Is U.S. real output growth really non-normal? Testing distributional assumptions in time-varying location-scale models*

Matei Demetrescu† TU Dortmund

Robinson Kruse-Becher‡ University of Hagen

September 29, 2023

Abstract

Testing distributional assumptions is an evergreen topic in econometrics. A key assumption is stationarity. Yet, under time-varying moments, the marginal distribution belongs to a mixture family. Therefore, tests consistently reject when stationarity assumptions are violated, even under correct specification of the baseline distribution. However, time-varying moments are common in economic data. We propose robust tests by means of local standardization. We demonstrate our approach in detail for normality, while our main results are extended to general location-scale models without essential modifications. Probability integral transforms accommodate a wide range of null distributions and imply simple raw moment restrictions. We use raw moments of probability integral transformations of locally standardized series (by flexible nonparametric estimators). Short-run dynamics are accounted for by the fixed-bandwidth approach which leads to robustness of the proposed test statistics to the estimation error induced by the local standardization. We propose a simple rule for choosing the tuning parameters and an effective finite-sample adjustment. Monte Carlo experiments show that the new tests perform well in terms of size and power and outperform alternative tests even under stationarity. We find - in contrast to other studies - no evidence against normality of U.S. real output growth after accounting for time-variation.

Key words: Distribution testing; Probability integral transformation; Local standardization; Nonparametric estimation; Heteroskedasticity and autocorrelation robust inference.

JEL classification: C12; C14; C22; E01; E32

1 Introduction

Testing distributional assumptions is an important aspect of applied work. In particular, normal distributions are often assumed in economics, and not only for simplicity or their analytical properties: for instance, the central limit theorem provides a justification for normally distributed

*The authors would like to thank Karim Abadir, Guiseppe Cavaliere, Juan Dolado, Peter Reinhard Hansen, Mehdi Hosseinkouchack, Sanne Kruse-Becher, Dominik Liebl, Andre Lucas, Nour Meddahi, Philipp Sibbertsen, George Tauchen, Robert Taylor, Dimitrios Thomakos and Michael Vogt for very helpful comments. We are grateful to Yuze Liu for providing excellent research assistance. Part of this research was carried out while the second author was visiting the European University Institute in Florence, whose kind hospitality is gratefully acknowledged.

†TU Dortmund, Faculty of Statistics, Germany. E-mail address: `mdeme@statistik.tu-dortmund.de`.

‡**Corresponding author:** University of Hagen, Faculty of Business Administration and Economics and Center for Economic and Statistical Analysis (CESA), Germany. E-mail address: `robinson.kruse-becher@fernuni-hagen.de`.

stochastic model components, which is of particular interest when working with macroeconomic aggregates. In many dynamic macroeconometric models, distributional assumptions are key in the modelling cycle of specification, estimation, inference and prediction.[1] The popular Kolmogorov-Smirnov statistic can be used to test hypotheses on distributions in an iid sampling situation only, and is not straightforwardly extended to serial dependence and parameter estimation error.[2] Bai (2003) employs the martingale transformation of Khmaladze (1981) to this end. In spite of its advantages, such an approach is quite demanding, though, and Bai and Ng (2005) decisively extend the work of Jarque and Bera (1980) to propose moment-based tests of normality under serial dependence; see also Lomnicki (1961) for an early discussion for linear processes or Bontemps and Meddahi (2005) for an ingenious choice of moment restrictions. While Bai and Ng (2005) address normality testing explicitly, moment-based testing can be extended to test other distributions as well.

But serial dependence and estimation uncertainty are not the only issues to be faced in applied work. Consider for instance the situation where a time series is marginally normal, but exhibits one (or more) shifts in the mean or the variance. The "pooled" distribution is mixed Gaussian with two (or more) components, which is not normal anymore; depending on the mixture parameters, the resulting distribution may exhibit skewness, excess kurtosis or even multimodality. Therefore, a normality test ignoring mean and variance changes tends to reject the true null hypothesis essentially more often than required by the nominal significance level. The reasoning extends to general patterns of time-variation in mean and variance, and also to other families of distributions.

Such implications of time-variability are not hypothetical in nature. Rather, economic data are in fact often found to exhibit time-varying moments, both in mean and variance. Both, non-constant means and non-constant variances have been widely documented. Empirical evidence for structural changes in trend in U.S. real gross domestic product [GDP] are provided in Perron and Wada (2009) and Luo and Startz (2014), and so are mean shifts in the corresponding output growth rate; see e.g. Morley and Piger (2012) and Antolin-Diaz et al. (2017). Such time-variation in the mean of GDP growth rates is typically associated with recessions and oil crises. Time-varying volatility can be found in economic growth and also in other macroeconomic variables such as inflation measures (see e.g. Stock and Watson, 2002; Sensier and van Dijk, 2004; Justiniano and Primiceri, 2008). Typical patterns are permanent changes or slowly evolving trends in the variance (with the "Great Moderation" as an example for a downward change; see e.g. McConnell and Perez-Quiros, 2000, Blanchard and Simon, 2001, Ramey and Vine, 2006 and Gadea et al., 2018).

Such time-varying means and variances may conveniently be captured by Markov switching models, case already made for the U.S. real output growth by Hamilton (1989). More generally, Markov switching models gained popularity in the field of identification of structural macroeconomic shocks involving output; see e.g. Lanne et al. (2010) and Herwartz and Lütkepohl (2014). Furthermore, time-heteroskedasticity is quite useful for identifying structural shocks; see

[1] In the same vein, non-normality of disturbances might be taken to indicate model misspecification in regression analysis; in duration models, deviations from the exponential distribution may again indicate misspecification.

[2] It has been known at least since Durbin (1973) that plugging in estimated parameters may have an effect on the relevant limiting null distributions; see also Andrews (1997) for a conditional Kolmogorov-Smirnov test in a parametric regression setup with independent observations.

e.g. Rigobon (2003), Lanne and Lütkepohl (2008), Lewis (2017) and Lütkepohl et al. (2021). In this context, the normal distribution plays an important role as elaborated by Lanne et al. (2017) and Gouriéroux et al. (2020).

Allowing for time-varying means and variances when testing for normality, say, is therefore of undeniable importance in practice. To account for possible time-variation of unknown shape in the location and the scale of the variable of interest, we introduce here tests based on *local* standardization using means and variances estimated in a flexible nonparametric fashion. A nonparametric approach allows for various forms and types of changes in the mean or the variance by using rolling windows of observations for standardization rather than the full sample. The distribution tests build on moments of probability integral transformations [PIT]s of the locally standardized time series. PITs have already been used successfully by Diebold et al. (1998) in the context of forecast density evaluation.[3] The use of raw moments as suggested in Knüppel (2015) is advantageous as they are almost unbiased even in smaller samples, while centered moments like sample skewness or kurtosis can be severely biased in the presence of serial dependence (Bao, 2013), leading to distorted moment-based tests for normality.

To assess the statistical significance of the deviations of the empirical raw PIT moments from the theoretical ones we follow Bai and Ng (2005) and rely on long-run variance estimation for the PITs of the locally standardized series. We however adopt here the fixed-b asymptotic framework of Kiefer and Vogelsang (2005), and do so for two reasons. First, fixed-b asymptotics provide more accurate finite-sample inference by holding the required bandwidth fixed as a linear fraction of the sample size T, i.e. $B = bT$ with $b \in (0, 1]$. This way, the resulting asymptotic limiting null distributions of test statistics (like t, Wald and F) reflect the choice of bandwidth B and kernel, say κ, even as $T \to \infty$.[4] Indeed, the fixed-b approach has been successfully employed by Wang and Sun (2022) to test the null of zero autocorrelation. Perhaps more importantly in the economy of this paper, the use of the fixed-b framework has a second key motivation beyond possible finite-sample improvements. Namely, due to the nature of fixed-b asymptotics relying on partial sums, the cumulated estimation effect cancels out in the limiting distributions of our proposed test statistics, as opposed to the usual small-b asymptotics. This is quite useful in practice as only one set of fixed-b critical values are then required.

Concretely, we show here that the mean and variance functions may be estimated using non-parametric methods without altering the limiting distributions of the fixed-b test statistics based on locally standardized series (under certain conditions on the involved window widths). This asymptotic result implies that practitioners do not need to specify a particular model for the mean and the variance explicitly. Since the tests are applicable just as well when the mean and variance functions are known as when they are locally estimated, this leads to a straightforward implementation of the discussed procedures. Furthermore, we propose a simple selection rule for the window width required for local standardization and discuss a simple yet effective finite-sample correction. We provide extensive Monte Carlo evidence that the performance of the proposed tests does not hinge on a particular choice of kernel for nonparametric estimators.

[3] See also Rossi and Sekhposyan (2014) for a discussion of the normality assumption in the context of predictive density evaluation for U.S. output growth.

[4] See Yang and Vogelsang (2011), Vogelsang and Wagner (2013) or Sun (2014a,b) for contributions to this field, inter alia.

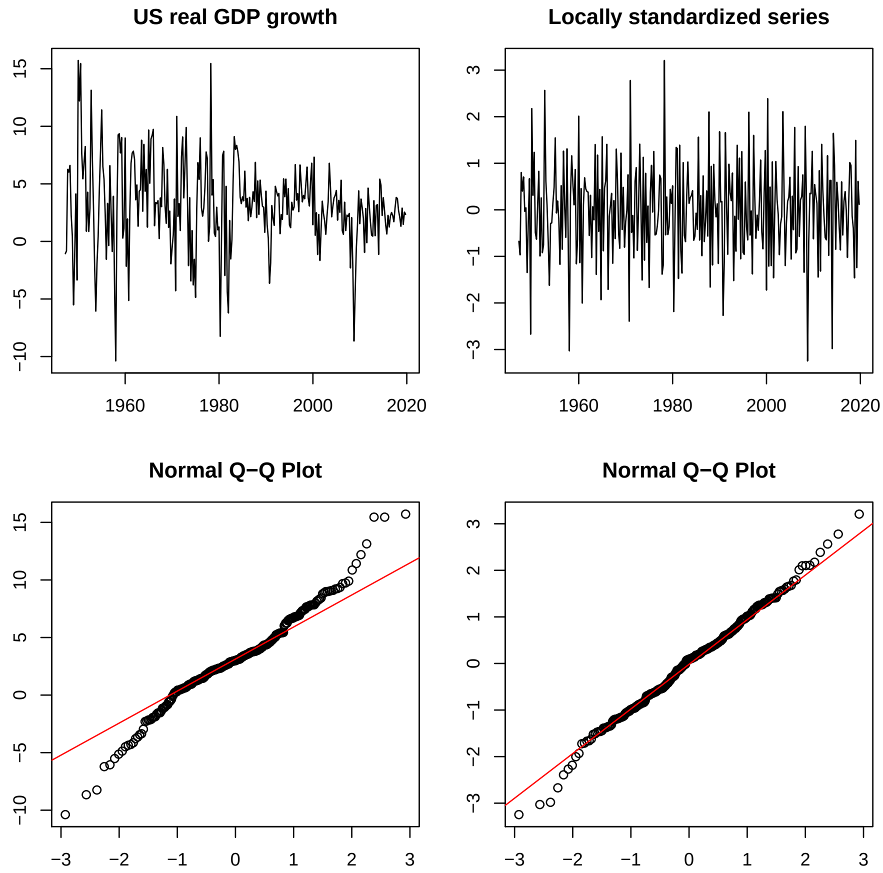

Figure 1: U.S. real GDP growth rate (1947Q2-2019Q4). Top left: time series plot of the raw time series x_t; top right: nonparametric locally standardized time series ($\hat{z}_t$); bottom left: normal QQ-plot for x_t and bottom right: normal QQ-plot for $\hat{z}_t$.

Building on the proposed robust tests, we then investigate the non-normality of U.S. real output growth series. In the context and motivation of macroeconomic tail risks (see e.g. Acemoglu et al., 2017), deviations from normality due to large economic movements during booms and recessions play a key role. This hypothetical empirical regularity also serves as the basis for a behavioral macroeconomic model as in De Grauwe (2012). Fagiolo et al. (2008) acknowledge the presence of conditional heteroskedasticity in real output growth rates, but the slow variation of unconditional variances due to the Great Moderation and the Great Recession is overlooked; see also Acemoglu et al. (2017). These major and long-run changes in volatility affect extant distributional tests like the Bai and Ng (2005) test in a severe way, as our theoretical framework and simulation results show.

In contrast to the popular view that U.S. real output growth rates are non-Gaussian, our results suggest that the opposite is actually the case. The newly proposed tests, which account for time-varying means and variances, do not reject the null of Gaussianity, while the Bai and Ng (2005) test rejects – quite likely specifically due to neglected time-variation of location and scale. This finding is summarized by the visualization in Figure 1. In a 2×2 plot, we show U.S. real GDP growth rate (left), the locally standardized time series (right) together with their respective quantile-quantile (QQ) plots. While the raw real output growth rate seems non-normal (mainly due to its striking tail behaviour), the normal distribution provides a much better approximation of the distribution of the locally standardized series.

The remainder of the paper is structured as follows. In Section 2, the setup is described and newly proposed test statistics are introduced for the important particular case of normality. The case of uncertainty induced by nonparametrically estimated standardization is located in Section 3, followed by the extension to other distributions. Our Monte Carlo simulation study is included in Section 4. Section 5 discusses the normality hypothesis of U.S. real output growth rates from the perspective of possible time-variation of location and scale. Section 6 summarizes, while additional results, technical proofs, response curves for critical values and a summary of the popular Bai and Ng (2005) test procedure are given in the Appendix.

2 Model and baseline test

To fix ideas and notation, we first describe the proposed procedure for the null hypothesis of normality and assuming that the mean and variance functions of the time series of interest are known. Section 3 discusses the feasible version with nonparametric local standardization together with the application to other null distributions besides the normal.

2.1 Technical assumptions

As the leading example, the null hypothesis of interest is marginal normality of the time series x_t under consideration. In the following, x_t exhibits time-varying mean and variance behavior as given by the following simple component model

$$x_t = \mu_t + \sigma_t z_t, \qquad t = 1, 2, \ldots, T,$$

where z_t is unconditionally homoskedastic and otherwise short-range dependent, while the time-varying mean and variance are allowed to have triangular array structures, $\mu_t = \mu_{t,T}$ and $\sigma_t = \sigma_{t,T}$, allowing e.g. for slowly varying functions of time. The following assumptions make the notions of short-run dependence and time-variation precise.

Assumption 1 *Let z_t be a marginally standard normal, strictly stationary time series with strong mixing and mixing coefficients $\alpha(j)$ for which $\alpha(j) < Aj^{-b}$ for some $b > 10/3$. Assume that the long-run covariance matrix of z_t and z_t^2 is positive definite. Finally, assume absolutely summable 8th-order cumulants of z_t.*

Together with regularity conditions on the mean and variance functions μ_t and σ_t^2 in Assumption 2, Assumption 1 specifies a particular case of a so-called *locally stationary* process; see Dahlhaus (2012) for a general introduction. Such models are well-suited to capture stylized facts of economic time series such as time-varying volatility. The strong mixing condition is a standard way of controlling for the persistence of stochastic processes and ensures z_t to have short memory. Given unit variance, z_t is integrated of order zero and σ_t^2 is interpreted as the *local* variance. The mixing coefficients $\alpha(j)$ are only mildly restricted, given that normality of z_t ensures finiteness of moments of any order and the typical trade-off between serial dependence and finiteness of higher-order moments is only present via the cumulant summability condition.[5] The contemporaneous covariance of z_t and z_t^2 is zero under marginal normality, but the long-run covariance need not be zero due to serial dependence in z_t. The condition also allows for mild forms of conditional heteroskedasticity, so the observed time series x_t may exhibit both conditional and unconditional heteroskedasticity.

For later reference, note that Assumption 1 ensures e.g. weak convergence of the suitably normalized partial sums of z_t and z_t^2,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\begin{pmatrix} z_t \\ z_t^2-1 \end{pmatrix} \Rightarrow \begin{pmatrix} W_1(s) \\ W_2(s) \end{pmatrix}, \tag{1}$$

where $(W_1, W_2)'$ is a bivariate Brownian motion and "$\Rightarrow$" stands for weak convergence in a space of càdlàg functions endowed with a suitable norm (see e.g. Davidson, 1994, Chapter 29). Again, due to the fact that z_t is only marginally standard normal, the long-run covariance matrix of z_t and z_t^2 is not necessarily diagonal (unlike their covariance matrix), which reflects on the dependence of W_1 and W_2.

Strict stationarity is a more restrictive condition than needed for the convergence in (1), for which weak stationarity would have sufficed in addition to the mixing property and uniform boundedness of higher-order moments. We however consider nonlinear transformations of z_t for our tests, and strict stationarity of z_t ensures that the *transformed* time series have constant variance. Moreover, strict stationarity is a reasonable assumption once the time-varying mean and variance have been accounted for. Finally, strict stationarity of z_t also separates the variance fluctuations from the serial dependence properties. The mean and variance functions themselves are taken to satisfy smoothness conditions:

Assumption 2 *The triangular arrays $\mu_{t,T}$ and $\sigma_{t,T}$ are given as $\mu_{t,T} = \mu(t/T)$ and $\sigma_{t,T} = \sigma(t/T)$, where both $\mu(\cdot)$ and $\sigma(\cdot)$ are Lipschitz-continuous on $[0,1]$, and $\sigma(\cdot)$ is bounded away from zero on $[0,1]$. Let $\sigma''(\cdot)$ exist and be bounded on $[0,1]$.*

The Assumption allows e.g. for trends in the mean and the variance of various shapes and directions, but also for cyclic behavior, provided the frequency of the cycle is not too high. In particular, Assumption 2 accommodates phenomena like the Great Moderation with changes occurring over several subperiods. Importantly, we make no parametric assumptions on the nature of these changes in mean and variance, such that our distribution tests are robust against generic time-variation in location and scale. Formally allowing for sudden breaks (captured by a discontinuity in the mean or the variance function) requires more involved proofs and is left for further research. We show in the analysis of the finite-sample properties of our tests that abrupt breaks barely deteriorate their performance; see Section 4 for details.

[5]This cumulant condition reinforces the short memory requirement expressed by the exponential decay of the mixing coefficients, but in a manner specific to z_t; we impose it to expedite some technical arguments leading to the uniform convergence of the local standardization procedure, and could be relaxed at the expense of lengthening the proofs.

2.2 Infeasible PIT-based tests

We base our test of the null hypothesis on moments of transformed time series rather than the original time series z_t, more precisely on the probability integral transform of z_t. We shall motivate our procedures by first assuming the mean and variance components to be known. This will be relaxed in Section 3, where we consider flexible nonparametric estimation of μ_t and σ_t. Concretely, with Φ being the cumulative distribution function (and φ denoting the density function) of the standard normal distribution, the PIT $p_t = \Phi(z_t)$ is marginally uniform on $[0,1]$ under the null. It then holds under the null of uniformly distributed PITs that

$$\mathrm{E}\left(p_t^k\right) = \frac{1}{k+1}\,; \qquad k \in \mathbb{N} \tag{2}$$

such that, jointly with the convergence in (1),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left(p_t - \frac{1}{2}, \cdots, p_t^K - \frac{1}{K+1}\right)' \Rightarrow \left(B_1(s), \cdots, B_K(s)\right)' \tag{3}$$

under Assumption 1, where $\boldsymbol{B}(s) = (B_1(s), \ldots, B_K(s))'$ is a K-variate Brownian motion with covariance matrix denoted by $\boldsymbol{\Omega} = \mathrm{E}\left(\boldsymbol{B}(1)\,\boldsymbol{B}'(1)\right)$ which is also the long-run covariance matrix of $\boldsymbol{p}_t = (p_t, p_t^2, \ldots, p_t^K)'$. One may resort to an estimate thereof (based on the usual spectral density based approach; see Newey and West, 1987; Andrews, 1991; Andrews and Monahan, 1992) to build Wald test statistics of the moment restrictions in (2), so it is not required to know $\boldsymbol{\Omega}$. This follows the approach of Bai and Ng (2005) or Bontemps and Meddahi (2005) to deal with serial dependence of unknown form.

Compared with relying on z_t directly, PITs have several advantages. For instance, PITs are bounded such that their higher-order cumulants are smaller than those of the standard normal; therefore, the variability of the long-run covariance matrix estimators is smaller and the χ^2 asymptotic approximation is typically more accurate. The bias of the sample moments of PITs is also typically smaller than those of the untransformed time series; see e.g. Knüppel (2015) for evidence in this respect. At the same time, PITs still allow to distinguish between skewness and kurtosis as causes of non-normality: since the cumulative distribution function of the standard normal is symmetric about the point $x = 0, y = 0.5$, the first raw moment of the PITs captures distributional asymmetry, but not skewness alone. So a rejection of the null which is not driven by the first raw moment is clearly not due to skewness; see Knüppel (2015) again.

Suppose for now that the test can be based directly on empirical moments of p_t (i.e. under known functions μ_t and σ_t). With $m_k = \frac{1}{T}\sum_{t=1}^{T} p_t^k$, a simple t-statistic for a single restriction on the k-th moment is given by

$$t_k = \sqrt{T}\,\frac{m_k - \frac{1}{k+1}}{\hat{\omega}_k} \tag{4}$$

with ω_k^2 being the k-th diagonal element of $\boldsymbol{\Omega}$ (i.e. the long-run variance of p_t^k). Concretely, let

$$\hat{\boldsymbol{\Omega}} = \sum_{j=-T+1}^{T-1} \kappa\left(\frac{j}{B}\right) \hat{\boldsymbol{\Gamma}}_j \tag{5}$$

denote an estimator of $\mathbf{\Omega}$ with proportional bandwidth $B = bT$, $b > 0$, where $\hat{\mathbf{\Gamma}}_j$ is the usual autocovariance matrix estimator at lag j,

$$\hat{\mathbf{\Gamma}}_j = \frac{1}{T} \sum_{t=j+1}^{T} \left(\boldsymbol{p}_t - \bar{\boldsymbol{p}}\right) \left(\boldsymbol{p}_{t-j} - \bar{\boldsymbol{p}}\right)', \ j \geq 0, \ \text{and } \hat{\mathbf{\Gamma}}_j = \hat{\mathbf{\Gamma}}_{-|j|}, \ j < 0, \tag{6}$$

with $\boldsymbol{p}_t = (p_t, p_t^2, \dots, p_t^K)'$. Given the weak convergence in (3) with assumed positive definite $\mathbf{\Omega}$, we have from Kiefer and Vogelsang (2005) the following limit result for $b \in (0, 1]$:

$$t_k^2 \Rightarrow \frac{\tilde{W}^2(1)}{\mathcal{Q}_{b,\kappa}},$$

where $\tilde{W}$ represents a standard Wiener process and the functional $\mathcal{Q}_{b,\kappa}$ is given in terms of the Brownian bridge $\tilde{W}(s) - s\tilde{W}(1)$ and depends explicitly on the choice of kernel κ and bandwidth B. For simplicity we work with the two most popular kernels in applied time series analysis, a) the quadratic spectral [QS] kernel of Andrews (1991) with $\kappa(s) = \frac{25}{12\pi^2 s^2}\left(\frac{\sin(6\pi s/5)}{6\pi s/5} - \cos(6\pi s/5)\right)$ and b) the Bartlett kernel $\kappa(s) = (1 - |s|)\,\mathbf{1}\,(|s| \leq 1)$ with $\mathbf{1}(\cdot)$ being the indicator function. For kernels with smooth second-order derivative (e.g. the QS kernel) it holds under fixed-b that

$$\mathcal{Q}_{b,\kappa} = -\int_0^1 \int_0^1 \frac{1}{b^2}\kappa''\left(\frac{r-s}{b}\right)\left(\tilde{W}(r) - r\tilde{W}(1)\right)\left(\tilde{W}(s) - s\tilde{W}(1)\right) \mathrm{d}r\mathrm{d}s \ ,$$

while, for the Bartlett kernel,

$$\mathcal{Q}_{b,\kappa} = \frac{2}{b}\int_0^1 \left(\tilde{W}(r) - r\tilde{W}(1)\right)^2 \mathrm{d}r - \frac{2}{b}\int_0^{1-b}\left(\tilde{W}(r+b) - (r+b)\tilde{W}(1)\right)\left(W(r) - rW(1)\right) \mathrm{d}r.$$

For both kernels, the standard limit result, i.e. $t_k^2 \Rightarrow \chi_1^2$, is recovered when $b \to 0$ at a suitable rate ($\mathcal{Q}_{b,\kappa} \stackrel{d}{\to} 1$ for $b \to 0$; c.f. Kiefer and Vogelsang, 2005). It should be noted that such HAR covariance matrix estimation does not work with a different estimator than HAC, but rather provides an asymptotic framework which results in more accurate asymptotic approximations.

Working with several raw moments (in effect a portmanteau test), we construct

$$\mathcal{T}_K = T\left(m_1 - \frac{1}{2}, \dots, m_K - \frac{1}{K+1}\right)\hat{\mathbf{\Omega}}^{-1}\left(m_1 - \frac{1}{2}, \dots, m_K - \frac{1}{K+1}\right)'. \tag{7}$$

Under the null, it follows similarly that

$$\mathcal{T}_K \Rightarrow \tilde{\boldsymbol{W}}_K'(1)\boldsymbol{\mathcal{Q}}_{K,b,\kappa}^{-1}\tilde{\boldsymbol{W}}_K(1),$$

where $\tilde{\boldsymbol{W}}_K(s)$ represents a K-dimensional vector of independent standard Wiener processes, $\boldsymbol{\mathcal{Q}}_{K,b,\kappa}$ is the $K \times K$ matrix variant of the above functionals relying on the Brownian bridges $\tilde{\boldsymbol{W}}_K(s) - s\tilde{\boldsymbol{W}}_K(1)$; see Kiefer and Vogelsang (2005) for details.[6]

To take advantage of the properties of the PITs based test it is however imperative to correctly

[6]It should be noted that the standard implementations for long-run variance estimators build on sample-demeaned series. In a testing context like ours, one may center p_t^k at the theoretical expectation $1/(k+1)$; this would however change the limiting distributions.

standardize the time series prior to applying the PIT, so the unknown μ_t and σ_t have to be estimated. We address this issue in the following Section.

3 Dealing with estimation uncertainty

3.1 Local standardization

Under the particular circumstance that the location and scale functions of time series under test are known, the tests can be applied directly. In most cases, however, these functions are unknown and need to be estimated. An exception would be for instance the case of density forecast evaluation; see e.g. Diebold et al. (1998) and Knüppel, 2015. Let $\hat{z}_t$ denote the locally standardized time series and

$$\hat{p}_t = \Phi\left(\hat{z}_t\right) = \Phi\left(\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}\right) \tag{8}$$

denote the PIT applied to $\hat{z}_t$. In (8), $\hat{\mu}_t$ and $\hat{\sigma}_t$ are generic estimators of the time-varying location and scale functions μ_t and σ_t. Let then $\hat{m}_k = \frac{1}{T}\sum_{t=1}^{T}\hat{p}_t^k$denote the sample average of $\hat{p}_t^k$.

The use of $\hat{p}_t$ instead of p_t for computing a feasible statistic, say $\hat{t}_k$, typically affects the limiting distributions and requires corrections. This is known in the literature as the Durbin problem; see Durbin (1973). In previous work, Bai and Ng (2005) show how to robustify against estimating (constant) mean and variance, while Bontemps and Meddahi (2012) derive conditions under which more general parametric standardization does not affect the limiting distribution. Alternatively, Bai (2003) uses the Khmaladze transform to tackle this issue. These approaches have, however, only been discussed under stationarity assumptions, and it is not clear how they are applicable under our conditions – or whether this is possible at all.

To account for time-variation, we employ a local standardization to match the local stationarity properties of the model. This requires a localized estimation of the unknown functions $\mu(\cdot)$ and $\sigma(\cdot)$, on which we now become more precise. Consider the Nadaraya-Watson estimator for $\mu(\cdot)$ and $\sigma(\cdot)$, i.e. the local constant regressions of x_t and $(x_t - \mathrm{E}(x_t))^2$ on the relative time t/T:

$$\hat{\mu}\left(\frac{t}{T}\right) = \hat{\mu}_t = \frac{\sum_{j=1}^{T} K\left(\frac{t/T - j/T}{h}\right) x_j}{\sum_{j=1}^{T} K\left(\frac{t/T - j/T}{h}\right)}$$

and

$$\hat{\sigma}^2\left(\frac{t}{T}\right) = \hat{\sigma}_t = \frac{\sum_{j=1}^{T} K\left(\frac{t/T - j/T}{h}\right)\left(x_j - \hat{\mu}_j\right)^2}{\sum_{j=1}^{T} K\left(\frac{t/T - j/T}{h}\right)}$$

where we assume for simplicity that $x_{0,-1,\ldots}$ and $x_{T+1,\ldots}$ are observed.[7]

These estimators are actually quite familiar in "classical" time series analysis: using the uniform kernel, $K(s) = 1/2\,\mathbf{1}(|s| \leq 1)$, we obtain the more familiar centered moving averages

$$\hat{\mu}_t = \frac{1}{2\tau + 1}\sum_{j=t-\tau}^{t+\tau} x_j \quad \text{and} \quad \hat{\sigma}_t^2 = \frac{1}{2\tau + 1}\sum_{j=t-\tau}^{t+\tau}\left(x_j - \hat{\mu}_j\right)^2,$$

[7] In the Nadaraya-Watson regression framework, we effectively treat the relative time t/T as a fixed-design regressor, as for instance Vogt (2012) who considers nonparametric multivariate regressions with time as one of the regressors.

where the window width τ is obtained from the bandwidth h by multiplication with T. The window width τ is substantially smaller than T, hence ensuring that x_t is approximately standardized in finite samples, and letting $\tau \to \infty$ at a rate lower than T ensures that, asymptotically, each x_t is standardized correctly. In fact, we may allow for different window widths τ_μ and τ_σ to allow for a more flexible choice of these tuning parameters; see also Remark 1 below. This is simply local standardizing instead of global standardizing using the sample mean $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$ and the sample standard deviation $\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2}$, as would have been sufficient for the case of strict stationarity of x_t. The key step in analyzing the feasible statistic based on $\hat{p}_t$ is to note that the weak convergence in (3) is – conveniently – replaced by the following limiting behavior.

Lemma 1 *Let $\tau_\mu, \tau_\sigma, T \to \infty$ such that $\frac{T^{\nu_1}}{\tau_\varrho} + \frac{\tau_\varrho}{T^{\nu_2}} \to 0$ for $2/3 < \nu_1 < \nu_2 < 3/4$ and $\varrho = \{\mu, \sigma\}$. Then, under Assumptions 1 and 2 and the uniform kernel, we have jointly for all $k = 1, \ldots, K$ that*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left(\hat{p}_t^k - \frac{1}{k+1} \right) \Rightarrow B_k(s) - k\vartheta_{k-1} W_1(s) - \frac{k}{2}\varpi_{k-1} W_2(s)$$

with W_1 and W_2 from (1) and B_k from (3), $\vartheta_{k-1} = \mathrm{E}\left(p_t^{k-1}\varphi(z_t)\right)$ and $\varpi_{k-1} = \mathrm{E}\left(p_t^{k-1} z_t \varphi(z_t)\right)$.

Remark 1 *The choice of the window widths τ_μ and τ_σ is critical for the performance of the smoothers. Note that the imposed restrictions imply* undersmoothing. *This is explained by the nature of the desired result: while classical nonparametric regression focusses on minimizing the MSE of the estimated curve, we need to reduce estimation bias in $\hat{\mu}(\cdot)$ and $\hat{\sigma}^2(\cdot)$ to a minimum, since the effect of the bias on the partial sums of $\hat{p}_t$ cumulates in t. Such bias effectively induces undesired trends in the partial sums of $\hat{p}_t^k$. For weak convergence to Wiener processes it is however required that such trends are of negligible magnitude. As a consequence, if one were to use standard procedures for the selection of the window width τ (such as cross-validation), one would tend to select window widths (or equivalently bandwidths h in the local regression framework) that are quite likely* too large *for our testing situation. In the Monte Carlo study, we advocate a simple recommendation regarding the choice of the window width τ in practice, based on downscaling the cross-validated window width.*

Remark 2 *We state Lemma 1 for the case of the uniform kernel to simplify the proofs and to keep the notation simple. Other kernel choices, e.g. the popular Gaussian or Epanechnikov kernel, plausibly lead to analogous results, as would the use of the local linear (or polynomial) regression estimator. We provide extensive simulation evidence which demonstrates that the particular kernel choice and the employed estimator have minor impact on the size and power of the tests in various situations; see Section 4.*

Remark 3 *Note that $\vartheta_0 = \mathrm{E}(\varphi_t) = \int_{-\infty}^{\infty} \varphi^2(x)\,\mathrm{d}x = \frac{1}{2\sqrt{\pi}}$. Using power series expansions one may show that $\vartheta_1 = \frac{1}{4\sqrt{\pi}}$, but the higher-order expectations (for ϑ_k, $k \geq 2$) do not seem to have a closed-form expression. We simulate the expectations $\vartheta_{k-1} = \mathrm{E}(p_t^{k-1}\varphi(z_t))$ via Monte Carlo simulation for $k = 1, 2, 3, 4$ with 1,000,000 observations and 10,000 replications; the resulting* *values are as follows:* $\boldsymbol{\vartheta} = (0.2820948, 0.1410473, 0.0857805, 0.0581472)$. *The simulated values for* $k = 1$ *and* $k = 2$ *match their theoretical counterpart to seven exact digits. Therefore, we expect that the Monte Carlo simulation precision of the higher-order terms (i.e.* $k \geq 3$*) is high enough.*

Feasible test statistics are based on $\hat{p}_t$. For a single moment restriction, say on the kth power of p_t, we therefore consider

$$\hat{t}_k = \sqrt{T}\,\frac{\hat{m}_k - \frac{1}{k+1}}{\hat{\omega}_k} \tag{9}$$

and, for a portmanteau-type test considering several moments, we take

$$\hat{\mathcal{T}}_K = T\left(\hat{m}_1 - \frac{1}{2}, \dots, \hat{m}_K - \frac{1}{K+1}\right)\hat{\boldsymbol{\Omega}}^{-1}\left(\hat{m}_1 - \frac{1}{2}, \dots, \hat{m}_K - \frac{1}{K+1}\right)' \tag{10}$$

with $\hat{\boldsymbol{\Omega}}$ being here a feasible long-run covariance matrix estimator computed on the basis of $\hat{p}_t^k$ (i.e. with $\hat{p}_t^k$ replacing p_t^k in the expression from equations (5) and (6)) and $\hat{\omega}_k^2$ its kth diagonal element.

By Lemma 1 we have that the K normalized partial sums $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{p}_t^k - \frac{1}{k+1}\right)$ still converge weakly to a K-dimensional Brownian motion, albeit with a different long-run covariance matrix than $\boldsymbol{\Omega}$, namely $\boldsymbol{\Psi} = \mathbf{V}\boldsymbol{\Xi}\mathbf{V}'$ .Here, $\boldsymbol{\Xi}$ is the long-run covariance matrix of $\left(p_t, \dots, p_t^K, z_t, z_t^2 - 1\right)'$ (such that $\boldsymbol{\Omega}$ is just the upper $K \times K$ diagonal block of $\boldsymbol{\Xi}$), and

$$\mathbf{V} = \begin{pmatrix} \mathbf{I}_K; & \boldsymbol{\Upsilon}_K \end{pmatrix} \quad \text{with} \quad \boldsymbol{\Upsilon}_K = -\begin{pmatrix} \vartheta_0 & \cdots & K\vartheta_{K-1} \\ \frac{1}{2}\varpi_0 & \cdots & \frac{K}{2}\varpi_{K-1} \end{pmatrix}'. \tag{11}$$

The coefficients are given as $\vartheta_j = \mathrm{E}\left(p_t^j\varphi(z_t)\right)$ and $\varpi_j = \mathrm{E}\left(p_t^j z_t\varphi(z_t)\right)$. We shall bow impose an additional technical assumption.

Assumption 3 *Let* $\boldsymbol{\Xi}$ *be positive definite.*

We require Assumption 3 to ensure that the covariance matrix $\boldsymbol{\Psi}$ of the K Brownian motions $B_k(s) - k\vartheta_{k-1}W_1(s) - \frac{k}{2}\varpi_{k-1}W_2(s)$ is nonsingular. (We hold it for obvious that $\mathbf{V}$ is of full rank.) The positive definiteness of $\boldsymbol{\Xi}$ is in fact plausible given that p_t^k are different nonlinear transformations of z_t.

Since the fixed-b asymptotics leads to partial-sums asymptotics for both $\hat{m}_k$ and $\hat{\boldsymbol{\Omega}}$, where the relevant long-run (co)variance matrix $\boldsymbol{\Psi}$ simply cancels out in (12) and (10), Lemma 1 ultimately implies that no explicit correction for the estimation error is required. This is formalized in the following proposition.

Proposition 1 *Under the Assumptions of Lemma 1 together with Assumption 3, it holds that*

$$\hat{t}_k \Rightarrow \frac{\tilde{W}(1)}{\sqrt{\mathcal{Q}_{b,\kappa}}} \qquad \text{and} \qquad \hat{\mathcal{T}}_K \Rightarrow \tilde{\boldsymbol{W}}_K'(1)\boldsymbol{\mathcal{Q}}_{K,b,\kappa}^{-1}\tilde{\boldsymbol{W}}_K(1),$$

where $\tilde{\boldsymbol{W}}_K(s)$ *represents a* K*-dimensional vector of independent standard Wiener processes.*

Remark 4 *Positive definiteness of* $\mathbf{\Xi}$ *implies positive definiteness of* $\mathbf{\Psi}$ *which is required for the result (see the proof for details); one could directly require invertibility of* $\mathbf{\Psi}$*, but this would depend on the null of interest via* $\mathbf{V}$*, while restricting* $\mathbf{\Xi}$ *to be of full rank does not.*

Remark 5 *It is questionable whether small-b asymptotics (e.g. Newey and West, 1987) are worth pursuing. One reason not to do so is technical. Due to the local nature of the standardization, the convergence of* $\hat{p}_t$ *to* p_t *is quite slow compared to the parametric case where typically* $\hat{p}_t - p_t = O_p(T^{-1/2})$*. Showing that* $\hat{\mathbf{\Omega}}$ *converges in probability for* $b \to 0$ *is therefore more difficult for local standardization, even if nonparametric convergence rates of* $\hat{p}_t^k$ *could ultimately be accommodated. Since such convergence is not required for the fixed-b approach, this actually delivers a further argument in favor of resorting to fixed-b asymptotics. The second reason is rather of practical nature, namely that the finite-sample properties of test statistics based on small-b asymptotics is suboptimal compared to that of fixed-b approaches.*

Remark 6 *It should be noted that the partial sums of* $\hat{p}_t^k - 1/(k+1)$ *converge to Brownian motion under local standardization, but, under a global standardization, they converge to Brownian-bridge type processes. Under global, we understand here adjustments made under parametric specifications of the mean and the variance of the observed series. As a comparison, Appendix C provides a discussion of parametric standardization. Parametric approaches require models for the mean* μ_t *and the variance* σ_t^2 *which are prone to misspecification. Moreover, there are essential differences between the implications of the two approaches. While parametric adjustment typically leads to bridge-type processes (see Lemma A.2 in Appendix C for details), we obtain Brownian motions as limits in the case of local adjustment. As a consequence, parametric standardization leads to the need of an explicit correction for the estimation error; see Proposition A.1 in Appendix C. As argued there, a further disadvantage of the parametric approach is that the required correction depends on the shape of the parametric mean or variance component adjusted for.*

3.2 Other null distributions and a finite-sample correction

Our framework allows testing other null distributions in location-scale models, since PITs apply to any continuous distribution. Concretely, compute $\hat{p}_t = F_0(\hat{z}_t)$ where F_0 is the cumulative distribution function of the *standardized* null distribution and $\hat{z}_t$ is the locally standardized time series.

In particular, it is straightforward to show that Lemma 1 and Proposition 1 hold under mild regularity conditions, yet with e.g. $\vartheta_k = \mathrm{E}\left(p_t^k f_0(z_t)\right)$ where f_0 is the density function of the null distribution of z_t. One further advantage of the local standardization approach is that the expectations $\vartheta_k = \mathrm{E}\left(p_t^k f_0(z_t)\right)$ and $\varpi_k = \mathrm{E}\left(p_t^k z_t f_0\left(z_t\right)\right)$ need not to be computed explicitly, such that the test is immediately applicable for any continuous null shape in location-scale families.

There is a minor exception, namely the case $k = 1$ when testing the null of a uniform distribution; this results in a linear F_0, so $\hat{p}_t$ is just a rescaled $\hat{z}_t$; it is argued in Appendix B (cf. the proof of Proposition 2) that the normalized partial sums of $\hat{z}_t$ converge to zero rather than Brownian motion, so we cannot base inference on them. Still, all other powers of the feasible PITs are available for testing: individual and joint statistics based on $\hat{m}_k$ are computed the same way, but excluding $k = 1$.

After having discussed the asymptotic properties, we now turn to a simple finite-sample adjustment procedure. The reason such a correction may be necessary is that the asymptotic approximation tends to have reduced quality in samples of smaller size. Intuitively, the problem with local standardization is that, although the resulting $\hat{z}_t$ are standardized in the limit, they may in fact still exhibit a non-negligible deviation from zero regarding the sample mean and a non-unit sample variance in small samples. In order to ensure a zero mean and a unit variance, we propose a simple finite-sample correction based on additional, "global" standardization of the filtered time series $\hat{z}_t$. Concretely, a full-sample adjustment step is applied onto the locally standardized series before applying the PIT. Thus, rather than $\hat{m}_k$, use

$$\tilde{m}_k = \frac{1}{T}\sum_{t=1}^{T} \tilde{p}_t^k \qquad \text{with} \qquad \tilde{p}_t = \Phi\left(\tilde{z}_t\right),$$

where the doubly standardized $\tilde{z}_t$ are simply obtained as

$$\tilde{z}_t = \frac{\hat{z}_t - \bar{\hat{z}}}{\tilde{\sigma}_{\hat{z}}} \quad \text{with} \quad \bar{\hat{z}} = \frac{1}{T}\sum_{t=1}^{T}\hat{z}_t \quad \text{and} \quad \tilde{\sigma}_{\hat{z}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{z}_t - \bar{\hat{z}}\right)^2}.$$

When doubly standardizing, however, a correction may be required for the asymptotic pivotality of the test statistics; cf. Bai and Ng (2005) and Appendix C. This leads to the use of a modified test statistic. In particular, we now build

$$\tilde{t}_k = \sqrt{T}\,\frac{\tilde{m}_k - \frac{1}{k+1}}{\hat{\psi}_k} \tag{12}$$

and, for the portmanteau-type test, take analogously

$$\tilde{\mathcal{T}}_K = T\left(\tilde{m}_1 - \frac{1}{2}, \ldots, \tilde{m}_K - \frac{1}{K+1}\right)\hat{\boldsymbol{\Psi}}^{-1}\left(\tilde{m}_1 - \frac{1}{2}, \ldots, \tilde{m}_K - \frac{1}{K+1}\right)' \tag{13}$$

with $\hat{\boldsymbol{\Psi}}$ being a long-run variance estimator replacing $\hat{\boldsymbol{\Omega}}$ in (10) by $\hat{\boldsymbol{\Psi}} = \mathbf{V}\hat{\boldsymbol{\Xi}}\mathbf{V}'$, where $\hat{\boldsymbol{\Xi}}$ is computed analogously to $\hat{\boldsymbol{\Omega}}$ but based on $(\tilde{p}_t, \ldots, \tilde{p}_t^k, \tilde{z}_t, \tilde{z}_t^2 - 1)'$. The matrix $\mathbf{V}$ depends on the distribution under test; see (11). We provide values of $\mathbf{V}$ for selected distributions in Appendix G.

Proposition 2 *Under the Assumptions of Proposition 1, we have that*

$$\tilde{t}_k \Rightarrow \frac{\tilde{W}(1)}{\sqrt{\mathcal{Q}_{b,\kappa}}} \qquad \text{and} \qquad \tilde{\mathcal{T}}_K \Rightarrow \tilde{\boldsymbol{W}}'_K(1)\boldsymbol{\mathcal{Q}}_{K,b,\kappa}^{-1}\tilde{\boldsymbol{W}}_K(1).$$

The proposed double standardization is just a finite-sample correction not depending on the actual shape of $\mu(\cdot)$ and $\sigma(\cdot)$, as local standardization already removes the time-variation in mean and variance, and the global standardization in the second step only removes some of the "left-overs" from the first step. The performance is investigated in the following Monte Carlo simulation study.

4 Simulation evidence

In our Monte Carlo simulation study we consider the case of testing for a Gaussian distribution due to the wide practical interest. We compare the newly proposed test statistics $\tilde{t}_k$ and $\tilde{\mathcal{T}}_K$ to the procedure of Bai and Ng (2005).[8] Regarding the alternative non-Gaussian distributions, we consider the log-normal, $\chi^2(3)$- and $t(3)$-distribution. The general form of the data generating process [DGP] for the time series of interest x_t is given by

$$\begin{aligned} x_t &= \mu_t + \sigma_t z_t \\ \mu_t &= \mu_1 + (\mu_2 - \mu_1) \cdot \mathbf{1}(t \geq \lfloor T_B \rfloor) \\ \sigma_t &= \sigma_1 + (\sigma_2 - \sigma_1) \cdot \mathbf{1}(t \geq \lfloor T_B \rfloor) \\ z_t &= \phi z_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \\ \varepsilon_t &\sim iid(0,1) \,. \end{aligned}$$

Here, the mean and variance functions may exhibit structural shifts from $\mu_1 = 0$ to $\mu_2 = 3$ and from $\sigma_1 = 1$ to $\sigma_2 = 3$.[9] We consider four different possibilities: DGP I: no shifts; DGP II: mean shift; DGP III: variance shift and DGP IV: mean and variance shift. Regarding autocorrelation, we specify, besides an independent white noise [IID] series, a strongly autocorrelated, but stationary and invertible ARMA(1,1) process [ARMA] with AR and MA parameter $\phi = 0.85$ and $\theta = -0.45$, respectively. Since all procedures are scale-invariant, there is no need to normalize the long-run variance of z_t to unity. However, the short-run variance is normalized to unity to make the mean and variance functions comparable for the cases of a white noise and an ARMA process. We use sample sizes of $T = \{250, 1000\}$ matching typical situations in macroeconomics (e.g. data with a quarterly frequency) and a large-sample setup. The breakpoint is set in the middle of the sample such that $T_B = T/2$.

We consider the first four individual raw moments ($k = 1, 2, 3, 4$) and the expanding set of multiple moments $\{1, 2\}$, $\{1, 2, 3\}$ and $\{1, 2, 3, 4\}$ for investigating (i) the relevance of single moments to detect specific alternatives, and (ii) possible advantages of using multiple moments jointly.

Regarding the heteroskedasticity and autocorrelation robust covariance matrix estimator, the fixed-bandwidth parameter b is specified as $b = 0.1$, as preliminary simulation experiments showed that the size of the tests is quite stable for different values of b, but that power is higher for smaller values of b; see also Kiefer and Vogelsang (2005). Results are presented for the Bartlett kernel with linearly decaying weights $\kappa(j/B)$. The nominal significance level equals 5% and the number of Monte Carlo replications is set to 2000 for each single experiment. In what concerns

[8] To keep the paper self-contained, details on the test proposed by Bai and Ng (2005) can be found in Appendix D, where we also formally discuss the impact of time-varying mean and volatility.

[9] This setup is characterized by adverse conditions for the nonparametric estimators as the breaks are abrupt and take place in a single time period. However, such a setup is realistic as there is ample evidence for time-variation in the mean and the variance of U.S. real output growth series. Most strikingly is the impact of the Great Moderation which manifested itself as a significant reduction in macroeconomic volatility. In several related studies (see e.g. Campbell, 2007), this feature is even modelled as a one-time structural change, and our simulation setup examines the effect of such changes; the nonparametric estimators and related tests perform even better under more rosy conditions, as is confirmed in a number of additional unreported simulation results with smooth shifts which are available upon request.

critical values for the test statistics based on fixed-b inference, we provide them on the basis of the limiting results with 1000 observations and 50000 replications for $K = \{2, 3, 4\}$. For the case of testing a single moment restriction (via the t-statistic), we employ critical values reported in Kiefer and Vogelsang (2005). For $k \geq 2$, estimated cubic response curves $cv(b)$ are reported in Table A.2 in Appendix F together with an R^2 measure for the precision of approximation; as can be seen from the least squares estimation results, the fit of the estimated response curves is very well and comparable to Kiefer and Vogelsang (2005).

When computing $\hat{\mu}_t$ and $\hat{\sigma}_t$, the employed nonparametric estimator is either the local constant Nadaraya-Watson [LC] or the local linear [LL] estimator. Possible kernel [$K(\cdot)$] choices are Uniform [Unif], Epanechnikov [Epa] and Gaussian [Gauss]. The chosen bandwidth h (or equivalently window width $\tau = hT$) takes into account the fact that the asymptotic results require undersmoothing (i.e. estimation bias negligible w.r.t. estimation standard deviation). To ensure that our window/bandwidth restrictions are satisfied in practice, we resort to the following empirical strategy. Bearing in mind that the MSE-optimal window width is known to be of order $T^{4/5}$ in our case, it suggests itself to use it as a reference point for the desired window width for our procedure, not least because the MSE-optimal window width is readily estimable using standard cross-validation approaches. The theoretical rate required by our results has a magnitude order lying between $T^{2/3}$ and $T^{3/4}$, where one may meaningfully choose a rate close to the upper bound such that the variance be small and the bias still dominated as required by our theoretical results. Summing up, we propose to scale down the cross-validated MSE-optimal rate by $T^{0.051}$ which delivers a magnitude order of $T^{0.749}$ close to the upper bound but not attaining it. All computations are carried out in R on the basis of the np, gplm and sandwich packages.

Results are reported in Tables 1–5. Size results are provided in Tables 1 and 2 for $T = 250$ and $T = 1000$, respectively. In these cases, the distribution of ε_t is Gaussian. Power results are reported in Tables 3 (log-normal distribution), 4 (Chi-squared distribution) and 5 (Student-t distribution). For power, we consider weighted averages of a Gaussian and a non-Gaussian random variable in the form of $\varepsilon_t = (1 - c) \cdot \varepsilon_{1t} + c \cdot \varepsilon_{2t}$, where ε_{1t} is standard Gaussian and ε_{2t} is standard non-Gaussian. When $c = 0$, full weight is put on the Gaussian distribution, while the case of $c = 1$ reflects a fully non-Gaussian situation. For intermediate values of c between zero and unity, the distribution of ε_t is still non-Gaussian and the null hypothesis is false, where the power of normality tests should increase with c with more weight given to the non-Gaussian distribution. We consider the values $c = \{0, 1/3, 3/8, 5/12, 1/2, 1\}$ in our simulations.

For the size results in Tables 1 and 2, we see that the Bai and Ng (2005) [BN] test is generally somewhat oversized for DGP I (excluding any shifts), while the raw moment-based tests are closer to the nominal significance level of 5%. In some cases we observe that they are marginally over- or undersized, while, for the larger sample size of $T = 1000$ (including short-run dynamics), most of them are quite close to the desired frequency of rejections. Overall, the employed fixed-b approach accounts fairly well for the quite strong autocorrelation in the ARMA DGP. The particular choice of the kernel function for the nonparametric estimator does not appear to play an important role. Clearly, the BN test is massively oversized when there are mean and/or volatility shifts (DGPs II, III and IV). Even for $T = 250$, we find rejection rates between 0.671 and 0.998. For $T = 1000$, these rejection rates approach unity. On the contrary, the newly proposed

Table 1: Size, $T = 250$

DGP	$\hat{\mu}(\cdot), \hat{\sigma}(\cdot)$	z_t	$K(\cdot)$	$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$	$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN
I	LC	IID	Unif	0.047	0.054	0.053	0.053	0.051	0.044	0.044	0.114
			Epa	0.049	0.051	0.046	0.045	0.049	0.046	0.038	
			Gauss	0.040	0.048	0.045	0.036	0.039	0.035	0.029	
		ARMA	Unif	0.051	0.057	0.070	0.072	0.061	0.061	0.057	0.090
			Epa	0.055	0.054	0.071	0.069	0.070	0.059	0.055	
			Gauss	0.045	0.053	0.058	0.061	0.058	0.055	0.060	
	LL	IID	Unif	0.054	0.049	0.049	0.054	0.059	0.052	0.051	
			Epa	0.050	0.045	0.042	0.046	0.043	0.043	0.040	
			Gauss	0.047	0.050	0.045	0.047	0.045	0.052	0.043	
		ARMA	Unif	0.059	0.050	0.061	0.060	0.063	0.054	0.063	
			Epa	0.050	0.052	0.054	0.057	0.057	0.056	0.054	
			Gauss	0.044	0.055	0.063	0.067	0.062	0.058	0.056	
II	LC	IID	Unif	0.046	0.052	0.071	0.070	0.063	0.056	0.047	0.774
			Epa	0.052	0.050	0.068	0.075	0.061	0.054	0.049	
			Gauss	0.046	0.051	0.056	0.065	0.061	0.053	0.050	
		ARMA	Unif	0.047	0.067	0.083	0.078	0.070	0.065	0.073	0.671
			Epa	0.044	0.067	0.080	0.087	0.075	0.064	0.066	
			Gauss	0.049	0.064	0.080	0.088	0.073	0.068	0.073	
	LL	IID	Unif	0.049	0.107	0.137	0.138	0.135	0.113	0.112	
			Epa	0.055	0.091	0.132	0.145	0.126	0.111	0.101	
			Gauss	0.047	0.097	0.136	0.137	0.127	0.119	0.100	
		ARMA	Unif	0.047	0.088	0.125	0.135	0.119	0.105	0.104	
			Epa	0.050	0.094	0.125	0.129	0.119	0.106	0.102	
			Gauss	0.043	0.090	0.103	0.111	0.106	0.095	0.091	
III	LC	IID	Unif	0.042	0.049	0.052	0.053	0.062	0.050	0.045	0.998
			Epa	0.049	0.057	0.057	0.052	0.053	0.056	0.047	
			Gauss	0.046	0.052	0.046	0.050	0.047	0.050	0.043	
		ARMA	Unif	0.041	0.091	0.123	0.127	0.108	0.102	0.125	0.992
			Epa	0.040	0.086	0.102	0.112	0.099	0.089	0.114	
			Gauss	0.052	0.085	0.106	0.115	0.091	0.093	0.110	
	LL	IID	Unif	0.048	0.053	0.051	0.049	0.045	0.040	0.043	
			Epa	0.050	0.042	0.041	0.042	0.046	0.040	0.039	
			Gauss	0.051	0.051	0.046	0.049	0.051	0.058	0.054	
		ARMA	Unif	0.047	0.084	0.112	0.114	0.088	0.095	0.102	
			Epa	0.051	0.085	0.103	0.104	0.092	0.092	0.103	
			Gauss	0.054	0.089	0.102	0.117	0.102	0.097	0.112	
IV	LC	IID	Unif	0.063	0.069	0.076	0.071	0.071	0.050	0.048	0.995
			Epa	0.066	0.075	0.079	0.083	0.080	0.059	0.056	
			Gauss	0.056	0.055	0.067	0.073	0.060	0.053	0.043	
		ARMA	Unif	0.045	0.067	0.083	0.091	0.078	0.069	0.080	0.992
			Epa	0.044	0.072	0.079	0.088	0.078	0.070	0.083	
			Gauss	0.048	0.076	0.090	0.093	0.086	0.084	0.090	
	LL	IID	Unif	0.064	0.108	0.130	0.133	0.125	0.107	0.099	
			Epa	0.061	0.125	0.145	0.146	0.134	0.121	0.111	
			Gauss	0.062	0.120	0.144	0.152	0.135	0.108	0.095	
		ARMA	Unif	0.034	0.090	0.125	0.132	0.112	0.118	0.119	
			Epa	0.048	0.099	0.129	0.140	0.127	0.102	0.102	
			Gauss	0.046	0.101	0.130	0.134	0.126	0.112	0.112	

Table 2: Size, $T = 1000$

DGP	$\hat{\mu}(\cdot), \hat{\sigma}(\cdot)$	z_t	$K(\cdot)$	$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$	$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN
I	LC	IID	Unif	0.043	0.046	0.040	0.041	0.046	0.049	0.041	0.090
			Epa	0.049	0.048	0.053	0.053	0.056	0.054	0.049	
			Gauss	0.056	0.049	0.047	0.048	0.052	0.045	0.039	
		ARMA	Unif	0.055	0.057	0.062	0.061	0.058	0.058	0.062	0.098
			Epa	0.045	0.058	0.060	0.056	0.059	0.058	0.055	
			Gauss	0.037	0.059	0.066	0.062	0.059	0.052	0.052	
	LL	IID	Unif	0.043	0.042	0.041	0.045	0.043	0.040	0.037	
			Epa	0.051	0.046	0.044	0.047	0.047	0.040	0.031	
			Gauss	0.045	0.043	0.037	0.040	0.048	0.043	0.041	
		ARMA	Unif	0.044	0.054	0.053	0.052	0.051	0.047	0.051	
			Epa	0.041	0.052	0.057	0.056	0.054	0.046	0.048	
			Gauss	0.036	0.049	0.056	0.057	0.044	0.046	0.040	
II	LC	IID	Unif	0.044	0.054	0.065	0.079	0.055	0.046	0.043	1.000
			Epa	0.038	0.055	0.067	0.073	0.056	0.051	0.049	
			Gauss	0.042	0.058	0.070	0.080	0.070	0.059	0.051	
		ARMA	Unif	0.048	0.066	0.072	0.082	0.073	0.073	0.070	0.989
			Epa	0.049	0.080	0.085	0.089	0.088	0.077	0.072	
			Gauss	0.041	0.062	0.070	0.074	0.066	0.060	0.061	
	LL	IID	Unif	0.047	0.058	0.076	0.074	0.062	0.056	0.044	
			Epa	0.047	0.055	0.067	0.075	0.057	0.044	0.039	
			Gauss	0.046	0.059	0.070	0.070	0.057	0.048	0.041	
		ARMA	Unif	0.053	0.072	0.077	0.081	0.077	0.062	0.071	
			Epa	0.047	0.060	0.065	0.067	0.069	0.059	0.060	
			Gauss	0.052	0.068	0.083	0.081	0.075	0.073	0.067	
III	LC	IID	Unif	0.049	0.046	0.045	0.047	0.050	0.043	0.040	1.000
			Epa	0.053	0.054	0.050	0.048	0.054	0.050	0.041	
			Gauss	0.044	0.039	0.042	0.043	0.045	0.038	0.039	
		ARMA	Unif	0.046	0.068	0.072	0.074	0.077	0.071	0.075	1.000
			Epa	0.045	0.062	0.064	0.062	0.057	0.050	0.057	
			Gauss	0.047	0.061	0.066	0.065	0.066	0.064	0.060	
	LL	IID	Unif	0.044	0.046	0.045	0.046	0.048	0.040	0.045	
			Epa	0.040	0.038	0.042	0.050	0.044	0.035	0.032	
			Gauss	0.054	0.051	0.043	0.042	0.049	0.049	0.039	
		ARMA	Unif	0.048	0.068	0.076	0.069	0.072	0.069	0.057	
			Epa	0.037	0.055	0.067	0.066	0.057	0.054	0.052	
			Gauss	0.055	0.067	0.073	0.071	0.072	0.061	0.061	
IV	LC	IID	Unif	0.059	0.067	0.066	0.064	0.070	0.053	0.041	1.000
			Epa	0.061	0.070	0.072	0.065	0.070	0.056	0.043	
			Gauss	0.054	0.059	0.068	0.071	0.064	0.049	0.040	
		ARMA	Unif	0.047	0.073	0.079	0.089	0.073	0.068	0.073	1.000
			Epa	0.047	0.078	0.081	0.084	0.076	0.072	0.070	
			Gauss	0.041	0.065	0.072	0.081	0.069	0.066	0.069	
	LL	IID	Unif	0.058	0.065	0.075	0.072	0.068	0.056	0.046	
			Epa	0.054	0.068	0.070	0.071	0.058	0.046	0.038	
			Gauss	0.049	0.059	0.067	0.072	0.055	0.048	0.040	
		ARMA	Unif	0.045	0.060	0.061	0.071	0.067	0.063	0.063	
			Epa	0.046	0.071	0.078	0.081	0.067	0.063	0.064	
			Gauss	0.052	0.066	0.073	0.072	0.073	0.059	0.060	

Table 3: Power against a log-normal distribution

DGP	$\hat{\mu}(\cdot), \hat{\sigma}(\cdot)$	T	c	$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$	$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN
I	LC	250	1/3	0.142	0.113	0.095	0.084	0.095	0.075	0.067	0.126
			3/8	0.229	0.205	0.164	0.122	0.153	0.117	0.104	0.175
			5/12	0.334	0.288	0.226	0.165	0.211	0.170	0.141	0.234
			1/2	0.681	0.610	0.484	0.366	0.490	0.414	0.381	0.563
			1	0.975	0.940	0.883	0.784	0.976	1.000	1.000	0.952
		1000	1/3	0.531	0.567	0.520	0.429	0.413	0.317	0.263	0.445
			3/8	0.800	0.784	0.704	0.607	0.666	0.535	0.448	0.731
			5/12	0.942	0.930	0.874	0.797	0.874	0.775	0.693	0.915
			1/2	0.998	0.976	0.952	0.917	0.996	0.984	0.970	0.991
			1	0.999	0.990	0.980	0.959	0.999	1.000	1.000	0.993
II	LC	250	1/3	0.115	0.100	0.088	0.078	0.078	0.068	0.063	
			3/8	0.182	0.170	0.138	0.108	0.118	0.086	0.088	
			5/12	0.282	0.267	0.221	0.173	0.176	0.153	0.128	
			1/2	0.549	0.528	0.455	0.354	0.383	0.319	0.280	
			1	0.959	0.932	0.889	0.804	0.922	1.000	1.000	
		1000	1/3	0.482	0.531	0.497	0.423	0.378	0.287	0.230	
			3/8	0.733	0.751	0.705	0.617	0.603	0.478	0.398	
			5/12	0.914	0.914	0.864	0.781	0.824	0.723	0.622	
			1/2	0.993	0.971	0.952	0.926	0.991	0.969	0.955	
			1	0.996	0.987	0.980	0.961	0.997	1.000	1.000	
III	LC	250	1/3	0.139	0.178	0.181	0.169	0.129	0.100	0.085	
			3/8	0.228	0.279	0.267	0.221	0.196	0.147	0.138	
			5/12	0.307	0.387	0.362	0.307	0.239	0.210	0.187	
			1/2	0.611	0.635	0.566	0.462	0.463	0.425	0.422	
			1	0.963	0.931	0.870	0.772	0.946	1.000	1.000	
		1000	1/3	0.514	0.655	0.639	0.568	0.471	0.344	0.273	
			3/8	0.746	0.835	0.810	0.750	0.695	0.549	0.488	
			5/12	0.922	0.932	0.910	0.864	0.879	0.759	0.721	
			1/2	0.996	0.978	0.960	0.938	0.993	0.980	0.976	
			1	0.994	0.987	0.981	0.966	0.994	1.000	1.000	
IV	LC	250	1/3	0.213	0.235	0.203	0.167	0.165	0.123	0.104	
			3/8	0.314	0.329	0.287	0.216	0.229	0.182	0.155	
			5/12	0.415	0.433	0.373	0.285	0.314	0.252	0.222	
			1/2	0.693	0.659	0.557	0.440	0.522	0.449	0.439	
			1	0.950	0.934	0.887	0.805	0.931	1.000	1.000	
		1000	1/3	0.630	0.697	0.652	0.551	0.517	0.386	0.314	
			3/8	0.818	0.855	0.819	0.746	0.736	0.600	0.506	
			5/12	0.962	0.956	0.920	0.868	0.899	0.808	0.733	
			1/2	0.996	0.981	0.966	0.943	0.994	0.988	0.975	
			1	0.996	0.992	0.985	0.968	0.997	1.000	1.000	

Table 4: Power against a $\chi^2(3)$-distribution

DGP	$\hat{\mu}(\cdot), \hat{\sigma}(\cdot)$	T	c	$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$	$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN
I	LC	250	1/3	0.100	0.083	0.060	0.051	0.085	0.077	0.062	0.148
			3/8	0.176	0.135	0.098	0.075	0.117	0.095	0.082	0.198
			5/12	0.294	0.239	0.172	0.110	0.194	0.146	0.129	0.289
			1/2	0.682	0.587	0.405	0.224	0.518	0.403	0.328	0.659
			1	1.000	0.999	0.956	0.654	1.000	1.000	1.000	1.000
		1000	1/3	0.300	0.290	0.205	0.124	0.215	0.169	0.132	0.277
			3/8	0.561	0.533	0.411	0.271	0.420	0.352	0.276	0.538
			5/12	0.846	0.822	0.693	0.489	0.741	0.646	0.543	0.838
			1/2	0.998	0.997	0.984	0.886	0.993	0.983	0.959	1.000
			1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
II	LC	250	1/3	0.089	0.078	0.071	0.069	0.079	0.074	0.068	
			3/8	0.156	0.131	0.102	0.071	0.113	0.093	0.068	
			5/12	0.236	0.206	0.151	0.103	0.162	0.137	0.105	
			1/2	0.608	0.627	0.525	0.361	0.464	0.344	0.289	
			1	0.998	0.996	0.953	0.719	1.000	1.000	0.998	
		1000	1/3	0.251	0.242	0.202	0.141	0.193	0.159	0.121	
			3/8	0.524	0.527	0.410	0.271	0.417	0.337	0.255	
			5/12	0.791	0.796	0.675	0.479	0.674	0.585	0.475	
			1/2	0.996	1.000	0.991	0.939	0.990	0.971	0.943	
			1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
III	LC	250	1/3	0.091	0.126	0.129	0.117	0.101	0.079	0.070	
			3/8	0.146	0.195	0.195	0.166	0.142	0.099	0.077	
			5/12	0.261	0.329	0.295	0.235	0.221	0.142	0.115	
			1/2	0.561	0.489	0.355	0.213	0.416	0.314	0.256	
			1	1.000	0.996	0.941	0.672	0.999	0.999	0.996	
		1000	1/3	0.289	0.396	0.363	0.267	0.276	0.189	0.141	
			3/8	0.502	0.623	0.576	0.446	0.459	0.355	0.282	
			5/12	0.815	0.881	0.806	0.667	0.750	0.636	0.532	
			1/2	0.997	0.999	0.981	0.874	0.992	0.971	0.926	
			1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
IV	LC	250	1/3	0.165	0.180	0.149	0.117	0.123	0.087	0.078	
			3/8	0.243	0.251	0.206	0.147	0.182	0.134	0.110	
			5/12	0.376	0.373	0.298	0.203	0.284	0.202	0.161	
			1/2	0.699	0.681	0.532	0.342	0.572	0.455	0.371	
			1	0.996	0.9945	0.9635	0.749	0.9985	0.9985	0.9985	
		1000	1/3	0.395	0.435	0.381	0.277	0.312	0.246	0.193	
			3/8	0.647	0.671	0.590	0.449	0.533	0.425	0.331	
			5/12	0.877	0.892	0.802	0.639	0.798	0.690	0.592	
			1/2	0.999	0.997	0.989	0.943	0.997	0.985	0.959	
			1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

Table 5: Power against a $t(3)$-distribution

DGP	$\hat{\mu}(\cdot), \hat{\sigma}(\cdot)$	T	c	$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$	$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN
I	LC	250	1/3	0.050	0.057	0.051	0.048	0.056	0.046	0.040	0.073
			3/8	0.051	0.044	0.053	0.063	0.058	0.049	0.041	0.058
			5/12	0.053	0.063	0.080	0.083	0.063	0.060	0.046	0.043
			1/2	0.055	0.077	0.121	0.162	0.103	0.101	0.086	0.049
			1	0.064	0.451	0.739	0.775	0.719	0.695	0.775	0.201
		1000	1/3	0.040	0.055	0.078	0.103	0.076	0.052	0.048	0.038
			3/8	0.050	0.086	0.151	0.191	0.130	0.106	0.090	0.052
			5/12	0.048	0.148	0.261	0.334	0.222	0.166	0.126	0.096
			1/2	0.051	0.338	0.600	0.674	0.528	0.392	0.333	0.222
			1	0.066	0.917	0.935	0.928	0.932	0.955	1.000	0.391
II	LC	250	1/3	0.041	0.042	0.049	0.054	0.056	0.050	0.038	
			3/8	0.041	0.049	0.052	0.058	0.053	0.045	0.038	
			5/12	0.040	0.051	0.051	0.068	0.063	0.065	0.051	
			1/2	0.046	0.140	0.255	0.316	0.223	0.178	0.143	
			1	0.061	0.524	0.776	0.793	0.750	0.741	0.817	
		1000	1/3	0.040	0.068	0.094	0.118	0.084	0.064	0.053	
			3/8	0.051	0.093	0.156	0.189	0.136	0.102	0.087	
			5/12	0.049	0.139	0.259	0.318	0.215	0.167	0.128	
			1/2	0.051	0.466	0.706	0.768	0.661	0.534	0.433	
			1	0.076	0.934	0.946	0.944	0.949	0.969	1.000	
III	LC	250	1/3	0.048	0.053	0.088	0.106	0.073	0.056	0.056	
			3/8	0.055	0.064	0.104	0.129	0.089	0.078	0.058	
			5/12	0.048	0.087	0.148	0.177	0.127	0.101	0.084	
			1/2	0.044	0.070	0.119	0.150	0.112	0.092	0.083	
			1	0.041	0.326	0.640	0.710	0.664	0.590	0.629	
		1000	1/3	0.056	0.122	0.208	0.250	0.182	0.119	0.095	
			3/8	0.057	0.146	0.275	0.338	0.250	0.181	0.139	
			5/12	0.043	0.230	0.414	0.490	0.339	0.247	0.191	
			1/2	0.039	0.303	0.569	0.643	0.511	0.371	0.313	
			1	0.042	0.911	0.938	0.933	0.958	0.953	0.998	
IV	LC	250	1/3	0.049	0.075	0.099	0.102	0.075	0.066	0.053	
			3/8	0.050	0.097	0.123	0.138	0.102	0.077	0.061	
			5/12	0.051	0.116	0.164	0.161	0.125	0.102	0.090	
			1/2	0.049	0.145	0.227	0.261	0.179	0.125	0.121	
			1	0.053	0.507	0.763	0.786	0.724	0.678	0.731	
		1000	1/3	0.056	0.148	0.205	0.231	0.160	0.118	0.088	
			3/8	0.052	0.184	0.300	0.342	0.234	0.159	0.115	
			5/12	0.046	0.273	0.428	0.498	0.345	0.239	0.178	
			1/2	0.044	0.471	0.716	0.758	0.624	0.482	0.396	
			1	0.056	0.934	0.952	0.950	0.950	0.958	1.000	

statistics properly account for such shifts and deliver relatively accurate empirical sizes. Some minor upward size distortions are observed for the smaller sample size and tests using the third and fourth moment (including combination tests). The local constant estimator is performing better than the local linear estimator in smaller samples. The distortions almost vanish for the larger sample size, reflecting the asymptotic validity of our nonparametric estimation approach.

For power simulations, we consider the IID case only to save space. Results for the case of ARMA innovations are quite similar and therefore not reported. Moreover, given that the kernel choice does not matter much, we focus on the commonly used Gaussian kernel in the following. Also, as the results for the local linear and the local constant estimator are quite similar, we only report power for the latter one. We distinguish DGP I (no shifts) from the remaining ones exhibiting mean and/or variance shifts. The BN test performs reasonably only in the absence of mean and variance shifts, as it was not intended to account for such structural change. For DGP I, we are able to compare the power of the BN test to the newly proposed ones in a meaningful way, of course while keeping in mind that the BN test may be somewhat over-sized under the null (see Tables 1 and 2). Power results are reported in the top panel of Tables 3, 4 and 5.

As expected, power increases as more weight is given to the non-Gaussian distribution c and also as T increases. For a given value of c and T, it is interesting to compare the empirical power of the $\tilde{t}$- and $\tilde{\mathcal{T}}$-tests to the BN test. In DGP I, there are no shifts present, but the mean and variance function are estimated nonparametrically anyway for the new tests. In twenty-seven out of thirty cases (three non-Gaussian distributions, two sample sizes and five values of c each), either $\tilde{t}_1$ (for log-normal and χ^2-distribution) or $\tilde{t}_4$ (for Student-t-distribution) outperforms the BN test even under stationarity.[10] This finding is remarkable since the mean and volatility functions are constant over time and their nonparametric estimation is unnecessary. Moreover, it reflects the benefits of using raw moments and a fixed-b approach even in the absence of time-varying location and scale. A combination of moments does not generally pay off in terms of power. Therefore, we recommend the use of the test statistics $\tilde{t}_1$ (for log-normal and χ^2-distribution) and $\tilde{t}_4$ (for Student-t-distribution).

For the remaining DGPs II, III and IV, a direct comparison of the BN test to the new proposed test statistics is not meaningful due to the massive over-rejections of the BN test under the null. These additional experiments allow us to compare the power across different scenarios regarding the breaks in comparison to the benchmark case of no shifts. The results reveal that power is somewhat lowered in the presence of mean shifts, but a bit higher when pure volatility shifts are present. In case of a combination of both effects being present, power is almost similar to the benchmark case in many cases. Overall, the results underline the robustness and empirical usefulness of the newly suggested test statistics.

5 Testing normality of U.S. real output growth rates

When looking at the empirical distribution of real output growth series, it might seem at a first sight that Gaussianity is a poor assumption, particularly in the tails of the distribution.

[10] We can see that the first moment cannot detect excess kurtosis given symmetry as in the $t(3)$-distribution, but the second, third and fourth moment tests are successful and all deliver higher power than the BN test.

Sometimes, large economic movements due to booms and recessions are seen to occur more frequently than the normal distribution would predict, while the center of the distribution is typically well-captured by a Gaussian distribution. This observation motivates researchers to provide explanations for the non-Gaussian tail behavior of real output growth, coined as macroeconomic tail risk (Acemoglu et al., 2017). Although most studies recognize the importance of heteroskedasticity (especially in the context of the Great Moderation and the Great Recession), the key feature of time-varying volatility is not seldom overlooked in practice; see e.g. Fagiolo et al. (2008), Ascari et al. (2015), Acemoglu et al. (2017) and Kiss and Österholm (2020). While e.g. standard GARCH and related models account for conditionally time-varying volatility, the implied unconditional variance is still constant in such models. Inference regarding Gaussianity as in the aforementioned studies may therefore be unreliable.

We reconsider the question whether U.S. real output growth is non-normally distributed. In particular, we account for time-varying mean and variance by estimating these nonparametrically. Indeed, we find that real output growth is rather marginally normally distributed and non-normality as a popular stylized notion cannot be maintained after accounting for time-varying first and second moments. As an important implication, this finding challenges the empirical relevance of macroeconomic tail risks which may have been overstated for the U.S..

Another implication of practical relevance of normality refers to the measurement of real output growth. While real GDP growth is the obvious candidate to proxy real output growth, similar measures exist: e.g. real gross domestic income [GDI] growth quantifies (at least in theory) the same variable. In practice, these measures rarely take the same value, and their spread is called "statistical discrepancy". Both variables are measures for real aggregate production, but they are constructed in a quite distinct manner and rely on different sources; see (Landefeld et al., 2008). In light of the Great Recession, there is a renewed interest in the macroeconomic literature to obtain an improved output measure by averaging these real GDP and real GDI growth in an optimal way. Typically, this is done in a state-space framework due the latency of the "true GDP"; see e.g. Aruoba et al. (2016). In this Kalman filter-based framework, normality is a key assumption; see e.g. Jacobs and Van Norden (2011) and more recently Almuzara et al. (2019) and Jacobs et al. (2020).[11]

We apply the newly proposed tests in comparison to the BN test to real GDP growth and real GDI growth. The data is sampled quarterly from 1947Q2 to 2019Q4 yielding $T = 291$ observations and retrieved from the Federal Reserve Bank of Philadelphia.[12] This sample size matches our simulation setup. Accordingly, we apply the finite-sample correction procedure as outlined in Section 3.2.

In Figures 1 (c.f. the Introduction) and 2, we show in 2×2 plots the two time series under test $x_t = 400\Delta \log GDP_t$ and $x_t = 400\Delta \log GDI_t$ (upper left), the locally standardized time series $\hat{z}_t^{GDP}$ and $\hat{z}_t^{GDI}$ (upper right) together with their respective quantile-quantile (QQ) plots (lower left for x_t and lower right for $\hat{z}_t$). Indeed, the QQ-plots suggest the presence of deviations from

[11] In fact, our testing approach is related to the one in Almuzara et al. (2019) in the sense that both focus on filtered series; however, Almuzara et al. work in a parametric time-invariant unobserved components model and derive analytic expressions for the autocovariances, while we explicitly allow for time-varying location and scale.

[12] See https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/gdpplus

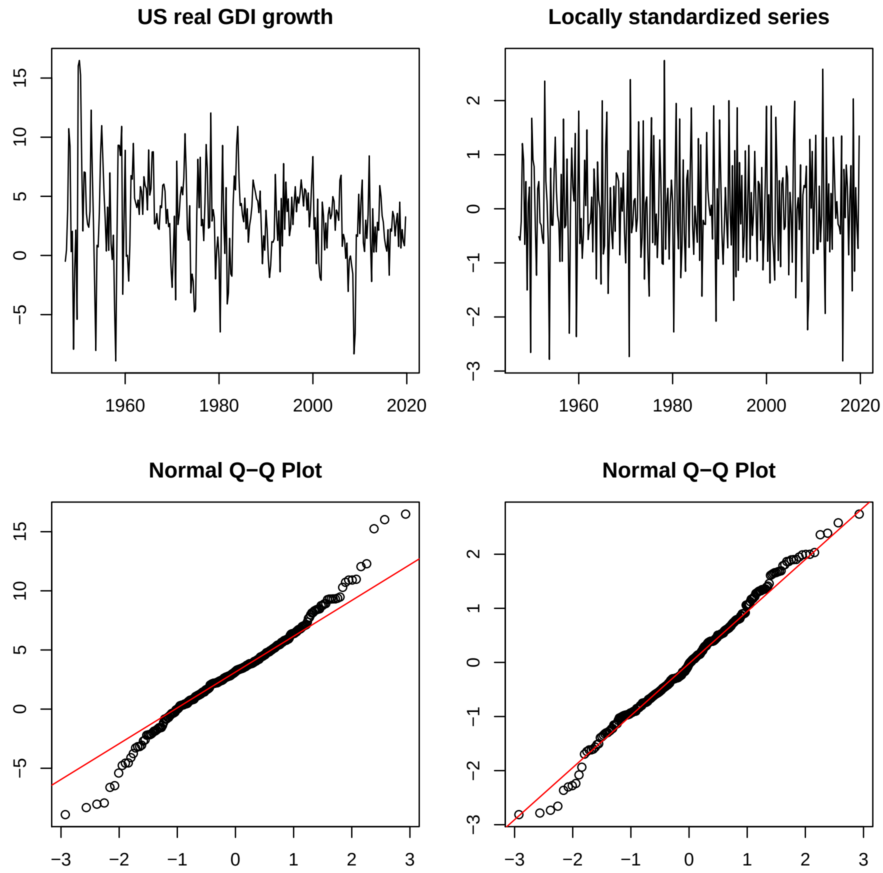

Figure 2: U.S. real GDI growth rate (1947Q2-2019Q4). Top left: time series plot of the raw time series x_t; top right: nonparametric locally standardized time series ($\hat{z}_t$); bottom left: normal QQ-plot for x_t and bottom right: normal QQ-plot for $\hat{z}_t$.

normality in the lower and upper tail for the raw series; this is however not the case for the standardized series.

To assess the quality of the local standardization, we provide empirical variance profiles of x_t and $\hat{z}_t$ for both time series in Figure 3. The sample variance profile (see e.g. Cavaliere and Taylor, 2007) for x_t is computed as

$$\eta_t = \frac{\sum_{j=1}^{t}(x_t - \hat{\mu}_t)^2}{\sum_{j=1}^{T}(x_t - \hat{\mu}_t)^2}$$

and relates the quadratic variation up to observation t to the total variation. By construction, it takes values between 0 and 1 in relative time (which also ranges between 0 and 1) and is compared to a 45°-line. Time-variation in the variance would result in deviations from the 45°-line. Clearly, both growth rates exhibit heteroskedasticity, while it appears that U.S. real GDP growth has some stronger volatility (reduction) in the mid-80s in comparison to U.S. real GDI growth. Their locally standardized counterparts (with sample variance profile computed as $\sum_{j=1}^{t} \hat{z}_t^2 / \sum_{j=1}^{T} \hat{z}_t^2$) do not share this property. This indicates that the nonparametric estimator successfully captures the time-varying volatility as their variance profiles are tracking the 45°-line very closely.[13]

[13] The conclusions from this descriptive analysis are further supported by inference based on the Deng and Perron (2008) test. The CUSUM of squares test rejects the null of constant variances against a one-time structural change in the raw series, while it does not for the locally standardized series. This finding holds for both, real GDP and

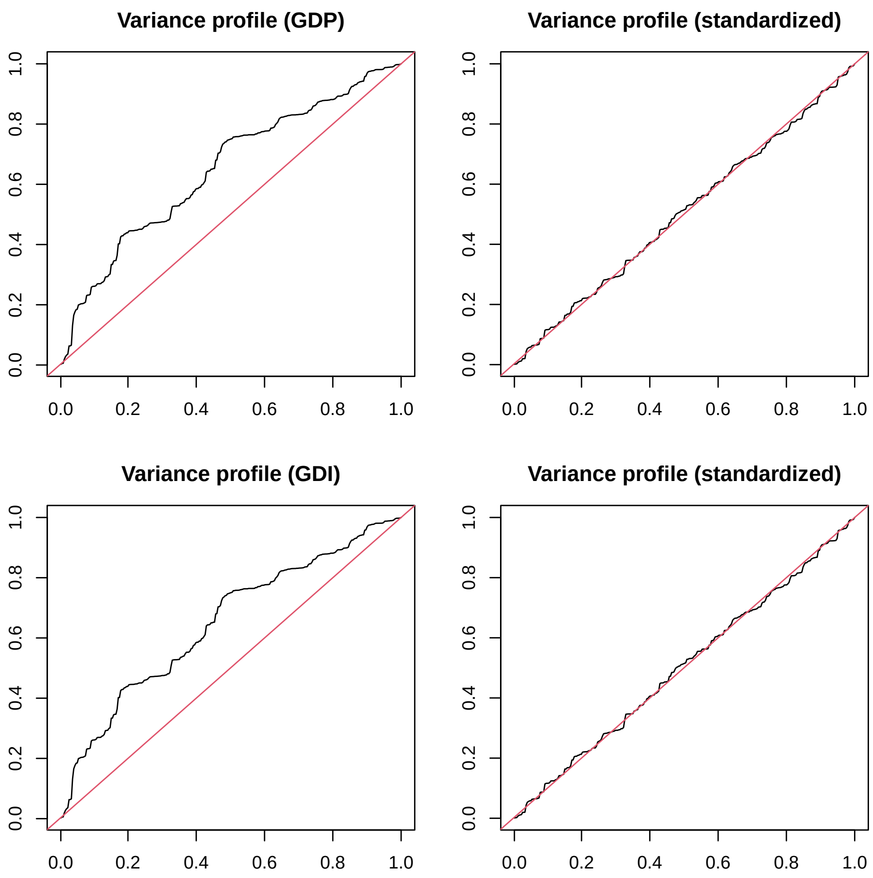

Figure 3: U.S. real GDP (top) and GDI (bottom) growth rate (1947Q2-2019Q4). Left: Variance profile for the raw time series x_t; right: Variance profile for the nonparametric locally standardized $\hat{z}_t$.

Table 6: Normality testing results for U.S. real GDP and GDI growth rates (1947Q2-2019Q4). $\tilde{m}_i$ denotes the estimated i-th raw moment and $\tilde{t}_i$ is the t-statistic for testing the individual i-th moment restriction ($i = 1, 2, 3, 4$) based on the nonparametric local standardization. The corresponding joint statistics $\tilde{\mathcal{T}}_{1 \ldots j}$ test the first j moment restrictions ($j = 2, 3, 4$); BN labels the Bai and Ng (2005) statistic. Critical values for the nominal significance level of 5% are provided in parentheses below the statistics.

GDP				GDI			
$\tilde{m}_1$	$\tilde{m}_2$	$\tilde{m}_3$	$\tilde{m}_4$	$\tilde{m}_1$	$\tilde{m}_2$	$\tilde{m}_3$	$\tilde{m}_4$
0.501	0.332	0.247	0.196	0.497	0.329	0.247	0.199
$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$	$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$
0.259	-0.369	-0.910	-1.261	-0.850	-1.594	-1.289	-0.528
(2.261)	(2.261)	(2.261)	(2.261)	(2.261)	(2.261)	(2.261)	(2.261)
$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN	$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN
1.765	1.760	4.112	8.972	2.931	5.855	8.729	9.892
(8.872)	(13.200)	(18.258)	(5.991)	(8.872)	(13.200)	(18.258)	(5.991)

Formal test results on the null of normality are provided in Table 6. First, we report the first four estimated raw moments of the PIT. Under the null (of normality), they are equal to 1/2, 1/3, 1/4 and 1/5, respectively. As can be seen, the point estimates are quite close to the theoretical null values. This holds for both measures. None of the individual raw moment-based test statistics ($\tilde{t}_1, \tilde{t}_2, \tilde{t}_3, \tilde{t}_4$) and also none of the joint versions ($\tilde{\mathcal{T}}_{12}, \tilde{\mathcal{T}}_{123}, \tilde{\mathcal{T}}_{1234}$) lead to a rejection of normality for GDP and GDI growth rates at the five percent level.This result clearly points towards marginal normality of the two time series under test.

On the contrary, the Bai and Ng (2005) test rejects the null of normality. This result, in combination with the obvious time-heteroskedasticity and the non-rejections of our newly proposed tests, is plausible and can be explained by our simulation evidence and also our theoretical findings.

To provide further insight, we repeat the tests for a sub-sample excluding the Great Moderation period, i.e. we examine the sample starting in 1984Q3 instead. Thereby, we exclude the downward volatility break in the mid-80s. As demonstrated in Figure A.3 (see Appendix E), the empirical variance profiles suggest only very little deviations from the 45°-line and confirm the absence of large volatility changes in this sub-sample for both time series. This finding is in line with those obtained in Gadea et al. (2018).[14] This is also visible from the plots in Figures A.1 and A.2 (see Appendix E) which display the time series, the locally standardized version and their QQ-plots.

We re-run tests for normality and find none of the employed tests ($\tilde{t}_k$, $\tilde{\mathcal{T}}_K$ and BN) to reject the null of normality; see Table A.1 (located in Appendix E as well). This confirms the previous finding that neglected time-variation presumably led to a rejection of normality by the BN test, while the newly proposed tests are robust. For this sub-sample, our results are in line with those provided in Almuzara et al. (2019). But, due to the strong effects of heteroskedasticity, the test decisions differ for the full sample. While we find support for the normality hypothesis over the whole post-war period, normality is rejected in Almuzara et al. (2019) in the full sample.

As a concluding robustness check, we follow a popular empirical modelling strategy (cf. e.g. Fagiolo et al., 2008; Acemoglu et al., 2017) and fit an ARMA-GARCH-type model to the U.S. real output growth series, and test the *standardized residuals* for normality. In particular, we fit an autoregressive model including an intercept (for which the Schwarz information criterion indicates a first-order lag structure) with GARCH(1,1) disturbances, and estimate the model by quasi maximum likelihood. The resulting standardized output series are then tested for normality by applying the Bai and Ng (2005) for illustrative purposes. This test is not designed for use with such residuals; however, its outcomes can still be quite informative. For GDP, we still find evidence of non-normality (BN= 6.917); regarding GDI, there is some evidence against normality as well, albeit weaker than for GDP (BN= 4.731).[15]

The estimated persistence of the GARCH component (as quantified by the sum of the ARCH and the GARCH parameters) equals 1.003 (GDP) and 0.957 (GDI). Given the low frequency of

GDI growth rates.

[14]Similar to our previous analysis, we test for structural changes in the variance in the subsample as well. When excluding the Great Moderation, no such structural break is found in the raw series further supporting our interpretation that the structural change due to the Great Moderation may be the main driver behind the rejection of the Bai and Ng (2005) test.

[15]When applying our PIT-based tests to these residuals (without local standardization), see (4) and (7), we come to a similar conclusion.

the quarterly recorded observations, such strong persistence is rather indicative of the typical IGARCH effect occurring when deterministic changes in the unconditional variance are ignored; see Lamoureux and Lastrapes (1990), Mikosch and Stărică (2004) and Hillebrand (2005) inter alia. Therefore, the fitted AR-GARCH filters presumably do not remove the unconditional variation of the mean and variance satisfactorily – thus providing a plausible explanation of the observed residual non-normality.

This is further substantiated by simple trend tests applied to detect remaining time-variation.[16] We compare three sets of residuals. First, we examine the AR-GARCH standardized residuals. Second, we extend the standard GARCH component and fit a time-varying (TV-)GARCH (Amado and Teräsvirta, 2013) variance equation to account for potential (smoothly) changing unconditional variances. Third, we examine the locally standardized series $\hat{z}_t$. The t-statistics (with HAC standard errors) for the absence of linear trends are used as a diagnostic tool to check whether some time-variation is remaining. Overall, the results suggest that the time-variation in the mean is not well captured by the AR-GARCH (with a t-statistic of -2.02 for GDP and -1.82 for GDI) and AR-TV-GARCH (GDP: -2.49; GDI: -1.91) specifications, while nonparametric standardization is essentially more successful in this respect (GDP: 0.07; GDI: 0.24). To check for remaining variance variation, we detrend and square the three sets of residuals, and examine the squares by means of a subsequent trend test. The corresponding t-statistics for AR-TV-GARCH models (GDP: -0.58; GDI: -0.88) are lower in absolute value than those for standard AR-GARCH models (GDP: -1.62; GDI: -1.85) indicating that the former pick up more time-variation in variance, but still less than for the nonparametrically standardized series (GDP: -0.24; GDI: -0.02).

Summing up, evidence on non-normality of U.S. real output growth vanishes after accounting for unconditional time-variation in mean and variance.

6 Concluding remarks

This work considers the long-standing issue of testing distributional assumptions for dependent time series processes with structural instabilities in view. A case of leading interest is the one of normality of U.S. output growth rates, whose time-varying mean and volatility are well-documented.

We cope with time-varying unconditional means and variances by flexible nonparametric local standardization. These types of structural shifts are omnipresent in many economic and financial time series such as economic growth and stock market volatility. In a first step, the time series under consideration is locally standardized. Second, the applied probability integral transformation produces a uniformly distributed random process under the null hypothesis of a certain distribution to be tested for. The advantage of applying the PIT stems from the fact that a wide range of distributions (besides the normal) can be taken as the null distribution. The newly proposed tests are based on simple raw moment conditions which are very simple to obtain for

[16] As shown in Gadea and Gonzalo (2020) such a least squares-based t-test is consistent and powerful for a wide range of possible kinds of trends.

the uniform distribution. In addition, they can be estimated precisely even in small samples and the usage of PITs leads to tests which are quite sensitive towards deviations from theoretical values under the null restrictions.

The framework we provide makes use of the so-called fixed-bandwidth approach for the estimation of long-run covariance matrices of different raw moments in order to account for the serial dependence in the time series. Importantly, when accounting for the estimation effect arising from the nonparametric estimation of the first and second moment of the raw series, it turns out that the estimation error is automatically taken into account in the fixed-b framework in the limit. This is due to the involved fixed-bandwidth asymptotics in connection to the asymptotic behaviour of partial sums on the PITs. Therefore, no explicit asymptotic correction for the estimation error is needed. When turning to finite-samples, we furthermore advocate a simple adjustment consisting of a double standardization of the time series prior to computing the PITs and adjusting the long-run variance estimator accordingly. As shown, the limiting behaviour is unaffected by this correction.

We demonstrate in a simulation study that the suggested tests perform well in finite samples. The proposed tests outperform a benchmark test in the case of testing for normality against several alternative distributions. While the choice of the kernel for the nonparametric estimator of time-varying means and variances and the estimator itself only plays a subordinate role, the window width needs to be chosen in a way that the theoretical requirement of undersmoothing is reflected in practice. To this end, we advise in favor of an easy-to-use downscaled cross-validation rule and demonstrate its effectiveness in finite samples.

In our analysis of U.S. output growth rates, we demonstrate the merits and limitations of the robust raw moment-based statistics. The evidence provided here does not support the claim that U.S. aggregated output growth rates follow a non-normal distribution. In particular, we find U.S. output growth rates to be normally distributed after accounting for time-varying means and variances in a nonparametric way. Ignoring such instabilities (e.g. the Great Moderation) or improperly accounting for them, may lead to the biased conclusion that U.S. output growth rates are in fact non-normally distributed. The property of normality has important consequences for the understanding of "macroeconomic tail risks" whose concept builds on non-normal sectoral shocks leading to non-normal aggregate output growth as in e.g. Acemoglu et al. (2017). Moreover, our finding has direct implications for the related literature on modelling and predicting "Growth-at-Risk"; see e.g. Adrian et al. (2019) and Brownlees and Souza (2021).

References

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2017). Microeconomic origins of macroeconomic tail risks. *American Economic Review 107*(1), 54–108.

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Almuzara, M., D. Amengual, and E. Sentana (2019). Normality tests for latent variables. *Quantitative Economics 10*(3), 981–1017.

Amado, C. and T. Teräsvirta (2013). Modelling volatility by variance decomposition. *Journal of Econometrics 175*(2), 142–153.

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

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica 65*(5), 1097–1128.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), 953–966.

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2017). Tracking the slowdown in long-run GDP growth. *The Review of Economics and Statistics 99*(2), 343–356.

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics 191*(2), 384–397.

Ascari, G., G. Fagiolo, and A. Roventini (2015). Fat-tail distributions and business-cycle models. *Macroeconomic Dynamics 19*(2), 465–476.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics 85*(3), 531–549.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Bao, Y. (2013). On sample skewness and kurtosis. *Econometric Reviews 32*(4), 415–448.

Blanchard, O. and J. Simon (2001). The long and large decline in u.s. output volatility. *Brookings Papers on Economic Activity 2001*(1), 135–174.

Bontemps, C. and N. Meddahi (2005). Testing normality: A GMM approach. *Journal of Econometrics 124*(1), 149–186.

Bontemps, C. and N. Meddahi (2012). Testing distributional assumptions: A GMM aproach. *Journal of Applied Econometrics 27*(6), 978–1012.

Brownlees, C. and A. B. Souza (2021). Backtesting global growth-at-risk. *Journal of Monetary Economics 118*, 312–330.

Campbell, S. D. (2007). Macroeconomic volatility, predictability, and uncertainty in the Great Moderation: Evidence from the Survey of Professional Forecasters. *Journal of Business & Economic Statistics 25*(2), 191–200.

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

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics*, Volume 30, pp. 351–413. Elsevier.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

De Grauwe, P. (2012). Booms and busts in economic activity: A behavioral explanation. *Journal of Economic Behavior & Organization 83*(3), 484–501.

Deng, A. and P. Perron (2008). The limit distribution of the CUSUM of squares test under general mixing conditions. *Econometric Theory 24*(3), 809–822.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–883.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, 279–290.

Fagiolo, G., M. Napoletano, and A. Roventini (2008). Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal of Applied Econometrics 23*(5), 639–669.

Gadea, M. D., A. Gómez-Loscos, and G. Pérez-Quirós (2018). Great Moderation and Great Recession: From plain sailing to stormy seas? *International Economic Review 59*(4), 2297–2321.

Gadea, M. D. and J. Gonzalo (2020). Trends in distributional characteristics: Existence of global warming. *Journal of Econometrics 214*(1), 153–174.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2020). Identification and estimation in non-fundamental structural VARMA models. *The Review of Economic Studies 87*(4), 1915–1953.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, 357–384.

Herwartz, H. and H. Lütkepohl (2014). Structural vector autoregressions with markov switching: Combining conventional with statistical identification of shocks. *Journal of Econometrics 183*(1), 104–116.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics 129*(1-2), 121–138.

Jacobs, J. P. A. M., S. Sarferaz, J.-E. Sturm, and S. van Norden (2020). Can GDP measurement be further improved? Data revision and reconciliation. *Journal of Business & Economic Statistics*, forthcoming.

Jacobs, J. P. A. M. and S. Van Norden (2011). Modeling data revisions: Measurement error and dynamics of "true" values. *Journal of Econometrics 161*(2), 101–109.

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters 6*(3), 255–259.

Justiniano, A. and G. Primiceri (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review 98*(3), 604–641.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability & Its Applications 26*(2), 240–257.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 1130–1164.

Kiss, T. and P. Österholm (2020). Fat tails in leading indicators. *Economics Letters 193*, 109317.

Knüppel, M. (2015). Evaluating the calibration of multi-step-ahead density forecasts using raw moments. *Journal of Business & Economic Statistics 33*(2), 270–281.

Lamoureux, C. G. and W. D. Lastrapes (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics 8*(2), 225–234.

Landefeld, J. S., E. P. Seskin, and B. M. Fraumeni (2008). Taking the pulse of the economy: Measuring GDP. *Journal of Economic Perspectives 22*(2), 193–216.

Lanne, M. and H. Lütkepohl (2008). Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking 40*(6), 1131–1149.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010). Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control 34*(2), 121–131.

Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics 196*(2), 288–304.

Lewis, D. J. (2017). Robust inference in models identified via heteroskedasticity. *The Review of Economics and Statistics*, 1–45.

Lomnicki, Z. A. (1961). Tests for departure from normality in the case of linear stochastic processes. *Metrika 4*(1), 37–62.

Luo, S. and R. Startz (2014). Is it one break or ongoing permanent shocks that explains US real GDP? *Journal of Monetary Economics 66*, 155–163.

Lütkepohl, H., M. Meitz, A. Netšunajev, and P. Saikkonen (2021). Testing identification via heteroskedasticity in structural vector autoregressive models. *The Econometrics Journal 24*(1), 1–22.

MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics 6*(2), 422–433.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review 90*(5), 1464–1476.

Mikosch, T. and C. Stărică (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics 86*(1), 378–390.

Morley, J. and J. Piger (2012). The asymmetric business cycle. *The Review of Economics and Statistics 94*(1), 208–221.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Perron, P. and T. Wada (2009). Let's take a break: Trends and cycles in US real GDP. *Journal of Monetary Economics 56*(6), 749–765.

Ramey, V. A. and D. J. Vine (2006). Declining volatility in the US automobile industry. *American Economic Review 96*(5), 1876–1889.

Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics 85*(4), 777–792.

Rossi, B. and T. Sekhposyan (2014). Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting 30*(3), 662–682.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics 86*(3), 833–839.

Stock, J. H. and M. W. Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*(1), 159–218.

Sun, Y. (2014a). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica 82*(6), 2327–2370.

Sun, Y. (2014b). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics 178*(3), 659–677.

Vogelsang, T. J. and M. Wagner (2013). A fixed-b perspective on the Phillips-Perron unit root tests. *Econometric Theory 29*, 609–628.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics 40*(5), 2601–2633.

Wang, X. and Y. Sun (2022). A simple asymptotically F-distributed portmanteau test for diagnostic checking of time series models with uncorrelated innovations. *Journal of Business & Economic Statistics 40*(2), 505–521.

Yang, J. and T. J. Vogelsang (2011). Fixed-b analysis of LM-type tests for a shift in mean. *The Econometrics Journal 14*(3), 438–456.

Appendix

To simplify notation in the proofs, we let w.l.o.g. $\tau_\mu = \tau_\sigma = \tau$. In terms of notation, C stands for a generic constant whose value may change from one occurrence to another and the probabilistic Landau symbols O_p and o_p have their usual meaning.

A Preliminary results

Lemma A.1 *Let g, h be two functions such that $\mathrm{E}(g(z_t)) = \mathrm{E}(h(z_t)) = 0$ and $g(x) = h(x) = O(x^2)$ as $x \to \pm\infty$. Under the Assumptions of Lemma 1, we have that*

1. $\sup_{t=1,\ldots,T} \left| \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} w_j g(z_j) \right| = O_p\left(T^{-\delta}\right)$ *for some* $0 < \delta < 1/8$ *and any bounded sequence* w_j*;*
2. $\sup_{t=1,\ldots,T} |\hat{\mu}_t - \mu_t| = O_p\left(T^{-\delta}\right)$ *and* $\sup_{t=1,\ldots,T} |\hat{\sigma}_t - \sigma_t| = O_p\left(T^{-\delta}\right)$ *for some* $0 < \delta < 1/8$*;*
3. $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} g(z_t) \left(\frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} h(z_j) \right) = o_p(1)$*;*
4. $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} g(z_t) \left(\frac{\sigma_t}{\hat{\sigma}_t} - 1 \right) = o_p(1)$*;*
5. $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} (\hat{z}_t - z_t)^2 = o_p(1)$*;*
6. $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left(\varphi(z_t)(\hat{z}_t - z_t) + \varphi'(\xi_t)(\hat{z}_t - z_t)^2 \right)^2 = o_p(1)$*, where ξ_t lies between z_t and $\hat{z}_t$ for any $1 \le t \le T$;*
7. $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} p_t^{k-1} \varphi(z_t) \left(\frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} \frac{\sigma_j}{\hat{\sigma}_t} z_j \right) = \mathrm{E}\left(p_t^{k-1} \varphi(z_t) \right) \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} z_t + o_p(1)$*;*
8. $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} p_t^{k-1} \varphi(z_t) z_t \left(\frac{\sigma_t}{\hat{\sigma}_t} - 1 \right) = -\frac{1}{2} \mathrm{E}\left(p_t^{k-1} z_t \varphi(z_t) \right) \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left(z_t^2 - 1 \right) + o_p(1)$*,*

where the $o_p(1)$ terms are uniform in $s \in [0,1]$.

B Proofs

Proof of Lemma A.1

Lemma A.1 contains eight individual results and they are proven item by item in the following.

Proof of item 1

We first show that $\frac{1}{\sqrt{\tau}} \sum_{j=t-\tau}^{t+\tau} w_j g(z_j)$ is uniformly L_4-bounded in $t = 1, \ldots, T$. We have

$$
\begin{aligned}
\left\| \frac{1}{\sqrt{\tau}} \sum_{j=t-\tau}^{t+\tau} w_j g(z_j) \right\|_4^4 &= \mathrm{E}\left(\left(\frac{1}{\sqrt{\tau}} \sum_{j=t-\tau}^{t+\tau} w_j g(z_j) \right)^4 \right) \\
&= \frac{1}{\tau^2} \sum_{j_1=t-\tau}^{t+\tau} \sum_{j_2=t-\tau}^{t+\tau} \sum_{j_3=t-\tau}^{t+\tau} \sum_{j_4=t-\tau}^{t+\tau} w_{j_1} w_{j_2} w_{j_3} w_{j_4} \mathrm{E}\left(g(z_{j_1}) g(z_{j_2}) g(z_{j_3}) g(z_{j_4}) \right).
\end{aligned}
$$

Now, an upper bound is given by

$$\left\| \frac{1}{\sqrt{\tau}} \sum_{j=t-\tau}^{t+\tau} w_j g(z_j) \right\|_4^4 \leq \frac{C}{\tau^2} \sum_{j_1=t-\tau}^{t+\tau} \sum_{j_2=t-\tau}^{t+\tau} \sum_{j_3=t-\tau}^{t+\tau} \sum_{j_4=t-\tau}^{t+\tau} \mathrm{E}\left(z_{j_1}^2 z_{j_2}^2 z_{j_3}^2 z_{j_4}^2\right) < \infty$$

where the finiteness of this upper bound follows with standard arguments from the absolute summability of the 8th-order cumulants of z_t.

Then, the maximum over T elements of a positive, uniformly L_4-bounded sequence is known to be $O_p\left(T^{1/4}\right)$, so

$$\sup_{t=1,\dots,T} \left| \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} w_j g(z_j) \right| = O_p\left(\frac{\sqrt[4]{T}}{\sqrt{\tau}}\right),$$

from which the desired result follows given the rate restrictions on τ.

Proof of item 2

Let us examine the properties of $\hat{\mu}_t$ first. We have that

$$\hat{\mu}_t - \mu_t = \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} (x_j - \mu_t) = \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} (\mu_j - \mu_t) + \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} \sigma_j z_j.$$

Thanks to the Lipschitz property of $\mu(\cdot)$, the first summand on the r.h.s. is $O_p\left(\frac{\tau}{T}\right)$ uniformly in $t = 1, \dots, T$. Item 1 can be used to derive the uniform behavior of the second summand, such that $\sup_{t=1,\dots,T} |\hat{\mu}_t - \mu_t| = O_p\left(T^{-\delta}\right)$ for some $0 < \delta < 1/8$ as required. The local variance estimator is given by

$$\hat{\sigma}_t^2 = \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} (x_j - \hat{\mu}_j)^2 = \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} (\sigma_j z_j - (\hat{\mu}_j - \mu_j))^2$$

so

$$\hat{\sigma}_t^2 - \sigma_t^2 = \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} \left(\sigma_j^2 z_j^2 - \sigma_t^2\right) - \frac{2}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} \sigma_j z_j (\hat{\mu}_j - \mu_j) + \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} (\hat{\mu}_j - \mu_j)^2.$$

Now,

$$\sup_{t=1,\dots,T} \left| \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} \sigma_j z_j (\hat{\mu}_j - \mu_j) \right| \leq \sup_{t=1,\dots,T} |\hat{\mu}_j - \mu_j| \sup_{t=1,\dots,T} \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} |\sigma_j z_j|$$

where

$$0 \leq \sup_{t=1,\dots,T} \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} |\sigma_j z_j| \leq \mathrm{E}(|z_t|) \sup_{t=1,\dots,T} \sigma_t + \sup_{t=1,\dots,T} \frac{1}{2\tau+1} \sum_{j=t-\tau}^{t+\tau} \sigma_j (|z_j| - \mathrm{E}(|z_j|)) = O_p(1)$$

with the same arguments used in the proof of item 1. Furthermore, for all $1 \leq t \leq T$,

$$\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\hat{\mu}_j-\mu_j\right)^2 \leq \left(\sup_{t=1,\ldots,T}\left|\hat{\mu}_j-\mu_j\right|\right)^2 = o_p\left(1\right)$$

so, after using item 1 again to conclude that $\sup_{t=1,\ldots,T}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j^2\left(z_j^2-1\right)=O_p\left(T^{-\delta}\right)$ for some $0<\delta<1/8$, we have that

$$\begin{aligned}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\sigma_j^2 z_j^2-\sigma_t^2\right) &= \frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j^2\left(z_j^2-1\right)+\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\sigma_j^2-\sigma_t^2\right)\\ &= O_p\left(T^{-\delta}\right)+O\left(\frac{\tau}{T}\right)\end{aligned}$$

uniformly in $t=1,\ldots,T$ and thus $\sup_{t=1,\ldots,T}\left|\hat{\sigma}_t^2-\sigma_t\right|=O_p\left(T^{-\delta}\right)$ as well.

Note that uniform consistency of $\hat{\sigma}_t$ implies, thanks to the properties of σ_t, $\sup_{t=1,\ldots,T}\hat{\sigma}_t=O_p\left(1\right)$ and $\sup_{t=1,\ldots,T}\hat{\sigma}_t^{-1}=O_p\left(1\right)$.

Proof of item 3

Split the sample in B disjoint blocks of length M and assume that $T=MB$ and $[sT]=M[sB]$ for the sake of the exposition. Then

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}g\left(z_t\right)\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}h\left(z_j\right)\right)=\\ &=\frac{1}{\sqrt{T}}\sum_{b=1}^{[sB]}\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\left(\frac{1}{2\tau+1}\left(\sum_{j=M(b-1)+m-\tau}^{M(b-1)+m+\tau}h\left(z_j\right)-\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right)\right)\\ &\quad+\frac{1}{\sqrt{T}}\sum_{b=1}^{[sB]}\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\left(\frac{1}{2\tau+1}\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right).\end{aligned}$$

The first summand on the r.h.s. is easily shown to be $O_p\left(\frac{\sqrt{TM}}{\tau}\sup_{t\in1,\ldots,T}\left|g\left(z_t\right)\right|\right)$. Now, given the finiteness of moments of any order of z_t and thus of z_t^2 and $g\left(z_t\right)$, we have $\sup_t\left|g\left(z_t\right)\right|=O_p\left(T^{\gamma}\right)=\sup_{t=1,\ldots,T}z_t^2$ for any $\gamma>0$, such that the first term on the r.h.s. is $O_p\left(\frac{T^{1+\gamma}}{\tau\sqrt{B}}\right)$. For the second, note that

$$\begin{aligned}&\left|\frac{1}{\sqrt{T}}\sum_{b=1}^{[sB]}\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\left(\frac{1}{2\tau+1}\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right)\right|\\ &\quad\leq\frac{1}{\sqrt{T}}\sum_{b=1}^{B}\left|\left(\frac{1}{2\tau+1}\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right)\left(\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\right)\right|.\end{aligned}$$

The expectation of the r.h.s. is given by

$$\begin{aligned}
&\sum_{b=1}^{B}\mathrm{E}\left(\left|\left(\frac{1}{2\tau+1}\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right)\left(\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\right)\right|\right)\\
&\quad\leq\sqrt{\mathrm{E}\left(\left(\frac{1}{2\tau+1}\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right)^2\right)\mathrm{E}\left(\left(\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\right)^2\right)}
\end{aligned}$$

where

$$\mathrm{E}\left(\left(\frac{1}{2\tau+1}\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right)^2\right)=O\left(\frac{1}{\tau}\right)$$

and

$$\mathrm{E}\left(\left(\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\right)^2\right)=O\left(M\right).$$

Hence

$$\frac{1}{\sqrt{T}}\sum_{b=1}^{[sB]}\sum_{m=1}^{M}g\left(z_{M(b-1)+m}\right)\left(\frac{1}{2\tau+1}\sum_{j=M(b-1)-\tau}^{M(b-1)+\tau}h\left(z_j\right)\right)=O_p\left(\frac{B\sqrt{M}}{\sqrt{\tau T}}\right)=O_p\left(\sqrt{\frac{B}{\tau}}\right)$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}g\left(z_t\right)\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}h\left(z_j\right)\right)=O_p\left(\max\left\{\frac{\sqrt{M}}{\tau}T^{1/2+\gamma};\sqrt{\frac{B}{\tau}}\right\}\right);$$

since γ may be chosen arbitrarily small, picking $B=T^{\eta}$ such that $2/3<\eta<\nu_1$ leads to the desired result.

Proof of item 4

Use a Taylor series expansion for $x^{-1/2}$ about $x_0=1$ with rest term in differential form to obtain

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}g\left(z_t\right)\left(\frac{\sigma_t}{\hat{\sigma}_t}-1\right) &= -\frac{1}{2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}g\left(z_t\right)\left(\frac{\hat{\sigma}_t^2}{\sigma_t^2}-1\right)+\frac{3}{8}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}g\left(z_t\right)\xi_t^{-5/2}\left(\frac{\hat{\sigma}_t^2}{\sigma_t^2}-1\right)^2\\
&= A_{1T}+A_{2T}
\end{aligned}$$

with ξ_t between $\frac{\hat{\sigma}_t^2}{\sigma_t^2}$ and unity for all $t = 1, \ldots, T$. Now, for A_{1T}, write

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} g(z_t)\left(\frac{\hat{\sigma}_t^2}{\sigma_t^2}-1\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} g(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\frac{1}{\sigma_t^2}\left(\sigma_j z_j + (\mu_j-\hat{\mu}_j)\right)^2-1\right) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \frac{g(z_t)}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\sigma_j^2 z_j^2-\sigma_t^2}{\sigma_t^2} + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \frac{g(z_t)}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{(\hat{\mu}_j-\mu_j)^2}{\sigma_t^2} \\
&\quad + \frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]} g(z_t)\frac{1}{\sigma_t^2}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j(\mu_j-\hat{\mu}_j) \\
&= B_{1T}+B_{2T}+B_{3T}.
\end{aligned}$$

For B_{1T}, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} g(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\sigma_j^2 z_j^2-\sigma_t^2}{\sigma_t^2} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} g(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(z_j^2-1\right) \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} g(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left(\sigma_j^2-\sigma_t^2\right)}{\sigma_t^2}z_j^2,
\end{aligned}$$

where the first summand on the r.h.s. vanishes thanks to item 3, while for the second we employ a Taylor series approximation of $\sigma^2(\cdot)$ about t/T to obtain

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \frac{g(z_t)}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left(\sigma_j^2-\sigma_t^2\right)z_j^2}{\sigma_t^2} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \frac{g(z_t)}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left.\frac{\partial\sigma^2}{\partial s}\right|_{s=\frac{t}{T}}\frac{j-t}{T}\left(z_j^2-1\right)}{\sigma_t^2} \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \frac{g(z_t)}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left.\frac{\partial\sigma^2}{\partial s}\right|_{s=\frac{t}{T}}\frac{j-t}{T}}{\sigma_t^2} \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \frac{g(z_t)}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left.\frac{\partial^2\sigma^2}{\partial s^2}\right|_{s=\xi_{t,j}}\frac{(j-t)^2}{T^2}z_j^2}{\sigma_t^2} \\
&= C_{1T}+C_{2T}+C_{3T}
\end{aligned}$$

for suitable $\xi_{t,j}$ between t/T and $j/T - t/T$. Here, C_{1T} vanishes along the lines of item 3 by noting that deterministic weights do not affect the result, $C_{2T} = 0$ and

$$|C_{3T}| \leq \frac{sC\tau^2}{T\sqrt{T}}\sup_{t=1,\ldots,T}|g(z_t)|\sup_{t=1,\ldots,T} z_t^2;$$

this is seen to vanish too uniformly in $s \in [0,1]$ since, given the finiteness of moments of any order of z_t and thus of z_t^2 and $g(z_t)$, we have $\sup_t |g(z_t)| = O_p(T^\gamma) = \sup_{t=1,\ldots,T} z_t^2$ for any $\gamma > 0$, and γ can then be chosen arbitrarily close to 0 to make the r.h.s. $o_p(1)$.

For B_{2T}, we have

$$\begin{aligned}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} g\left(z_t\right)\frac{1}{\sigma_t^2}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\hat{\mu}_j-\mu_j\right)^2\right| &\leq C\sup_t\left|g\left(z_t\right)\right|\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\hat{\mu}_j-\mu_j\right)^2 \\ &\leq C\sup_t\left|g\left(z_t\right)\right|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mu_t-\hat{\mu}_t\right)^2+o_p(1)\end{aligned}$$

with $|g\left(z_t\right)| = O_p\left(T^{\gamma}\right)$ for any $\gamma > 0$. We show in the following that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mu_t-\hat{\mu}_t\right)^2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\left(\mu_t-\mu_j\right)-\sigma_j z_j\right)\right)^2 = O_p\left(T^{-\delta}\right) \tag{14}$$

for some $0 < \delta < \min\{\nu_1 - 1/2; 3/4 - \nu_2\}$, and simply pick $\gamma < \delta$ for our purposes. With the help of the Cauchy-Schwarz inequality, the term is easily seen to vanish when the terms $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\mu_t-\mu_j\right)\right)^2$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\right)^2$ both vanish themselves. This is indeed the case under our rate restrictions considering that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\mu_t-\mu_j\right)\right)^2 = O_p\left(\sqrt{T}\frac{\tau^2}{T^2}\right)$$

and

$$\mathrm{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\right)^2\right) \leq \sqrt{T}\,\mathrm{E}\left(\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\right)^2\right) = C\frac{\sqrt{T}}{\tau}$$

thanks to the uniform L_4-boundedness of normalized running averages of z_t; see the proof of item 1. Thus, B_{2T} vanishes at the required rate.

Moving on, we have

$$\begin{aligned}B_{3T} &= \frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{g\left(z_t\right)}{\sigma_t^2}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\left(\mu_j-\frac{1}{2\tau+1}\sum_{k=j-\tau}^{j+\tau}\sigma_k z_k-\frac{1}{2\tau+1}\sum_{k=j-\tau}^{j+\tau}\mu_k\right) \\ &= -\frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{g\left(z_t\right)}{\sigma_t^2}\frac{1}{\left(2\tau+1\right)^2}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\sum_{k=j-\tau}^{j+\tau}\left(\mu_k-\mu_j\right) \\ &\quad -\frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{g\left(z_t\right)}{\sigma_t^2}\frac{1}{\left(2\tau+1\right)^2}\left(\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\sum_{k=j-\tau}^{j+\tau}\sigma_k z_k\right),\end{aligned}$$

where the first summand on the r.h.s. is immediately shown to vanish thanks to item 3 after noting that deterministic weights of uniform order $O\left(\tau/T\right)$ do not affect the arguments there. A tedious, yet straightforward application of the blocking arguments from the proof of item 3 shows the second summand to vanish in probability as well.

We have $\sup_{s\in[0,1]}|A_{1T}| \xrightarrow{p} 0$ and in addition we note that

$$0 \le \sup_{s\in[0,1]} |A_{2T}| \le C \sup_{t=1,\dots,T} \left|\xi_t^{-5/2}\right| \sup_{t=1,\dots,T} |g(z_t)| \sup_{s\in[0,1]} \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left(\frac{\hat{\sigma}_t^2}{\sigma_t^2} - 1\right)^2,$$

where the r.h.s., and thus A_{2T}, vanishes since $\sup_{t=1,\dots,T}\left|\xi_t^{-5/2}\right| = O_p(1)$, $\sup_{t=1,\dots,T}|g(z_t)| = O_p(T^\gamma)$ for any positive γ and $\sup_{s\in[0,1]} \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\hat{\sigma}_t^2}{\sigma_t^2} - 1\right)^2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_t^2}{\sigma_t^2} - 1\right)^2 = O_p(T^{-\delta})$, analogously to Equation (14), so the desired result follows after choosing $\gamma < \delta$.

Proof of item 5

We have that

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} - z_t\right)^2 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(z_t\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right) + \frac{\mu_t - \hat{\mu}_t}{\hat{\sigma}_t}\right)^2 \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} z_t^2\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right)^2 + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{\hat{\sigma}_t^2}(\mu_t - \hat{\mu}_t)^2 \\ &\quad + \frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]} z_t\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right)\frac{1}{\hat{\sigma}_t}(\mu_t - \hat{\mu}_t). \end{aligned}$$

Noting that $\sup_{t=1,\dots,T}\hat{\sigma}_t$ is bounded in probability, an application of the Cauchy-Schwarz inequality for the third summand on the r.h.s. shows that the result follows given that the two terms $\sup_{s\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} z_t^2\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right)^2$ and $\sup_{s\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{\hat{\sigma}_t^2}(\mu_t - \hat{\mu}_t)^2$ vanish in probability. To show this, we have like in the proof of item 4 that

$$\begin{aligned} \left|\sup_{s\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} z_t^2\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right)^2\right| &\le \sup_{t=1,\dots,T} z_t^2 \sup_{s\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right)^2 = \sup_{t=1,\dots,T} z_t^2 \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right)^2 \\ &= o_p(1) \end{aligned}$$

since $\sup_{t=1,\dots,T} z_t^2 = O_p(T^\gamma)$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_t}{\sigma_t} - 1\right)^2 = O_p(T^{-\delta})$, where $\gamma < \delta$ may be picked, and similarly

$$0 \le \sup_{s\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{\hat{\sigma}_t^2}(\mu_t - \hat{\mu}_t)^2 \le \sup_{t=1,\dots,T}\frac{1}{\hat{\sigma}_t^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mu_t - \hat{\mu}_t)^2 = o_p(1),$$

since $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mu_t - \hat{\mu}_t)^2 = o_p(1)$, again like in the proof of item 4.

Proof of item 6

Note that $r_t = \left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} - z_t\right)\left(\varphi(z_t) + \varphi'(\xi_t)\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} - z_t\right)\right)$ where φ and φ' are bounded. The result follows with item 5 if $\sup_t\left|\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} - z_t\right| = O_p(1)$. This is indeed the case, since

$$\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} - z_t = \left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right) z_t + \frac{\mu_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

where $\hat{\mu}_t$ and $\hat{\sigma}_t$ converge uniformly at some rate $O_p\left(T^{\delta}\right)$, see item 2, and $\sup_t |z_t| = O_p\left(T^{\gamma}\right)$ for any $\gamma > 0$ such that choosing $\gamma < \delta$ leads to the desired result.

Proof of item 7

Begin by writing

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\sigma_j}{\hat{\sigma}_t}z_j &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\frac{\sigma_j}{\hat{\sigma}_t} - 1\right)z_j \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau} z_j \\
&= A_{1T} + A_{2T}.
\end{aligned}$$

We now show the first summand (A_{1T}) to vanish and resort to this end to the Taylor series approximation of $x^{-1/2}$ employed in the proof of item 4 to obtain analogously

$$\begin{aligned}
A_{1T} &= -\frac{1}{2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\frac{\hat{\sigma}_t^2}{\sigma_j^2} - 1\right)z_j \\
&\quad + \frac{3}{8}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\xi_{t,j}^{-5/2}\left(\frac{\hat{\sigma}_t^2}{\sigma_j^2} - 1\right)^2 z_j
\end{aligned}$$

where $\xi_{t,j}$ lies between $\frac{\sigma_j}{\hat{\sigma}_t}$ and unity for all $t = 1, \ldots, T$, being hence uniformly bounded. The first summand of A_{1T} can be shown to be negligible by writing

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\frac{\hat{\sigma}_t^2}{\sigma_j^2} - 1\right)z_j \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(p_t^{k-1}\varphi(z_t) - \mathrm{E}\left(p_t^{k-1}\varphi(z_t)\right)\right)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\frac{\hat{\sigma}_t^2}{\sigma_j^2} - 1\right)z_j \\
&\quad + \mathrm{E}\left(p_t^{k-1}\varphi(z_t)\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\hat{\sigma}_t^2}{\sigma_j^2} - 1\right)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau} z_j;
\end{aligned}$$

and noting that arguments analog to those in the proof of item 3 apply.

For the second summand of A_{1T}, with $\varphi(\cdot)$ being bounded on $\mathbb{R}$, we have

$$\begin{aligned} 0 &\leq \sup_{s\in[0,1]}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\xi_{t,j}^{-5/2}\left(\frac{\hat\sigma_t^2}{\sigma_j^2}-1\right)^2 z_j\right| \\ &\leq \max_{x\in\mathbb{R}}\varphi(x)\sup_t|z_t|\sup_{t,j}\left(\xi_{t,j}^{-5/2}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\frac{\hat\sigma_t^2}{\sigma_j^2}-1\right)^2, \end{aligned}$$

where $\frac{\hat\sigma_t^2}{\sigma_j^2}=\frac{\hat\sigma_t^2}{\sigma_t^2}\frac{\sigma_t^2}{\sigma_j^2}=(1+O(\tau/T))\frac{\hat\sigma_t^2}{\sigma_t^2}$. Then, with $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\hat\sigma_t^2}{\sigma_t^2}-1\right)^2$ vanishing like in the proof of item 4 and $\sup_t|z_t|=O_p(T^\gamma)$ for positive γ arbitrarily close to zero, it is not difficult to show that the second summand of A_{1T} vanishes.

To complete the result, write

$$\begin{aligned} A_{2T} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(p_t^{k-1}\varphi(z_t)-\mathrm{E}\left(p_t^{k-1}\varphi(z_t)\right)\right)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}z_j \\ &\quad +\mathrm{E}\left(p_t^{k-1}\varphi(z_t)\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}z_j, \end{aligned}$$

where the first summand on the r.h.s. vanishes thanks to item 3, while the second delivers the desired approximation upon re-arranging its sum elements.

Proof of item 8

Write

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}p_t^{k-1}\varphi(z_t)z_t\left(\frac{\sigma_t}{\hat\sigma_t}-1\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(p_t^{k-1}\varphi(z_t)z_t-\mathrm{E}\left(p_t^{k-1}\varphi(z_t)z_t\right)\right)\left(\frac{\sigma_t}{\hat\sigma_t}-1\right) \\ &\quad +\mathrm{E}\left(p_t^{k-1}\varphi(z_t)z_t\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\sigma_t}{\hat\sigma_t}-1\right). \end{aligned}$$

The first summand on the r.h.s. vanishes, see item 4, and, with the same Taylor series expansion of $x^{-1/2}$ employed there, we have for the second summand that

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\sigma_t}{\hat\sigma_t}-1\right) &= -\frac{1}{2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\hat\sigma_t^2}{\sigma_t^2}-1\right)+\frac{3}{8}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\xi_t^{-5/2}\left(\frac{\hat\sigma_t^2}{\sigma_t^2}-1\right)^2 \\ &= A_{1T}+A_{2T} \end{aligned}$$

with ξ_t lying between $\frac{\hat\sigma_t^2}{\sigma_t^2}$ and unity for all $t=1,\ldots,T$. The arguments in the proof of item 4 apply directly, with the exception of the analogues of B_{1T} and B_{3T}. For the analogue of B_{1T} from the proof of item 4 we write

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\sigma_j^2z_j^2-\sigma_t^2}{\sigma_t^2}=\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}(z_j^2-1)+\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left(\sigma_j^2-\sigma_t^2\right)z_j^2}{\sigma_t^2}$$

where the summands of the first term on the r.h.s. are re-arranged to give the desired approximation, and the second term is given, similarly to the proof of item 4, by

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left(\sigma_j^2-\sigma_t^2\right)z_j^2}{\sigma_t^2} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left.\frac{\partial\sigma^2}{\partial s}\right|_{s=\frac{t}{T}}\frac{j-t}{T}\left(z_j^2-1\right)}{\sigma_t^2}\\ &\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left.\frac{\partial\sigma^2}{\partial s}\right|_{s=\frac{t}{T}}\frac{j-t}{T}}{\sigma_t^2}\\ &\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\left.\frac{\partial^2\sigma^2}{\partial s^2}\right|_{s=\xi_{t,j}}\frac{(j-t)^2}{T^2}z_j^2}{\sigma_t^2}\\ &= C_{1T}+C_{2T}+C_{3T}\end{aligned}$$

for suitable $\xi_{t,j}$ between t/T and $j/T - t/T$. To analyze C_{1T}, re-arrange sum terms to obtain

$$C_{1T}=\frac{C}{\sqrt{T}}\frac{1}{(2\tau+1)\,T}\left(\sum_{t=1}^{[sT]}\left(z_t^2-1\right)\tau(\tau+1)+O_p\left(\tau^2\right)\right)=o_p(1)$$

uniformly in $s\in[0,1]$,

$$C_{2T}=0,$$

and, for all $s\in[0,1]$,

$$0\leq C_{3T}\leq\frac{C\tau^2}{T^2\sqrt{T}}\sum_{t=1}^{[sT]}z_t^2\leq\frac{C\tau^2}{T^2\sqrt{T}}\sum_{t=1}^{T}z_t^2=O_p\left(\frac{\tau^2}{T\sqrt{T}}\right)=o_p(1)\,.$$

For the analog of B_{3T} from the proof of item 4, we re-arrange sum terms to obtain

$$\frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{\sigma_t^2}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\left(\mu_j-\hat{\mu}_j\right)=\frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]}\sigma_t z_t\left(\mu_t-\hat{\mu}_t\right)\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{1}{\sigma_j^2}+o_p(1)\,.$$

To complete the result, write

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\sigma_t z_t\left(\mu_t-\hat{\mu}_t\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\sigma_t z_t\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\mu_t-\mu_j\right)\right)-\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\sigma_t z_t\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\sigma_j z_j\right)$$

where both summands on the r.h.s. can be shown to vanish uniformly in s using e.g. item 3.

Proof of Lemma 1

We let w.l.o.g. $\tau_\mu=\tau_\sigma=\tau$. Write with a Taylor expansion

$$\begin{aligned}\hat{p}_t &= p_t+\varphi(z_t)\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t}-z_t\right)+\varphi'(\xi_t)\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t}-z_t\right)^2\\ &= p_t+r_t,\end{aligned}$$

where ξ_t lies between $\frac{x_t-\mu}{\sigma_t} = z_t$ and $\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} = \hat{z}_t$; note that $\varphi'(\cdot)$ is bounded on $\mathbb{R}$. Then,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{p}_t^k - \frac{1}{k+1}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(p_t^k - \frac{1}{k+1}\right) + \frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1} r_t + \frac{k(k-1)}{2\sqrt{T}}\sum_{t=1}^{[sT]}\tilde{p}_t^{k-1} r_t^2$$

where $\tilde{p}_t$ lies between p_t and $\hat{p}_t$. Since $\tilde{p}_t \in [0,1]\ \forall t$, like p_t and $\hat{p}_t$, we have for the third term

$$0 \leq \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\tilde{p}_t^{k-1} r_t^2 \leq \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} r_t^2 \xrightarrow{p} 0$$

uniformly in s, thanks to Lemma A.1 item 6.

We may then focus on the second term and obtain

$$\frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1} r_t = \frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} - z_t\right) + \frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi'(\xi_t)\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} - z_t\right)^2.$$

Here, the second summand vanishes uniformly in s since

$$\left|\frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi'(\xi_t)\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t} - z_t\right)^2\right| \leq \frac{C}{\sqrt{T}}\sum_{t=1}^{[sT]}(\hat{z}_t - z_t)^2$$

due to the boundedness of φ' and p_t, and Lemma A.1 item 5 applies. Now,

$$\hat{z}_t - z_t = \frac{\sigma_t z_t + \mu_t - \hat{\mu}_t}{\hat{\sigma}_t} - z_t = z_t\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right) - \frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\sigma_j}{\hat{\sigma}_t} z_j + \frac{1}{\hat{\sigma}_t}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}(\mu_t - \mu_j),$$

such that the leading term of $\frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1} r_t$ is given by

$$\begin{aligned}\frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)(\hat{z}_t - z_t) &= \frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t) z_t\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right) \\ &\quad - \frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]} p_t^{k-1}\varphi(z_t)\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\sigma_j}{\hat{\sigma}_t} z_j\right) \\ &\quad + \frac{k}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{p_t^{k-1}\varphi(z_t)}{\hat{\sigma}_t}\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}(\mu_t - \mu_j)\right)\end{aligned}$$

where

$$\begin{aligned}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{\hat{\sigma}_t} p_t^{k-1}\varphi(z_t)\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}(\mu_t - \mu_j)\right)\right| &\leq \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{\hat{\sigma}_t} p_t^{k-1}\varphi(z_t)\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}|\mu_t - \mu_j|\right) \\ &= O_p\left(\frac{\tau}{\sqrt{T}}\right)\end{aligned}$$

with p_t and $\phi(z_t)$ being bounded and positive, and $\sup_t \hat{\sigma}_t$ and $\inf_t \hat{\sigma}_t$ bounded in probability and non-negative (due to $\hat{\sigma}_t$ being uniformly consistent for σ_t which is bounded and bounded away from 0).

Using Lemma A.1 again, items 7 and 8, we obtain uniformly in $s \in [0,1]$ that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{p}_t^k - \frac{1}{k+1}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(p_t^k - \frac{1}{k+1}\right) - k\,\mathrm{E}\left(p_t^{k-1}\varphi\left(z_t\right)\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} z_t - \frac{k}{2}\,\mathrm{E}\left(p_t^{k-1} z_t \varphi\left(z_t\right)\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(z_t^2 - 1\right) + o_p\left(1\right)$$

as required for the result which follows with a multivariate invariance principle for strongly mixing sequences (see e.g. Davidson, 1994, Chapter 29).

Proof of Proposition 1

To simplify notation we provide the arguments for $\hat{t}_k$ only; the extension for $K > 1$ is trivial.

The arguments in the proof of Theorem 2 in Kiefer and Vogelsang (2005) indicate that

$$\hat{t}_k = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{p}_t^k - \frac{1}{k+1}\right)}{\sqrt{-\frac{1}{T^2}\sum_{i=1}^{T-1}\sum_{j=1}^{T-1}\frac{T^2}{B^2}k''\left(\frac{i-j}{B}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{i}\left(\hat{p}_t^k - \overline{\hat{p}^k}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{j}\left(\hat{p}_t^k - \overline{\hat{p}^k}\right)}} + o_p\left(1\right)$$

for kernels with smooth derivatives, or

$$\hat{t}_k = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{p}_t^k - \frac{1}{k+1}\right)}{\sqrt{\frac{2}{bT}\sum_{i=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{i}\left(\hat{p}_t^k - \overline{\hat{p}^k}\right)\right)^2 - \frac{2}{bT}\sum_{i=1}^{[(1-b)T]}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{i}\left(\hat{p}_t^k - \overline{\hat{p}^k}\right)\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{i+[bT]}\left(\hat{p}_t^k - \overline{\hat{p}^k}\right)\right)}} + o_p\left(1\right)$$

for the Bartlett kernel. From Lemma 1, we know that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{p}_t^k - \frac{1}{k+1}\right) \Rightarrow B_k\left(s\right) - k\vartheta_{k-1}W_1\left(s\right) - \frac{k}{2}\varpi_{k-1}W_2\left(s\right),$$

where the process on the r.h.s. is Brownian motion. The vector stacking the K individual Brownian motions is itself a K-dimensional Brownian motion with covariance matrix $\mathbf{V\Xi V}'$, whose kth diagonal element, say $\tilde{\omega}_k^2$, is the variance of $\tilde{B}_k(s)$. Positive definiteness of $\mathbf{\Xi}$ ensures that $\tilde{\omega}_k^2 > 0$ which then cancels out, so the continuous mapping theorem [CMT] then establishes the desired limiting null distribution.

Proof of Proposition 2

We first have that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\hat{z}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{\sigma_t z_t - \hat{\mu}_t}{\hat{\sigma}_t} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} z_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} z_t\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right) - \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{\hat{\mu}_t}{\hat{\sigma}_t}.$$

Item 4 of Lemma A.1 shows the second term on the r.h.s. to vanish uniformly, while the arguments in the proof of item 7 imply for the last term that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{\hat{\mu}_t}{\hat{\sigma}_t}=\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\frac{\sigma_t}{\hat{\sigma}_t}z_t=\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}z_t+o_p(1)$$

uniformly in s, and therefore

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\hat{z}_t\Rightarrow 0.$$

Examine then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{z}_t^2-1\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(z_t^2-1\right)+\frac{2}{\sqrt{T}}\sum_{t=1}^{[sT]}z_t\left(\hat{z}_t-z_t\right)+\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{z}_t-z_t\right)^2,$$

where item 5 of Lemma A.1 indicates the third term on the r.h.s. to vanish uniformly in s. For analyzing the second term, write

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}z_t\left(\hat{z}_t-z_t\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}z_t^2\left(\frac{\sigma_t}{\hat{\sigma}_t}-1\right)-\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}z_t\frac{\hat{\mu}_t}{\hat{\sigma}_t}$$

where the first term on the r.h.s. is treated like in the proof of item 8 (using $\mathrm{E}\left(z_t^2\right)=1$) and the second vanishes uniformly in s along the lines of item 7, such that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}z_t\left(\hat{z}_t-z_t\right)=-\frac{1}{2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(z_t^2-1\right)+o_p(1)$$

uniformly in s. Summing up,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{z}_t^2-1\right)\Rightarrow 0.$$

This leads immediately to

$$\sqrt{T}\bar{\hat{z}}\xrightarrow{p}0\qquad\text{and}\qquad\sqrt{T}\left(\frac{1}{\tilde{\sigma}_{\hat{z}}}-1\right)\xrightarrow{p}0$$

as well as

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\tilde{z}_t=\frac{1}{\tilde{\sigma}_{\hat{z}}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\hat{z}_t-\frac{[sT]}{T}\sqrt{T}\bar{\hat{z}}\right)\Rightarrow 0$$

and, analogously,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\tilde{z}_t^2-1\right)\Rightarrow 0.$$

Use now the mean value theorem to conclude that, for suitable ζ_t between $\tilde{z}_t$ and $\hat{z}_t$,

$$\Phi^k\left(\tilde{z}_t\right)=\Phi^k\left(\hat{z}_t\right)+k\Phi^{k-1}\left(\zeta_t\right)\varphi\left(\zeta_t\right)\left(\tilde{z}_t-\hat{z}_t\right);$$

therefore,

$$\begin{aligned}\tilde{p}_t^k &= \hat{p}_t^k - \frac{1}{\tilde{\sigma}_{\hat{z}}}\sqrt{T}\bar{\hat{z}}\frac{k}{\sqrt{T}}\Phi^{k-1}(\zeta_t)\varphi(\zeta_t) + \sqrt{T}\left(\frac{1}{\tilde{\sigma}_{\hat{z}}}-1\right)\frac{k}{\sqrt{T}}\Phi^{k-1}(\zeta_t)\varphi(\zeta_t)\zeta_t \\ &\quad +\sqrt{T}\left(\frac{1}{\tilde{\sigma}_{\hat{z}}}-1\right)\frac{k}{\sqrt{T}}\Phi^{k-1}(\zeta_t)\varphi(\zeta_t)(\hat{z}_t-\zeta_t).\end{aligned}$$

Upon building partial sums of $\tilde{p}_t^k$, notice that $\sum_{t=1}^{[sT]}\left|\Phi^{k-1}(\zeta_t)\varphi(\zeta_t)\right| \leq CT$ and $\sum_{t=1}^{[sT]}\left|\Phi^{k-1}(\zeta_t)\varphi(\zeta_t)\zeta_t\right| \leq CT$ for any $s\in[0,1]$; also, it is easily seen that $\hat{z}_t-\zeta_t = o_p(1)$ uniformly in t, so, summing up,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\tilde{p}_t^k = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\hat{p}_t^k + o_p(1)$$

uniformly in s. To derive the long-run behavior of the fixed-b variance estimator $\hat{\mathbf{\Omega}}$, we finally note that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\begin{pmatrix}\tilde{p}_t-\frac{1}{2}\\ \vdots\\ \tilde{p}_t^K-\frac{1}{K+1}\\ \tilde{z}_t\\ \tilde{z}_t^2-1\end{pmatrix} \Rightarrow \begin{pmatrix}B_1(s)-\vartheta_0W_1(s)-\frac{1}{2}\varpi_1W_2(s)\\ \vdots\\ B_K(s)-K\vartheta_{K-1}W_1(s)-\frac{K}{2}\varpi_{K-1}W_2(s)\\ 0\\ 0\end{pmatrix}$$

and the result immediately follows along the lines of the proof of Proposition 1 by noting that premultiplying $\mathbf{V}$ to the above r.h.s. leads to the same weak limit indicated in Lemma 1.

C More on parametric mean adjustment

Since this Section only serves the purpose of illustrating the influence the specific choice of model has on the feasible PITs $\hat{p}_t$, we treat σ_t as known and set it to unity; effects similar to those highlighted here for the estimation of a parametric mean arise when σ_t is to be modeled as well. Consider therefore a parametric model for the mean of the observed time series x_t such that

$$x_t = \mu\left({}^t\!/\!_T, \boldsymbol{\theta}\right) + \sigma_t z_t \ .$$

Note that normalizing the time is not restrictive, since one may e.g. redefine a classical linear trend model $\mu_t = \theta_1 + \theta_2 t$ as $\mu_t = \theta_1 + (T\theta_2)\, {}^t\!/\!_T$ without loss of generality. We take the mean component to satisfy the following requirements.

Assumption A.1 *Let $\mu(s, \boldsymbol{\theta})$ have uniformly continuous second-order partial derivatives. The first and second order partial derivatives w.r.t. $\boldsymbol{\theta}$ are weakly bounded uniformly in s, in the sense that there exists a nondecreasing function f such that $\max\left\{\left\|\frac{\partial\mu(s,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|; \left\|\frac{\partial^2\mu(s,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|\right\} \leq f(\|\boldsymbol{\theta}\|)$ for all $s \in [0,1]$.*

This Assumption allows for polynomial trend models, $\mu(s, \boldsymbol{\theta}) = \sum_{j=1}^{p+1} s^{j-1}\theta_j$, for breaks in the mean, $\mu(s, \boldsymbol{\theta}) = \theta_1 + \theta_2 I(s \geq \tau)$, for smooth mean changes, e.g. $\mu(s, \boldsymbol{\theta}) = \frac{1}{1+\exp(\theta_3(s-\theta_4))}\theta_1 + \frac{\exp(\theta_3(s-\theta_4))}{1+\exp(\theta_3(s-\theta_4))}\theta_2$, or for $\mu(s, \boldsymbol{\theta}) = \theta_1 + \sum_{j=1}^{p}(\theta_{2j}\sin 2\pi js + \theta_{2j+1}\cos 2\pi js)$ motivated by approximations via Fourier sums.

Based on this model, one obtains

$$\hat{p}_t^{(\boldsymbol{\theta})} = \Phi\left(\hat{z}_t^{(\boldsymbol{\theta})}\right) = \Phi\left(x_t - \mu\left({}^t\!/\!_T, \hat{\boldsymbol{\theta}}\right)\right)$$

by plugging in an estimator $\hat{\boldsymbol{\theta}}$ which is taken to be $\sqrt{T}$-consistent. The straightforward choice we employ in the following is the NLS estimator. Some of the requirements of Assumption A.1 help to establish the limiting behavior of the NLS estimator. Then,

$$\hat{p}_t^{(\boldsymbol{\theta})} = \Phi\left(z_t - \left(\mu\left({}^t\!/\!_T, \hat{\boldsymbol{\theta}}\right) - \mu\left({}^t\!/\!_T, \boldsymbol{\theta}\right)\right)\right) \tag{15}$$

to demonstrate the estimation effect. The following Lemma provides the precise result when x_t is adjusted for nonzero mean in a parametric way.

Lemma A.2 *Under the additional Assumption A.1, it holds as $T \to \infty$ that*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\left(\hat{p}_t^{(\boldsymbol{\theta})}\right)^k - \frac{1}{k+1}\right) \Rightarrow B_k(s) - k\vartheta_{k-1}\boldsymbol{\delta}'(s, \boldsymbol{\theta})\,\boldsymbol{\Theta}(1) \tag{16}$$

where $\boldsymbol{\Theta}(1) = \left(\int_0^1 \frac{\partial\mu(s,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\mu(s,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}'\mathrm{d}s\right)^{-1}\int_0^1 \frac{\partial\mu(s,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathrm{d}W_1(s)$, $\boldsymbol{\delta}(s, \boldsymbol{\theta}) = \int_0^s \frac{\partial\mu(r,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathrm{d}r$ and $\vartheta_k = \mathrm{E}\left(p_t^k\varphi(z_t)\right)$ as before.

Proof of Lemma A.2

We begin by discussing the limiting behavior of the NLS estimators $\hat{\boldsymbol{\theta}}$. We have under Assumptions 1 and A.1 that

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \Rightarrow \left(\int_0^1 \frac{\partial \mu(s,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \mu(s,\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \mathrm{d}s\right)^{-1}\int_0^1 \frac{\partial \mu(s,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathrm{d}W_1(s) \ .$$

This is a standard application of extremum estimator theory and we omit the details.

With the application of the mean value theorem when $k=1$ (or Taylor series expansion with remainder term in differential form) we obtain

$$\hat{p}_t^{(\boldsymbol{\theta})} = p_t + \varphi(z_t)\left(\mu(t/T,\boldsymbol{\theta}) - \mu\left(t/T,\hat{\boldsymbol{\theta}}\right)\right) + \varphi'(\xi_t)\left(\mu(t/T,\boldsymbol{\theta}) - \mu\left(t/T,\hat{\boldsymbol{\theta}}\right)\right)^2$$

where ξ_t lies between z_t and $z_t - \mu\left(t/T,\hat{\boldsymbol{\theta}}\right) + \mu(t/T,\boldsymbol{\theta})$ for all t. The exact values for ξ_t do not matter since φ' is bounded. A second expansion, here about $\boldsymbol{\theta}$, is required for the trend function μ:

$$\mu(t/T,\boldsymbol{\theta}) - \mu\left(t/T,\hat{\boldsymbol{\theta}}\right) = -\frac{\partial \mu(t/T,\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) - \left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)'\left.\frac{\partial^2 \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)$$

again with $\boldsymbol{\vartheta}_t$ between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ (note that since t is an argument of μ, $\boldsymbol{\vartheta}$ also depends on t – hence the notation). Putting the two together we obtain

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{p}_t^{(\boldsymbol{\theta})}-\frac{1}{2}\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(p_t-\frac{1}{2}\right) - \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\varphi(z_t)\frac{\partial \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \\ &\quad - \left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\varphi(z_t)\left.\frac{\partial^2 \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) + R_{s,T}\end{aligned}$$

where $R_{s,T}$ is the normalized partial sum of $\varphi'(\xi_t)\left(\mu(t/T,\boldsymbol{\theta}) - \mu\left(t/T,\hat{\boldsymbol{\theta}}\right)\right)^2$.

Examining the third summand on the r.h.s., we note that the boundedness of φ and the fact that $\left|\left.\frac{\partial^2 \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\right| \leq f(\|\boldsymbol{\vartheta}_t\|) \leq f\left(\max\left\{\|\boldsymbol{\theta}\|;\left\|\hat{\boldsymbol{\theta}}\right\|\right\}\right)$ make the partial sums of order $O_p(T)$, but $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta} = O_p\left(1/\sqrt{T}\right)$ and the normalization with $\sqrt{T}$ lets the entire summand vanish.

For the fourth summand, $R_{s,T}$, we have with a first-order Taylor expansion, $\mu(t/T,\boldsymbol{\theta}) - \mu\left(t/T,\hat{\boldsymbol{\theta}}\right) = \left.\frac{\partial \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|'_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)$ with $\boldsymbol{\vartheta}_t$ between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ for each t, that

$$R_{s,T} = \left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\varphi'(\xi_t)\left.\frac{\partial \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\left.\frac{\partial \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|'_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right).$$

Similarly, φ' is bounded and $\left|\left.\frac{\partial \mu(t/T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\right| \leq f(\|\boldsymbol{\vartheta}_t\|) \leq f\left(\max\left\{\|\boldsymbol{\theta}\|;\left\|\hat{\boldsymbol{\theta}}\right\|\right\}\right)$ for all t, it follows that $\sup_s R_{s,T} = O_p\left(T^{-1/2}\right)$.

Summing up, we are left with the first two summands,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{p}_t^{(\boldsymbol{\theta})}-\frac{1}{2}\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(p_t-\frac{1}{2}\right)-\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\varphi\left(z_t\right)\frac{\partial\mu\left(t/T,\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)'\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)+o_p\left(1\right)\ .$$

The same arguments show that analogous relations hold for $\left(\hat{p}_t^{(\boldsymbol{\theta})}\right)^k$. With $\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\Rightarrow\boldsymbol{\Theta}\left(1\right)$ and $\frac{1}{T}\sum_{t=1}^{[sT]}p_t^{k-1}\varphi\left(z_t\right)\frac{\partial\mu(t/T,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Rightarrow\mathrm{E}\left(p_t^{k-1}\varphi\left(z_t\right)\right)\int_0^s\frac{\partial\mu(r,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathrm{d}r=\vartheta_{k-1}\boldsymbol{\delta}\left(s,\boldsymbol{\theta}\right)$, the desired result follows.

Remark A.1 *Bai and Ng (2005) show in their Theorem 5 that regressing x_t on a set of regressors has no effect on the limiting distributions beyond that of the intercept. There is no contradiction between their result and our Lemma A.2, since the result we give in (16) applies in the case where the regressors are deterministic. For a comparison with Theorem 5 in Bai and Ng (2005), take one stochastic regressor and a linear model $x_t=\theta w_t$ such that $\frac{\partial\mu(t/T,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=w_t$. We obtain for stationary regressors that $\frac{1}{T}\sum_{t=1}^{[sT]}\varphi\left(z_t\right)w_t\Rightarrow s\,\mathrm{E}\left(\varphi\left(z_t\right)w_t\right)$. Now, Bai and Ng (2005) assume that an intercept is always present in the regression which is equivalent to setting $\mathrm{E}\left(w_t\right)=0$; they also assume the regressors to be independent of z_t, hence $\mathrm{E}\left(\varphi\left(z_t\right)w_t\right)=0$ and correspondingly $\mu\left(s\right)=0$. This is not the case when w_t is deterministic, say an intercept or a trend, and the limiting distribution of $\hat{\theta}$ needs to be taken into account.*

Clearly, the estimation effect described by Equation (16) will affect the limiting distribution of a fixed-b statistic based on a parametric estimated standardization. The effect is different from that derived in Lemma 1, since the presence of $\boldsymbol{\Theta}\left(1\right)$ (as opposed to $W_1\left(s\right)$) indicates a bridge-type behavior of the limit process of the relevant partial sums. Moreover, the components $\boldsymbol{\Theta}$ and $\boldsymbol{\delta}$ depend on the specific model chosen for μ. The statistics can be made pivotal, see below, but the limiting distributions differ from those obtained commonly in the fixed-b framework, except in the case of an intercept. The bottom line is that different deterministic components will lead to different distributions (with the exception of the small-b case, where χ^2 asymptotics may be recovered for all consistent choices of HAC covariance matrix estimator). This implies the need to simulate the distributions for each specific type of deterministic component accounted for in the data (similar to typical unit root testing situation). While this can be done in advance for some popular combinations (see below for the case of intercept and trend, where the generalized Brownian bridge plays a role; cf. MacNeill, 1978), one solution for a generic mean function m is to resort to some form of bootstrap. Since z_t is strictly stationary and mixing, the residual-based iid or wild bootstrap is likely valid, but we do not pursue the topic here.

We now highlight the concrete difference between nonparametric and parametric mean adjustment for the cases of a constant and of a linear trend. Considering constant variance for simplicity, we have the following procedure simplified by the linearity of the mean function. Detrend x_t using OLS regression and standardize the detrended time series with σ to obtain $\hat{z}_t^{(\boldsymbol{\theta})}$. With $\left(\hat{p}_t^{(\boldsymbol{\theta})},\ldots,\left(\hat{p}_t^{(\boldsymbol{\theta})}\right)^K,\hat{z}_t^{(\boldsymbol{\theta})}\right)'$, compute a fixed-$b$ estimate of the long-run covariance matrix of $\left(p_t,\ldots,p_t^K,z_t\right)'$, say $\hat{\boldsymbol{\Lambda}}$, and, based on it, the scaling matrix $\hat{\boldsymbol{\Psi}}^{(\boldsymbol{\theta})}=\mathbf{V}^{(\boldsymbol{\theta})}\hat{\boldsymbol{\Lambda}}\left(\mathbf{V}^{(\boldsymbol{\theta})}\right)'$ with $\mathbf{V}^{(\boldsymbol{\theta})}$ constructed like in (11) but using just the first column of $\boldsymbol{\Upsilon}_K$, and then $\hat{\mathcal{T}}$ from (10). Then,

Proposition A.1 *Under Assumptions 1 and 2, it holds as $T \to \infty$ that*

$$\hat{\mathcal{T}}_K^{(\boldsymbol{\theta})} \Rightarrow \tilde{\boldsymbol{W}}'_K(1)\boldsymbol{\mathcal{Q}}^{-1}_{K,b,\kappa}\tilde{\boldsymbol{W}}_K(1).$$

with

$$\boldsymbol{\mathcal{Q}}_{K,b,\kappa} = -\int_0^1\int_0^1 \frac{1}{b^2}\kappa''\left(\frac{r-s}{b}\right)\boldsymbol{V}(r)\boldsymbol{V}'(s)\,\mathrm{d}r\mathrm{d}s$$

for smooth kernels and

$$\boldsymbol{\mathcal{Q}}_{K,b,\kappa} = \frac{2}{b}\int_0^1 \boldsymbol{V}(r)\boldsymbol{V}'(r)\mathrm{d}r - \frac{1}{b}\int_0^{1-b}\boldsymbol{V}(r+b)\boldsymbol{V}'(r)\,\mathrm{d}r - \frac{1}{b}\int_0^{1-b}\boldsymbol{V}(r)\boldsymbol{V}'(r+b)\,\mathrm{d}r$$

for the Bartlett kernel, where $\boldsymbol{V}(s)$ is, for demeaning, the first-order Brownian bridge

$$\boldsymbol{V}(s) = \tilde{\boldsymbol{W}}_K(s) - s\tilde{\boldsymbol{W}}_K(1)$$

with $\tilde{\boldsymbol{W}}$ a vector of independent standard Wiener processes; for detrending, $\boldsymbol{V}(s)$ is the second-level Brownian bridge

$$\boldsymbol{V}(s) = \tilde{\boldsymbol{W}}_K(s) + (2s-3s^2)\tilde{\boldsymbol{W}}_K(1) - 6s(1-s)\int_0^1 \tilde{\boldsymbol{W}}_K(s)\mathrm{d}s.$$

Proof of Proposition A.1

To deal with demeaning, let $\mu = \theta_1$ in Lemma A.2 to obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\left(\hat{p}_t^{(\boldsymbol{\theta})}\right)^k - \frac{1}{k+1}\right) \Rightarrow B_k(s) - k\vartheta_{k-1}W_1(1).$$

We then need to examine the limiting behavior of the suitably normalized partial sums of $\hat{z}_t^{(\boldsymbol{\theta})}$. Then (having assumed for simplicity σ_t to be known)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\hat{z}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}(z_t - \bar{z}) \Rightarrow W_1(s) - sW_1(1).$$

Let

$$\bar{\boldsymbol{B}}(s) = \left(B_1(s), \ldots, B_K(s), W_1(s)\right)'$$

and

$$\tilde{\boldsymbol{B}} = \begin{pmatrix} B_1(s) - s\,\vartheta_0 W_1(1) \\ \vdots \\ B_K(s) - s\,K\vartheta_{K-1}W_1(1) \\ W_1(s) - sW_1(1) \\ W_2(s) - sW_2(1) \end{pmatrix};$$

using the arguments of the proof of Theorem 2 in Kiefer and Vogelsang (2005) together with the Lemma A.2, we obtain e.g. for smooth kernels

$$\hat{\mathcal{T}}_K^{(\boldsymbol{\theta})} \Rightarrow \left(\mathbf{V}^{(\boldsymbol{\theta})}\bar{\boldsymbol{B}}(1)\right)' \times \left(\mathbf{V}^{(\boldsymbol{\theta})}\left(-\int_0^1\int_0^1 \frac{1}{b^2}\kappa''\left(\frac{r-s}{b}\right)\left(\tilde{\boldsymbol{B}}(r)-r\tilde{\boldsymbol{B}}(1)\right)\left(\tilde{\boldsymbol{B}}(s)-s\tilde{\boldsymbol{B}}(1)\right)' \mathrm{d}r\mathrm{d}s\right)\left(\mathbf{V}^{(\boldsymbol{\theta})}\right)'\right)^{-1}\mathbf{V}^{(\boldsymbol{\theta})}\bar{\boldsymbol{B}}(1).$$

Note further that

$$\mathbf{V}^{(\boldsymbol{\theta})}\left(\tilde{\boldsymbol{B}}(s)-s\tilde{\boldsymbol{B}}(1)\right) = \mathbf{V}^{(\boldsymbol{\theta})}\left(\bar{\boldsymbol{B}}(s)-s\bar{\boldsymbol{B}}(1)\right),$$

and let $\boldsymbol{Y} = \mathbf{V}^{(\boldsymbol{\theta})}\bar{\boldsymbol{B}}$ such that

$$\hat{\mathcal{T}}_K \Rightarrow \boldsymbol{Y}'(1)\left(-\int_0^1\int_0^1 \frac{1}{b^2}\kappa''\left(\frac{r-s}{b}\right)\left(\boldsymbol{Y}(r)-r\boldsymbol{Y}(1)\right)\left(\boldsymbol{Y}(s)-s\boldsymbol{Y}(1)\right)' \mathrm{d}r\mathrm{d}s\right)^{-1}\boldsymbol{Y}(1)$$

where $\boldsymbol{Y}$ is a multivariate Brownian motion with covariance matrix $\mathbf{\Psi}^{(\boldsymbol{\theta})} = \mathbf{V}^{(\boldsymbol{\theta})}\mathbf{\Lambda}\left(\mathbf{V}^{(\boldsymbol{\theta})}\right)'$. To obtain the required distribution, let $\tilde{\boldsymbol{W}} = \mathbf{\Psi}^{(\boldsymbol{\theta})-1/2}\boldsymbol{Y}(s)$, and note that $\mathbf{\Psi}^{(\boldsymbol{\theta})}$ cancels out and the desired result follows. The result for the Bartlett kernel follows analogously.

To deal with detrending, let $\mu = \theta_1 + \theta_2 s$ in Lemma A.2 to obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\left(\hat{p}_t^{(\boldsymbol{\theta})}\right)^k - \frac{1}{k+1}\right) \Rightarrow B_k(s) - k\vartheta_{k-1}\left(4sW_1(1) - 3s^2W_1(1) - 6s(1-s)\int_0^1 s\mathrm{d}W_1(s)\right).$$

Note that $\int_0^1 s\mathrm{d}W_1(s) = W_1(1) - \int_0^1 W_1(s)\mathrm{d}s$; use then the same steps as for demeaning to arrive at the desired result.

D The Bai and Ng (2005) test procedure

The test statistic suggested by Bai and Ng (2005) is given by

$$\hat{\mu}_{34} = Y_T'(\hat{\gamma}\hat{\Phi}\hat{\gamma}')^{-1}Y_T$$

where

$$Y_T = \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\bar{x})^3 \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(x_t-\bar{x})^4 - 3\hat{\sigma}^4] \end{bmatrix}$$

and

$$\hat{\gamma} = \begin{bmatrix} -3\hat{\sigma}^2 & 0 & 1 & 0 \\ 0 & -6\hat{\sigma}^2 & 0 & 1 \end{bmatrix}$$

$\bar{x}$, $\hat{\sigma}^2$ and $\hat{\Phi}$ are consistent estimators under constant mean and variance. Denoting $Z_t' = \begin{bmatrix} x_t-\mu, & (x_t-\mu)^2-\sigma^2, & (x_t-\mu)^3, & (x_t-\mu)^4-3\sigma^4 \end{bmatrix}$ with $\bar{Z}$ being the sample mean of Z_t, the long-run covariance matrix Φ is given by $\Phi = \lim_{T\to\infty} T\,\mathrm{E}(\bar{Z}\bar{Z}')$. The limiting distribution of $\hat{\mu}_{34}$ is $\chi^2(2)$. This result is motivated by the fact that under normality, one obtains $Y_T = \gamma\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t + o_p(1)$ with $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \Rightarrow N(0,\Phi)$. We follow Bai and Ng (2005) and consider the Newey and West (1987) estimator.

Under our Assumptions, however, it is a standard exercise to show that

$$\bar{x} := \frac{1}{T}\sum_{t=1}^{T} x_t \xrightarrow{p} \int_0^1 \mu(s)\mathrm{d}s =: \bar{\mu},$$

$$\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^2 \xrightarrow{p} \int_0^1 \sigma^2(s)\mathrm{d}s + \int_0^1 (\mu(s)-\bar{\mu})^2\,\mathrm{d}s,$$

$$\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^3 \xrightarrow{p} 3\int_0^1 \sigma^2(s)\,(\mu(s)-\bar{\mu})\,\mathrm{d}s + \int_0^1 (\mu(s)-\bar{\mu})^3\,\mathrm{d}s$$

and

$$\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^4 \xrightarrow{p} 3\int_0^1 \sigma^4(s)\mathrm{d}s + 6\int_0^1 \sigma^2(s)\,(\mu(s)-\bar{\mu})^2\,\mathrm{d}s + \int_0^1 (\mu(s)-\bar{\mu})^4\,\mathrm{d}s.$$

We note that the sample skewness converges to 0 if the mean function is constant, but not in general – as would have been expected for a normal population. Furthermore, the sample excess kurtosis may have a limit different from 3 (corresponding to the standard normal distribution) even under a constant mean, as long as the variance function changes over sets of nonzero Lebesgue measure.

E Additional empirical results and graphs

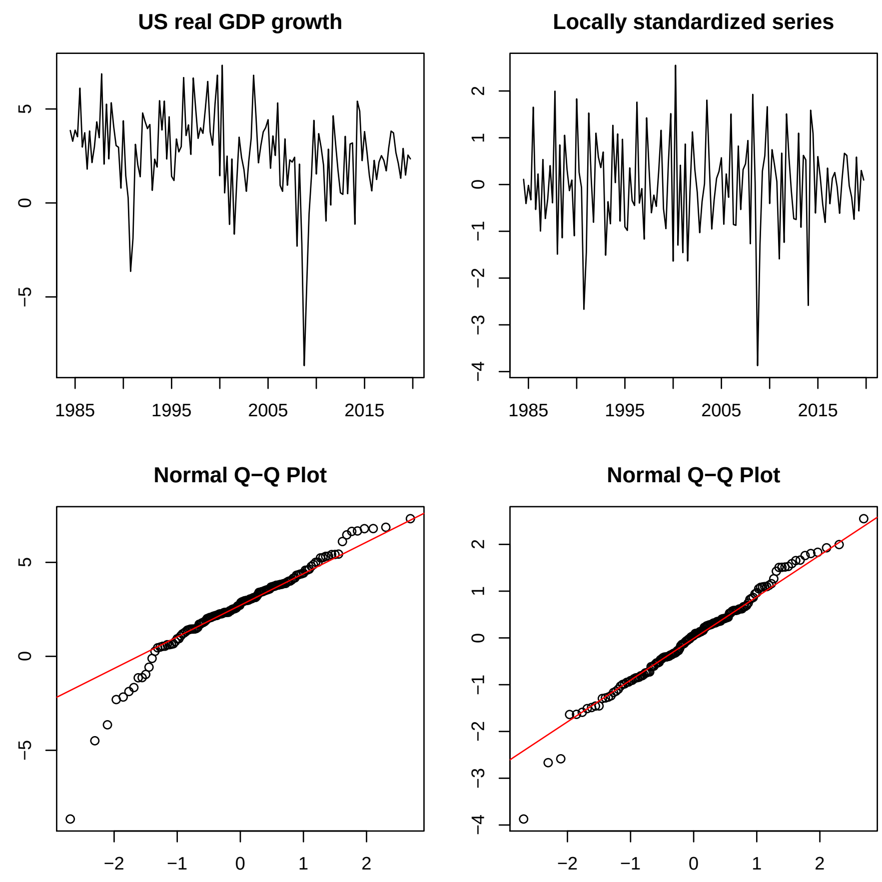

Figure A.1: U.S. real GDP growth rate (1984Q3-2019Q4). Top left: time series plot of the raw time series x_t; top right: nonparametric locally standardized time series $(\hat{z}_t)$; bottom left: normal QQ-plot for x_t and bottom right: normal QQ-plot for $\hat{z}_t$.

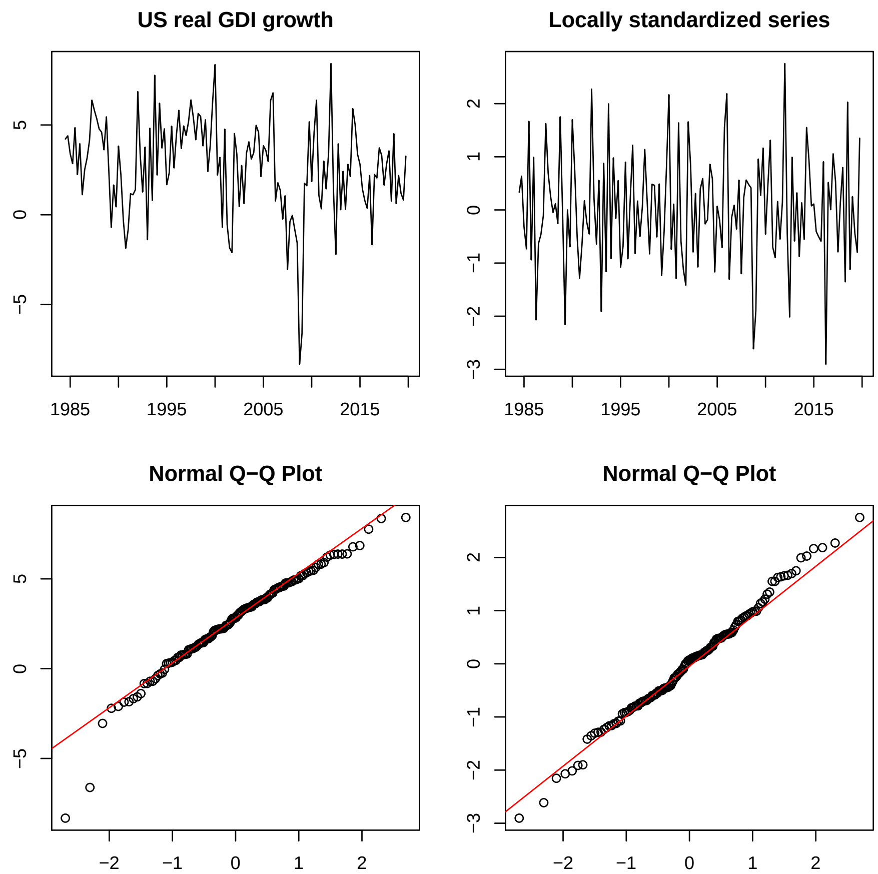

Figure A.2: U.S. real GDI growth rate (1984Q3-2019Q4). Top left: time series plot of the raw time series x_t; top right: nonparametric locally standardized time series ($\hat{z}_t$); bottom left: normal QQ-plot for x_t and bottom right: normal QQ-plot for $\hat{z}_t$.

Table A.1: Normality testing results for U.S. real GDP and GDI growth rates (1984Q3-2019Q4). $\tilde{m}_i$ denotes the estimated i-th raw moment and $\tilde{t}_i$ is the t-statistic for testing the individual i-th moment restriction ($i = 1, 2, 3, 4$) based on the nonparametric local standardization. The corresponding joint statistics $\tilde{\mathcal{T}}_{1\ldots j}$ test the first j moment restrictions ($j = 2, 3, 4$); BN labels the Bai and Ng (2005) statistic. Critical values for the nominal significance level of 5% are provided in parentheses below the statistics.

GDP				GDI			
$\tilde{m}_1$	$\tilde{m}_2$	$\tilde{m}_3$	$\tilde{m}_4$	$\tilde{m}_1$	$\tilde{m}_2$	$\tilde{m}_3$	$\tilde{m}_4$
0.503	0.332	0.246	0.195	0.498	0.329	0.245	0.196
$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$	$\tilde{t}_1$	$\tilde{t}_2$	$\tilde{t}_3$	$\tilde{t}_4$
0.499	-0.314	-0.877	-1.033	-0.388	-0.955	-1.166	-0.964
(2.261)	(2.261)	(2.261)	(2.261)	(2.261)	(2.261)	(2.261)	(2.261)
$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN	$\tilde{\mathcal{T}}_{12}$	$\tilde{\mathcal{T}}_{123}$	$\tilde{\mathcal{T}}_{1234}$	BN
1.670	2.144	2.351	1.880	1.711	2.341	3.877	1.633
(8.872)	(13.200)	(18.258)	(5.991)	(8.872)	(13.200)	(18.258)	(5.991)

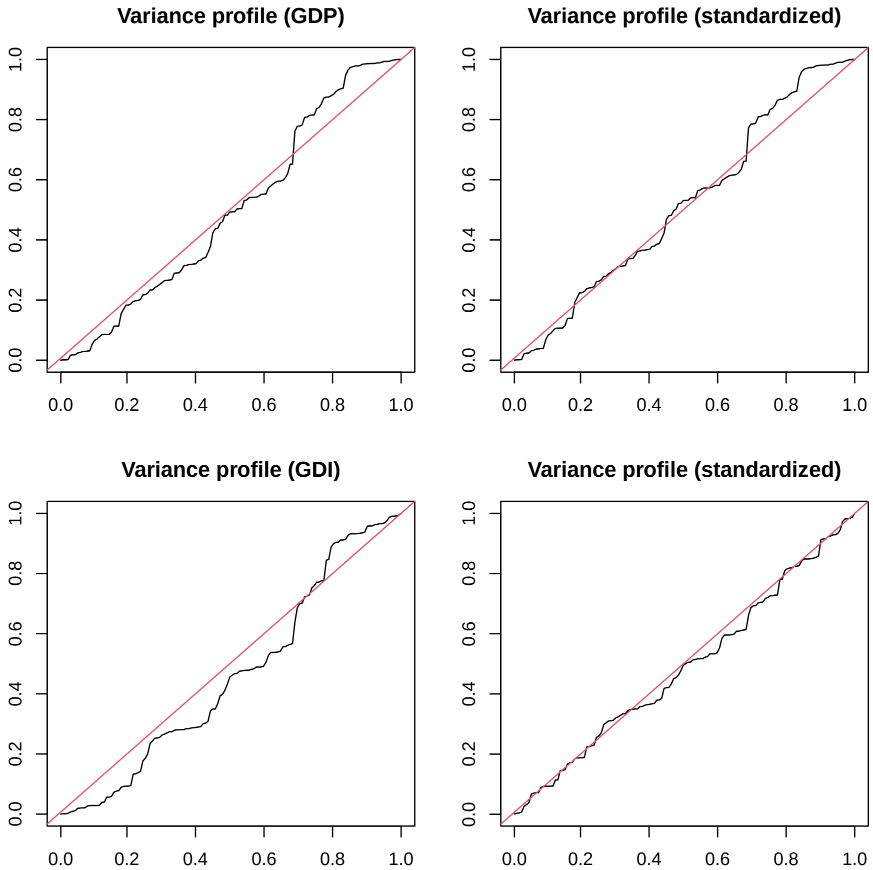

Figure A.3: U.S. real GDP (top) and GDI (bottom) growth rate (1984Q3-2019Q4). Left: Variance profile for the raw time series x_t; right: Variance profile for the nonparametric locally standardized $\hat{z}_t$.

F Critical values

Table A.2: Critical values (cv) and fitted least squares response curves for the limiting distribution of the $\mathcal{T}_K$ statistic for $K = \{2, 3, 4\}$ when the Bartlett kernel is entertained. The least squares regression is given by $cv(b) = a_0 + a_1 b + a_2 b^2 + a_3 b^3 + error$, estimated parameters are provided together with the corresponding R^2 (last column) is given in the last column. Nominal significance levels are 0.9, 0.95, 0.975 and 0.99. For the case of a single moment restriction, we reproduce the estimated response curves provided in Table 1 from Kiefer and Vogelsang (2005) for the t-statistic.

	a_0	a_1	a_2	a_3	R^2
t					
0.9	1.2816	1.3040	0.5135	-0.2286	0.9995
0.95	1.6449	2.1859	0.3142	-0.3427	0.9991
0.975	1.9600	2.9694	0.4160	-0.5324	0.9980
0.99	2.3263	4.1618	0.5368	-0.9060	0.9957
$K = 2$					
0.9	4.6052	15.5300	33.0455	-18.0050	0.9998
0.95	5.9915	24.2350	48.4528	-27.7431	0.9998
0.975	7.3778	35.6889	62.8696	-36.8917	0.9997
0.99	9.2103	53.2832	88.7896	-55.9722	0.9996
$K = 3$					
0.9	6.2514	30.2793	67.5629	-42.2680	0.9998
0.95	7.8147	45.5956	88.1783	-56.1070	0.9997
0.975	9.3484	63.5918	109.2760	-70.7583	0.9997
0.99	11.3449	94.2752	127.9765	-84.0108	0.9996
$K = 4$					
0.9	7.7794	54.1072	94.7069	-61.0147	0.9997
0.95	9.4877	76.3485	121.5104	-79.8180	0.9997
0.975	11.1433	102.1803	145.6040	-97.0618	0.9997
0.99	13.2767	142.5323	169.0490	-113.2457	0.9997

G Details on V matrices for different distributions

In this appendix, we report simulated entries for the matrix $\boldsymbol{\Upsilon}_K$ (see equation (11)) as in

$$\mathbf{V} = \left(\mathbf{I}_K; \quad \boldsymbol{\Upsilon}_K \right) \quad \text{with} \quad \boldsymbol{\Upsilon}_K = - \begin{pmatrix} \vartheta_0 & \cdots & K\vartheta_{K-1} \\ \frac{1}{2}\varpi_0 & \cdots & \frac{K}{2}\varpi_{K-1} \end{pmatrix}'.$$

The coefficients are given as $\vartheta_j = \mathrm{E}\left(p_t^j f_0(z_t)\right)$ and $\varpi_j = \mathrm{E}\left(p_t^j z_t f_0(z_t)\right)$, where f_0 denotes the density function of the standardized null distribution, $z_t \sim f_0$.

Table A.3: Simulated matrix entries for $\boldsymbol{\Upsilon}_K$ are reported for four different standardized distributions: Normal, Log-normal, Exponential and Uniform.

Normal				Log-normal			
-0.2822	-0.2821	-0.2573	-0.2326	-0.7834	-0.5341	-0.3833	-0.2920
0.0000	-0.0459	-0.0689	-0.0800	0.1579	0.0629	0.0177	-0.0048
Exponential				**Uniform**			
-0.4998	-0.3333	-0.2501	-0.2001	-0.2887	-0.2887	-0.2887	-0.2887
0.1249	0.0277	-0.0105	-0.0284	0.0000	-0.0833	-0.1250	-0.1500