) = \begin{cases} S^t & \text{if } B < cv_B \\ S^{tb}(\tilde{\lambda}) & \text{if } B \geq cv_B \end{cases}, \tag{8}$$

where t_{PY} test proposed by Perron and Yabu (2009) or t_λ test proposed by Harvey *et al.* (2009)[5] can be used as B pre-test.

This testing strategy, however, has an important drawback. The break magnitude may be too small to be reliably detected (and the S^t test without break will tend to be used), but it can be sufficiently large to radically increase the size of S^t (and, therefore, the overall strategy including this test) as this test ignores a break. We call these serious size distortions in the intermediate range of small κ size “hills” (similar to power “valleys” that arise in the unit root testing context, see HLT12). In simulation analysis of Section 5 the effect of size “hills” will be shown. Notably, this effect is undetectable considering the fixed break magnitude, as the test in (8) asymptotically selects the correct statistics, S^t or S^{tb}.

Similar to analysis in HLT12, there are several approaches to smooth the effect of size “hills”. One approach is to always implement a test with break $S^{tb}(\tilde{\lambda})$, as in this case the size “hills” do not arise because the test without break is not used. Then, for $\kappa = 0$ the size of test is larger than the nominal one[6], while the size is close to the nominal one for large κ. Therefore, for better size control over all κ this test with conservative critical values (obtained for $\kappa = 0$) can be implemented. However, for large κ the power of $S^{tb}(\tilde{\lambda})$ with conservative critical values will be lower than the power of this test with typical critical values.

The first method proposed by HLT in the unit root testing context, provides control on size through use of conservative critical values cv_{tb}^{consv} (obtained for $\kappa = 0$), if the trend break is not detected by some pre-test[7]. Simultaneously, if the break is detected by a pre-test (indicating a

[5] We do not give exact formulas for t_{PY} and t_λ tests to preserve brevity. For a brief description see HLT2012, Section 3.

[6] Preliminary simulations show that the asymptotic size of the $S^{tb}(\tilde{\lambda})$ for $c = \bar{c}_\lambda = 17.6$ is equal to 0.12 in case of break absence.

[7] At the nominal 10, 5 and 1 percent significance level, cv_{tb}^{consv} are 0.092, 0.115 and 0.173, respectively.

fairly large value of κ), the conventional critical values $cv_{tb}^{\tilde{\lambda}}$ (associated with a known break fraction, obtained for $\lambda = \lambda_0$) can be used. This alternative procedure can be written as

$$AS(B) = S^{tb}(\tilde{\lambda}) \text{ with critical values } \begin{cases} cv_{tb}^{consv} & \text{if } B < cv_B \\ cv_{tb}^{\tilde{\lambda}} & \text{if } B \geq cv_B \end{cases}. \tag{9}$$

This procedure allows to avoid power losses for large breaks. Similar to test $S(B)$ in (8), one of two pre-tests, t_{PY} or t_λ, can be implemented in this case as B test.

The second method is the simultaneous use of two tests, S^t and $S^{tb}(\tilde{\lambda})$ (the latter with conservative critical values), when the break is not detected. When there is clear evidence of a break in trend there is no need to use the S^t test (its size tends to unity), and only the $S^{tb}(\tilde{\lambda})$ statistic should be implemented (with a critical value associated with a known break fraction, $\lambda = \lambda_0$). This strategy of intersection of rejection[8] can be written as follows:

$$IR(B) = \begin{cases} \text{Reject } H_0 \text{ if } \{S^t > m_\xi cv_t \text{ and } S^{tb}(\tilde{\lambda}) > m_\xi cv_{tb}^{consv}\} & \text{if } B \leq cv_B \\ \text{Reject } H_0 \text{ if } \{S^{tb}(\tilde{\lambda}) > cv_{tb}^{\tilde{\lambda}}\} & \text{if } B > cv_B \end{cases}, \tag{10}$$

where m_ξ is some scaling constant ensuring that the asymptotic size equals ξ for a given value c in the joint implementation of S^t and $S^{tb}(\tilde{\lambda})$ tests[9]. In case of scaling absence, the size and power of tests decreases, so we refer to the decision rule using the scaling as liberal. We use the cv_t and cv_{tb}^{consv} to control the size for small break magnitudes.

The third way follows the approach of Busetti and Harvey (2001) and HLT2013, where the test statistic is minimized over all possible break dates. More precisely, this statistic is constructed as

$$MS = \inf_{\lambda \in \Lambda} S^{tb}(\lambda). \tag{11}$$

It should be noted, that while for infimum-based KPSS test in Busetti and Harvey (2001) an additional assumption is needed so that the magnitude of trend break converges to zero at a faster rate than $T^{-3/2}$, for present GLS-based modification this assumption is not necessary in our local to unity asymptotic framework.

The limiting distributions of all tests, (8), (9), (10) and (11), directly follow from the results of Theorems 1 and 2 and applications of the CMT (the results for MS test also follow from the arguments proved by (Zivot and Andrews, 1992)), and we omit the exact formulas of these distributions to save space.

5 Simulation analysis

5.1 Asymptotic size

In this section we consider the asymptotic size of procedures proposed in the previous section under local break in trend[10]. As the $IR(B)$ procedure in (10) uses both the S^t test for $\bar{c} = 13.5$

[8] In stationarity testing context we reject the null hypothesis if all of the tests reject it, and we call this strategy the intersection of rejections in contrast to union of rejection strategy in HLT12 for unit root testing.

[9] Corresponding scaling constants at 10, 5 and 1 percent significance level are 0.661, 0.633 and 0.620, respectively.

[10] Results are obtained by simulations of the limiting distributions of test statistics, approximating the Wiener process by $i.i.d.N(0,1)$ random variates and with integrals approximated by normalized sums of 2,000 steps, with 50,000 replications.

and $S^{tb}(\tilde{\lambda})$ test for $\bar{c} = 17.6$, and that, for given $c = \bar{c}$ each of these two tests has an asymptotic size equal to nominal one, it becomes problematic to compare all of the tests by fixing the size at $c = 17.6$ or $c = 13.5$. Therefore, similar to HLM, (Müller, 2005) and (Skrobotov, 2013), we compare size ($c > 0$) of all tests by fixing power ($c = 0$) at given level.

For a given break fraction λ_0 we calculate power ($c = 0$, critical values are obtained so that the power is 0.70) of every test, considered in previous sections, over all $\kappa = \{0, \ldots, 15\}$. Let the κ^* parameter denotes κ, for which the specific test has minimum power. The (power-adjusted) size curves are calculated by scaling the critical values of a particular procedure so that the power is 0.70 for $\kappa = \kappa^*$ (i.e., power is never below 0.70). The same scaling applies for all κ. It should be noted that the choice of setting power at a value of 0.70 is not crucial, and this value is chosen for visualization convenience.

Importantly, we note the following for the power of the $S^{tb}(\tilde{\lambda})$ test. Preliminary simulations show that for $\kappa = 0$ the power of this test (with using of 70% points) calculated for $\tilde{\lambda} = \lambda_0$ (known break date) is equal to 0.66 (lower than 0.70), i.e., the liberal critical value should be used instead of conservative critical value (e.g. for $AS(B)$ test in order for the power to equal 0.70 for $\kappa = 0$). However, for 5% point and $c = \bar{c}_\lambda$ the results will be opposite, i.e. the size for large κ will be higher than for small κ. However, this does not affect the final results, as we are interested in the trade-off between size and power of tests by using multiple tests simultaneously.

Figures 1(a)-(c), 2(a)-(c) and 3(a)-(c) show the asymptotic power-adjusted size of S^t, $S^{tb}(\tilde{\lambda})$, $S(t_{PY})$, $S(t_\lambda)$ tests for $c = 10, 20$ and 30, respectively. It is seen that the S^t test without break has a lower size across all tests, but with increasing κ its size increases rapidly to unity. The $S^{tb}(\tilde{\lambda})$ test is more robust to κ, but its size is much higher than the size of the S^t test for small κ. The $S(t_{PY})$ and $S(t_\lambda)$ tests with pre-testing for trend coefficient have a size close to that of S^t for small κ. This is due to the fact that for small values of the break the pre-tests fail to detect the presence of the break. However, for intermediate κ these pre-tests suffer from serious size distortions approaching to unity for every reasonable value of c, because the break is too small to be detected by pre-tests and the $S(t_{PY})$ and $S(t_\lambda)$ tests inherit the properties of S^t for moderate κ, i.e. the size "hills" effect occurs. For large κ the break will almost always be detected by pre-tests, thus the size will be close to the size of a with-break test.

Figures 1(d)-(f), 2(d)-(f) and 3(d)-(f) show the asymptotic power-adjusted size of $S^{tb}(\tilde{\lambda})$, $AS(t_{PY})$, $AS(t_\lambda)$, $IS(t_{PY})$, $IS(t_\lambda)$ and MS tests for $c = 10, 20$ and 30, respectively. The $AS(t_{PY})$ and $AS(t_\lambda)$ tests do not show the size "hills", for small κ the size of their tests is slightly higher than that of $S^{tb}(\tilde{\lambda})$, while for large κ the size reduces in comparison to $S^{tb}(\tilde{\lambda})$ (especially for the $\lambda_0 = 0.3$ case, less pronounced in case of $\lambda_0 = 0.7$). The size of $IS(t_{PY})$ and $IS(t_\lambda)$ tests for small κ considerably gains in comparison to $S^{tb}(\tilde{\lambda})$ and $AS(\cdot)$ due to inclusion of S^t in testing strategy. However, for moderate κ the size "hills" are still observed (due to scaling of critical values), although much less pronounced compared to $S(t_{PY})$ and $S(t_\lambda)$.

Interesting features are observed in the MS test that minimizes the sequence of GLS-detrended KPSS statistics over all possible break date. For $c = 10$ the size curve of MS is close to $S^{tb}(\tilde{\lambda})$ across all κ. Increasing c, for $c = 20$, the size curve of MS is lower than that of $S^{tb}(\tilde{\lambda})$ almost everywhere. For $\lambda_0 = 0.3$ for moderate and large κ the size curves of these two tests are nearly identical, but for small κ the MS test outperforms $S^{tb}(\tilde{\lambda})$. Finally, for $c = 30$ the size of MS is significantly lower for all tests for $\lambda_0 = 0.5$ and $\lambda_0 = 0.7$, except for a small range at small κ, where $IS(t_{PY})$ and $IS(t_\lambda)$ have minimal size across all tests as they use the S^t without break in construction. For $\lambda_0 = 0.3$ the size of MS is lower than the size of $S^{tb}(\tilde{\lambda})$ for all κ, however, it has higher size than $AS(t_{PY})$, $AS(t_\lambda)$, $IS(t_{PY})$, $IS(t_\lambda)$ for large κ and $IS(t_{PY})$ and $IS(t_\lambda)$ for small κ. In general, the MS test shows the best asymptotic properties across all considered tests and better robustness over all κ.

5.2 Finite sample evidence

In this section we investigate the size of all considered tests, S^t, $S^{tb}(\tilde{\lambda})$, $S(t_\lambda)$, $S(t_{PY})$, $AS(t_\lambda)$, $AS(t_{PY})$, $IS(t_\lambda)$, $IS(t_{PY})$ и MS, in finite samples by using the sample size $T = 150$ and 50,000 replications. We calculate power adjusted size ($c > 0$) similar to Section 5.1. The break magnitude is considered equal to $\gamma_T = \kappa T^{-1/2}$ with $\kappa = \{0, \ldots, 15\}$, $\lambda_0 = 0.5$. It should be noted that while nonparametric long-run variance estimator $\hat{\omega}_\varepsilon^2$ (based on corresponding GLS-detrended residuals and by using quadratic spectral (QS) kernel with automatic choice of bandwidth proposed by Newey and West (1994)) performs quite well in finite samples for all considered tests for $i.i.d.$ and AR(1) errors it is still poorly implemented in case of a negative MA(1) component. The use of autoregressive estimator as in Ng and Perron (2001) and Perron and Qu (2007) leads to significant improvement for MA(1) errors, and for $i.i.d.$ and AR(1) errors properties comparable to nonparametric estimation of long-run variance. However, for $T = 150$ the autoregressive long-run variance estimator for S^t test performs worse than the non-parametric estimator, so in simulations we calculate the S^t test by using the latter (with finite sample critical values, otherwise the power of test will be lower than 0.7; for autoregressive estimator in $S^{tb}(\tilde{\lambda})$ tests the asymptotic critical values lead to a power that is very close 0.70). Unreported results for lager T approach to asymptotic.

Figures 4(a)-(c), 5(a)-(c) and 6(a)-(c) show the finite sample power-adjusted size of S^t, $S^{tb}(\tilde{\lambda})$, $S(t_{PY})$, $S(t_\lambda)$ for $c = 10, 20$ and 30, respectively, and Figures 4(d)-(f), 5(d)-(f) and 6(d)-(f) show the finite sample power-adjusted size of $S^{tb}(\tilde{\lambda})$, $AS(t_{PY})$, $AS(t_\lambda)$, $IS(t_{PY})$, $IS(t_\lambda)$ and MS for $c = 10, 20$ and 30, respectively. First, we consider the behavior of the tests under DGP (1) and (2), when the errors $\varepsilon_t \sim i.i.d.N(0,1)$. Results are qualitatively similar to asymptotic except the finite sample size is higher than the asymptotic size, and the size "hills" are somewhat less pronounced (even for $S(t_{PY})$ and $S(t_\lambda)$ tests). The size of MS is also below all others except for a small range of small κ, where this tests is dominated by $S(t_{PY})$ and $S(t_\lambda)$. This range decreases with increasing of c. Also with increasing c the size of MS is improved in comparison to other tests.

Next, we consider the possibility of serial correlation of errors when ε_t follows AR(1) or MA(1) processes. More precisely, for AR(1) process $\varepsilon_t = 0.5\varepsilon_{t-1} + e_t$, for MA(1) process $\varepsilon_t = e_t - 0.5e_{t-1}$, where $e_t \sim i.i.d.N(0,1)$. The results for AR(1) errors are similar to $i.i.d.$ case except that the size "hill" appears to the right in comparison with the $i.i.d$ case, and the size is higher than in $i.i.d.$ case. For MA(1) errors the size of all tests is almost the same (except for $S(t_{PY})$ и $S(t_\lambda)$), but slightly higher than in the AR(1) case; the size "hill" now appears to the left of the $i.i.d$ case. However, for $c = 20$ and $\kappa > 1$ the MS test outperforms all other tests, and for $c = 30$ the MS test is the best for all κ. Additionally, the size of MS changes are small enough with κ, as in asymptotic simulations.

6 Conclusion

In this paper different procedures for stationarity testing (local to unit root) with structural break in trend against alternative hypothesis about the presence of a unit root are considered. The GLS-based stationarity test proposed by HLM was extended to the case of structural break. For this test we proposed testing strategies similar to HLT12 based on pre-testing for the trend break parameter.

We investigated the effect of size "hills" (considerable increases in size for a small break), when procedures for break detection are used in the testing strategy. Asymptotic behavior of all procedures is analyzed under local break in trend and provides a good approximation of the finite sample behavior. Results show that considered procedures allow to smooth the effect of size "hills". Moreover, we investigated the behavior of infimum-test for GLS-detrended data. This test shows greater size robustness across break magnitude and reveals better properties in many cases. Thus, the proposed procedures are useful in empirical applications as a complement to the unit root tests HLT12 and HLT13 for confirmatory analysis.

Generalization of testing strategies to the case of multiple structural changes (similar to HLT13) should be noted as an avenue for future studies. For $AS(t_{PY})$, $AS(t_\lambda)$, $IS(t_{PY})$, $IS(t_\lambda)$ tests this can be nontrivial, but the extension of infimum-test MS should be straightforward taking into account the very attractive properties this test. Additionally, proposed tests can be extended allowing a break in level to occur at the same time as the break in trend, as in contrast to the unit root test, the break in level will no longer be asymptotically negligible and will have an impact on the size and power of the stationarity tests.

Appendix

Proof of Theorem 1.

Consider the estimates $\tilde{\mu}$ and $\tilde{\beta}$:

$$\begin{bmatrix} \tilde{\mu} \\ \tilde{\beta} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \tag{12}$$

where

$$\begin{aligned}
g_{11} &= (1-\bar{\rho})^2(T-1), \\
g_{12} &= (1-\bar{\rho})\sum_{t=2}^{T}\{t-\bar{\rho}(t-1)\}, \\
g_{22} &= \sum_{t=2}^{T}\{t-\bar{\rho}(t-1)\}^2, \\
h_1 &= (1-\bar{\rho})\sum_{t=2}^{T}(y_t-\bar{\rho}y_{t-1}), \\
h_2 &= \sum_{t=2}^{T}(y_t-\bar{\rho}y_{t-1})\{t-\bar{\rho}(t-1)\}.
\end{aligned}$$

The limits included in matrix 2×2 are the following: $Tg_{11} \to \bar{c}^2$, $g_{12} \to \bar{c} + \bar{c}^2/2$, $T^{-1}g_{22} = 1 + \bar{c} + \bar{c}^2/3$. For 2×1 vector the corresponding limits are the following:

$$\begin{aligned} T^{1/2}h_1 &= \bar{c}T^{-1/2}(y_T - y_1) + \bar{c}^2 T^{-3/2}\sum_{t=2}^{T} y_{t-1} \\ &= \bar{c}T^{-1/2}u_T + \bar{c}^2T^{-3/2}\sum_{t=2}^{T} u_{t-1} + \bar{c}\kappa + \bar{c}^2T^{-3/2}\sum_{t=2}^{T} u_{t-1} \\ &\quad + T^{-1}(T - \lfloor \lambda_0 T \rfloor) + \bar{c}^2 T^{-3/2}\sum_{t=2}^{T} DT_{t-1}(\lambda_0) \\ &\Rightarrow \bar{c}W_c(1) + \bar{c}^2\int_0^1 W_c(s)ds + \kappa[\bar{c}(1-\lambda_0) + \bar{c}^2(1-\lambda_0)^2/2], \end{aligned}$$

$$T^{-1/2}h_2 \Rightarrow b_{c,\bar{c}} + \kappa f_{c,\bar{c}}(\lambda_0),$$

where obtaining the h_2 is similar to HLT.

Thus,

$$\begin{bmatrix} T^{-1/2}\tilde{\mu} \\ T^{1/2}\tilde{\beta} \end{bmatrix} = \begin{bmatrix} Tg_{11} & g_{12} \\ g_{12} & T^{-1}g_{22} \end{bmatrix}^{-1} \begin{bmatrix} T^{1/2}h_1 \\ T^{-1/2}h_2 \end{bmatrix} \tag{13}$$

where the rate of convergence $T^{-1/2}$ before $\tilde{\mu}$ coefficient differs from HLT, because when later finding GLS-residuals the component depending on от $\tilde{\mu}$ is asymptotically negligible, but in our case after the summation (see (18) and (19)) it has a nondegenerate limiting distribution.

Consider the first 2×2 matrix from the equation (13):

$$\begin{aligned} \begin{bmatrix} Tg_{11} & g_{12} \\ g_{12} & T^{-1}g_{22} \end{bmatrix}^{-1} &= \begin{bmatrix} \bar{c}^2 & \bar{c} + \bar{c}^2/2 \\ \bar{c} + \bar{c}^2/2 & (1 + \bar{c} + \bar{c}^2/3) \end{bmatrix}^{-1} \\ &= \frac{12}{\bar{c}^4}\begin{bmatrix} 1 + \bar{c} + \bar{c}^2/3 & -\bar{c} - \bar{c}^2/2 \\ -\bar{c} - \bar{c}^2/2 & \bar{c}^2 \end{bmatrix}. \end{aligned} \tag{14}$$

Multiplying it by 2×1 vector from (13), we obtain the following estimates of the parameters:

$$\begin{aligned} T^{-1/2}\tilde{\mu} &= -\frac{2}{\bar{c}}W_c(1) - \frac{6}{\bar{c}^2}W_c(1) + \frac{12}{\bar{c}^2}\int_0^1 W_c(s)ds + \frac{12}{\bar{c}}\int_0^1 W_c(s)ds \\ &\quad -\frac{12}{\bar{c}}\int_0^1 sW_c(s)ds + 4\int_0^1 W_c(s)ds - 6\int_0^1 sW_c(s)ds \\ &\quad +\frac{\kappa}{\bar{c}^2}\lambda_0(1-\lambda_0)[(\bar{c}^2 + 2\bar{c})\lambda_0 - \bar{c}^2 - 4\bar{c} - 6], \end{aligned} \tag{15}$$

$$T^{1/2}\tilde{\beta} = \frac{12}{\bar{c}}\left[\frac{1}{2}W(1) - \int_0^1 W_c(s)ds + \bar{c}\left\{\int_0^1 sW_c(s)ds - \frac{1}{2}\int_0^1 W_c(s)ds\right\}\right] \tag{16}$$

$$-\frac{\kappa}{\bar{c}}(1-\lambda_0)(2\bar{c}\lambda_0^2 - \lambda_0(\bar{c}+6) - \bar{c}) \tag{17}$$

Thus, residuals $\tilde{u}_t$ is written as

$$\tilde{u}_t = (y_t - \bar{\rho}y_{t-1}) - \tilde{\mu}(1-\bar{\rho}) - \tilde{\beta}\{t - \bar{\rho}(t-1)\} \tag{18}$$

Then

$$\begin{aligned} T^{-1/2}\sum_{i=2}^{t}\tilde{u}_i &= T^{-1/2}\sum_{i=2}^{t}(u_i-\bar{\rho}u_{i-1})+\kappa\left[T^{-1}\sum_{i=2}^{t}\mathbb{I}(r>\lambda_0)+\bar{c}T^{-2}\sum_{i=2}^{t}DT_{i-1}\right] \\ & -\bar{c}tT^{-1}T^{-1/2}\tilde{\mu}-T^{-1/2}\tilde{\beta}\sum_{i=2}^{t}(\bar{c}i/T+1-\bar{c}/T). \end{aligned} \quad (19)$$

As $T^{-1/2}\sum_{i=2}^{\lfloor rT\rfloor}(u_i-\bar{\rho}u_{i-1})\Rightarrow W_c(r)+\bar{c}\int_0^r W_c(r)dr$ (see HLM), then

$$\begin{aligned} T^{-1/2}\sum_{i=2}^{\lfloor rT\rfloor}\tilde{u}_i &\Rightarrow W_c(r)+\bar{c}\int_0^r W_c(r)dr-\bar{c}r\left\{T^{-1/2}\tilde{\mu}\right\}-\left(\bar{c}\frac{r^2}{2}+r\right)\left\{T^{1/2}\tilde{\beta}\right\} \\ & +\kappa\left[(r-\lambda_0)+\frac{\bar{c}}{2}(r-\lambda_0)^2\right]\mathbb{I}(r>\lambda_0) \\ &= W_c(r)+\bar{c}\int_0^r W_c(r)dr-r\left(W_c(1)+\bar{c}\int_0^1 W_c(s)ds\right) \\ & +\kappa\left[\left((r-\lambda_0)+\frac{\bar{c}}{2}(r-\lambda_0)^2\right)\mathbb{I}(r>\lambda_0)-r\left((1-\lambda_0)+\frac{\bar{c}}{2}(1-\lambda_0)^2\right)\right] \\ & -6r^2\left[\frac{1}{2}W_c(1)-\int_0^1 W_c(s)ds+\bar{c}\int_0^1 sW_c(s)ds-\frac{1}{2}\bar{c}\int_0^1 W_c(s)ds\right] \\ & +6r\left[\frac{1}{2}W_c(1)-\int_0^1 W_c(s)ds+\bar{c}\int_0^1 sW_c(s)ds-\frac{1}{2}\bar{c}\int_0^1 W_c(s)ds\right] \\ & +\kappa\left[6r(1-r)(1-\lambda_0)\left(\frac{\bar{c}}{12}+\frac{\lambda_0}{2}+\frac{c\lambda_0}{12}-\frac{\bar{c}\lambda_0}{6}\right)\right] \\ &= H_{c,\bar{c}}(r,\lambda_0,\kappa)-6r(1-r)\int_0^1 H_{c,\bar{c}}(s,\lambda_0,\kappa)ds. \end{aligned}$$

The last equality follows from the fact that

$$\begin{aligned} \int_0^1 H_{c,\bar{c}}(s)ds &= \int_0^1 W_c(s)ds+\bar{c}\int_0^1\int_0^s W(l)dl-\frac{1}{2}W_c(1)-\frac{\bar{c}}{2}\int_0^1 W_c(s)ds \\ &= -\left[\frac{1}{2}W_c(1)-\int_0^1 W_c(s)ds+\bar{c}\int_0^1 sW_c(s)ds-\frac{1}{2}\bar{c}\int_0^1 {}_c(s)ds\right], \end{aligned}$$

because $\int_0^1 sW(s)ds=\int_0^1 W(s)ds-\int_0^1\left(\int_0^s W(l)dl\right)dr$ (see Tanaka (1996, Ch. 2)), and

$$\begin{aligned} &\int_0^1\left[\left((r-\lambda_0)+\frac{\bar{c}}{2}(r-\lambda_0)^2\right)\mathbb{I}(r>\lambda_0)-r\left((1-\lambda_0)+\frac{\bar{c}}{2}(1-\lambda_0)^2\right)\right]dr \\ &\qquad = -(1-\lambda_0)\left(\frac{\bar{c}}{12}+\frac{\lambda_0}{2}+\frac{c\lambda_0}{12}-\frac{\bar{c}\lambda_0}{6}\right). \end{aligned} \quad (20)$$

Proof of Theorem 2.

Consider the estimates $\tilde{\mu}$, $\tilde{\beta}$ и $\tilde{\gamma}$:

$$\begin{bmatrix}\tilde{\mu}\\ \tilde{\beta}\\ \tilde{\gamma}\end{bmatrix}=\begin{bmatrix}g_{11}&g_{12}&g_{13}\\ g_{12}&g_{22}&g_{23}\\ g_{13}&g_{23}&g_{33}\end{bmatrix}^{-1}\begin{bmatrix}h_1\\ h_2\\ h_3\end{bmatrix}, \quad (21)$$

where

$$
\begin{aligned}
g_{13} &= (1-\bar{\rho}) \sum_{t=\lfloor \lambda T \rfloor+1}^{T} \{t - \lfloor \lambda T \rfloor - \bar{\rho}(t - \lfloor \lambda T \rfloor - 1)\}, \\
g_{23} &= \sum_{t=\lfloor \lambda T \rfloor+1}^{T} \{t - \bar{\rho}(t-1)\}\{t - \lfloor \lambda T \rfloor - \bar{\rho}(t - \lfloor \lambda T \rfloor - 1)\}, \\
g_{33} &= \sum_{t=\lfloor \lambda T \rfloor+1}^{T} \{t - \lfloor \lambda T \rfloor - \bar{\rho}(t - \lfloor \lambda T \rfloor - 1)\}^2, \\
h_3 &= \sum_{t=\lfloor \lambda T \rfloor+1}^{T} (y_t - \bar{\rho} y_{t-1})\{t - \lfloor \lambda T \rfloor - \bar{\rho}(t - \lfloor \lambda T \rfloor - 1)\}.
\end{aligned}
$$

Similarly to the proof of Theorem 1 and HLT:

$$
\begin{aligned}
g_{13} &\to k_{\bar{c}_\lambda}(\lambda) \\
T^{-1} g_{23} &\to m_{\bar{c}_\lambda}(\lambda) \\
T^{-1} g_{33} &\to d_{\bar{c}_\lambda}(\lambda) \\
T^{-1/2} h_3 &\Rightarrow b_{c,\bar{c}_\lambda}(\lambda) + \kappa f_{c,\bar{c}_\lambda}(\lambda_0, \lambda)
\end{aligned}
$$

Then

$$
\begin{aligned}
\begin{bmatrix} T^{-1/2}\tilde{\mu} \\ T^{1/2}\tilde{\beta} \\ T^{1/2}\tilde{\gamma} \end{bmatrix}
&= \begin{bmatrix} T g_{11} & g_{12} & g_{13} \\ g_{12} & T^{-1} g_{22} & T^{-1} g_{23} \\ g_{13} & T^{-1} g_{23} & T^{-1} g_{33} \end{bmatrix}^{-1}
\begin{bmatrix} T^{1/2} h_1 \\ T^{-1/2} h_2 \\ T^{-1/2} h_3 \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} \bar{c}_\lambda^2 & \bar{c}_\lambda + \bar{c}_\lambda^2/2 & k_{\bar{c}_\lambda}(\lambda) \\ \bar{c}_\lambda + \bar{c}_\lambda^2/2 & 1 + \bar{c}_\lambda + \bar{c}_\lambda^2/3 & m_{\bar{c}_\lambda}(\lambda) \\ k_{\bar{c}_\lambda}(\lambda) & m_{\bar{c}_\lambda}(\lambda) & d_{\bar{c}_\lambda}(\lambda) \end{bmatrix}^{-1}
\begin{bmatrix} a_{c,\bar{c}_\lambda} + \kappa q_{c,\bar{c}_\lambda}(\lambda_0) \\ b_{c,\bar{c}_\lambda} + \kappa f_{c,\bar{c}_\lambda}(\lambda_0) \\ b_{c,\bar{c}_\lambda}(\lambda) + \kappa f_{c,\bar{c}_\lambda}(\lambda_0, \lambda) \end{bmatrix}
\end{aligned}
$$

Here again, the estimate $\tilde{\mu}$ is not asymptotically negligible, in contrast to HLT.

Thus, the residuals $\tilde{u}_t$ is written as

$$
\begin{aligned}
\tilde{u}_t &= (y_t - \bar{\rho} y_{t-1}) - \tilde{\mu}(1-\bar{\rho}) - \tilde{\beta}\{t - \bar{\rho}(t-1)\} \\
&\quad - \tilde{\gamma}\{t - \lfloor \lambda T \rfloor - \bar{\rho}(t - \lfloor \lambda T \rfloor - 1)\}\mathbb{I}(t > T_1).
\end{aligned}
$$

Then

$$
\begin{aligned}
T^{-1/2}\sum_{i=2}^{t}\tilde{u}_i &= T^{-1/2}\sum_{i=2}^{t}(u_i-\bar{\rho}u_{i-1})+\kappa\left[T^{-1}\sum_{i=2}^{t}\mathbb{I}(r>\lambda_0)+\bar{c}_\lambda T^{-2}\sum_{i=2}^{t}DT_{i-1}\right] \\
&\quad -\bar{c}_\lambda tT^{-3/2}\tilde{\mu}-T^{-1/2}\tilde{\beta}\sum_{i=2}^{t}(\bar{c}_\lambda i/T+1-\bar{c}_\lambda/T) \\
&\quad -T^{-1/2}\tilde{\gamma}\sum_{i=\lfloor\lambda T\rfloor+1}^{t}(\bar{c}_\lambda i/T+1-\bar{c}_\lambda/T-\bar{c}_\lambda T_1/T) \\
&\Rightarrow W_c(r)+\bar{c}_\lambda\int_0^r W_c(r)dr+\kappa\left[(r-\lambda_0)+\frac{\bar{c}_\lambda}{2}(r-\lambda_0)^2\right]\mathbb{I}(r>\lambda_0) \\
&\quad -\bar{c}_\lambda r\left\{T^{-1/2}\tilde{\mu}\right\}-\left(\bar{c}_\lambda\frac{r^2}{2}+r\right)\left\{T^{1/2}\tilde{\beta}\right\} \\
&\quad -\left(\bar{c}_\lambda\frac{r^2}{2}+r+\bar{c}_\lambda\frac{\lambda^2}{2}-\lambda-\bar{c}_\lambda\lambda r\right)\mathbb{I}(r>\lambda_0)\{T^{1/2}\tilde{\gamma}\}
\end{aligned}
$$

References

Busetti, F. and Harvey, A.C. (2001). Testing for the Presence of a Random Walk in Series with Structural Breaks. *Journal of Time Series Analysis*, **22**, 127–150.

Busetti, F. and Harvey, A.C. (2003). Further Comments On Stationarity Tests In Series With Structural Breaks At Unknown Points. *Journal of Time Series Analysis*, **24**, 137–140.

Carrion-i-Silvestre, J.L., Kim, D., and Perron, P. (2009). GLS-Based Unit Root Tests with Multiple Structural Breaks Both under the Null and the Alternative Hypotheses. *Econometric Theory*, **25**, 1754–1792.

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

Harris, D., Harvey, D.I., Leybourne, S.J., and Taylor, A.M.R. (2009). Testing for a Unit Root in the Presence of a Possible Break in Trend. *Econometric Theory*, **25**, 1545–1588.

Harris, D., Leybourne, S., and McCabe, B. (2007). Modified KPSS Tests for Near Integration. *Econometric Theory*, **23**, 355–363.

Harvey, D.I., Leybourne, S.J., and Taylor, A.M.R. (2009). Simple, Robust and Powerful Tests of the Breaking Trend Hypothesis. *Econometric Theory*, **25**, 995–1029.

Harvey, D.I., Leybourne, S.J., and Taylor, A.M.R. (2012). Unit Root Testing under a Local Break in Trend. *Journal of Econometrics*, **167**, 140–167.

Harvey, D.I., Leybourne, S.J., and Taylor, A.M.R. (2013). Testing for Unit Roots in the Possible Presence of Multiple Trend Breaks using Minimum Dickey-Fuller Statistics. *Journal of Econometrics*, forthcoming.

Harvey, D.I. and Mills, T.C. (2003). A Note on Busetti-Harvey Tests for Stationarity in Series with Structural Breaks. *Journal of Time Series Analysis*, **24**, 159–164.

Kurozumi, E. (2002). Testing for Stationarity with a Break. *Journal of Econometrics*, **108**, 63–99.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Econometrics*, **54**, 159Ц178.

Müller, U. (2005). Size and Power of Tests for Stationarity in Highly Autocorrelated Time Series. *Journal of Econometrics*, **128**, 195–213.

Newey, W.K. and West, K.D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, **61**, 631–653.

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

Perron, P. (1989). The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Econometrica*, **57**, 1361–1401.

Perron, P. and Qu, Z. (2007). A Simple Modification to Improve the Finite Sample Properties of Ng and PerronŠs Unit Root Tests. *Economics Letters*, **94**, 12–19.

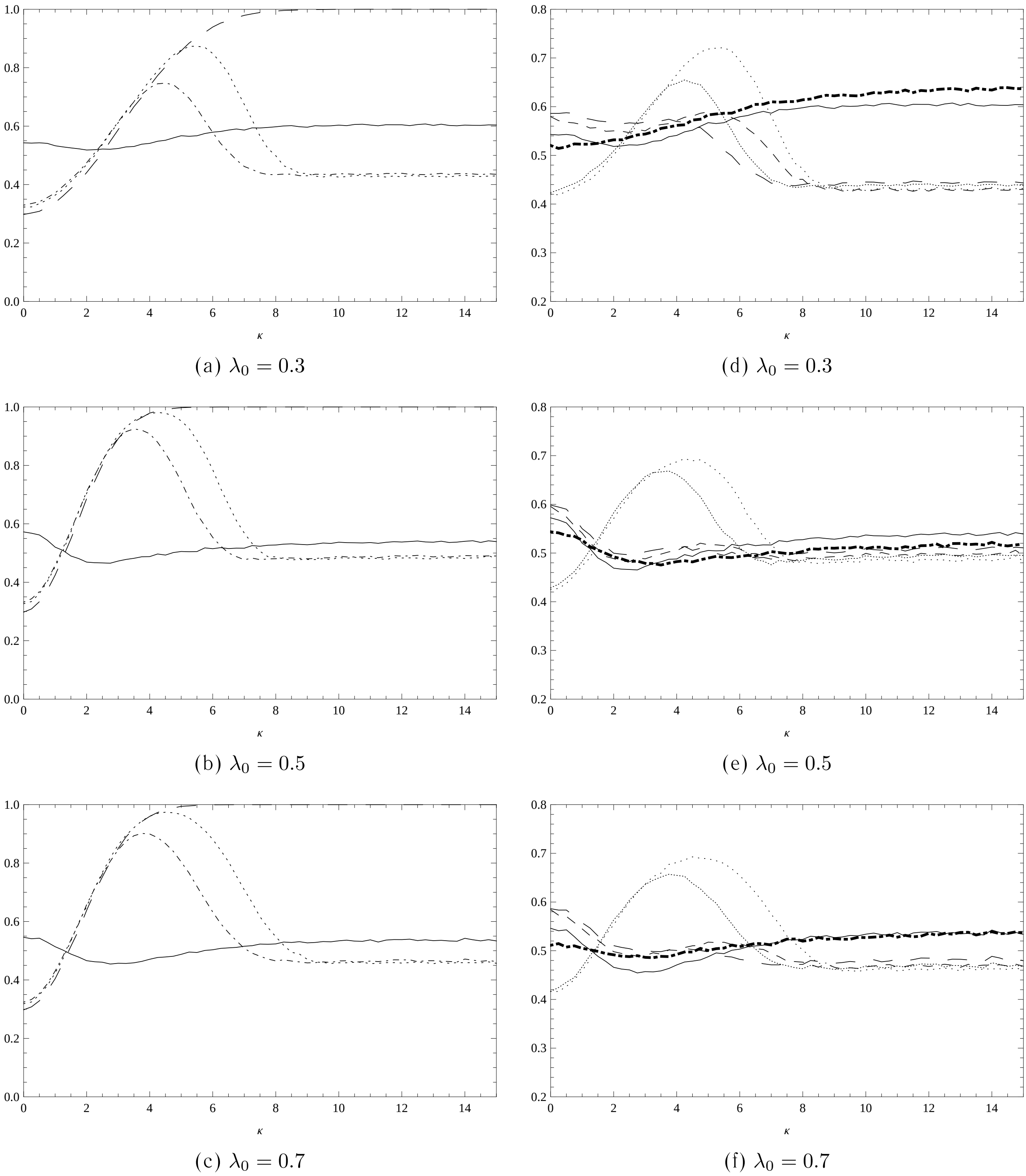
Perron, P. and Yabu, T. (2009). Testing for Shifts in Trend with an Integrated or Stationary Noise Component. *Journal of Business and Economic Statistics*, **27**, 369–396.

Phillips, P.C.B. and Solo, V. (1992). Asymptotics for Linear Processes. *Annals of Statistics*, **20**, 971–1001.

Skrobotov, A. (2013). Trend and Initial Value in Stationarity Tests: The Asymptotic Analysis. Unpublished Manuscript.

Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. Wiley, New York.

Zivot, E. and Andrews, D.W.K. (1992). Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Journal of Business and Economic Statistics*, **10**, 251–270.



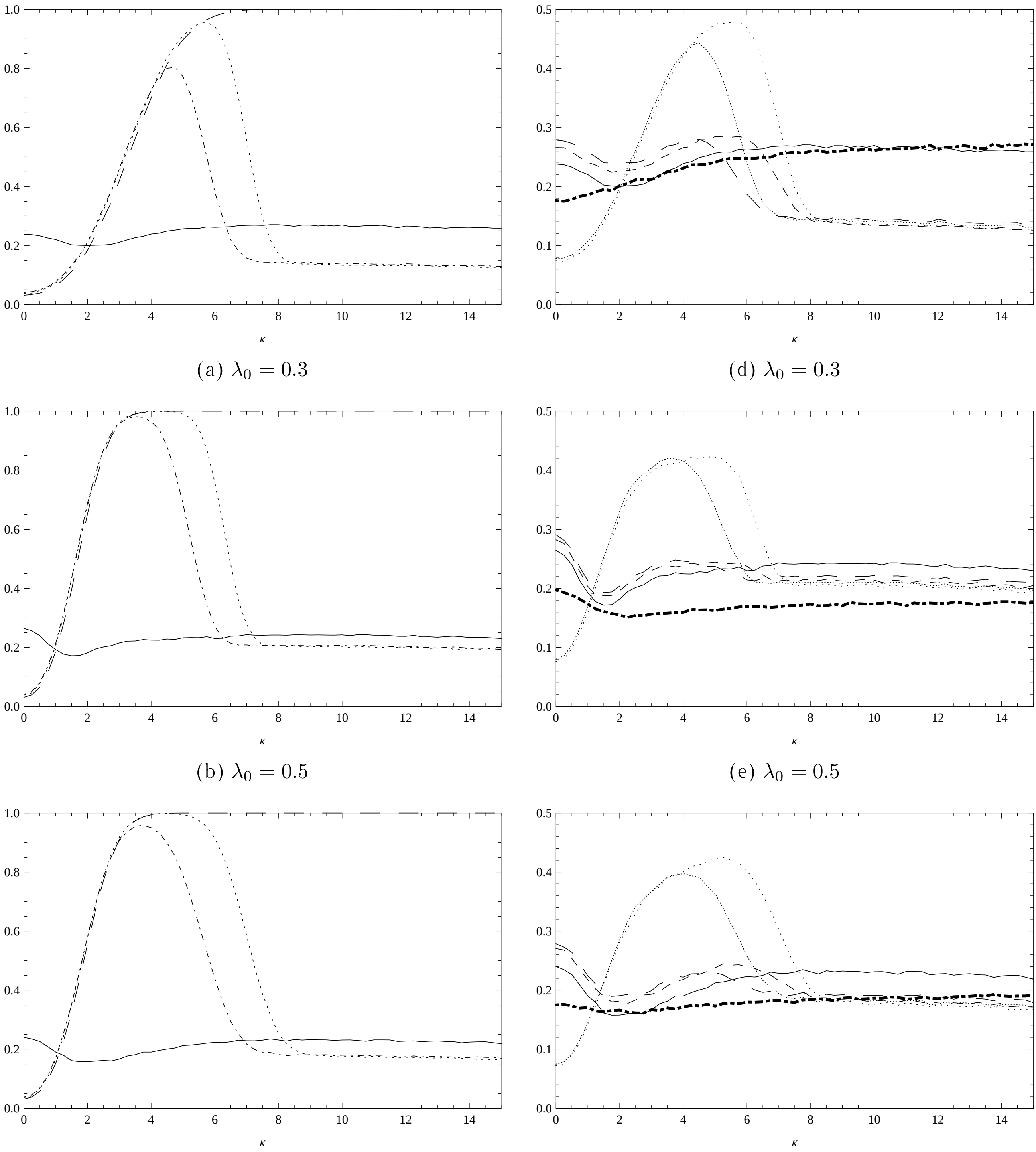
(a) $\lambda_0 = 0.3$ (d) $\lambda_0 = 0.3$

(b) $\lambda_0 = 0.5$ (e) $\lambda_0 = 0.5$

(c) $\lambda_0 = 0.7$ (f) $\lambda_0 = 0.7$

Figure 1: Asymptotic power-adjusted size: $c = 10$

$$S^t : — —, S^{tb} : ——, S(t_{PY}) : \cdot\ \cdot\ \cdot, S(t_\lambda) : — \cdot —$$
$$S^{tb} : ——, AS(t_\lambda) : — —, AS(t_{PY}) : - -, IS(t_\lambda) : \cdots, IS(t_{PY}) : \cdot\ \cdot\ \cdot, MS : \mathbf{— \cdot —}$$



(a) $\lambda_0 = 0.3$ (d) $\lambda_0 = 0.3$

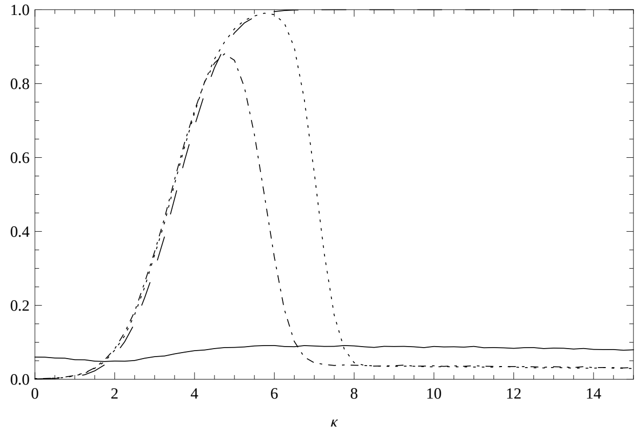
(b) $\lambda_0 = 0.5$ (e) $\lambda_0 = 0.5$

(c) $\lambda_0 = 0.7$ (f) $\lambda_0 = 0.7$

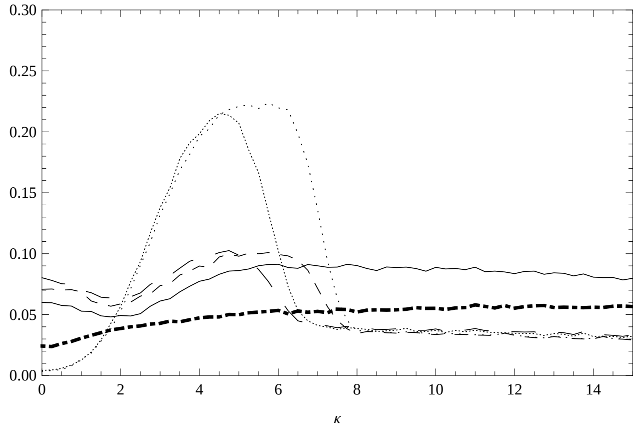
Figure 2: Asymptotic power-adjusted size: $c = 20$

S^t : — —, S^{tb} : ——, $S(t_{PY})$: · · · ·, $S(t_\lambda)$: — · —

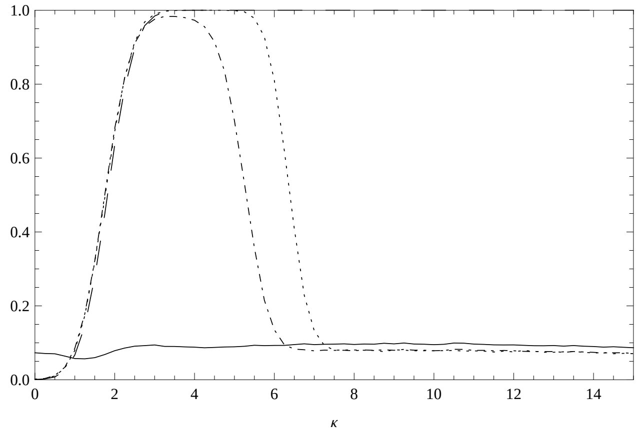
S^{tb} : ——, $AS(t_\lambda)$: — —, $AS(t_{PY})$: - -, $IS(t_\lambda)$: · · ·, $IS(t_{PY})$: · · · ·, MS : **— · —**



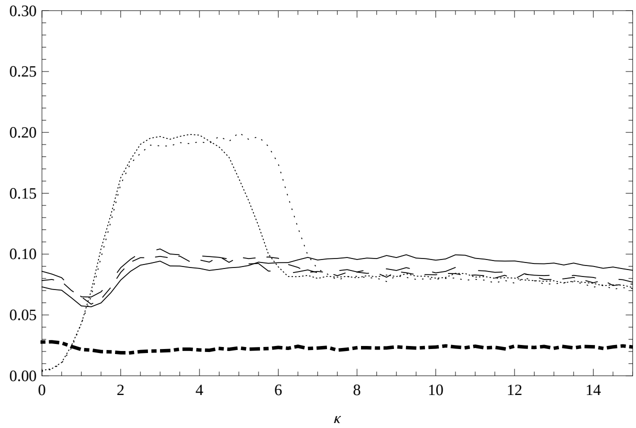
(a) $\lambda_0 = 0.3$



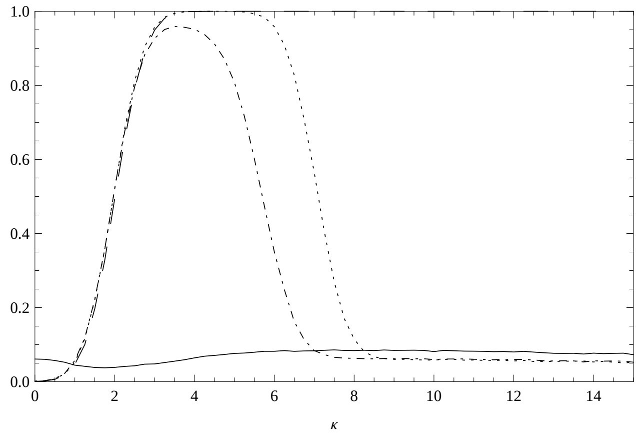
(d) $\lambda_0 = 0.3$



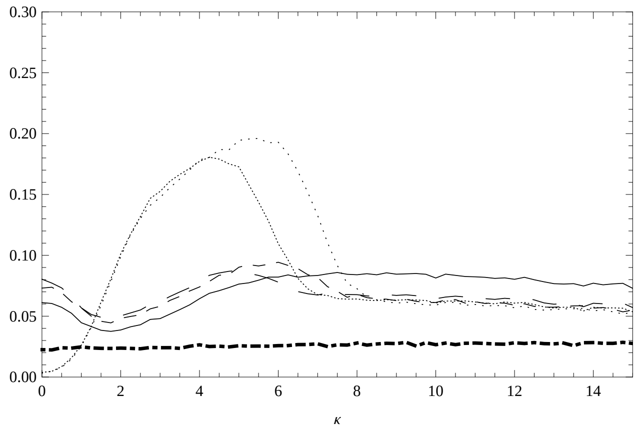
(b) $\lambda_0 = 0.5$



(e) $\lambda_0 = 0.5$



(c) $\lambda_0 = 0.7$

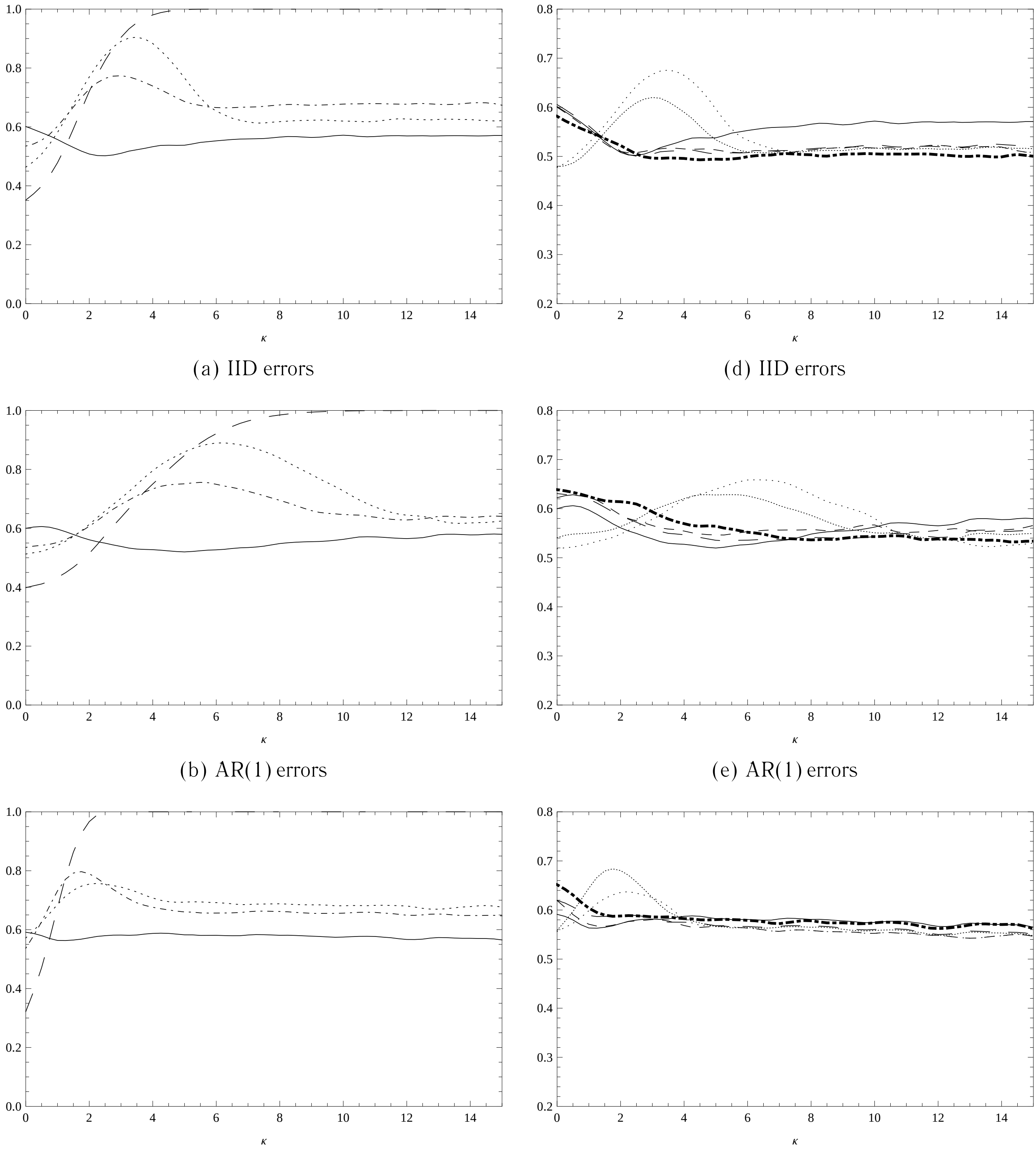


(f) $\lambda_0 = 0.7$

Figure 3: Asymptotic power-adjusted size: $c = 30$

S^t : — — , S^{tb} : —— , $S(t_{PY})$: · · · , $S(t_\lambda)$: — · —

S^{tb} : —— , $AS(t_\lambda)$: — — , $AS(t_{PY})$: - - , $IS(t_\lambda)$: ··· , $IS(t_{PY})$: · · · , MS : **— · —**



(a) IID errors

(d) IID errors

(b) AR(1) errors

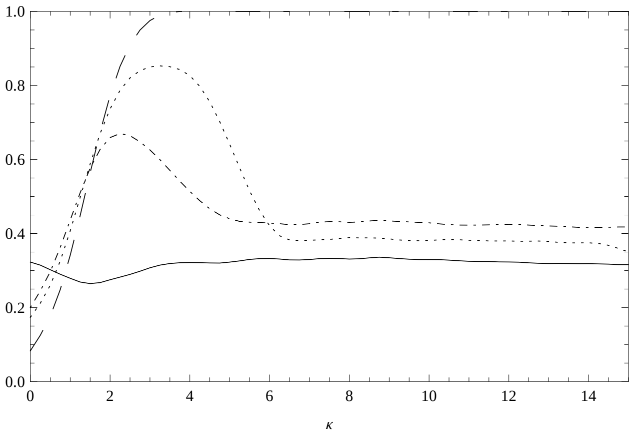
(e) AR(1) errors

(c) MA(1) errors

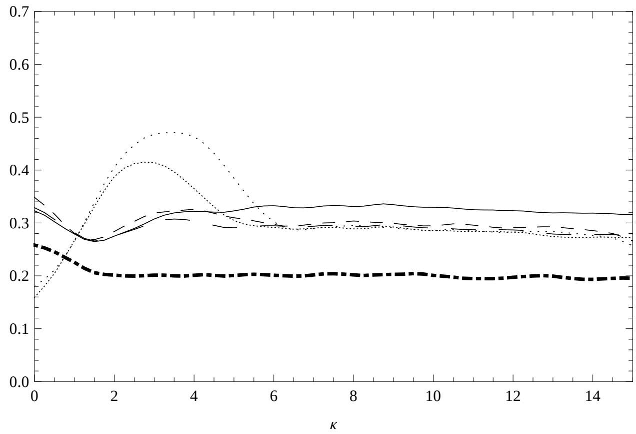
(f) MA(1) errors

Figure 4: Finite sample power-adjusted size, $T = 150$: $c = 10$

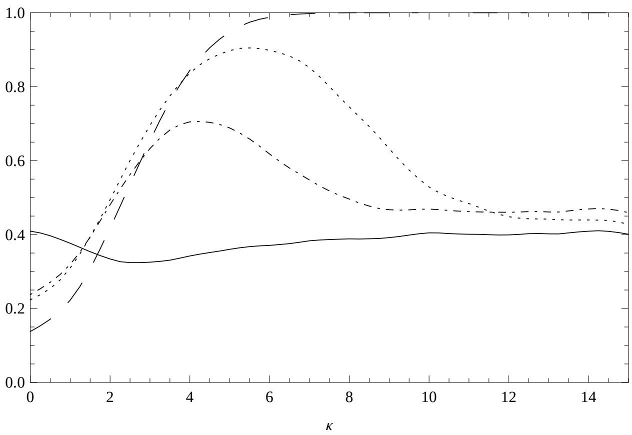
$$S^t : — — , S^{tb} : —— , S(t_{PY}) : \cdot\ \cdot\ \cdot , S(t_\lambda) : — \cdot —$$
$$S^{tb} : —— , AS(t_\lambda) : — — , AS(t_{PY}) : - - , IS(t_\lambda) : \cdots , IS(t_{PY}) : \cdot\ \cdot\ \cdot , MS : \mathbf{— \cdot —}$$



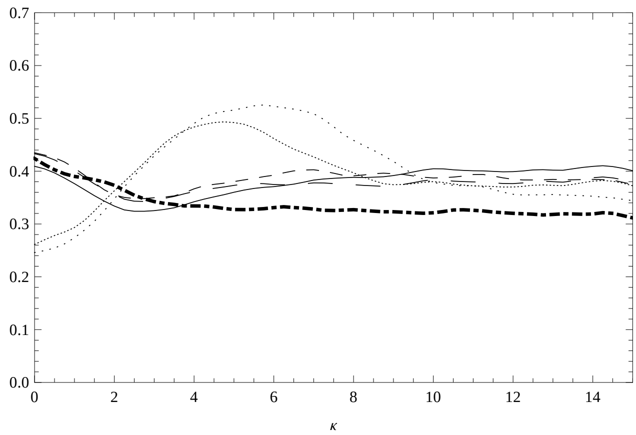
(a) IID errors



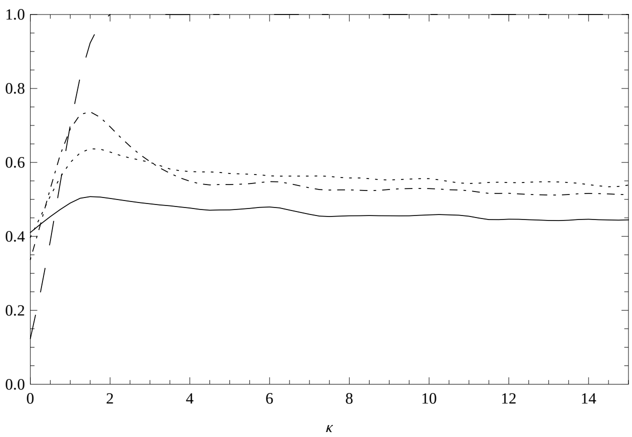
(d) IID errors



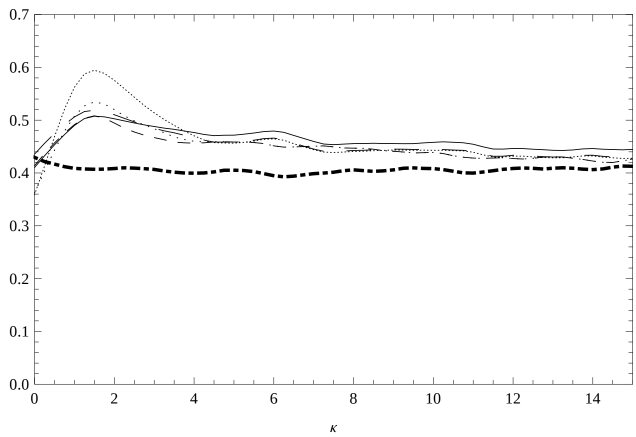
(b) AR(1) errors



(e) AR(1) errors



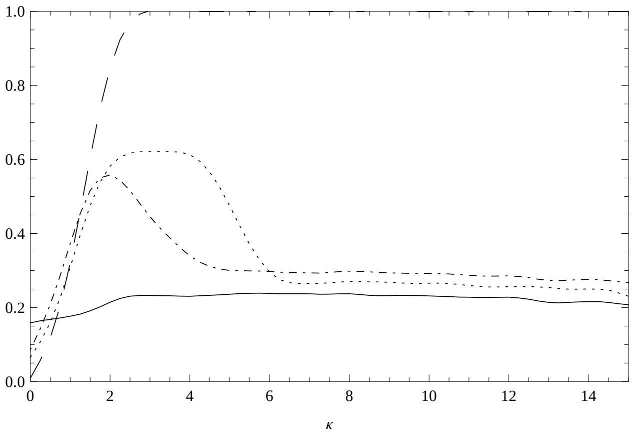
(c) MA(1) errors



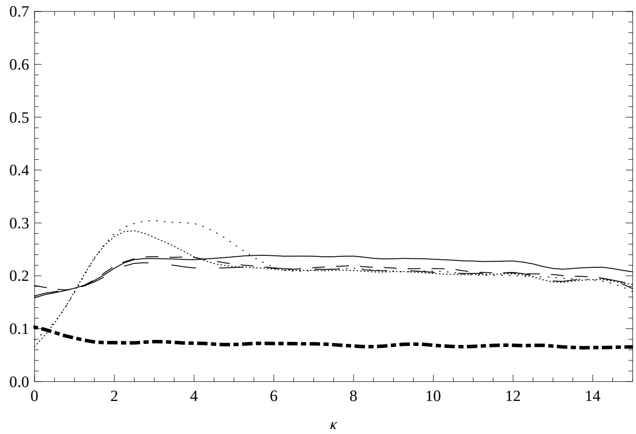
(f) MA(1) errors

Figure 5: Finite sample power-adjusted size, $T = 150$: $c = 20$

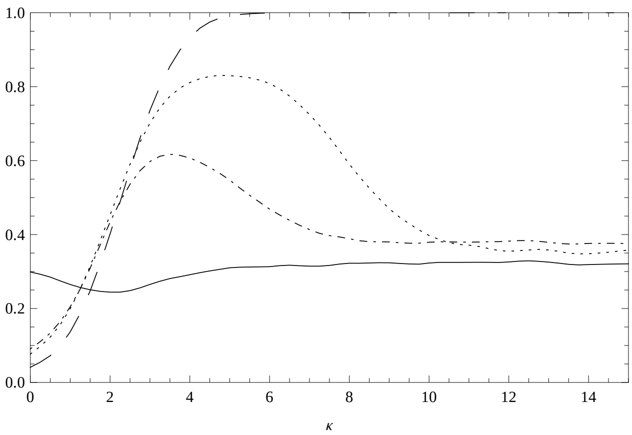
S^t : — — , S^{tb} : —— , $S(t_{PY})$: · · · , $S(t_\lambda)$: — · —
S^{tb} : —— , $AS(t_\lambda)$: — — , $AS(t_{PY})$: - - , $IS(t_\lambda)$: ··· , $IS(t_{PY})$: · · · , MS : **— · —**



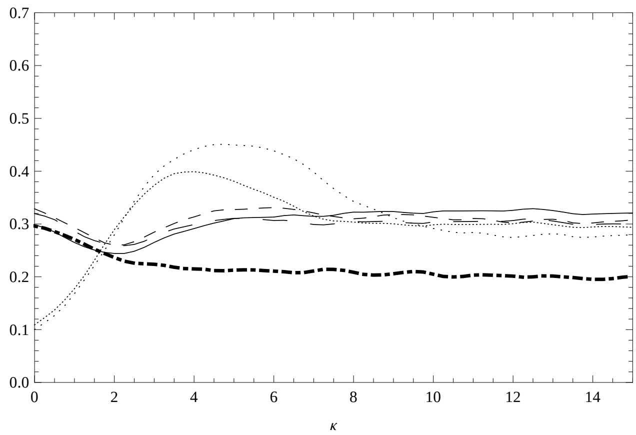
(a) IID errors



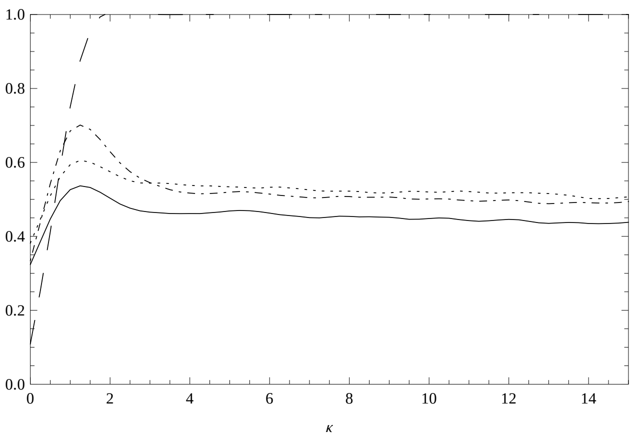
(d) IID errors



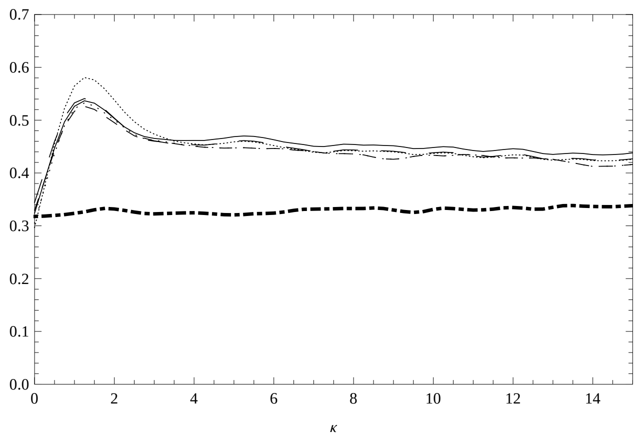
(b) AR(1) errors



(e) AR(1) errors



(c) MA(1) errors



(f) MA(1) errors

Figure 6: Finite sample power-adjusted size, $T = 150$: $c = 30$

S^t : — — , S^{tb} : —— , $S(t_{PY})$: · · · , $S(t_\lambda)$: — · —
S^{tb} : —— , $AS(t_\lambda)$: — — , $AS(t_{PY})$: - - , $IS(t_\lambda)$: ··· , $IS(t_{PY})$: · · · , MS : **— · —**