Unit Root Testing with Slowly Varying Trends*

Sven Otto**
University of Cologne

September 20, 2016

Abstract

We propose a unit root test for time series with a nonparametric trend component using a rolling window scheme with overlapping blocks. For both fixed-b and small-b blocklength asymptotics we derive limiting distributions that are robust under heteroskedasticity. By pre-whitening the series we account for autocorrelation. In a Monte Carlo simulation with slowly varying trend components both the fixed-b and small-b tests yield good size and higher power for near integrated alternatives than conventional unit root tests. An application to quarterly inflation rates of 22 countries provides evidence that inflation rates are stationary around a slowly varying trend.

Keywords: Unit root tests, nonlinear trends, heteroskedasticity, inflation

JEL Classification: C12, C14, C22, E31

*I would like to thank my supervisor Jörg Breitung as well as Matei Demetrescu for their extensive advise and support. My thanks also go to Hans Manner and Markus Kösler for many helpful comments and useful discussions.

**Corresponding address: Institute of Econometrics and Statistics, Universität zu Köln, Albertus-Magnus-Platz, D-50923 Köln, Germany. +49-221-470-6186. sven.otto@statistik.uni-koeln.de.

1 Introduction

In the time series literature there has been a long debate about whether economic variables such as GDP, inflation and interest rates are $I(1)$ or $I(0)$ around a deterministic trend. The standard unit root tests of Dickey and Fuller (1979) often fail to reject the null hypothesis for these time series. Trend stationary processes that are nearly integrated may be hardly distinguishable from unit root processes, especially when the parametric model for the deterministic trend component is not correctly specified. In conventional unit root testing the trend term is often treated to be known up to some parameter vector. However, misspecified trend models may lead to high power losses. Many authors recognized this fact and considered various generalizations of the simple trend models of Dickey and Fuller (1979). Perron (1989) as well as Rappoport and Reichlin (1989) show that under a broken trend model with one-time changes in mean or slope the unit root hypothesis could be rejected for many macroeconomic variables. Leybourne et al. (1998) and Kapetanios et al. (2003) allow for smooth transitions from one trend regime to another. Bierens (1997) uses Chebyshev polynomials to approximate a possibly nonlinear trend. His results indicate that the US GNP deflator, consumer price index and interest rate are stationary around a nonlinear trend. The unit root test of Enders and Lee (2012) is designed for unknown dates and forms of the changes in the trend and is based on Fourier series approximations. Furthermore, Gao and Robinson (2016) propose a test against a fractional integration alternative around an unknown, smooth and nonparametric trend component.

In this paper we propose a unit root test against the $I(0)$ alternative in the presence of an unknown nonparametric slowly varying trend. Our estimator for the autoregressive parameter ρ is based on the idea that in a small window slowly varying trends are approximately constant. We divide the series into overlapping blocks and then estimate ρ locally under a constant trend specification in each block. The blocks can be interpreted as units of a panel. Following the panel unit root test approach of Levin et al. (2002), we obtain a pooled estimator for ρ that is consistent under very general conditions on the unknown trend function. The unit root test statistic is then given by a pseudo t-statistic. Its limiting distribution under fixed-b asymptotics is a functional of Brownian motions, which is typical for unit root tests. Interestingly, under small-b asymptotics the limiting null-distribution is standard normal.

The paper is organized as follows. In Section 2 we analyze the asymptotic behavior of the pooled estimator in the presence of a general nonparametric trend model. Under both fixed-b and small-b asymptotics the limiting distributions for pseudo t-statistics are derived under the unit root hypothesis. Under heteroskedasticity nuisance parameters appear in the limiting distributions. The estimation of these parameters is discussed in Section 3 and we present heteroskedasticity robust test statistics. In Section 4 we apply a pre-whitening procedure in order to apply the tests in the presence of short run dynamics. Monte Carlo simulations are performed in Section 5. The tests are sized correctly in finite samples. In the presence of slowly varying trends the pooled tests tend to yield higher power than conventional unit root tests. In Section 6 we apply these tests to quarterly inflation rates of 22 countries and provide evidence that inflation rates are trend stationary around a slowly varying deterministic component. Finally, Section 7 concludes.

2 The model, estimation and testing

We are interested in inference concerning the autoregressive parameter ρ in the model

$$y_t = d_t + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \quad t = 1, \dots, T, \tag{1}$$

where ρ is close or equal to one. The deterministic term $d_t = d(t/T)$ is nonparametric and unknown and the initial value u_0 is random with $E[u_0^2] < \infty$. We establish asymptotic results under the following assumptions on the error term ϵ_t and the deterministic term d_t.

Assumption 1. *The process $\{\epsilon_t\}_{t\in\mathbb{N}}$ is independent with $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = \sigma^2$ and $E[\epsilon_t^4] < \infty$.*

Assumption 2. *The trend component $d_t = d(t/T)$ is Hölder continuous[1] of order $\alpha > 0.5$.*

The error terms in Assumption 1 are not necessarily identically distributed. Assumption 2 is a quite weak assumption on the deterministic component. It allows for twice boundedly differentiable trends on $[0, 1]$ as considered in Gao and Robinson (2016). The assumption also includes any Lipschitz continuous trend function, which is the case if $\alpha = 1$. Note that $\alpha > 1$ implies that the trend component is constant.[2] Additionally, some types of nondifferentiable functions that are ragged at some points are allowed.[3] Hölder coefficients can be interpreted as a measure for the

[1]A function d on $[0, 1]$ is Hölder continuous of order α, if there exists a constant $C < \infty$ such that $|d(r) - d(s)| \leq C|r - s|^\alpha$ for all $r, s \in [0, 1]$.

[2]If $\alpha > 1$ then for any $h > 0$ we have $|d(r) - d(r + h)|/h \leq Ch^{\alpha-1} \to 0$ as $h \to 0$ which implies that the derivative of d exists and is equal to zero on $[0, 1]$.

[3]Consider for instance the nondifferentiable trend $d_t = |t/T - 0.5|^\alpha$.

amount of variation at ragged nondifferentiable points of the trend component. The smaller the Hölder coefficient α, the more variation we allow at these ragged points.

Our basic idea for dealing with a nonparametric slowly varying trend is to approximate the unknown trend locally by a constant and apply a rolling window scheme. We divide the time series into $T-B$ overlapping blocks of length B and then blockwise estimate ρ via OLS under a constant trend specification. By subtracting the first observation in each block we correct for the constant.[4] Analogous to the LLC unit root test approach for panel data (see Levin et al. 2002) we sum up the numerators and denominators of the $T-B$ estimators for ρ separately and obtain the pooled estimator

$$\hat{\rho} = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j} - y_j)(y_{t+j-1} - y_j)}{\sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2}. \tag{2}$$

The asymptotic distribution of the estimator depends on the choice of the blocklength B relative to the sample size T. We consider both a fixed-b and a small-b approach. In the fixed-b approach we assume that the relative blocklength B/T converges to some positive value b. The small-b approach considers a relative blocklength that converges to zero. The right hand side of (2) is called fixed-b pooled estimator if $B/T \to b$ with $0 < b < 1$ and is denoted by $\hat{\rho}^{\text{fb}}$. If the relative blocklength satisfies $B/T \to 0$ as T increases, then (2) is called small-b pooled estimator and is denoted by $\hat{\rho}^{\text{sb}}$.

Unit root tests with the null hypothesis $\rho = 1$ can be developed analogously to the Dickey-Fuller τ-statistics.[5] We consider the pseudo t-statistics

$$\tau\text{-FB} = \frac{\hat{\rho}^{\text{fb}} - 1}{s_{\text{fb}}}, \quad \tau\text{-SB} = \sqrt{3/2}\, \frac{\hat{\rho}^{\text{sb}} - 1}{s_{\text{sb}}},$$

where the denominators are given by

$$s_{\text{fb}}^2 = \hat{\sigma}_{\text{b}}^2 \left(\frac{1}{T} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2 \right)^{-1}$$

[4] Here we follow the idea of Schmidt and Phillips (1992). Another alternative to approximate the trend locally is to subtract the mean or higher order polynomials in each block. This leads to different asymptotic behavior but it does not improve the finite sample power of the corresponding unit root tests. Therefore we omit these estimation variants.

[5] The corresponding Dickey-Fuller ρ-statistics perform slightly worse than the τ-statistics in finite sample simulations and are therefore omitted.

and

$$s_{\mathrm{sb}}^2 = \hat{\sigma}_{\mathrm{b}}^2 \left(\frac{1}{B} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2 \right)^{-1}.$$

Note that in τ-SB we multiply the pseudo t-statistic by $\sqrt{3/2}$. This factor arises from reorganizing the summands in the double sums of $\hat{\rho}^{\mathrm{sb}}$ (see Appendix A.4 for more details).

The error variance is also estimated by a rolling window scheme. We consider

$$\hat{\sigma}_{\mathrm{b}}^2 = \frac{1}{T-B} \sum_{j=1}^{T-B} \frac{1}{B-1} \sum_{t=1}^{B} (\hat{y}_{j+t} - \bar{\hat{y}}_j)^2,$$

with $\hat{y}_t = y_t - \hat{\rho} y_{t-1}$ for $t = 2, \ldots, T$ and $\bar{\hat{y}}_j = B^{-1} \sum_{t=1}^{B} \hat{y}_{j+t}$. Here $\hat{\rho}$ is either the fixed-b or the small-b pooled estimator. The variance estimator is consistent for σ^2.

Lemma 1. *Let Assumption 1 and 2 hold and let $|\rho| \leq 1$. Then*

$$\hat{\sigma}_b^2 \xrightarrow{p} \sigma^2$$

as $T \to \infty$.

The asymptotic distributions of the fixed-b pooled estimator and the corresponding pseudo t-statistic are functionals of standard Brownian motions on the unit interval.

Theorem 1. *Let Assumption 1 and 2 hold and let $\rho = 1$. Then as $T \to \infty$*

(i)

$$T(\hat{\rho}^{fb} - 1) \xrightarrow{d} \frac{\int_0^{1-b} (W(b+r) - W(r))^2 \, dr - b(1-b)}{2 \int_0^{1-b} \int_r^{b+r} (W(s) - W(r))^2 ds dr},$$

(ii)

$$\tau\text{-}FB \xrightarrow{d} \frac{\int_0^{1-b} (W(b+r) - W(r))^2 \, dr - b(1-b)}{2 \sqrt{\int_0^{1-b} \int_r^{b+r} (W(s) - W(r))^2 ds dr}},$$

where W is a standard Brownian motion on the unit interval.

In order to obtain critical values for the fixed-b statistic τ-FB, we simulate its asymptotic null distribution. Standard Brownian motions are generated by the cumulated sum of $T = 5000$ independent random numbers from a normal distribution with mean zero and variance T^{-1}. The critical values are then obtained from the empirical 0.01, 0.05 and 0.1-quantiles of the functional obtained by 1,000,000 Monte Carlo repetitions. The results are presented in Table 1.

Table 1: Critical values for the τ-FB statistic

	significance level		
b	0.01	0.05	0.1
0.1	-0.5987	-0.4412	-0.3547
0.2	-0.8183	-0.6145	-0.5022
0.3	-0.9569	-0.7270	-0.6017
0.4	-1.0372	-0.7961	-0.6662
0.5	-1.0695	-0.8278	-0.6970
0.6	-1.0555	-0.8219	-0.6956
0.7	-1.0013	-0.7799	-0.6628
0.8	-0.8933	-0.6967	-0.5920
0.9	-0.6948	-0.5409	-0.4597

Theorem 2. *Let Assumption 1 and 2 hold and let* $\rho = 1$. *Then*

(i)

$$\sqrt{TB}(\hat{\rho}^{sb} - 1) \xrightarrow{d} \mathcal{N}\left(0, \frac{4}{3}\right),$$

(ii)

$$\tau\text{-}SB \xrightarrow{d} \mathcal{N}(0, 1).$$

as $T \to \infty$.

The small-b pooled estimator and the τ-SB statistic are asymptotically normally distributed. After reorganizing the summands in $\hat{\rho}^{\mathrm{sb}}$ we obtain elements of a martingale difference array which yields a Central Limit Theorem. This is quite remarkable since many standard unit root tests have nonstandard limiting distributions.

Remark 1. The asymptotic distributions under the unit root hypothesis for the small-b pooled estimator and the τ-SB statistic are still valid under a slightly more general setting. We also allow for Hölder coefficients that satisfy $\alpha > (3 - \sqrt{5})/4$, where $(3 - \sqrt{5})/4 \approx 0.191$. This yields a broader class of trend functions with higher variation in ragged nondifferentiable points. For instance the sample path of a Brownian motion is almost surely Hölder continuous for $\alpha < 0.5$. The blocklength needs to be restricted for these highly varying trends. Setting $B = \mathcal{O}(T^{\gamma})$ such that $(3 - \sqrt{5})/2 < \gamma < \min(2\alpha, 1)$ yields the same distributions as in Theorem 2 for $(3 - \sqrt{5})/4 < \alpha \leq 0.5$. For more details see Appendix A.5.

3 Testing under heteroskedasticity

While stationary time-varying conditional variances such as ARCH processes do not affect unit root testing, permanent changes in volatility alter the limiting distributions of unit root tests dramatically (see e.g. Hamori and Tokihisa 1997). Cavaliere (2005) shows that nonstationary volatility induces a time-transformation in the Brownian process that appears in the limiting distributions of many unit root test statistics. By means of transforming the data by the inverse variance profile (see Cavaliere and Taylor 2007), standard critical values can be retained. The rolling window scheme of the small-b estimator avoids this problem in a natural fashion since the error variance is estimated locally in a window with a relative blocklength that diminishes asymptotically. We consider the same model (1) but now with heteroskedastic errors.

Assumption 3. *The process* $\{\epsilon_t\}_{t\in\mathbb{N}}$ *is independent with* $E[\epsilon_t] = 0$*,* $E[\epsilon_t^2] = \sigma_t^2$ *and* $E[\epsilon_t^4] < \infty$*, where* $\sigma_t = \sigma(t/T)$ *with* $\sigma(\cdot) \in D[0,1]$ *is non-stochastic and strictly positive.*[6]

The asymptotic null-distribution of the small-b test statistic is discussed in the following theorem.

Theorem 3. *Let Assumption 2 and 3 hold and let* $\rho = 1$*. Then as* $T \to \infty$

$$\tau\text{-}SB \xrightarrow{d} \mathcal{N}\left(0, \sigma_*^2\right),$$

where $\sigma_*^2 = \int_0^1 \sigma(r)^4 dr / (\int_0^1 \sigma(r)^2 dr)^2$.

The distribution depends on the parameter σ_*^2, which is equal to 1 in the case of homoskedasticity. We estimate this parameter by its rolling window sample counterpart

$$\hat{\sigma}_*^2 = (T-B)(B-1) \cdot \frac{\sum_{j=1}^{T-B} \sum_{t=1}^{B} (\hat{y}_j - \bar{y})^2 (\hat{y}_{j+t} - \bar{\hat{y}}_j)^2}{\left(\sum_{j=1}^{T-B} \sum_{t=1}^{B} (\hat{y}_{j+t} - \bar{\hat{y}}_j)^2\right)^2},$$

where $\hat{y}_t = y_t - \hat{\rho}^{\text{sb}} y_{t-1}$ for $t = 2, \ldots, T$, $\bar{\hat{y}}_j = B^{-1} \sum_{t=1}^{B} \hat{y}_{j+t}$, and $\bar{y} = T^{-1} \sum_{t=1}^{T} y_t$. The following Lemma shows the consistency of $\hat{\sigma}_*^2$.

[6] $D[0,1]$ is the space of càdlàg functions together with the uniform metric.

Lemma 2. *Let Assumption 2 and 3 hold and let* $|\rho| \leq 1$*. Then*

$$\hat{\sigma}_*^2 \xrightarrow{p} \sigma_*^2$$

as $T \to \infty$*.*

The heteroskedasticity-robust test statistic is given by

$$\tau\text{-SB}^{HC} = \frac{\tau\text{-SB}}{\hat{\sigma}_*}.$$

Now we consider the fixed-b case. The limiting distribution under the unit root hypothesis depends on the variance profile $\eta(s) = (\int_0^1 \sigma(r)^2 dr)^{-2} \int_0^s \sigma(r)^2 dr$.

Theorem 4. *Let Assumption 2 and 3 hold and let* $\rho = 1$*. Then as* $T \to \infty$

$$\tau\text{-}FB \xrightarrow{d} \frac{\int_0^{1-b} \left(W_\eta(b+r) - W_\eta(r)\right)^2 dr - b(1-b)}{2\sqrt{\int_0^{1-b} \int_r^{b+r} (W_\eta(s) - W_\eta(r))^2 ds dr}},$$

where $W_\eta = (W \circ \eta)$ *is a time-transformed Brownian motion with variance profile* η*.*

Cavaliere and Taylor (2007) propose to transform the data with its inverse variance profile in order to deal with the presence of time-transformed Brownian motions. The idea is to find a continuous transformation $\tilde{y}_t$ of the process y_t such that the asymptotic null distribution of the τ-FB-statistic applied to the transformed data yields the same distribution as in Theorem 1 that we already have tabulated. We consider the time series

$$\tilde{y}_t := y_{\lfloor \hat{\eta}^{-1}(t/\tilde{T})\tilde{T} \rfloor}, \quad t = 1, \ldots, \tilde{T}. \tag{3}$$

Here $\hat{\eta}^{-1}(s)$ is the inverse of the estimator

$$\hat{\eta}(s) = \frac{\sum_{t=1}^{\lfloor sT \rfloor} \hat{e}_t^2 + (sT - \lfloor sT \rfloor)\hat{e}_{\lfloor sT \rfloor + 1}^2}{\sum_{t=1}^{T} \hat{e}_t^2},$$

where $\hat{e}_t$ are the residuals of a linear regression of y_t on y_{t-1} (see Cavaliere and Taylor 2007). In practice, we alter the observed time series in the way, that we insert copies of adjacent observations between the sample points in high volatility periods according to transformation (3). Note that we do not loose any observations since we can set $\tilde{T} > T$ arbitrarily high. Applying the τ-FB statistic to the transformed time series $\{\tilde{y}_t\}_{t=1}^{\tilde{T}}$ yields the heteroskedasticity robust test statistic τ-FB^{HC}.

In the following, we show that τ-FB^{HC} and τ-SB^{HC} have the same limiting distributions under heteroskedasticity as their homoskedastic counterparts in Theorems 1 and 2.

Theorem 5. *Let Assumption 2 and 3 hold and let $\rho = 1$. Then as $T \to \infty$*

(i)

$$\tau\text{-}FB^{HC} \xrightarrow{d} \frac{\int_0^{1-b} \left(W(b+r) - W(r)\right)^2 dr - b(1-b)}{2\sqrt{\int_0^{1-b}\int_r^{b+r}(W(s)-W(r))^2 ds dr}},$$

where W is a standard Brownian motion on the unit interval.

(ii)

$$\tau\text{-}SB^{HC} \xrightarrow{d} \mathcal{N}(0,1).$$

4 Testing under short run dynamics

The assumption that the error terms $\{\epsilon_t\}_{t\in\mathbb{N}}$ are uncorrelated may be too restrictive. Macroeconomic variables often involve short run dynamics where higher order autocovariances are not necessarily equal to zero. If the long run variance $\omega^2 = \gamma(0) + 2\sum_{j=1}^{\infty}\gamma(j)$ with the j-th order autocovariance $\gamma(j)$ does not coincide with the error variance σ^2, then the asymptotic distribution of the test statistics are shifted and standard critical values are not valid any more.

Assumption 4. *The process $\{\epsilon_t\}_{t\in\mathbb{N}}$ is covariance stationary with $E[\epsilon_t^4] < \infty$ and $T^{-0.5}S_T \Rightarrow \omega W$ with $\omega < \infty$, where $S_T(r) = \sum_{k=1}^{\lfloor rT \rfloor}\epsilon_t$, $0 \le r \le 1$.*[7]

If we impose for instance the α-mixing assumption given in Phillips (1987) then this assumption is satisfied. For a discussion on sufficient conditions for the FCLT in Assumption 4 see Davidson (2002). Under the null hypothesis and under Assumption 4 the asymptotic distribution of the Dickey-Fuller test statistic is given by

$$\frac{\sum_{t=2}^{T}\Delta y_t y_{t-1}}{\hat{\sigma}\sqrt{\sum_{t=2}^{T} y_{t-1}^2}} \xrightarrow{d} \frac{\int_0^1 W dW(r) + \frac{1-\kappa}{2}}{\sqrt{\int_0^1 W^2(r) dr}}$$

[7]Here "$\Rightarrow$" denotes weak convergence in the càdlàg space $D[0,1]$ together with the uniform metric.

(see Phillips 1987). This involves the additional term $(1-\kappa)/2$ where $\kappa = \gamma(0)/\omega^2$, which vanishes if all higher order autocovariances are zero. The asymptotic null distribution is shifted and the standard Dickey-Fuller critical values are not valid any more. For the fixed-b statistic we encounter a similar phenomenon.

Theorem 6. *Let Assumption 2 and 4 hold and let $\rho = 1$. Then*

$$T(\hat{\rho}^{fb} - 1) \xrightarrow{d} \frac{\frac{1}{2}\int_0^{1-b} \left(W(b+r) - W(r)\right)^2 dr - b(1-b)\frac{\kappa}{2}}{\int_0^{1-b}\int_r^{b+r}(W(s)-W(r))^2 ds dr}.$$

One solution to solve this problem is to estimate ω directly following Phillips and Perron (1988) using for instance the HAC-estimator of Andrews (1991). Due to estimation errors and resulting size distortions and power losses in small samples it is often not a good idea to estimate ω directly. We prefer a pre-whitening procedure. We fix some lag length p and estimate the model

$$\Delta y_t = \phi y_{t-1} + \alpha_1 \Delta y_{t-1} + \ldots + \alpha_p \Delta y_{t-p} + e_t, \tag{4}$$

where under the unit root hypothesis we have $\phi = 0$. Let $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ and $\widehat{\phi}$ be the OLS estimators of the parameters in model (4). Then the pre-whitened time series is given by

$$\widehat{Z}_t = y_t - \widehat{\alpha}_1 y_{t-1} - \ldots - \widehat{\alpha}_p y_{t-p}.$$

The lag order p of the regression model (4) can be chosen in a data-driven way using an information criterion or Schwert's rule of thumb, which is given by

$$p = \left\lfloor 12 \cdot \left(\frac{T}{100}\right)^{1/4} \right\rfloor$$

(see Schwert 1989).

5 Monte Carlo analysis

We evaluate the finite sample performance of the unit root tests by means of a Monte Carlo simulation. The autoregressive model (1) is simulated with 100,000 repetitions for $T = 100$ and $T = 300$ under both the null hypothesis $\rho = 1$ and the alternative hypotheses $\rho = 0.9$ and $\rho = 0.8$. We consider several trend functions which are presented in Table 2 and Figure 1. Trend function (zer) is the zero-trend and considered as a benchmark. The trend in (lin) is linear with a small positive slope. For logarithmic data with $T = 100$ this corresponds to a growth rate of 5% and for data with $T = 300$ to a growth rate of 1.7%. The three subsequent trends (fo1), (fo2) and (fo3) are different types of Fourier trends. Trend function (bre) represents a structural break and the logistic trends (lo1) and (lo2) are smooth transitions from one regime to another.

Table 2: Trend functions for the Monte Carlo simulations

(zer)	$d_t = 0$	(fo3)	$d_t = 4\sin(2\pi t/T) - 4\cos(\pi t/T)$
(lin)	$d_t = 10 + 5t/T$	(bre)	$d_t = 6 \cdot 1_{\{t \leq T/2\}}$
(fo1)	$d_t = -4\cos(2\pi t/T)$	(lo1)	$d_t = 6/(1 + \exp(0.1(t - 0.5T)))$
(fo2)	$d_t = 2\cos(2\pi t/T) + \sin(4\pi t/T)$	(lo2)	$d_t = 6/(1 + \exp(-0.1(t - 0.75T)))$

The innovations ϵ_t for the homoskedastic framework under Assumption 1 are simulated by $\mathcal{N}(0, 1)$ random variables independently for $t = 1, \ldots, T$. In order to capture potential starting value problems, we simulate some independent standard normal lagged innovations $\tilde{\epsilon}_k$ and set the initial value to $y_0 = d_0 + \sum_{k=1}^{100} \rho^{100-k} \tilde{\epsilon}_k$.

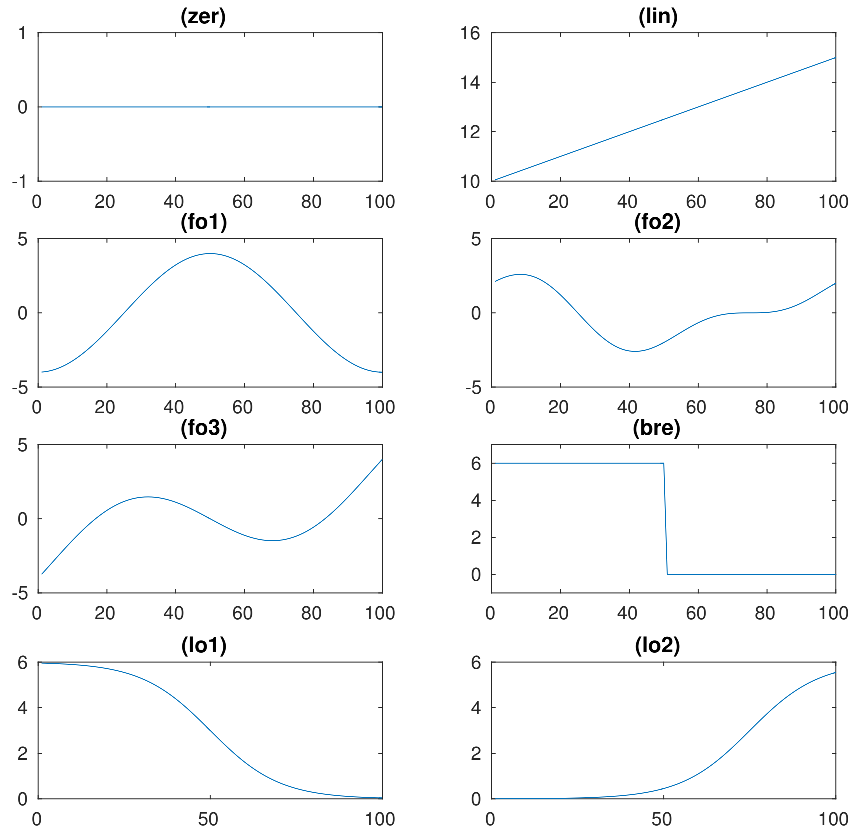
In Tables 3 and 4 we present the empirical size and power results for τ-FB, τ-FB^{HC}, τ-SB and τ-SB^{HC} using the asymptotic 5% critical values. The performance of the fixed-b and small-b statistics depends on the choice of the blocklength B. We have achieved good results with $B = T/10$ for fixed-b and $B = T^{0.6}$ for small-b. The blocklength B is a tuning parameter and the optimal choice depends on the underlying trend model. Tables A1 and A2 show some size and power results for different bandwidth choices for both the fixed-b and small-b statistics. Note that the empirical power of τ-SB does not exceed the empirical power of its heteroskedasticity robust counterpart in the case of homoskedasticity. Hence, we advise to use always the τ-SB^{HC} statistic in favour of τ-SB.

In order to emphasize the advantage of the pooled estimators and its unit root tests, we compare their small sample results to the performance of some conventional unit root tests. The tests by Dickey and Fuller (1979) with constant and linear trend specification are denoted DF$^{\mu}$ and DF* respectively. The statistic by Schmidt and Phillips (1992) with constant trend specification can be written as

$$\text{SP} = \frac{\sum_{t=2}^{T} \Delta y_t (y_{t-1} - y_1)}{\hat{\sigma}\sqrt{\sum_{t=2}^{T} (y_{t-1} - y_1)^2}}.$$

This test accounts for a constant trend specification by detrending with the first observation and is therefore a natural benchmark for the fixed-b and small-b statistics.

Figure 1: Plots of the trend functions in Table 2 with $T = 100$



Another approach that does not need a precise model for the trend component is the test by Enders and Lee (2012). A flexible Fourier form is used to approximate smooth breaks in the trend function. Structural changes can be captured by the low frequency components of a series. Enders and Lee (2012) consider the parametric trend model[8]

$$d_t = \alpha_0 + \gamma \cdot t + \alpha_1 \sin(2\pi t/T) + \beta_1 \cos(2\pi t/T).$$

[8]In general, more frequencies could be included but this leads to an over-fitting problem. Enders and Lee (2012) show that for more than one frequency the test looses power and that even a single low frequency can often approximate breaks reasonably well.

Table 3: Size and power for $T = 100$ using 5% critical values

d_t	ρ	τ-FB	τ-FB^{HC}	τ-SB	τ-SB^{HC}	DF$^{\mu}$	DF*	SP	EL
	1	0.042	0.022	0.052	0.058	0.057	0.051	0.050	0.054
(zer)	0.9	0.256	0.076	0.342	0.372	0.368	0.227	0.433	0.125
	0.8	0.683	0.239	0.815	0.837	0.902	0.717	0.767	0.402
	1	0.040	0.022	0.048	0.052	0.053	0.050	0.045	0.056
(lin)	0.9	0.219	0.130	0.281	0.300	0.173	0.212	0.163	0.124
	0.8	0.632	0.455	0.753	0.771	0.330	0.691	0.212	0.403
	1	0.014	0.007	0.015	0.017	0.029	0.013	0.014	0.057
(fo1)	0.9	0.031	0.008	0.023	0.028	0.015	0.004	0.015	0.124
	0.8	0.086	0.019	0.035	0.043	0.004	0.000	0.006	0.403
	1	0.026	0.013	0.030	0.033	0.036	0.027	0.033	0.040
(fo2)	0.9	0.120	0.038	0.139	0.155	0.117	0.049	0.183	0.080
	0.8	0.374	0.137	0.381	0.411	0.241	0.067	0.317	0.220
	1	0.020	0.010	0.024	0.026	0.054	0.033	0.020	0.056
(fo3)	0.9	0.087	0.025	0.108	0.123	0.150	0.073	0.052	0.125
	0.8	0.266	0.070	0.285	0.314	0.289	0.128	0.045	0.401
	1	0.040	0.011	0.049	0.062	0.040	0.039	0.050	0.051
(bre)	0.9	0.127	0.064	0.147	0.179	0.023	0.098	0.050	0.091
	0.8	0.311	0.266	0.329	0.387	0.005	0.192	0.029	0.201
	1	0.035	0.020	0.041	0.045	0.044	0.045	0.040	0.055
(lo1)	0.9	0.169	0.129	0.192	0.211	0.048	0.166	0.056	0.123
	0.8	0.544	0.444	0.607	0.630	0.028	0.530	0.043	0.397
	1	0.035	0.018	0.042	0.046	0.044	0.045	0.038	0.056
(lo2)	0.9	0.177	0.059	0.205	0.228	0.063	0.149	0.107	0.125
	0.8	0.549	0.234	0.619	0.648	0.059	0.414	0.154	0.395

Note: Simulation results are reported for 100,000 replications with $T = 100$ and $\epsilon_t \sim \mathcal{N}(0, 1)$. For the specifications of the trend function d_t see Table 2 and Figure 1. The bandwidth for fixed-b is $B = T/10$ and for the small-b statistics $B = T^{0.6}$. DF$^{\mu}$ and DF* are the Dickey-Fuller tests with constant and linear trend specification, SP is the test by Schmidt and Phillips (1992) and EL is the test by Enders and Lee (2012).

The test works as follows. Firstly, the auxiliary regression

$$\Delta y_t = \delta_0 + \delta_1 \Delta \sin(2\pi t/T) + \delta_2 \Delta \cos(2\pi t/T) + \tilde{u}_t$$

is performed with OLS-estimates $\widehat{\delta}_0$, $\widehat{\delta}_1$ and $\widehat{\delta}_2$. This yields the detrended series

$$\begin{aligned} \widetilde{S}_t = y_t &- \Big(y_1 - \widehat{\delta}_0 - \widehat{\delta}_1 \sin(2\pi/T) - \widehat{\delta}_2 \cos(2\pi/T)\Big) \\ &- \widehat{\delta}_0 t - \Big(\widehat{\delta}_1 \sin(2\pi t/T) + \widehat{\delta}_2 \cos(2\pi t/T)\Big). \end{aligned}$$

If y_t has a unit root then it must be the case that $\phi = 0$ in the test regression

$$\Delta y_t = \phi \widetilde{S}_{t-1} + \xi_0 + \xi_1 \Delta \sin(2\pi t/T) + \xi_2 \Delta \cos(2\pi t/T) + \tilde{v}_t.$$

Table 4: Size and power for $T = 300$ using 5% critical values

d_t	ρ	τ-FB	τ-FB^{HC}	τ-SB	τ-SB^{HC}	DF$^{\mu}$	DF*	SP	EL
	1	0.047	0.048	0.055	0.057	0.053	0.051	0.049	0.048
(zer)	0.9	0.946	0.846	0.957	0.959	0.997	0.958	0.887	0.729
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.985	0.999
	1	0.047	0.056	0.053	0.055	0.053	0.050	0.047	0.048
(lin)	0.9	0.940	0.923	0.952	0.955	0.887	0.957	0.590	0.731
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.730	0.999
	1	0.032	0.033	0.038	0.038	0.039	0.031	0.032	0.048
(fo1)	0.9	0.780	0.676	0.813	0.820	0.223	0.040	0.274	0.731
	0.8	1.000	0.998	1.000	1.000	0.480	0.048	0.339	1.000
	1	0.039	0.039	0.047	0.047	0.043	0.041	0.042	0.042
(fo2)	0.9	0.889	0.788	0.907	0.913	0.855	0.517	0.721	0.585
	0.8	1.000	0.999	1.000	1.000	0.999	0.961	0.874	0.985
	1	0.038	0.038	0.042	0.045	0.053	0.043	0.038	0.048
(fo3)	0.9	0.857	0.737	0.879	0.887	0.846	0.616	0.315	0.728
	0.8	1.000	0.999	1.000	1.000	0.998	0.983	0.347	0.999
	1	0.047	0.039	0.054	0.058	0.046	0.046	0.049	0.046
(bre)	0.9	0.821	0.845	0.848	0.854	0.232	0.705	0.373	0.540
	0.8	1.000	0.999	1.000	1.000	0.490	0.996	0.468	0.981
	1	0.038	0.042	0.045	0.047	0.043	0.043	0.041	0.044
(lo1)	0.9	0.875	0.848	0.893	0.899	0.231	0.763	0.343	0.627
	0.8	1.000	1.000	1.000	1.000	0.509	0.999	0.426	0.995
	1	0.039	0.039	0.045	0.046	0.042	0.043	0.042	0.044
(lo2)	0.9	0.874	0.772	0.894	0.899	0.347	0.651	0.555	0.626
	0.8	1.000	0.999	1.000	1.000	0.737	0.993	0.751	0.994

Note: Simulation results are reported for 100,000 replications with $T = 300$ and $\epsilon_t \sim \mathcal{N}(0,1)$. For the specifications of the trend function d_t see Table 2 and Figure 1. The bandwidth for fixed-b is $B = T/10$ and for the small-b statistics $B = T^{0.6}$. DF$^{\mu}$ and DF* are the Dickey-Fuller tests with constant and linear trend specification, SP is the test by Schmidt and Phillips (1992) and EL is the test by Enders and Lee (2012).

Finally, the LM test statistic is equal to the t-statistic for the null hypothesis $\phi = 0$, that is

$$\text{EL} = \frac{\widehat{\phi}}{se(\widehat{\phi})}.$$

Tables 3 and 4 show that the pooled tests as well as their benchmarks are sized correctly around 5% for sample sizes of $T = 300$. We observe that all tests are slightly undersized for $T = 100$ if the deterministic component involves high variation. The power for the small-b statistics under the zero-trend specification is similar to their benchmarks DF$^{\mu}$ and SP and the fixed-b statistics perform slightly worse. For any

Table 5: Size and power under heteroskedasticity using 5% critical values

d_t	ρ	τ-FB	τ-FB^{HC}	τ-SB	τ-SB^{HC}	DF$^{\mu}$	DF*	SP	EL
$T = 100$									
	1	0.075	0.014	0.080	0.041	0.145	0.108	0.062	0.080
(zer)	0.9	0.297	0.038	0.364	0.236	0.403	0.289	0.640	0.158
	0.8	0.659	0.127	0.773	0.627	0.884	0.718	0.952	0.434
	1	0.074	0.018	0.078	0.040	0.142	0.109	0.058	0.080
(lin)	0.9	0.286	0.059	0.347	0.217	0.317	0.281	0.381	0.159
	0.8	0.648	0.183	0.753	0.603	0.707	0.712	0.649	0.436
	1	0.058	0.011	0.062	0.031	0.125	0.087	0.048	0.079
(fo3)	0.9	0.225	0.027	0.269	0.161	0.276	0.183	0.223	0.158
	0.8	0.550	0.089	0.631	0.468	0.695	0.467	0.336	0.438
$T = 300$									
	1	0.081	0.036	0.092	0.045	0.136	0.103	0.061	0.072
(zer)	0.9	0.909	0.562	0.922	0.829	0.993	0.940	0.994	0.736
	0.8	1.000	0.964	1.000	1.000	1.000	1.000	1.000	1.000
	1	0.081	0.042	0.090	0.045	0.135	0.104	0.062	0.071
(lin)	0.9	0.907	0.615	0.920	0.827	0.977	0.940	0.946	0.736
	0.8	1.000	0.977	1.000	1.000	1.000	1.000	0.997	1.000
	1	0.075	0.034	0.085	0.040	0.129	0.096	0.056	0.071
(fo3)	0.9	0.884	0.547	0.902	0.792	0.977	0.874	0.821	0.733
	0.8	1.000	0.967	1.000	1.000	1.000	1.000	0.959	1.000

Note: Simulation results are reported for 100,000 replications with $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$. The variance function is given by $\sigma_t^2 = 6 \cdot 1_{\{t/T \leq 0.5\}} + 1_{\{t/T > 0.5\}}$. For the specifications of the trend function d_t see Table 2 and Figure 1. The bandwidth for fixed-b is $B = T/10$ and for the small-b statistics $B = T^{0.6}$. DF$^{\mu}$ and DF* are the Dickey-Fuller tests with constant and linear trend specification, SP is the test by Schmidt and Phillips (1992) and EL is the test by Enders and Lee (2012).

other trend specification the power of the fixed-b and small-b tests is much higher and exceeds the power of DF$^{\mu}$ and SP clearly. For instance under trends (bre) and (lo1) with $\rho = 0.8$ the power of the pooled tests is equal to 1 for $T = 300$ whereas the power for DF$^{\mu}$ and SP does not exceed 0.5. The statistic of Enders and Lee (2012) has an advantage if the trend model is correctly specified like in model (fo1). But when the trend slightly deviates from their Fourier trend-specification as in (fo2) their test performs worse than the small-b and fixed-b variants. For all other trend specifications the tests τ-SB, τ-SB^{HC} and τ-FB clearly outperform the EL statistic in terms of power. The DF* statistic performs well when the trend is approximately linear, but it never outperforms τ-SB^{HC}. The heteroskedasticity robust statistic τ-FB^{HC} looses power against τ-FB and is more or less on a par with the Dickey-Fuller benchmarks. To conclude, the statistics τ-SB and τ-SB^{HC} tend to outperform all benchmark statistics for the trend specifications in our analysis.

Table 6: Size and power under autocorrelation for $T = 300$ using 5% critical values

d_t	ρ	τ-FB	τ-FB^{HC}	τ-SB	τ-SB^{HC}	DF$^\mu$	DF*	SP	EL
AR-process (6)									
(zer)	1	0.021	0.023	0.024	0.025	0.050	0.035	0.052	0.015
	0.9	0.887	0.778	0.900	0.903	0.963	0.867	0.887	0.612
(fo1)	1	0.014	0.012	0.017	0.017	0.043	0.029	0.040	0.020
	0.9	0.598	0.501	0.632	0.642	0.401	0.168	0.413	0.421
(lo1)	1	0.020	0.022	0.023	0.024	0.043	0.032	0.043	0.017
	0.9	0.597	0.564	0.627	0.639	0.320	0.529	0.412	0.256
MA-process (7)									
(zer)	1	0.035	0.032	0.040	0.042	0.058	0.046	0.064	0.024
	0.9	0.934	0.852	0.943	0.946	0.981	0.923	0.909	0.717
(fo1)	1	0.029	0.026	0.035	0.037	0.053	0.038	0.059	0.027
	0.9	0.803	0.694	0.824	0.833	0.699	0.428	0.649	0.572
(lo1)	1	0.030	0.029	0.035	0.037	0.053	0.041	0.060	0.023
	0.9	0.772	0.692	0.797	0.805	0.635	0.733	0.637	0.417

Note: Simulation results are reported for 10,000 replications with $T = 300$. The innovation process is simulated under the AR-specification $\epsilon_t = 0.3\epsilon_{t-1} + e_t$ and under the MA-specification $\epsilon_t = e_t + 0.8e_{t-1}$, where $e_t \sim \mathcal{N}(0, 1)$. For the specifications of the trend function d_t see Table 2 and Figure 1. The bandwidth for fixed-b is $B = T/10$ and for the small-b statistics $B = T^{0.6}$. DF$^\mu$ and DF* are the Dickey-Fuller tests with constant and linear trend specification, SP is the test by Schmidt and Phillips (1992) and EL is the test by Enders and Lee (2012). The simulated series is pre-whitened using Schwert's rule.

The other statistics have their strengths and weaknesses depending on the trend specifications.

In order to evaluate the size and power under the heteroskedastic framework of Assumption 3 we simulate $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ independently with the variance function

$$\sigma(r)^2 = 6 \cdot 1_{\{r \leq 0.5\}} + 1_{\{r > 0.5\}}. \tag{5}$$

The results are presented in Table 5. The unrobust tests are clearly oversized. With a sample size of $T = 300$ we obtain size levels around 8-9% for τ-FB and τ-SB. The Dickey-Fuller statistics yield size levels up to 15%. With size levels around 4%, the heteroskedasticity robust variants are sized reasonably well. The power losses for the small-b test τ-SB^{HC} are moderate compared to its oversized counterpart. Further simulations show that in terms of size-corrected power the τ-SB^{HC} test yields the best results in the presence of nonzero trend functions.

Finally, we simulate the empirical size and power under the presence of autocorrelation. We simulate the innovations according to the AR-specification

$$\epsilon_t = 0.3\epsilon_{t-1} + e_t \tag{6}$$

and the MA-specification

$$\epsilon_t = e_t + 0.8e_{t-1}, \tag{7}$$

where $e_t \sim \mathcal{N}(0,1)$ independently. In the presence of autocorrelation, the limiting null-distributions of the test statistics are shifted to the right and unit root tests are clearly undersized. The simulation results in Table A3 illustrate the consequences ignoring this effect. Therefore we pre-whiten the series before the tests are applied. The lag length is chosen according to the rule of thumb by Schwert (1989). Simulation results are presented in Table 6. Pre-whitening captures the autocorrelation in the series reasonably well. In small samples the pooled tests are slightly undersized. We observe this effect for any trend specification (see Table A4). However, the pooled tests still yield better results in terms of power than the conventional Dickey-Fuller tests.

6 Empirical application: quarterly inflation rates

For monetary policy it is crucial to know whether shocks to the inflation rate will have permanent or transitory effect. From an econometric point of view, the integrational properties of inflation rates affect the choice of an appropriate model. There has been a long debate in the literature on whether inflation rates are $I(1)$ or $I(0)$. Early studies as in MacDonald and Murphy (1989) show that conventional Dickey-Fuller tests are often unable to reject the unit root hypothesis for quarterly inflation rates. Evans and Lewis (1995) as well as Ng and Perron (2001) also find strong evidence that inflation rates are nonstationary. The work of Hassler and Wolters (1995) shows mixed results using ARFIMA models for monthly data. On the other hand, Rose (1988) find that for 18 countries quarterly inflation rates are stationary. Using panel unit root tests Lee and Wu (2001) provide evidence that inflation rates of 13 OECD countries do not contain a unit root. Allowing for multiple breaks Narayan and Narayan (2010) also find strong evidence for stationarity.

We apply our tests to quarterly growth rates of the consumer price index (CPI).[9] The dataset includes 22 countries with 233 observations from $1958Q1$ to $2016Q1$. The data is first pre-whitened according to Section 4 using Schwert's rule and then several unit root tests are applied. The results for the fixed-b and small-b tests as well as some benchmark tests are presented in Table 7.

At the 10% significance level the τ-SB^{HC} test rejects the unit root hypothesis for 17 of 22 countries and the τ-FB test for 16 countries, whereas the DF$^{\mu}$ test can

[9]Source: https://data.oecd.org/

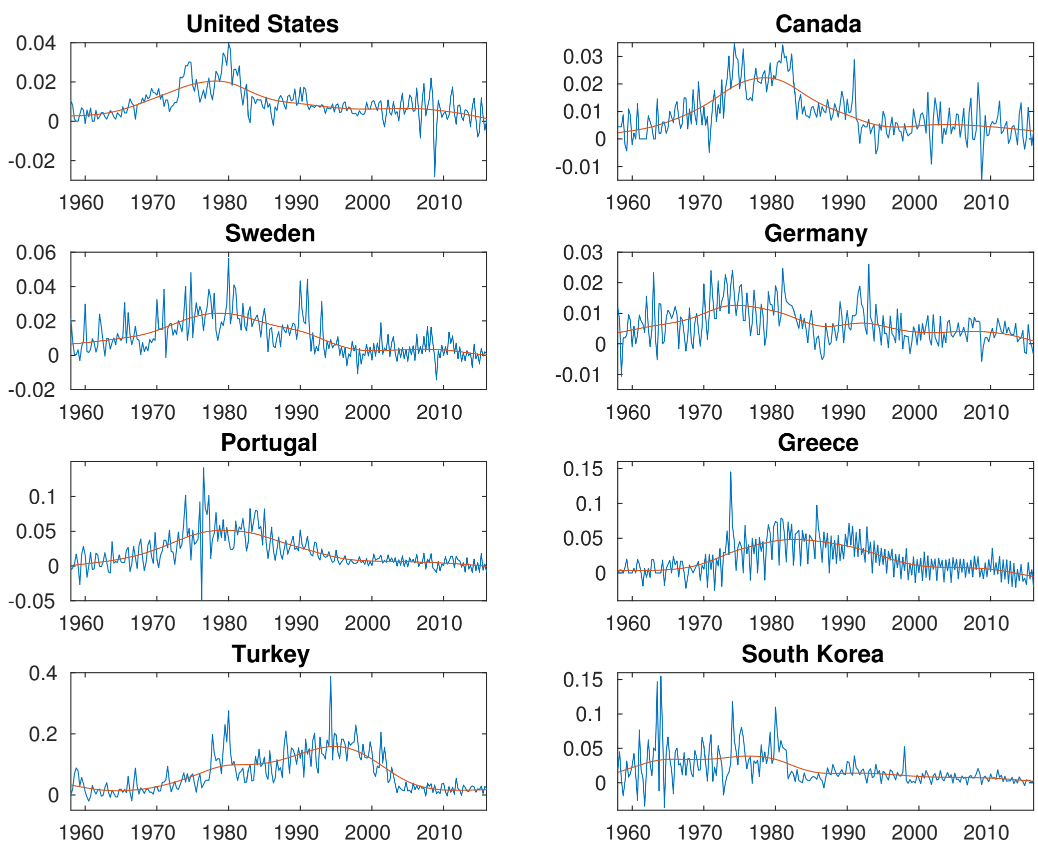


Figure 2: Quarterly CPI growth rates with HP trends

Table 7: Unit root tests applied to inflation data (1958Q1 - 2016Q1)

	τ-FB	τ-SB^{HC}	DF$^{\mu}$	DF*	SP	EL
BEL	−0.31	−1.39*	−2.50	−2.91	−1.87*	−3.43
CAN	−0.38*	−1.58*	−2.35	−2.78	−1.43	−3.61
FIN	−0.41*	−1.46*	−2.48	−3.40*	−1.72*	−3.94*
FRA	−0.18	−0.60	−1.51	−2.13	−1.36	−2.56
GER	−0.19	−0.98	−2.76*	−3.54**	−1.52	−3.48
GRC	−0.61***	−1.91**	−2.26	−2.42	−2.09**	−4.27**
ITA	−0.22	−0.59	−1.96	−2.60	−1.20	−3.15
JPN	−0.64***	−2.10**	−3.04**	−4.27***	−2.51**	−4.58**
KOR	−0.74***	−1.92**	−3.55***	−4.78***	−3.06***	−4.74***
LUX	−0.22	−1.12	−2.46	−2.90	−1.33	−3.17
NOR	−0.84***	−3.14***	−3.04**	−3.59**	−2.08**	−5.31***
PRT	−0.63***	−1.61*	−2.34	−2.84	−1.39	−4.58**
ESP	−0.33	−1.06	−1.84	−2.45	−1.77*	−3.00
SWE	−0.65***	−2.29**	−2.70*	−3.54**	−1.73*	−4.81***
CHE	−0.39*	−1.68**	−2.97**	−4.22***	−1.43	−3.74
TUR	−0.40*	−1.37*	−2.03	−1.98	−1.99**	−3.98*
GBR	−0.50**	−1.33*	−2.65*	−3.22*	−1.46	−3.78*
USA	−0.58**	−2.02**	−3.02**	−3.33*	−1.95**	−3.99*
AUS	−0.49**	−1.82**	−2.59*	−2.91	−1.82*	−3.71
NZL	−0.66***	−2.03**	−2.41	−2.79	−2.22**	−3.84*
IND	−0.99***	−4.11***	−6.11***	−6.11***	−5.30***	−6.09***
ZAF	−0.62***	−2.26**	−2.54	−2.50	−1.27	−4.78***

Note: Test statistics for 22 countries are reported. The bandwidth for τ-FB is $B = T/10$ and for the τ-SB^{HC} statistic $B = T^{0.6}$. DF$^{\mu}$ and DF* are the Dickey-Fuller tests with constant and linear trend specification, SP is the test by Schmidt and Phillips (1992) and EL is the test by Enders and Lee (2012). The asterisks *, ** and *** denote the significance at the 10%, 5% and 1% level, respectively.

only reject H_0 for 10 countries. At the 5% level the number of rejections for τ-SB^{HC} and τ-FB are 11 and 12 and for DF$^{\mu}$ only 6. The pooled tests seem to reject the unit root hypothesis more often than the classical tests of Dickey and Fuller (1979), Schmidt and Phillips (1992) and Enders and Lee (2012). This provides evidence that inflation rates are stationary around a slowly varying trend reflecting different regimes of monetary policy.

Figure 2 shows some plots of inflation rates and corresponding Hodrick-Prescott-Filter with $\lambda = 16000$. The trends typically show an inverted U-shaped behavior with peak in the mid-70s. For countries with a more pronounced inverted U-shape like Canada the pooled tests tend to reject the unit root hypothesis more clearly than the Dickey-Fuller counterparts. For countries with a quite flat trend in inflation rates like Germany the τ-FB and τ-SB^{HC} tests perform similarly to DF$^{\mu}$. The lower four pictures in Figure 2 indicate that varying volatility plays a role here for some countries. In South Korea the variance before the 1980s was clearly higher than in the 1990s and today. Due to size distortions under heteroskedasticity only the test results of τ-SB^{HC} are reliable here.

7 Conclusion

We have presented two variants of a unit root tests under an unknown trend specification that is robust under both heteroskedasticity and autocorrelation. In finite samples the tests show good size properties. The fixed-b pooled test statistic converges to a functional of Brownian motions under the unit root hypothesis. The small-b variant is standard normally distributed. By transforming the data by its inverse variance profile the fixed-b test is heteroskedasticity robust. For the small-b test we can estimate the heteroskedasticity nuisance parameter easily. By a pre-whitening scheme we correct for autocorrelation.

Monte Carlo simulations show that in terms of power, the τ-SB^{HC} statistic with a blocklength of $B = T^{0.6}$ and the τ-FB statistic with $B = T/10$ perform best. Under many trend specifications τ-SB^{HC} clearly outperforms the tests by Dickey and Fuller (1979), Schmidt and Phillips (1992) and Enders and Lee (2012). For quarterly inflation rates the unit root hypothesis is rejected by the pooled tests for more countries than by the conventional Dickey-Fuller tests. This provides evidence that inflation rates are stationary around a slowly varying deterministic component.

A Appendix

In order to show the results we rewrite the statistics as a functional of X_T, which is defined as $X_T(r) = \overline{\sigma}^{-1}T^{-0.5}\sum_{t=1}^{\lfloor rT \rfloor}\epsilon_t$, $0 \leq r \leq 1$, where $\overline{\sigma}^2 = \int_0^1 \sigma(r)^2 dr$ is the average variance. Note that in the case of homoskedastic errors as in Assumption 1 we have $\overline{\sigma}^2 = \sigma^2$. Let $\eta(s) = \overline{\sigma}^{-2}\int_0^s \sigma(r)^2 dr$ be the variance profile and $W_\eta = (W \circ \eta)$ be a corresponding time transformed Brownian Motion on the unit interval. From Lemma 4 in Cavaliere and Taylor (2007) we obtain the Functional Central Limit Theorem $X_T \Rightarrow W_\eta$ on the càdlàg space $D[0,1]$, where $\Rightarrow$ denotes weak convergence in this space together with the uniform metric. Under the unit root hypothesis $\rho = 1$, we have $u_t = \sum_{k=1}^t \epsilon_k + u_0$ and $X_T(r) = \overline{\sigma}^{-1}T^{-0.5}(u_{\lfloor rT \rfloor} - u_0)$.

For the deterministic trend we define $D_T(r) = T^\alpha \Delta d_{\lfloor rT \rfloor}$ where $0 \leq r \leq 1$. Then $\lim_{T\to\infty} D_T(r) = D(r)$ exist due to Hölder continuity, and D is continuous and bounded on $[0,1]$. From the Hölder continuity of d_t we also deduce that there exists a constant $C < \infty$ such that

$$|\Delta d_{t+j}| \leq \frac{C}{T^\alpha} \quad \text{and} \quad |d_{t+j-1} - d_j| \leq C\left(\frac{j-1}{T}\right)^\alpha \leq C\left(\frac{B}{T}\right)^\alpha.$$

A.1 Lemmas

For the proofs of the theorems the following lemmas will be useful.

Lemma 3. *Let $(a_t)_{t\in\mathbb{N}}$, $(A_{s,t})_{s,t\in\mathbb{N}}$, $n_1, n_2 \in \mathbb{N}$ and $m \geq 2n_1$. Then*

(a)

$$\sum_{k=n_1}^{n_2-1}\sum_{l=k+1}^{n_2} a_k a_l = \sum_{k=n_1+1}^{n_2}\sum_{l=1}^{k-n_1} a_k a_{k-l},$$

(b)

$$\sum_{t=2}^{n_1}\sum_{k=2}^{t-1}\sum_{l=1}^{k-1} a_{k+j}a_{k+j-l} = \sum_{k=2}^{n_1-1}\sum_{l=1}^{k}(n_1-k)a_{k+j}a_{k+j-l}, \quad j \in \mathbb{N},$$

(c)

$$\sum_{t=2}^{n_1}\sum_{k=1}^{t-1} a_{k+j} = \sum_{t=1}^{n_1-1}(n_1-t)a_{t+j}, \quad j \in \mathbb{N},$$

(d)

$$\sum_{j=1}^{m-n_1}\sum_{t=1}^{n_1-1} a_{t+j} = \sum_{t=2}^{n_1-1}(t-1)a_t + \sum_{t=n_1}^{m-n_1+1}(n_1-1)a_t + \sum_{t=m-n_1+2}^{m-1}(m-t)a_t,$$

(e)

$$2\sum_{j=1}^{m-n_1}\sum_{t=1}^{n_1-1} t a_{t+j} = \sum_{t=2}^{n_1-1} t(t-1)a_t + \sum_{t=n_1}^{m-n_1+1} n_1(n_1-1)a_t + \sum_{t=m-n_1+2}^{m-1}(m-t)(2n_1+t-m-1)a_t,$$

(f)

$$\sum_{t=2}^{n_1}\sum_{k=1}^{t-1} a_{j-k} = \sum_{k=1}^{n_1-1}(n_1-k)a_{j-k}, \quad j \in \mathbb{N}.$$

Proof. The identities follow by mathematical induction on n_1, n_2 and m. □

Lemma 4. *Let $\{\{q_{k,T}\}_{B+1\le k\le T-B}\}_{T\in\mathbb{N}}$ be a martingale difference array with*

$$q_{k,T} = \epsilon_k \sum_{l=1}^{B-1}\frac{B-l}{B}\epsilon_{k-l}, \quad V = \frac{1}{T-2B}\sum_{k=B+1}^{T-B} Var(q_{k,T}), \quad B+1 \le k \le T-B$$

and $B = o(T)$ and let ϵ_t satisfy Assumption 3. Then

$$\frac{1}{\sqrt{VT}}\sum_{k=B+1}^{T-B} q_{k,T} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. Consider

$$\tilde{q}_{k,T} = \frac{q_{k,T}}{\sqrt{(T-2B)V}}.$$

Clearly, $\tilde{q}_{k,T}$ is an element of a md-array with $\sum_{k=B+1}^{T-B} Var(\tilde{q}_{k,T}) = 1$ and by Chebyshev's inequality we obtain $\tilde{q}_{k,T} \xrightarrow{p} 0$. Thus

$$\max_{B+1\le k\le T-B} |\tilde{q}_{k,T}| \xrightarrow{p} 0. \tag{8}$$

Note that we have

$$
\begin{aligned}
Var(q_{k,T}) &= \frac{\sigma_k^4}{B^2}\sum_{l=1}^{B-1}\sigma_{k-l}^2(B-l)^2 = \sigma_k^2 B\int_0^1 \sigma^2(k/T-rB/T)(1-r)^2dr + o(B)\\
&= \sigma_k^4 B\int_0^1 (1-r)^2 dr + o(B) = \frac{\sigma_k^4 B}{3} + o(B). \qquad (9)
\end{aligned}
$$

Now, for $|r-k| \leq B$,

$$
\begin{aligned}
&E[\tilde{q}_{k,T}^2\tilde{q}_{r,T}^2] = E\left[\frac{1}{V^2(T-2B)^2}\left(\sum_{l=1}^{B-1}\frac{B-l}{B}\epsilon_k\epsilon_{k-l}\right)^2\left(\sum_{s=1}^{B-1}\frac{B-s}{B}\epsilon_r\epsilon_{r-s}\right)^2\right]\\
&= \sum_{l,m,s,t=1}^{B-1}\frac{(B-l)(B-m)(B-s)(B-t)}{V^2(T-2B)^2B^4}E[\epsilon_k^2\epsilon_{k-l}\epsilon_{k-m}\epsilon_r^2\epsilon_{r-s}\epsilon_{r-t}]\\
&\leq \frac{1}{(T-2B)^2V^2}\sum_{l,m,s,t=1}^{B-1}E[\epsilon_k^2\epsilon_{k-l}\epsilon_{k-m}\epsilon_r^2\epsilon_{r-s}\epsilon_{r-t}] = \mathcal{O}(T^{-2}),
\end{aligned}
$$

where the last equality follows from (9) and that $E[\epsilon_k^2\epsilon_{k-l}\epsilon_{k-m}\epsilon_r^2\epsilon_{r-s}\epsilon_{r-t}] \neq 0$ only if at least two of the restrictions

$$
\{l=m; s=t; k-l=r-s; k-l=r-t; k-m=r-s; k-m=r-t\}
$$

are fulfilled. Therefore at least two of the four sums indexed by l, m, s, t vanish. Furthermore we have $E[\epsilon_k^2\epsilon_{k-l}\epsilon_{k-m}\epsilon_r^2\epsilon_{r-s}\epsilon_{r-t}] < \infty$ since the fourth moments are bounded. For $|r-k| > B$ we have $E[\tilde{q}_{k,T}^2\tilde{q}_{r,T}^2] = (T-2B)^{-2}$. Then

$$
\begin{aligned}
&Var\left(\sum_{k=B+1}^{T-B}\tilde{q}_{k,T}^2\right) = \sum_{k,r=B+1}^{T-B}Cov\left(\tilde{q}_{k,T}^2,\tilde{q}_{r,T}^2\right) = \left(\sum_{k,r=B+1}^{T-B}E[\tilde{q}_{k,T}^2\tilde{q}_{r,T}^2]\right) - 1\\
&= \sum_{k=B+1}^{T-B}\sum_{r=k-B}^{k+B}E[\tilde{q}_{k,T}^2\tilde{q}_{r,T}^2] + o(1) = o(1)
\end{aligned}
$$

and by Chebyshev's inequality

$$
\sum_{k=B+1}^{T-B}\tilde{q}_{k,T}^2 \xrightarrow{p} 1. \qquad (10)
$$

With (8) and (10) we apply a Central Limit Theorem for martingale difference arrays (Theorem 24.3, Davidson (1994)) and obtain

$$\sum_{k=B+1}^{T-B} \tilde{q}_{k,T} \xrightarrow{d} \mathcal{N}(0,1).$$

Since

$$\left| \frac{1}{\sqrt{VT}} \sum_{k=B+1}^{T-B} q_{k,T} - \sum_{k=B+1}^{T-B} \tilde{q}_{k,T} \right| \longrightarrow 0,$$

the assertion follows. □

Lemma 5. *Let $d_t = d(t/T)$ be Hölder continuous, let Assumption 3 hold and let $\rho = 1$. For $B/T \to b$, $0 < b < 1$ we obtain that*

(i) $\frac{1}{T^2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta d_{t+j}(d_{t+j-1} - d_j) = \mathcal{O}(T^{1-2\alpha})$,

(ii) $\frac{1}{T^2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta u_{t+j}(d_{t+j-1} - d_j) = \mathcal{O}_P(T^{0.5-\alpha})$,

(iii) $\frac{1}{T^2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta d_{t+j}(u_{t+j-1} - u_j) = \mathcal{O}_P(T^{0.5-\alpha})$,

(iv) $\frac{1}{T^3} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (d_{t+j-1} - d_j)^2 = \mathcal{O}(T^{1-2\alpha})$,

(v) $\frac{1}{T^3} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (d_{t+j-1} - d_j)(u_{t+j-1} - u_j) = \mathcal{O}_P(T^{0.5-\alpha})$.

Proof. (i)

$$\left| \frac{T^{2\alpha-1}}{T^2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta d_{t+j}(d_{t+j-1} - d_j) \right| = \left| \frac{T^{2\alpha-1}}{T^2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \sum_{k=1}^{t-1} \Delta d_{t+j} \Delta d_{k+j} \right|$$

$$\leq \frac{T^{2\alpha-1}}{T^2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \sum_{k=1}^{t-1} \frac{C^2}{T^{2\alpha}} \leq \frac{C^2}{2} \frac{B^2}{T^2} = \mathcal{O}(1).$$

(ii)

$$\frac{T^{\alpha}}{T^{0.5}}\frac{1}{T^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta u_{t+j}(d_{t+j-1}-d_j)=\frac{T^{\alpha}}{T^{2.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\Delta u_{t+j}\Delta d_{k+j}$$
$$=\frac{1}{T^2}\overline{\sigma}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{j=t+1}^{T-B+t}(T^{\alpha}\Delta d_{j-t+k})(\overline{\sigma}^{-1}T^{-0.5}\Delta u_j)$$
$$=\overline{\sigma}\int_0^b\int_0^z\int_0^1 D_T(r)dX_T(r)dsdz+o_P(1)=\frac{b^2\overline{\sigma}}{2}\int_0^1 D_T(r)dX_T(r)+o_P(1)$$
$$=\mathcal{O}_P(1).$$

(iii)

$$\frac{T^{\alpha}}{T^{0.5}}\frac{1}{T^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta d_{t+j}(u_{t+j-1}-u_j)=\frac{T^{\alpha}}{T^{2.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\Delta d_{t+j}\Delta u_{k+j}$$
$$=\frac{1}{T^2}\overline{\sigma}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{j=k+1}^{T-B+k}(T^{\alpha}\Delta d_{t-k+j})(\overline{\sigma}^{-1}T^{-0.5}\Delta u_j)$$
$$=\overline{\sigma}\int_0^b\int_0^z\int_0^1 D_T(r)dX_T(r)dsdz+o_P(1)=\frac{b^2\overline{\sigma}}{2}\int_0^1 D_T(r)dX_T(r)+o_P(1)$$
$$=\mathcal{O}_P(1).$$

(iv)

$$\left|\frac{T^{2\alpha}}{T}\frac{1}{T^3}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(d_{t+j-1}-d_j)^2\right|=\left|\frac{T^{2\alpha}}{T^4}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\Delta d_{k+j}\Delta d_{l+j}\right|$$
$$\leq\frac{T^{2\alpha}}{T^4}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\sum_{j=1}^{T-B}\frac{C^2}{T^{2\alpha}}\leq C^2\frac{TB^3}{T^4}=\mathcal{O}(1).$$

(v)

$$
\begin{aligned}
&\frac{T^{\alpha}}{T^{0.5}}\frac{1}{T^{3}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(d_{t+j-1}-d_{j})(u_{t+j-1}-u_{j})\\
&=\frac{T^{\alpha}}{T^{3.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\Delta d_{k+j}\Delta u_{l+j}\\
&=\frac{1}{T^{3}}\overline{\sigma}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\sum_{j=l+1}^{T-B+l}(T^{\alpha}\Delta d_{j+k-l})(T^{-0.5}\overline{\sigma}^{-1}\Delta u_{j})\\
&=\overline{\sigma}\int_{0}^{b}\int_{0}^{z}\int_{0}^{z}\int_{0}^{1}D_{T}(r)dX_{T}(r)dsdxdz+o_{P}(1)\\
&=\frac{b^{3}\overline{\sigma}}{3}\int_{0}^{1}D_{T}(r)dX_{T}(r)+o_{P}(1)=\mathcal{O}_{P}(1).
\end{aligned}
$$

□

Lemma 6. *Let* $d_t = d(t/T)$ *be Hölder continuous, let Assumption 3 hold and let* $\rho = 1$*. For* $B/T \to 0$ *we obtain that*

(i) $\frac{1}{B^{1.5}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta d_{t+j}(d_{t+j-1}-d_j) = \mathcal{O}(T^{0.5-2\alpha}B^{\alpha-0.5})$,

(ii) $\frac{1}{B^{1.5}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta u_{t+j}(d_{t+j-1}-d_j) = \mathcal{O}_P(B^{0.5}T^{-\alpha})$,

(iii) $\frac{1}{B^{1.5}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta d_{t+j}(u_{t+j-1}-u_j) = \mathcal{O}_P(B^{0.5}T^{-\alpha})$,

(iv) $\frac{1}{B^{2}T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(d_{t+j-1}-d_j)^2 = \mathcal{O}(BT^{-2\alpha})$,

(v) $\frac{1}{B^{2}T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(d_{t+j-1}-d_j)(u_{t+j-1}-u_j) = \mathcal{O}_P(BT^{-0.5-\alpha})$.

Proof. (i)

$$
\begin{aligned}
&\left|\frac{B^{0.5-\alpha}}{T^{0.5-2\alpha}}\frac{1}{B^{1.5}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta d_{t+j}(d_{t+j-1}-d_j)\right|\\
&\le\frac{B^{0.5-\alpha}}{T^{0.5-2\alpha}}\frac{1}{B^{1.5}T^{0.5}}(T-B)BC^{2}\frac{B^{\alpha}}{T^{2\alpha}}\le C^{2}.
\end{aligned}
$$

(ii)

$$\frac{T^{\alpha}}{B^{0.5}}\frac{1}{B^{1.5}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta u_{t+j}(d_{t+j-1}-d_j) = \frac{T^{\alpha}}{B^{2}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\Delta u_{t+j}\Delta d_{k+j}$$
$$= \frac{1}{B^2}\overline{\sigma}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{j=t+1}^{T-B+t}(T^{\alpha}\Delta d_{j-t+k})(\overline{\sigma}^{-1}T^{-0.5}\Delta u_j)$$
$$= \overline{\sigma}\int_0^1\int_0^z\int_0^1 D_T(r)dX_T(r)dsdz + o_P(1) = \frac{\overline{\sigma}}{2}\int_0^1 D_T(r)dX_T(r) + o_P(1)$$
$$= \mathcal{O}_P(1).$$

(iii)

$$\frac{T^{\alpha}}{B^{0.5}}\frac{1}{B^{1.5}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta d_{t+j}(u_{t+j-1}-u_j) = \frac{T^{\alpha}}{B^{2}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\Delta d_{t+j}\Delta u_{k+j}$$
$$= \frac{1}{B^2}\overline{\sigma}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{j=k+1}^{T-B+k}(T^{\alpha}\Delta d_{t-k+j})(\overline{\sigma}^{-1}T^{-0.5}\Delta u_j)$$
$$= \overline{\sigma}\int_0^1\int_0^z\int_0^1 D_T(r)dX_T(r)dsdz + o_P(1) = \frac{\overline{\sigma}}{2}\int_0^1 D_T(r)dX_T(r) + o_P(1)$$
$$= \mathcal{O}_P(1).$$

(iv)

$$\left|\frac{T^{2\alpha}}{B}\frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(d_{t+j-1}-d_j)^2\right| = \left|\frac{T^{2\alpha}}{B^3T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\Delta d_{k+j}\Delta d_{l+j}\right|$$
$$\leq \frac{T^{2\alpha}}{B^3T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\frac{C^2}{T^{2\alpha}} = \mathcal{O}(1).$$

(v)

$$
\begin{aligned}
&\frac{T^{0.5+\alpha}}{B}\frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(d_{t+j-1}-d_j)(u_{t+j-1}-u_j)\\
&=\frac{T^{\alpha}}{B^3T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\Delta d_{k+j}\Delta u_{l+j}\\
&=\frac{1}{B^3}\overline{\sigma}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\sum_{j=l+1}^{T-B+l}(T^{\alpha}\Delta d_{j+k-l})(T^{-0.5}\overline{\sigma}^{-1}\Delta u_j)\\
&=\overline{\sigma}\int_0^1\int_0^z\int_0^z\int_0^1 D_T(r)dX_T(r)dsdxdz+o_P(1)\\
&=\frac{\overline{\sigma}}{3}\int_0^1 D_T(r)dX_T(r)+o_P(1)=\mathcal{O}_P(1).
\end{aligned}
$$

□

A.2 Proof of Lemma 1

We show the result under Assumption 2 and 3. Note that

$$
\hat{y}_t = y_t-\hat{\rho}y_{t-1}=\epsilon_t+d_t-\rho d_{t-1}+(\rho-\hat{\rho})y_{t-1}=\epsilon_t+\mathcal{O}(T^{-\alpha})+\mathcal{O}_P((TB)^{-1/2})
$$

for $|\rho|\leq 1$ due to Hölder continuity and Theorem 1(i). Then

$$
\hat{\sigma}_{\mathrm{b}}^2=\frac{1}{T-B}\sum_{j=1}^{T-B}\frac{1}{B-1}\sum_{t=1}^{B}\left(\hat{y}_{j+t}-\bar{\hat{y}}_j\right)^2=\frac{1}{T-B}\sum_{j=1}^{T-B}\frac{1}{B-1}\sum_{t=1}^{B}(\epsilon_{j+t}-\overline{\epsilon_j})^2+o_P(1)
$$

with $\overline{\epsilon_j}=\frac{1}{B}\sum_{k=1}^{B}\epsilon_{j+k}$. For the small-b statistic ($B/T\to 0$) we have

$$
\hat{\sigma}_{\mathrm{b}}^2=\frac{1}{T-B}\sum_{j=1}^{T-B}\sigma^2(j/T)+o_P(1)=\int_0^1\sigma^2(r)dr+o_P(1)
$$

and for the fixed-b statistic

$$
\hat{\sigma}_{\mathrm{b}}^2=\frac{1}{T-B}\sum_{j=1}^{T-B}\int_{\frac{j}{T}}^{\frac{j}{T}+b}\sigma^2(r)dr+o_P(1)=\int_0^1\sigma^2(r)dr+o_P(1).
$$

Note that under Assumption 1 we have $\overline{\sigma}=\sigma$.

A.3 Proof of Theorem 1

We show the results under Assumption 2 and 3. Then under Assumption 1, we have the special case $\eta(r) = r$, which yields $W_\eta = W$.
We first consider the term

$$
\begin{aligned}
&(y_{t+j-1} - y_j)\Delta y_{t+j} - (u_{t+j-1} - u_j)\Delta u_{t+j} \\
&= \Delta d_{t+j}(d_{t+j-1} - d_j) + \Delta u_{t+j}(d_{t+j-1} - d_j) + \Delta d_{t+j}(u_{t+j-1} - u_j)
\end{aligned}
$$

and we conclude from Lemma 5 and Assumption 2 that

$$
\frac{1}{T^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1} - y_j)\Delta y_{t+j} - (u_{t+j-1} - u_j)\Delta u_{t+j} = \mathcal{O}_P(T^{0.5-\alpha}) = o_P(1). \quad (11)
$$

Secondly, we have

$$
(y_{t+j-1} - y_j)^2 - (u_{t+j-1} - u_j)^2 = (d_{t+j-1} - d_j)^2 + 2(d_{t+j-1} - d_j)(u_{t+j-1} - u_j)
$$

and from Lemma 5 and Assumption 2 it follows that

$$
\frac{1}{T^3}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1} - y_j)^2 - (u_{t+j-1} - u_j)^2 = \mathcal{O}_P(T^{0.5-\alpha}) = o_P(1). \quad (12)
$$

Then with (11) and (12) we have

$$
\begin{aligned}
T(\hat{\rho}^{\text{fb}} - 1) &= T\left(\frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j} - y_j)(y_{t+j-1} - y_j)}{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1} - y_j)^2} - 1\right) \\
&= \frac{T^{-2}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1} - y_j)\Delta y_{t+j}}{T^{-3}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1} - y_j)^2} \\
&= \frac{T^{-2}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1} - u_j)\Delta u_{t+j} + o_P(1)}{T^{-3}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1} - u_j)^2 + o_P(1)}. \quad (13)
\end{aligned}
$$

Note that

$$\sum_{t=2}^{B} u_{t+j-1}\Delta u_{t+j} = \sum_{t=2}^{B} u_{t+j-1}\epsilon_{t+j} = \frac{1}{2}\sum_{t=2}^{B}\epsilon_{t+j}(2u_{t+j-1})$$
$$= \frac{1}{2}\sum_{t=2}^{B}\epsilon_{t+j}(u_{t+j} - \epsilon_{t+j} + u_{t+j-1}) = \frac{1}{2}\sum_{t=2}^{B}\epsilon_{t+j}(u_{t+j} + u_{t+j-1}) - \epsilon_{t+j}^2$$
$$= \frac{1}{2}\sum_{t=2}^{B}(u_{t+j} - u_{t+j-1})(u_{t+j} + u_{t+j-1}) - \epsilon_{t+j}^2 = \frac{1}{2}\sum_{t=2}^{B}(u_{t+j}^2 - u_{t+j-1}^2 - \epsilon_{t+j}^2)$$
$$= \frac{1}{2}(u_{B+j}^2 - u_{1+j}^2) - \frac{1}{2}\sum_{t=2}^{B}\epsilon_{t+j}^2.$$

For the numerator of (13) we have

$$\frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1} - u_j)\Delta u_{t+j}$$
$$= \frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}\frac{u_{B+j}^2 - u_{1+j}^2}{2} - \frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\frac{\epsilon_{t+j}^2}{2} - \frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}u_j\sum_{t=2}^{B}\Delta u_{t+j},$$

where

$$\frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}\frac{u_{B+j}^2 - u_{1+j}^2}{2} - \frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}u_j\sum_{t=2}^{B}\Delta u_{t+j}$$
$$= \int_0^{1-b}\frac{X_T^2(b+r) - X_T^2(r)}{2}dr - \int_0^{1-b}X_T(r)\left(X_T(b+r) - X_T(r)\right)dr + o_P(1)$$
$$= \int_0^{1-b}\frac{(X_T(b+r) - X_T(r))^2}{2}dr + o_P(1)$$

and

$$E\left[\frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\frac{\epsilon_{t+j}^2}{2}\right] = \frac{B(T-B)}{2\overline{\sigma}^2T^2}\int_0^1\sigma(r)^2dr + o(1) = \frac{b(1-b)}{2} + o(1).$$

Then with Chebyshev's inequality we have

$$\frac{1}{T^2\overline{\sigma}^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\frac{\epsilon_{t+j}^2}{2} \xrightarrow{p} \frac{b(1-b)}{2} \tag{14}$$

and the numerator of (13) is equal to

$$\int_0^{1-b} \frac{(X_T(b+r) - X_T(r))^2}{2} dr - \frac{b(1+b)}{2} + o_P(1). \tag{15}$$

For the denominator of (13) we have

$$\begin{aligned}
&\frac{1}{T^3\overline{\sigma}^2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (u_{t+j-1} - u_j)^2 = \frac{1}{T^2} \sum_{j=1}^{T-B} \sum_{t=2+j}^{B+j} \frac{1}{T\overline{\sigma}^2} (u_{t-1} - u_j)^2 \\
&= \int_0^{1-b} \int_r^{b+r} (X_T(s) - X_T(r))^2 ds dr + o_P(1).
\end{aligned} \tag{16}$$

Hence, by the CMT and the FCLT,

$$T(\hat{\rho}^{\text{fb}} - 1) \xrightarrow{d} \frac{\int_0^{1-b} \left(W_\eta(b+r) - W_\eta(r)\right)^2 dr - b(1-b)}{2\int_0^{1-b} \int_r^{b+r} (W_\eta(s) - W_\eta(r))^2 ds dr}.$$

Note that under Assumption 1 we have $W_\eta = W$.
For the pseudo t-statistic we have analogously to (i) that

$$\begin{aligned}
\tau\text{-FB} &= \frac{\sigma}{\hat{\sigma}_{\text{b}}} \cdot \frac{T^{-2}\sigma^{-2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)\Delta y_{t+j}}{\sqrt{T^{-3}\sigma^{-2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2}} \\
&= \frac{\sigma}{\hat{\sigma}_{\text{b}}} \cdot \frac{T^{-2}\sigma^{-2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (u_{t+j-1} - u_j)\Delta u_{t+j} + o_P(1)}{\sqrt{T^{-3}\sigma^{-2} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (u_{t+j-1} - u_j)^2} + o_P(1)}.
\end{aligned}$$

The result follows with (15), (16) and $\sigma/\hat{\sigma}_{\text{b}} \xrightarrow{p} 1$.

A.4 Proof of Theorem 2

We show the results under Assumption 2 and 3. Then under Assumption 1, we have the special case $\sigma_* = 1$.

We first consider consider the difference

$$\begin{aligned}
&(y_{t+j-1} - y_j)\Delta y_{t+j} - (u_{t+j-1} - u_j)\Delta u_{t+j} \\
&= \Delta d_{t+j}(d_{t+j-1} - d_j) + \Delta u_{t+j}(d_{t+j-1} - d_j) + \Delta d_{t+j}(u_{t+j-1} - u_j)
\end{aligned}$$

and we conclude from Lemma 6 that

$$\frac{1}{B^{1.5}T^{0.5}}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1}-y_j)\Delta y_{t+j}-(u_{t+j-1}-u_j)\Delta u_{t+j}$$
$$=\mathcal{O}(T^{0.5-2\alpha}B^{\alpha-0.5})+\mathcal{O}_P(B^{0.5}T^{-\alpha}).$$

Note that

$$T^{0.5-2\alpha}B^{\alpha-0.5}=\frac{T^{0.5}}{T^{\alpha}}\frac{B^{\alpha}}{T^{\alpha}}\frac{1}{B^{0.5}}$$

and for $\alpha>0.5$ we obtain

$$\mathcal{O}(T^{0.5-2\alpha}B^{\alpha-0.5})+\mathcal{O}_P(B^{0.5}T^{-\alpha})=o_P(1). \tag{17}$$

Secondly, we consider

$$(y_{t+j-1}-y_j)^2-(u_{t+j-1}-u_j)^2$$
$$=(d_{t+j-1}-d_j)^2+2(d_{t+j-1}-d_j)(u_{t+j-1}-u_j)$$

and from Lemma 6 it follows that

$$\frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1}-y_j)^2-(u_{t+j-1}-u_j)^2=\mathcal{O}_P(BT^{-0.5-\alpha})=o_P(1). \tag{18}$$

Then with (17) and (18) we have

$$\begin{aligned}\sqrt{BT}(\hat{\rho}^{\text{sb}}-1)&=\sqrt{BT}\left(\frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j}-y_j)(y_{t+j-1}-y_j)}{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1}-y_j)^2}-1\right)\\&=\frac{B^{-1.5}T^{-0.5}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1}-y_j)\Delta y_{t+j}}{B^{-2}T^{-1}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1}-y_j)^2}\\&=\frac{B^{-1.5}T^{-0.5}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j}+o_P(1)}{B^{-2}T^{-1}\overline{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)^2+o_P(1)}.\end{aligned} \tag{19}$$

We first consider the denominator of (19). Note that

$$(u_{t+j-1}-u_j)^2=\left(\sum_{k=j+1}^{t+j-1}\epsilon_k\right)^2=\sum_{k=j+1}^{t+j-1}\epsilon_k^2+2\sum_{k=j+1}^{t+j-2}\sum_{l=j+1}^{t+j-1}\epsilon_k\epsilon_l,$$

and therefore

$$\frac{1}{B^2T\sigma^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)^2$$

$$=\frac{1}{B^2T\sigma^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=j+1}^{t+j-1}\epsilon_k^2+\frac{2}{B^2T\sigma^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=j+1}^{t+j-2}\sum_{l=k+1}^{t+j-1}\epsilon_k\epsilon_l. \tag{20}$$

For the second summand of (20) it follows with Lemma 3(a) and (b) that

$$\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=j+1}^{t+j-2}\sum_{l=k+1}^{t+j-1}\epsilon_k\epsilon_l=\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=j+2}^{t+j-1}\sum_{l=1}^{k-j-1}\epsilon_k\epsilon_{k-l}$$

$$=\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=2}^{t-1}\sum_{l=1}^{k-1}\epsilon_{k+j}\epsilon_{k+j-l}=\sum_{j=1}^{T-B}\sum_{k=2}^{B-1}\sum_{l=1}^{k}(B-k)\epsilon_{k+j}\epsilon_{k+j-l}$$

$$=\sum_{k=2}^{B-1}(B-k)\sum_{j=1}^{T-B}\sum_{l=1}^{k}\epsilon_{k+j}\epsilon_{k+j-l}=\sum_{k=1}^{T}\sum_{l=1}^{B-1}N_{k,l}\epsilon_k\epsilon_{k-l},$$

where the numbers $N_{k,l}$ satisfy $0\leq N_{k,l}\leq B^2$. For convenience, we set $\epsilon_j=0$ for $j\leq 0$. Notice that for $k\neq r$ we have

$$Cov\left(\sum_{l=1}^{B-1}N_{k,l}\epsilon_k\epsilon_{k-l},\sum_{s=1}^{B-1}N_{r,s}\epsilon_r\epsilon_{r-s}\right)=0$$

such that

$$Var\left(\frac{2}{B^2T\overline{\sigma}^2}\sum_{k=1}^{T}\sum_{l=1}^{B-1}N_{k,l}\epsilon_k\epsilon_l\right)=\frac{4}{B^4T^2\overline{\sigma}^4}\sum_{k=1}^{T}\sum_{l=1}^{B-1}N_{k,l}^2\sigma_k^2\sigma_l^2=o(1). \tag{21}$$

By Chebyshev's inequality, we conclude that the second summand of (20) converges to zero in probability.
For the first summand of (20) it follows with Lemma 3(c) that

$$\sum_{t=2}^{B}\sum_{k=j+1}^{t+j-1}\epsilon_k^2=\sum_{t=2}^{B}\sum_{k=1}^{t-1}\epsilon_{k+j}^2=\sum_{t=1}^{B-1}(B-t)\epsilon_{t+j}^2$$

and by Lemma 3(d) and (e)

$$\begin{aligned}\sum_{j=1}^{T-B}\sum_{t=1}^{B-1}(B-t)\epsilon_{t+j}^2 &= B\sum_{j=1}^{T-B}\sum_{t=1}^{B-1}\epsilon_{t+j}^2 - \sum_{j=1}^{T-B}\sum_{t=1}^{B-1}t\epsilon_{t+j}^2 \\ &= B\sum_{t=B}^{T-B+1}(B-1)\epsilon_t^2 - \sum_{t=B}^{T-B+1}\frac{B(B-1)}{2}\epsilon_t^2 + o_P(B^2T) \\ &= \sum_{t=B}^{T-B+1}\frac{B(B-1)}{2}\epsilon_t^2 + o_P(B^2T).\end{aligned}$$

Hence, for the denominator of (19) we have

$$\frac{1}{B^2T\overline{\sigma}^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)^2 = \frac{1}{2T\overline{\sigma}^2}\sum_{t=B}^{T-B+1}\epsilon_t^2 + o_P(1) \xrightarrow{p} \frac{1}{2}. \qquad (22)$$

Now we consider the numerator of (19).

$$\begin{aligned}&\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j} = \sum_{j=1}^{T-B}\sum_{t=2}^{B}\epsilon_{t+j}\left(\sum_{k=1}^{t-1}\epsilon_{k+j}\right) \\ &= \sum_{t=2}^{B}\sum_{j=1}^{T-B}\epsilon_{t+j}\left(\sum_{k=1}^{t-1}\epsilon_{t+j-k}\right) = \sum_{t=2}^{B}\sum_{j=t+1}^{T-B+t}\epsilon_j\left(\sum_{k=1}^{t-1}\epsilon_{j-k}\right) \\ &= \sum_{t=2}^{B}\sum_{j=B+1}^{T-B}\epsilon_j\left(\sum_{k=1}^{t-1}\epsilon_{j-k}\right) + \sum_{t=2}^{B}\left(\sum_{j=t+1}^{B}\sum_{k=1}^{t-1}\epsilon_j\epsilon_{j-k} + \sum_{j=T-B+1}^{T-B+t}\sum_{k=1}^{t}\epsilon_j\epsilon_{j-k}\right) \\ &= \sum_{j=B+1}^{T-B}\epsilon_j\sum_{t=2}^{B}\sum_{k=1}^{t-1}\epsilon_{j-k} + \sum_{j=1}^{B}\sum_{k=1}^{j-1}M_{j,k}\epsilon_j\epsilon_{j-k} + \sum_{j=T-B+1}^{T}\sum_{k=1}^{B-1}M_{j,k}\epsilon_j\epsilon_{j-k},\end{aligned}$$

where the numbers $M_{j,k}$ satisfy $0 \leq M_{j,k} \leq B$. Analogously to (21), we have

$$\sum_{j=1}^{B}\sum_{k=1}^{j-1}M_{j,k}\epsilon_j\epsilon_{j-k} + \sum_{j=T-B+1}^{T}\sum_{k=1}^{B-1}M_{j,k}\epsilon_j\epsilon_{j-k} = o_P(B^{1.5}T^{0.5}).$$

Then by Lemma 3(f),

$$\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j} = \sum_{j=B+1}^{T-B}\epsilon_j\sum_{k=1}^{B-1}(B-k)\epsilon_{j-k} + o_P(B^{1.5}T^{0.5}).$$

Then the numerator of (19) is equal to

$$\frac{1}{B^{1.5}T^{0.5}\overline{\sigma}^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j}=\frac{V^{0.5}}{B^{0.5}\overline{\sigma}^2}\left(\frac{1}{\sqrt{VT}}\sum_{k=B+1}^{T-B}q_{k,T}\right)+o_P(1),$$

where $q_{t,T}$ is defined as in Lemma 4. From (9) we conclude that

$$V=\frac{1}{T-2B}\sum_{k=B+1}^{T-B}\frac{\sigma_k^4B}{3}+o(B)=\frac{B}{3}\int_0^1\sigma(r)^4dr+o(B)$$

and

$$\frac{V^{0.5}}{B^{0.5}\overline{\sigma}^2}=\frac{\sigma_*}{\sqrt{3}}+o(1), \tag{23}$$

where $\sigma_*=\sqrt{\int_0^1\sigma(r)^4dr}/(\int_0^1\sigma(r)^2dr)$. Finally, with (22), (23) and Lemma 4,

$$\sqrt{BT}(\hat{\rho}^{\text{sb}}-1)=\frac{\frac{1}{\sqrt{3}}\sigma_*\left(\frac{1}{\sqrt{VT}}\sum_{j=B+1}^{T-B}q_{k,T}\right)+o_P(1)}{\frac{1}{2}+o_P(1)}\xrightarrow{d}\mathcal{N}\left(0,\frac{4}{3}\sigma_*^2\right). \tag{24}$$

Note that under Assumption 1 we have $\sigma_*^2=1$.
For the pseudo t-statistic we have analogously to (19) that

$$\begin{aligned}\tau\text{-SB}&=\frac{\sigma}{\hat{\sigma}_{\text{b}}}\sqrt{\frac{3}{2}}\cdot\frac{B^{-1.5}T^{-0.5}\sigma^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1}-y_j)\Delta y_{t+j}}{\sqrt{B^{-2}T^{-1}\sigma^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1}-y_j)^2}}\\&=\frac{\sigma}{\hat{\sigma}_{\text{b}}}\sqrt{\frac{3}{2}}\cdot\frac{B^{-1.5}T^{-0.5}\sigma^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j}+o_P(1)}{\sqrt{B^{-2}T^{-1}\sigma^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)^2+o_P(1)}},\end{aligned}$$

where the denominator converges in probability to $1/\sqrt{2}$ (see (22)). The numerator converges in distribution to $\mathcal{N}(0,1/3)$ (see (24)) and $\sigma/\hat{\sigma}_{\text{b}}\xrightarrow{p}1$ (see Lemma 1). Thus, the assertion follows.

A.5 Proof of Remark 1

For $\alpha\leq 0.5$ and $B=\mathcal{O}(T^\gamma)$ we need to restrict γ in order to obtain expression (17) to converge to zero. Then $\mathcal{O}(T^{0.5-2\alpha}B^{\alpha-0.5})=\mathcal{O}(T^{0.5-2\alpha+\gamma(\alpha-0.5)})$ and $\mathcal{O}_P(B^{0.5}T^{-\alpha})=\mathcal{O}_P(T^{\gamma0.5-\alpha})$. From $0.5-2\alpha+\gamma(\alpha-0.5)<0$ and $0.5\gamma-\alpha<0$ it follows that $(0.5-2\alpha)(0.5-\alpha)<\gamma<2\alpha$ and therefore $\alpha>(3-\sqrt{5})/4$.

A.6 Proof of Theorem 3

From (19), (22), (23) and the Proof of Lemma 1 it follows that

$$\begin{aligned}
\tau\text{-SB} &= \frac{\bar{\sigma}}{\hat{\sigma}_{\mathrm{b}}}\sqrt{\frac{3}{2}} \cdot \frac{B^{-1.5}T^{-0.5}\bar{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j} + o_P(1)}{\sqrt{B^{-2}T^{-1}\bar{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)^2} + o_P(1)} \\
&= \frac{\bar{\sigma}}{\hat{\sigma}_{\mathrm{b}}}\sqrt{\frac{3}{2}} \cdot \frac{\frac{1}{\sqrt{3}}\sigma_* \left(\frac{1}{\sqrt{VT}}\sum_{j=B+1}^{T-B} q_{k,T}\right) + o_P(1)}{\frac{1}{\sqrt{2}} + o_P(1)} \\
&= \frac{\bar{\sigma}}{\hat{\sigma}_{\mathrm{b}}} \cdot \frac{\sigma_* \left(\frac{1}{\sqrt{VT}}\sum_{j=B+1}^{T-B} q_{k,T}\right) + o_P(1)}{1 + o_P(1)} \xrightarrow{d} \mathcal{N}\left(0, \sigma_*^2\right).
\end{aligned}$$

A.7 Proof of Lemma 2

From Lemma 1 we know that

$$\frac{1}{T-B}\sum_{j=1}^{T-B}\frac{1}{B-1}\sum_{t=1}^{B}\left(\hat{y}_{j+t}-\bar{\hat{y}}_j\right)^2 = \int_0^1 \sigma^2(r)dr + o_P(1).$$

Analogously to the proof of Lemma 1 we have

$$\begin{aligned}
&\frac{1}{T-B}\sum_{j=1}^{T-B}\left(\hat{y}_j-\bar{y}\right)^2\frac{1}{B-1}\sum_{t=1}^{B}\left(\hat{y}_{j+t}-\bar{\hat{y}}_j\right)^2 \\
&= \frac{1}{T-B}\sum_{j=1}^{T-B}\left(\epsilon_j-\bar{\epsilon}\right)^2\frac{1}{B-1}\sum_{t=1}^{B}\left(\epsilon_{j+t}-\bar{\epsilon}_j\right)^2 + o_P(1) \\
&= \frac{1}{T-B}\sum_{j=1}^{T-B}\sigma^4(j/T) + o_P(1) = \int_0^1 \sigma^4(r)dr + o_P(1).
\end{aligned}$$

The result follows from

$$\hat{\sigma}_*^2 = \frac{\frac{1}{T-B}\frac{1}{B-1}\sum_{j=1}^{T-B}\sum_{t=1}^{B}\left(\hat{y}_j-\bar{y}\right)^2\left(\hat{y}_{j+t}-\bar{\hat{y}}_j\right)^2}{\left(\frac{1}{T-B}\frac{1}{B-1}\sum_{j=1}^{T-B}\sum_{t=1}^{B}\left(\hat{y}_{j+t}-\bar{\hat{y}}_j\right)^2\right)^2} = \frac{\int_0^1\sigma^4(r)dr + o_P(1)}{\left(\int_0^1\sigma^2(r)dr\right)^2 + o_P(1)}.$$

A.8 Proof of Theorem 4

From 15 and 16, Lemma 1 and the FCLT $X_T \Rightarrow W_\eta$ it follows that

$$\begin{aligned}
\tau\text{-FB} &= \frac{\bar{\sigma}}{\hat{\sigma}_{\mathrm{b}}} \cdot \frac{T^{-2}\bar{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j} + o_P(1)}{\sqrt{t^{-3}\bar{\sigma}^{-2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)^2} + o_P(1)} \\
&= \frac{\bar{\sigma}}{\hat{\sigma}_{\mathrm{b}}} \cdot \frac{1/2\int_0^1 \left(X_T(b+r)-X_T(r)\right)^2 dr - b(1+b)/2 + o_P(1)}{\sqrt{\int_0^{1-b}\int_r^{b+r}\left(X_T(s)-X_T(r)\right)^2 dsdr} + o_P(1)} \\
&\xrightarrow{d} \frac{\int_0^1 \left(W_\eta(b+r)-W_\eta(r)\right)^2 dr - b(1+b)}{2\sqrt{\int_0^{1-b}\int_r^{b+r}\left(W_\eta(s)-W_\eta(r)\right)^2 dsdr}}.
\end{aligned}$$

A.9 Proof of Theorem 5

Let $\tilde{X}_T(r) = X_T(\hat{\eta}^{-1}(r))$ with $\tilde{X}_T \Rightarrow W$ on $D[0,1]$ (see Theorem 1 in Cavaliere and Taylor (2008)). Then analogously to Theorem 4

$$\begin{aligned}
\tau\text{-FB}^{HC} &= \frac{\bar{\sigma}}{\hat{\sigma}_{\mathrm{b}}} \cdot \frac{1/2\int_0^1 \left(\tilde{X}_T(b+r)-\tilde{X}_T(r)\right)^2 dr - b(1+b)/2 + o_P(1)}{\sqrt{\int_0^{1-b}\int_r^{b+r}\left(\tilde{X}_T(s)-\tilde{X}_T(r)\right)^2 dsdr} + o_P(1)} \\
&\xrightarrow{d} \frac{\int_0^1 \left(W(b+r)-W(r)\right)^2 dr - b(1+b)}{2\sqrt{\int_0^{1-b}\int_r^{b+r}\left(W(s)-W(r)\right)^2 dsdr}}.
\end{aligned}$$

The heteroskedasticity consistency result for the small-b statistic follows with Lemma 2.

A.10 Proof of Theorem 6

Let $Y_T(r) = \omega^{-1}T^{-0.5}\sum_{t=1}^{\lfloor rT \rfloor}\epsilon_t$, $0 \le r \le 1$ with $Y_T \Rightarrow W$ under Assumption 4. Then analogously to (14) we have

$$\frac{1}{T^2\omega^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\frac{\epsilon_{t+j}^2}{2} \xrightarrow{p} \frac{b(1-b)}{2}\frac{\gamma(0)}{\omega^2}.$$

Following (15) we obtain

$$\begin{aligned}
&\frac{1}{T^2\omega^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)\Delta u_{t+j} \\
&=\int_0^{1-b}\frac{(Y_T(b+r)-Y_T(r))^2}{2}dr-b(1+b)\frac{\kappa}{2}+o_P(1).
\end{aligned}$$

Furthermore

$$\frac{1}{T^3\omega^2}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(u_{t+j-1}-u_j)^2=\int_0^{1-b}\int_r^{b+r}(Y_T(s)-Y_T(r))^2dsdr+o_P(1)$$

which follows analogously to (16). Then following (13) we have

$$\begin{aligned}
T(\hat{\rho}^{\text{fb}}-1) &= \frac{\frac{1}{2}\int_0^{1-b}(Y_T(b+r)-Y_T(r))^2dr-b(1+b)\frac{\kappa}{2}+o_P(1)}{\int_0^{1-b}\int_r^{b+r}(Y_T(s)-Y_T(r))^2dsdr+o_P(1)} \\
&\xrightarrow{d}\frac{\frac{1}{2}\int_0^{1-b}\left(W(b+r)-W(r)\right)^2dr-b(1-b)\frac{\kappa}{2}}{\int_0^{1-b}\int_r^{b+r}(W(s)-W(r))^2dsdr}.
\end{aligned}$$

A.11 Additional Tables

In Tables A1 and A2 we present empirical size and power results for the fixed-b and small-b statistics with different bandwidth choices. For τ-FB and τ-FB^{HC} we consider $B = b \cdot T$ with $b = 0.1$ and $b = 0.5$. The heteroskedasticity robust small-b statistic τ-SB^{HC} is simulated for $B = T^{\gamma}$ with $\gamma \in \{0.4, 0.5, 0.6, 0.7, 0.8\}$.

Table A1: Size and power for $T = 100$ using 5% critical values

		τ-FB		τ-FB^{HC}				τ-SB^{HC}		
d_t	ρ	b=0.1	b=0.5	b=0.1	b=0.5	γ=0.4	γ=0.5	γ=0.6	γ=0.7	γ=0.8
	1	0.042	0.050	0.022	0.053	0.049	0.054	0.058	0.056	0.045
(zer)	0.9	0.256	0.424	0.076	0.244	0.232	0.310	0.372	0.422	0.386
	0.8	0.683	0.913	0.239	0.703	0.580	0.743	0.837	0.895	0.874
	1	0.040	0.045	0.022	0.047	0.044	0.050	0.052	0.050	0.040
(lin)	0.9	0.219	0.224	0.130	0.252	0.202	0.263	0.300	0.297	0.217
	0.8	0.632	0.505	0.455	0.396	0.536	0.689	0.771	0.774	0.572
	1	0.014	0.012	0.007	0.018	0.018	0.018	0.017	0.015	0.011
(fo1)	0.9	0.031	0.009	0.008	0.003	0.053	0.041	0.028	0.014	0.009
	0.8	0.086	0.001	0.019	0.000	0.183	0.116	0.043	0.005	0.001
	1	0.026	0.029	0.013	0.035	0.031	0.033	0.033	0.032	0.026
(fo2)	0.9	0.120	0.119	0.038	0.061	0.133	0.152	0.155	0.140	0.112
	0.8	0.374	0.180	0.137	0.086	0.386	0.441	0.411	0.285	0.176
	1	0.020	0.024	0.010	0.030	0.024	0.025	0.026	0.026	0.023
(fo3)	0.9	0.087	0.160	0.025	0.071	0.099	0.114	0.123	0.139	0.133
	0.8	0.266	0.386	0.070	0.172	0.302	0.329	0.314	0.292	0.294
	1	0.040	0.073	0.011	0.033	0.049	0.058	0.062	0.064	0.059
(bre)	0.9	0.127	0.038	0.064	0.012	0.144	0.174	0.179	0.142	0.071
	0.8	0.311	0.010	0.266	0.011	0.332	0.394	0.387	0.241	0.056
	1	0.035	0.038	0.020	0.042	0.041	0.044	0.045	0.043	0.034
(lo1)	0.9	0.169	0.060	0.129	0.050	0.169	0.206	0.211	0.155	0.074
	0.8	0.544	0.049	0.444	0.048	0.482	0.603	0.630	0.445	0.113
	1	0.035	0.039	0.018	0.044	0.041	0.045	0.046	0.043	0.034
(lo2)	0.9	0.177	0.151	0.059	0.077	0.175	0.215	0.228	0.200	0.141
	0.8	0.549	0.313	0.234	0.158	0.487	0.611	0.648	0.538	0.327

Table A2: Size and power for $T = 300$ using 5% critical values

		τ-FB		τ-FB^{HC}		τ-SB^{HC}				
d_t	ρ	b=0.1	b=0.5	b=0.1	b=0.5	γ=0.4	γ=0.5	γ=0.6	γ=0.7	γ=0.8
	1	0.047	0.049	0.048	0.082	0.047	0.054	0.057	0.056	0.047
(zer)	0.9	0.946	0.998	0.846	0.989	0.646	0.857	0.959	0.991	0.995
	0.8	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000
	1	0.047	0.048	0.056	0.109	0.047	0.052	0.055	0.056	0.047
(lin)	0.9	0.940	0.965	0.923	0.965	0.643	0.851	0.955	0.987	0.987
	0.8	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000
	1	0.032	0.031	0.033	0.061	0.036	0.039	0.038	0.036	0.030
(fo1)	0.9	0.780	0.188	0.676	0.124	0.556	0.742	0.820	0.659	0.243
	0.8	1.000	0.342	0.998	0.214	0.988	1.000	1.000	0.997	0.537
	1	0.039	0.040	0.039	0.071	0.042	0.044	0.047	0.046	0.038
(fo2)	0.9	0.889	0.804	0.788	0.714	0.607	0.809	0.913	0.928	0.825
	0.8	1.000	0.997	0.999	0.983	0.992	1.000	1.000	1.000	0.999
	1	0.038	0.040	0.038	0.072	0.040	0.043	0.045	0.044	0.039
(fo3)	0.9	0.857	0.937	0.737	0.816	0.580	0.781	0.887	0.902	0.872
	0.8	1.000	1.000	0.999	0.985	0.990	1.000	1.000	1.000	1.000
	1	0.047	0.061	0.039	0.084	0.047	0.053	0.058	0.058	0.051
(bre)	0.9	0.821	0.288	0.845	0.291	0.546	0.742	0.854	0.881	0.704
	0.8	1.000	0.538	0.999	0.578	0.973	0.998	1.000	1.000	0.987
	1	0.038	0.040	0.042	0.080	0.042	0.045	0.047	0.045	0.038
(lo1)	0.9	0.875	0.270	0.848	0.333	0.596	0.797	0.899	0.909	0.705
	0.8	1.000	0.550	1.000	0.667	0.991	1.000	1.000	1.000	0.993
	1	0.039	0.042	0.039	0.073	0.042	0.046	0.046	0.045	0.039
(lo2)	0.9	0.874	0.863	0.772	0.782	0.595	0.796	0.899	0.906	0.812
	0.8	1.000	1.000	0.999	0.992	0.991	1.000	1.000	1.000	0.999

Table A3: Size and power ignoring autocorrelation

d_t	ρ	τ-FB	τ-FB^{HC}	τ-SB	τ-SB^{HC}	DF$^{\mu}$	DF*	SP	EL
AR-process (6)									
(zer)	1	0.001	0.001	0.001	0.001	0.024	0.007	0.008	0.001
	0.9	0.266	0.170	0.299	0.311	0.671	0.287	0.810	0.051
(fo1)	1	0.000	0.001	0.000	0.000	0.027	0.007	0.005	0.000
	0.9	0.105	0.088	0.125	0.131	0.089	0.014	0.182	0.048
(lo1)	1	0.000	0.002	0.000	0.000	0.024	0.007	0.006	0.000
	0.9	0.172	0.199	0.202	0.211	0.078	0.165	0.234	0.038
MA-process (7)									
(zer)	1	0.000	0.001	0.000	0.000	0.023	0.006	0.007	0.000
	0.9	0.161	0.098	0.188	0.195	0.531	0.176	0.848	0.020
(fo1)	1	0.000	0.001	0.000	0.000	0.026	0.006	0.005	0.000
	0.9	0.084	0.065	0.099	0.104	0.126	0.021	0.294	0.021
(lo1)	1	0.000	0.000	0.000	0.000	0.022	0.007	0.005	0.000
	0.9	0.122	0.129	0.146	0.151	0.125	0.119	0.357	0.016

Note: Simulation results are reported for 10,000 replications with $T = 300$ using 5% critical values. The innovation process is simulated under the AR-specification $\epsilon_t = 0.3\epsilon_{t-1} + e_t$ and under the MA-specification $\epsilon_t = e_t + 0.8e_{t-1}$, where $e_t \sim \mathcal{N}(0, 1)$. For the specifications of the trend function d_t see Table 2 and Figure 1. The bandwidth for fixed-b is $B = T/10$ and for the small-b statistics $B = T^{0.6}$. DF$^{\mu}$ and DF* are the Dickey-Fuller tests with constant and linear trend specification, SP is the test by Schmidt and Phillips (1992) and EL is the test by Enders and Lee (2012).

Table A4: Size under autocorrelation with pre-whitening

	τ-FB	τ-FB^{HC}	τ-SB	τ-SB^{HC}	DF$^{\mu}$	DF*	SP	EL
AR-process (6)								
(zer)	0.021	0.023	0.024	0.025	0.050	0.035	0.052	0.015
(lin)	0.017	0.023	0.019	0.021	0.043	0.030	0.045	0.014
(fo1)	0.014	0.012	0.017	0.017	0.043	0.029	0.040	0.020
(fo2)	0.021	0.023	0.024	0.025	0.046	0.032	0.048	0.018
(fo3)	0.019	0.020	0.022	0.023	0.046	0.032	0.044	0.016
(bre)	0.020	0.017	0.024	0.026	0.045	0.031	0.050	0.017
(lo1)	0.020	0.022	0.023	0.024	0.043	0.032	0.043	0.017
(lo2)	0.017	0.018	0.019	0.020	0.042	0.030	0.042	0.015
MA-process (7)								
(zer)	0.035	0.032	0.040	0.042	0.058	0.046	0.064	0.024
(lin)	0.029	0.032	0.033	0.035	0.059	0.040	0.059	0.023
(fo1)	0.029	0.026	0.035	0.037	0.053	0.038	0.059	0.027
(fo2)	0.028	0.026	0.032	0.034	0.058	0.039	0.055	0.024
(fo3)	0.029	0.025	0.035	0.036	0.062	0.044	0.058	0.026
(bre)	0.028	0.017	0.032	0.031	0.048	0.036	0.052	0.022
(lo1)	0.030	0.029	0.035	0.037	0.053	0.041	0.060	0.023
(lo2)	0.030	0.025	0.035	0.037	0.054	0.039	0.060	0.024

Note: Simulation results are reported for 10,000 replications with $T = 300$ using 5% critical values. The innovation process is simulated under the AR-specification $\epsilon_t = 0.3\epsilon_{t-1} + e_t$ and under the MA-specification $\epsilon_t = e_t + 0.8e_{t-1}$, where $e_t \sim \mathcal{N}(0, 1)$. For the specifications of the trend function d_t see Table 2 and Figure 1. The bandwidth for fixed-b is $B = T/10$ and for the small-b statistics $B = T^{0.6}$. DF$^{\mu}$ and DF* are the Dickey-Fuller tests with constant and linear trend specification, SP is the test by Schmidt and Phillips (1992) and EL is the test by Enders and Lee (2012). The simulated series is pre-whitened using Schwert's rule.

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Bierens, H. J. (1997). Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the us price level and interest rate. *Journal of Econometrics*, 81(1):29–64.

Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews*, 23(3):259–292.

Cavaliere, G. and Taylor, A. R. (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, 140(2):919–947.

Cavaliere, G. and Taylor, A. R. (2008). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis*, 29(2):300–330.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. Oxford University Press, UK.

Davidson, J. (2002). Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics*, 106(2):243–269.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Enders, W. and Lee, J. (2012). A unit root test using a fourier series to approximate smooth breaks. *Oxford Bulletin of Economics and Statistics*, 74(4):574–599.

Evans, M. D. and Lewis, K. K. (1995). Do expected shifts in inflation affect estimates of the long-run fisher relation? *The Journal of Finance*, 50(1):225–253.

Gao, J. and Robinson, P. M. (2016). Inference on nonstationary time series with moving mean. *Econometric Theory*, 32(02):431–457.

Hamori, S. and Tokihisa, A. (1997). Testing for a unit root in the presence of a variance shift. *Economics Letters*, 57(3):245–253.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, 13(1):37–45.

Kapetanios, G., Shin, Y., and Snell, A. (2003). Testing for a unit root in the nonlinear star framework. *Journal of Econometrics*, 112(2):359–379.

Lee, H.-Y. and Wu, J.-L. (2001). Mean reversion of inflation rates: Evidence from 13 oecd countries. *Journal of Macroeconomics*, 23(3):477–487.

Levin, A., Lin, C.-F., and Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1):1–24.

Leybourne, S., Newbold, P., and Vougas, D. (1998). Unit roots and smooth transitions. *Journal of Time Series Analysis*, 19(1):83–97.

MacDonald, R. and Murphy, P. D. (1989). Testing for the long run relationship between nominal interest rates and inflation using cointegration techniques. *Applied Economics*, 21(4):439–447.

Narayan, P. K. and Narayan, S. (2010). Is there a unit root in the inflation rate? new evidence from panel data models with multiple structural breaks. *Applied economics*, 42(13):1661–1670.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6):1361–1401.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica*, 55(2):277–301.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Rappoport, P. and Reichlin, L. (1989). Segmented trends and non-stationary time series. *The Economic Journal*, 99(395):168–177.

Rose, A. K. (1988). Is the real interest rate stable? *The Journal of Finance*, 43(5):1095–1112.

Schmidt, P. and Phillips, P. C. (1992). Lm tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics*, 54(3):257–287.

Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 7(2):147–159.