Testing the marginal normality of heteroskedastic time series*

Matei Demetrescu†
Christian-Albrechts-University of Kiel

Robinson Kruse
Rijksuniversiteit Groningen and CREATES

Preliminary version: May 13, 2016
Please do not quote

Abstract

Normality testing is an evergreen topic in statistics, econometrics and other disciplines. The paper focuses on testing the normality of time series exhibiting features such as serial dependence and time-varying mean and volatility. If falsely assuming weak or strict stationarity, the marginal distribution of the series of interest may be mistaken for a mixture of normals and distribution tests may reject too often even when the null of normality is actually true. Here, we suggest tests based on raw moments of probability integral transform of the suitably standardized series. To standardize the series, nonparametric estimators of the mean and the variance functions may be used. The use of probability integral transforms is advantageous as they are quite sensitive to deviations from the null other than asymmetry and excess kurtosis. Short-run dynamics is taken into account using the (fixed-b) Heteroskedasticity and Autocorrelation Robust [HAR] approach of Kiefer and Vogelsang (2005, ET), which is suitably modified to capture the effect of estimation uncertainty arising from estimated standardization. The provided Monte Carlo experiments show that the new tests are performing well in terms of size (which is in part due to the adopted fixed-b framework for long-run covariance estimation), but also in terms of power, when compared to popular alternative procedures. An application to G7 industrial production growth rates sheds light on the empirical usefulness of the proposed test.

Key words: Normality testing; Probability integral transform; Estimated standardization; Nonparametric estimator; Robust testing.

JEL classification: C12; C14; C22

1 Introduction

Testing distributional assumptions, in particular normality, is an important aspect of applied work. For instance, nonnormality of disturbances may indicate a misspecification in regression

*The authors would like to thank Juan Dolado, Peter Reinhard Hansen, Mehdi Hosseinkouchack and Philipp Sibbertsen for helpful comments. Part of this research was carried out while the second author was visiting EUI Florence. Their kind hospitality is greatly appreciated.

†**Corresponding author:** Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany, email: `mdeme@stat-econ.uni-kiel.de`.

models. Nonnormality may also be a prerequisite of certain modelling approaches; see e.g. the analysis of non-causal time series models (Lanne and Saikkonen, 2011; Lanne et al., 2012; Lanne and Saikkonen, 2013). In an iid sampling situation, the Kolmogorov-Smirnov statistic is usually applied, but this is not straightforward to extend to serial dependence and the use of estimated parameters. For instance, Bai (2003) resorts to the martingale transformation of Khmaladze (1981); the martingale transform approach is quite demanding, though, so Bai and Ng (2005) follow Jarque and Bera (1980) and resort to moment-based testing; see Lomnicki (1961) for an early discussion for linear processes or Bontemps and Meddahi (2005) for an ingenious choice of moment restrictions.

But serial dependence and estimation uncertainty are not the only issues to be faced in econometric practice. Consider for instance the situation where a series is marginally normal, but exhibits one break in the mean or the variance. The "pooled" distribution is a mixture of two normals, which is nonnormal, so a normality test ignoring the break will reject the true null more often than the nominal level of the test requires. The reasoning extends to more general patterns of changes in mean or variance. And indeed, economic data are often found to exhibit time-varying moments. Even if arguing mean breaks away, examples of time-varying volatility can be found in the field of financial data such as asset returns (see among others Guidolin and Timmermann, 2006; Amado and Teräsvirta, 2014; Teräsvirta and Zhao, 2011; Amado and Teräsvirta, 2013) and also macroeconomic time series such as economic growth or price changes (see e.g. Stock and Watson, 2002; Sensier and van Dijk, 2004; Clark, 2009, 2011; Justiniano and Primiceri, 2008). Typical patterns are permanent breaks (like the "Great Moderation" as an example for a downward break) or trends in the variance. As a consequence, robust inference for time-heteroskedasticity with dependent data has received considerable attention in the last decade.[1]

We discuss in this paper tests based on moments of probability integral transforms [PIT]s of the standardized series using estimated mean and variances. The main reason to do so is that PITs may be more sensitive against nonormal alternatives with zero skewness or zero excess kurtosis since they take higher-order moments into account by construction. PITs have already been used successfully by Knüppel (2014), though without accounting for the estimated standardization. Therefore, one contribution of this paper is to show how to account for the effect of the estimated standardization such that pivotal inference is possible. Importantly, we show that the mean and variance functions may be estimated in a nonparametric fashion. As a consequence the practitioner does not have to specify a model for the mean and the variance explicitly.

Regarding robustness against serial dependence, we rely on long-run variance estimation following Bai and Ng (2005). We go one step further, though, and adopt the fixed-b asymptotic framework of Kiefer and Vogelsang (2005). The main feature of the fixed-b framework is that the bandwidth B used for long-run covariance estimation does not need to fulfill the standard assumption that $b = B/T \to 0$ as $T \to \infty$. On the contrary, the bandwidth is held fixed as a linear proportion of the sample size T, i.e. $B = [bT]$ with $b \in (0, 1]$. This leads to non-standard asymptotic limiting

[1] Phillips and Xu (2006) and Xu (2008) deal with stationary autoregressions, while, for unit root autoregressions, the reader is referred to Cavaliere and Taylor (2008) or Cavaliere and Taylor (2009). Time-varying volatility have even larger effects in panels of (nonstationary) series, prompting for suitable treatment; see e.g. Demetrescu and Hanck (2012) or Westerlund (2014).

distributions of tests statistics (like t, Wald and F). Importantly, the critical values obtained from such distributions reflect the choice of bandwidth and kernel even as $T \to \infty$, such that the fixed-b approach may provide much more accurate finite-sample inference.[2]

The remainder of the paper is structured as follows. In Section 2, the setup is described and newly proposed test statistics for normality are introduced. The case of uncertainty induced by nonparametrically estimated standardization is located in Section 3. Our Monte Carlo simulations study is included in Section 4. Section 5 provides an empirical application of normality tests to G7 industrial production growth rates. Section 6 concludes the study. Proofs, additional results, response curves for critical values and a description of the Bai and Ng (2005) test statistic are given in the Appendix.

In terms of notation, C stands for a generic constant whose value may change from one occurrence to another and "$\Rightarrow$" for weak convergence in a space of cadlag functions endowed with a suitable norm.

2 Model and test idea

The series of interest x_t is taken to be marginally normal under the null. It is taken to exhibit time-varying mean and variance behavior as given by the following 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 breaks.

The following assumptions make the notions of short-run dependence and time-varying moments precise.

Assumption 1 *Let z_t be a marginally standard normal, strictly stationary series with strong mixing with coefficients $\alpha(j)$ for which*

$$\alpha(j) < Aj^{-b} \qquad \text{for some} \qquad b > 10/3.$$

Assume furthermore that z_t has positive long-run variance, $\sum_{h=-\infty}^{\infty} \mathrm{E}(z_t z_{t+h}) > 0$.Assume absolutely summable 8th-order cumulants of z_t.

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 the non-zero long-run variance, z_t is integrated of order zero. 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 not relevant here. The condition also

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

allows for mild form of conditional heteroskedasticity, so the observed series x_t may exhibit both conditional and unconditional heteroskedasticity. 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 (see e.g. Davidson, 1994, Chapter 29). 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 I(0) property and uniform boundedness of higher-order moments. We shall consider nonlinear transformations of z_t, however, and strict stationarity of z_t ensures that the *transformed* series have constant variance; see below. Moreover, strict stationarity is quite plausible once the time-varying mean and variance have been accounted for.

Strict stationarity of z_t also separates the variance fluctuations from the serial dependence properties. The unity long-run variance assumption on z_t is an identifying restriction and allows for the interpretation of σ_t as marginal (long-run) standard deviation. The mean and variance functions themselves are taken to satisfy typical conditions in the literature (cf. Cavaliere, 2004):

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 σ'' exist and be bounded on $[0,1]$.*

To fix ideas, we describe the normality test on the (infeasible) basis of the unobserved z_t first. The following section shall discuss the feasible version of our test procedure with nonparametric standardization.

We base our test of the null hypothesis on moments of transformed series rather than the original series z_t. With Φ being the cdf (and φ denoting the pdf) of the standard normal distribution, the probability integral transform

$$p_t = \Phi(z_t)$$

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, under Assumption 1,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \begin{pmatrix} p_t - \frac{1}{2} \\ \vdots \\ p_t^K - \frac{1}{K+1} \end{pmatrix} \Rightarrow \begin{pmatrix} B_1(s) \\ \vdots \\ B_K(s) \end{pmatrix} \tag{3}$$

where $(B_1, \ldots B_K)'$ is a K-variate Brownian motion with covariance matrix denoted by $\Omega = \mathrm{E}\left((B_1(1), \ldots B_K(1))' (B_1(1), \ldots B_K(1))\right)$. Because p_t is only *marginally* uniform, Ω depends in general on the specific data generating process at hand. We shall 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 Ω. This follows the approach of Bai and Ng (2005) or Bontemps and Meddahi (2005) to deal with serial dependence of unknown form.

Suppose for now that the test can be based directly on empirical moments of p_t (i.e. under known parameters 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

$$t_k = \sqrt{T}\left(\frac{m_k - \frac{1}{k+1}}{\hat{\omega}_k}\right)$$

with ω_k^2 being the kth diagonal element of Ω (i.e. the long-run variance of p_t^k). Let

$$\hat{\Omega}_k^2 = \sum_{j=-T+1}^{T-1} \kappa\left(\frac{j}{B}\right)\hat{\Gamma}_j$$

denote an estimator of Ω with proportional bandwidth $B = [bT]$, $b > 0$, where the $\hat{\Gamma}_j$'s denote the usual autocovariance matrix estimator at lag j,

$$\hat{\Gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\left(\boldsymbol{p}_t - \bar{\boldsymbol{p}}\right)\left(\boldsymbol{p}_{t-j} - \bar{\boldsymbol{p}}\right)',$$

with $\boldsymbol{p}_t$ the vector stacking $p_t, p_t^2, \ldots, p_t^K$. For $b \in (0,1]$ we have from Kiefer and Vogelsang (2005) that

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

where W is a standard Wiener process and the functional $\mathcal{Q}_{b,\kappa}$ is given in terms of the Brownian bridge $W(s) - sW(1)$ and depends explicitly on the choice of kernel and bandwidth. 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 $\kappa(s) = (1 - |s|)\,\mathbf{1}\left(|s| \leq 1\right)$ with $\mathbf{1}$ the indicator function. For kernels with smooth 2nd order derivative, of which the QS kernel is one, it holds that

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

while, for the Bartlett kernel,

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

For both kernels, the standard asymptotic framework is recovered when $b \to 0$ at suitable rates and thus, $t_k^2 \Rightarrow \chi_1^2$; in fact $\mathcal{Q}_{b,\kappa} \xrightarrow{d} 1$ for $b \to 0$, c.f. (Kiefer and Vogelsang, 2005).

Since the test statistic is essentially the same,[3] and the practical difference between the two asymptotic frameworks lies in the resulting critical values, we shall work in the following with fixed-b asymptotics only, with the understanding that "small-b" (i.e. $b \to 0$) asymptotics are

[3]One difference is that small-b versions may resort to data-driven bandwidth selection.

encompassed.

Working with several raw moments (a portmanteau test so-to-say), we suggest to construct

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

Similarly,

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

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

Compared with relying on z_t directly, PITs have several advantages; see Knüppel (2014) again. Among others, PITs are bounded such that its higher-order cumulants are smaller than those of the standard normal such that the variability of the long-run covariance matrix estimators is smaller and the χ^2 asymptotic approximation is more accurate. The bias of the moments of PITs is also typically smaller than those of the untransformed series; see the Appendix for some evidence in this respect.At the same time, PITs still allow to distinguish between skewness and kurtosis as causes of nonnormality: since the cdf of the standard normal is symmetric about the point $(0, 0.5)$, 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.

To take advantage of the properties of the PITs based test, one must however standardize the series prior to applying the PIT. Let therefore

$$\hat{p}_t = \Phi\left(\hat{z}_t\right) = \Phi\left(\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}\right) \tag{4}$$

with $\hat{\mu}_t$ and $\hat{\sigma}_t$ being estimators of the (time-varying) mean and standard deviation μ_t and σ_t. Let also $\hat{m}_k = \frac{1}{T}\sum_{t=1}^{T}\hat{p}_t^k$ 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 $\hat{t}_k$, 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. Bai (2003) uses the Khmaladze transform to tackle this issue. Because we rely on sample moments to build the test, a simple adjustment of the covariance matrix estimator will suffice, unlike in Bai (2003). See Section 3 for the precise details. Before proceeding to evaluating the effects of estimation uncertainty, let us discuss the issue of time-varying mean and variance and the implied standardization.

3 Dealing with estimation uncertainty

Using the estimated $\hat{p}_t$ from (4) instead of the true yet unobservable PITs is not without asymptotic consequences, which we quantify below for our choice of standardization; see also Appendix C for a discussion of parametric standardization. As a consequence, the limiting distribution of the "naive" test statistics which simply replace m_k with $\hat{m}_k$ in the definition of t_k or $\mathcal{T}_K$ will not follow the asymptotic distributions posited in the previous section.

To analyze the Durbin effect in our case and discuss the necessary corrections, consider

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

(assuming for simplicity that $x_{0,-1,\ldots}$ and $x_{T+1,\ldots}$ are observed). The window width τ is smaller than T, hence ensuring that x_t is approximately standardized in finite samples, and letting $\tau \to \infty$ ensures that, asymptotically, x_t is standardized correctly. This is simply local standardizing instead of "global" as would have been sufficient for the case of strict stationarity of x_t. Clearly, this will have an effect; the key step in analyzing the feasible statistic based on $\hat{p}_t$ is to note that the weak convergence in (3) is replaced by the following limiting behavior.

Lemma 1 *Let $\tau, T \to \infty$ such that $\frac{T^{\kappa_1}}{\tau} + \frac{\tau}{T^{\kappa_2}} \to 0$ for $2/3 < \kappa_1 < \kappa_2 < 3/4$. Then, jointly for all $k = 1, \ldots, K$,*

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

with $\vartheta_{k-1} = \mathrm{E}\left(p_t^{k-1}\varphi(z_t)\right)$ and $\varpi_{k-1} = \mathrm{E}\left(p_t^{k-1}z_t\varphi(z_t)\right)$, as well as $W_{1,2}$ from (1) *and B_k from* (3)*.*

Proof: *see the Appendix.*

Remark 1 *The local averages $\hat{\mu}_t$ and $\hat{\sigma}_t$ are employed to simplify derivations; the result can plausibly be extended to Nadaraya-Watson type estimators. One may also allow for different bandwidth parameters τ for $\hat{\mu}_t$ and $\hat{\sigma}_t$, and also for a finite number of breaks. We provide simulation evidence in support of these claims, but choose not to follow through with the theory to focus on the main message.*

Remark 2 *Note that $\vartheta_0 = \mathrm{E}(\varphi_t) = \int_{-\infty}^{\infty}\varphi^2(x)\,\mathrm{d}x = \frac{1}{2\sqrt{\pi}}$; via the use of power series expansions one may show that $\vartheta_1 = \frac{1}{4\sqrt{\pi}}$, but the higher-order expectations (for $\vartheta_k,\ k \geq 2$) do not seem to have a closed-form expression. We computed 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: $\vartheta = (0.2820948, 0.1410473, 0.0857805, 0.0581472)$. Clearly, the simulated values for $k = 1$ and $k = 2$ match perfectly with their theoretical counterpart. We therefore expect that MC precision of the higher-order terms is quite reasonable.*

Remark 3 *Parametric approaches require a different asymptotic analysis. But the main disadvantage is that the corrections depend on the shape of the mean or variance component adjusted for. See Appendix C for details in the case of adjusting for nonzero mean in a parametric fashion.*

By Lemma 1 we have that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\hat{p}_t^k - \frac{1}{k+1}\right)$ still converges weakly to Brownian motion. So the non-pivotality is rather an issue of not adapting the long-run covariance matrix estimator to the fact that parameters have been estimated. Consider therefore the statistic

$$\hat{\mathcal{T}}_K = T\left(\hat{m}_1 - \frac{1}{2}, \ldots, \hat{m}_K - \frac{1}{K+1}\right)\hat{\hat{\Omega}}^{-1}\left(\hat{m}_1 - \frac{1}{2}, \ldots, \hat{m}_K - \frac{1}{K+1}\right)' \tag{5}$$

with

$$\hat{\hat{\Omega}} = V\hat{\Xi}V' \tag{6}$$

where $\hat{\Xi}$ is a long-run covariance matrix estimator of $\left(p_t, \ldots, p_t^K, z_t, z_t^2 - 1\right)'$ (based on the feasible $\left(\hat{p}_t, \ldots, \hat{p}_t^K, \hat{z}_t, \hat{z}_t^2 - 1\right)'$), and

$$V = \begin{pmatrix} I_K; & \iota_K \end{pmatrix} \quad \text{with} \quad \iota_K = -\begin{pmatrix} \vartheta_0 & \cdots & K\vartheta_{K-1} \\ \frac{1}{2}\varpi_0 & \cdots & \frac{K}{2}\varpi_{K-1} \end{pmatrix}'.$$

This is not an uncommon approach to ensure pivotality and has successfully been used before, e.g. by Bai and Ng (2005). Its validity here is established by the following

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

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

Proof: *see the Appendix.*

Remark 4 *Our framework allows testing other null distributions in location-scale models, since PITs apply to any continuous distribution. The discussion sofar involves however the expectations $\vartheta_k = \mathrm{E}\left(p_t^k\varphi(z_t)\right)$ and $\varpi_k = \mathrm{E}\left(p_t^k z_t\varphi\left(z_t\right)\right)$ which are specific to the normal distribution via $\varphi(\cdot)$, so the test statistic $\hat{\mathcal{T}}_K$ would have to be modified on a case-by-case basis when testing other null distribution. In particular, Lemma 1 and Proposition 1 can be shown to 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. This ultimately leads to test statistics with the same pivotal limiting distribution as in Proposition 1.*

Remark 5 *If e.g. the parameters of the distribution to be tested are known (or given to the researcher), it is possible to apply the original test without corrections as in* (6). *This is not just a hypothetical situation: for instance, the evaluation of density forecasts is often conducted under this assumption; see Knüppel (2014) and the referenced therein.*

Table 1: Empirical size results.

$T = 50$

	i.i.d.					ARMA(1,1)			
b	$\hat{\mathcal{T}}_1$	$\hat{\mathcal{T}}_2$	$\hat{\mathcal{T}}_3$	$\hat{\mathcal{T}}_4$	b	$\hat{\mathcal{T}}_1$	$\hat{\mathcal{T}}_2$	$\hat{\mathcal{T}}_3$	$\hat{\mathcal{T}}_4$
0.1	0.046	0.015	0.014	0.017	0.1	0.048	0.024	0.025	0.023
0.2	0.045	0.019	0.021	0.023	0.2	0.053	0.027	0.030	0.032
0.3	0.048	0.023	0.026	0.026	0.3	0.056	0.027	0.027	0.035
0.4	0.047	0.019	0.026	0.028	0.4	0.048	0.029	0.030	0.038
0.5	0.045	0.023	0.027	0.023	0.5	0.048	0.028	0.040	0.036
0.6	0.045	0.021	0.029	0.023	0.6	0.048	0.025	0.036	0.036
0.7	0.041	0.021	0.026	0.023	0.7	0.046	0.031	0.033	0.035
0.8	0.043	0.022	0.024	0.027	0.8	0.043	0.028	0.032	0.037
0.9	0.044	0.022	0.026	0.023	0.9	0.044	0.027	0.035	0.033
1	0.044	0.023	0.027	0.025	1	0.045	0.028	0.036	0.035
BN	0.097				BN	0.073			

$T = 250$

	i.i.d.					ARMA(1,1)			
b	$\hat{\mathcal{T}}_1$	$\hat{\mathcal{T}}_2$	$\hat{\mathcal{T}}_3$	$\hat{\mathcal{T}}_4$	b	$\hat{\mathcal{T}}_1$	$\hat{\mathcal{T}}_2$	$\hat{\mathcal{T}}_3$	$\hat{\mathcal{T}}_4$
0.1	0.054	0.018	0.027	0.024	0.1	0.063	0.032	0.042	0.036
0.2	0.051	0.022	0.031	0.019	0.2	0.064	0.030	0.046	0.044
0.3	0.053	0.021	0.035	0.030	0.3	0.059	0.035	0.047	0.040
0.4	0.058	0.021	0.031	0.024	0.4	0.060	0.033	0.047	0.042
0.5	0.057	0.020	0.035	0.029	0.5	0.065	0.034	0.050	0.040
0.6	0.054	0.022	0.031	0.031	0.6	0.061	0.033	0.050	0.042
0.7	0.052	0.018	0.027	0.023	0.7	0.059	0.033	0.045	0.038
0.8	0.056	0.019	0.030	0.031	0.8	0.059	0.032	0.048	0.041
0.9	0.055	0.020	0.026	0.027	0.9	0.056	0.032	0.049	0.039
1	0.056	0.020	0.029	0.028	1	0.057	0.032	0.052	0.044
BN	0.115				BN	0.096			

4 Monte Carlo study

In our Monte Carlo simulation study we compare the $\hat{\mathcal{T}}_K$ test to the procedure of Bai and Ng (2005).[4] The newly proposed test is carried out by using either the first one ($\hat{\mathcal{T}}_1$), two ($\hat{\mathcal{T}}_2$), three ($\hat{\mathcal{T}}_3$) or four moments ($\hat{\mathcal{T}}_4$). We use sample sizes of $T = \{50, 100, 250, 500\}$ and report results for $T = 50$ and $T = 250$ (the other results are similar and available upon request from the authors).

Regarding autocorrelation, we consider a causal and invertible ARMA(1,1) process with AR and MA parameter $\phi = \{0, 0.85\}$ and $\theta = \{0, -0.45\}$, respectively. The general form of the DGP is

[4]Details on the test proposed by Bai and Ng (2005) can be found in Appendix D.

given by

$$
\begin{aligned}
y_t &= \mu + \sigma z_t \\
z_t &= \phi z_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \\
\varepsilon_t &\overset{i.i.d.}{\sim} (0,1) \,.
\end{aligned}
$$

Since all procedures are scale-invariant, we do not normalize the long-run variance of z_t to unity. distribution of ε_t is specified as follows. Under H_0, innovations ε_t are standard normally distributed. Under the alternative, we consider two standardized non-normal (mixture) distributions with weights $c \in [0,1]$

1. CHI: Mixture of a normal and a $\chi^2(3)$-distribution,

2. LOGN: Mixture of a normal and a lognormal-distribution.

The fixed-bandwidth parameter b is specified on the grid $0.1, 0.2, ..., 1$. Results are presented for the Bartlett kernel with linearly decaying weights. The nominal significance level equals 5% and the number of Monte Carlo replications is set to 5,000 for each single experiment. In what concerns critical values for the fixed-b distributions, we provide them on the basis of the limiting results with 1,000 observations and 50,000 replications for $K = 1, 2, 3, 4$. Estimated cubic response curves $cv(b)$ are reported in Table 3 together with an R^2 measure for the precision of approximation.

Size results are reported in Table 1. While the Bai and Ng (2005) test is generally oversized (less for the ARMA(1,1) case), the raw moment-based tests are much closer to the nominal significance level of 5%. In some cases we observe that they are marginally undersized. But, for the larger sample size of $T = 250$ with short-run dynamics, most of them are pretty close to the desired frequency of rejections. It is of importance to note that the size does not vary much with the choice of the bandwidth parameter b. This will be of great advantage when it comes to the power of such tests which typically depend a lot on the bandwidth choice; cf. Kiefer and Vogelsang (2005). In this sense, we are not facing a size-power tradeoff as we can select the most suitable b in a way that power is maximized.

Power results are reported in Figures 1 to 4. We resort to the case with additional ARMA(1,1) short-run dynamics.[5] The weight $c \in [0,1]$ is located on the x-axis. For $c = 0$, full weight is given to the normal distribution so that a size experiment is conducted. For $c = 1$, full weight is given to the non-normal distribution (either $\chi^2(3)$ or log-normal). We present results for four different values of the fixed-bandwidth parameter: $b = \{0.1, 0.3, 0.6, 0.9\}$. In accordance with our size results, we are in the pleasant situation to select b on the basis of the performance under H_1 only. We observe a general monotonic behaviour of the power with respect to b which makes it a simple exercise: the lower b, the higher is the power. For $T = 50$, the newly suggested tests clearly outperform the benchmark (Bai and Ng, 2005, labeled as BN for short in the Figures). The cases where the BN test performs better (small values of c) are most likely due to its upward

[5] We also computed Monte Carlo averages of sample skewness and kurtosis to characterize the properties of simulated distributions. The numerical averages match very well with their theoretical counterparts. Results are available upon request from the authors.

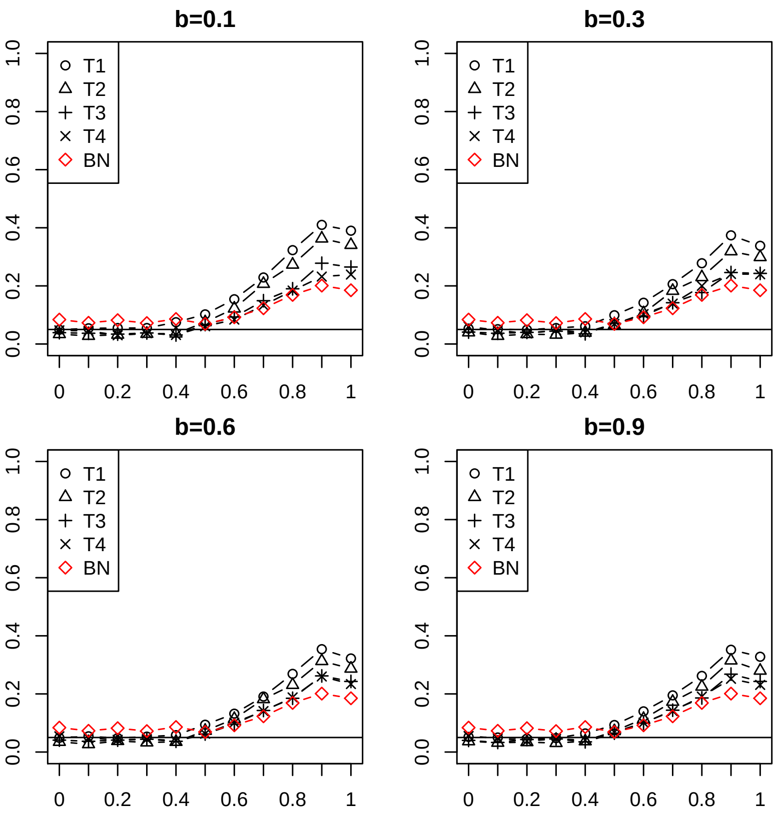

Figure 1: Rejection frequencies for mixed normal and $\chi^2(3)$ with weight $c \in [0, 1]$ (on x-axis), $T = 50$.

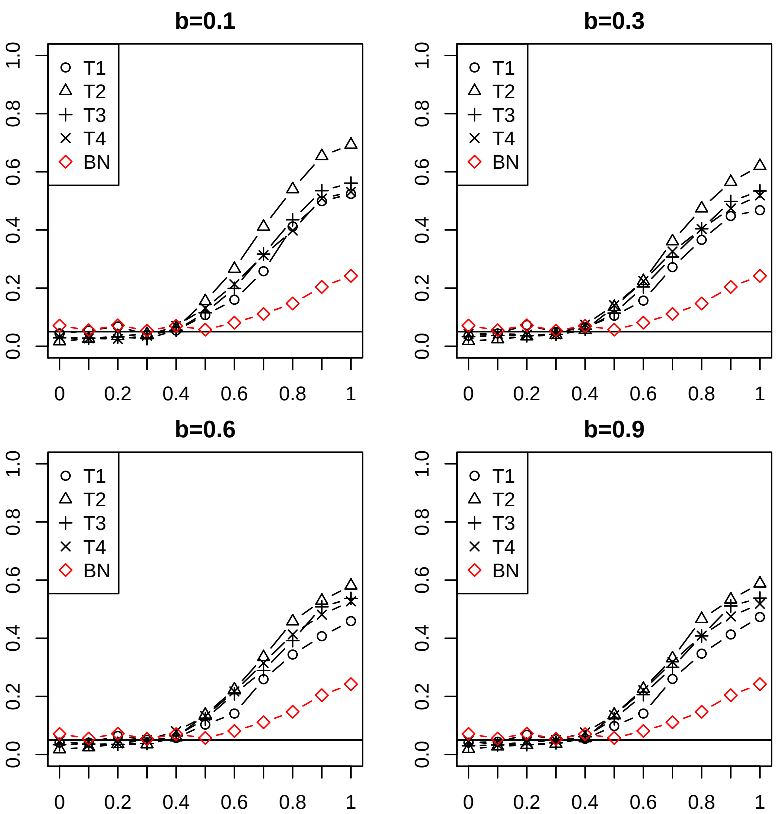

Figure 2: Rejection frequencies for mixed normal and log-normal with weight $c \in [0, 1]$ (on x-axis), $T = 50$.

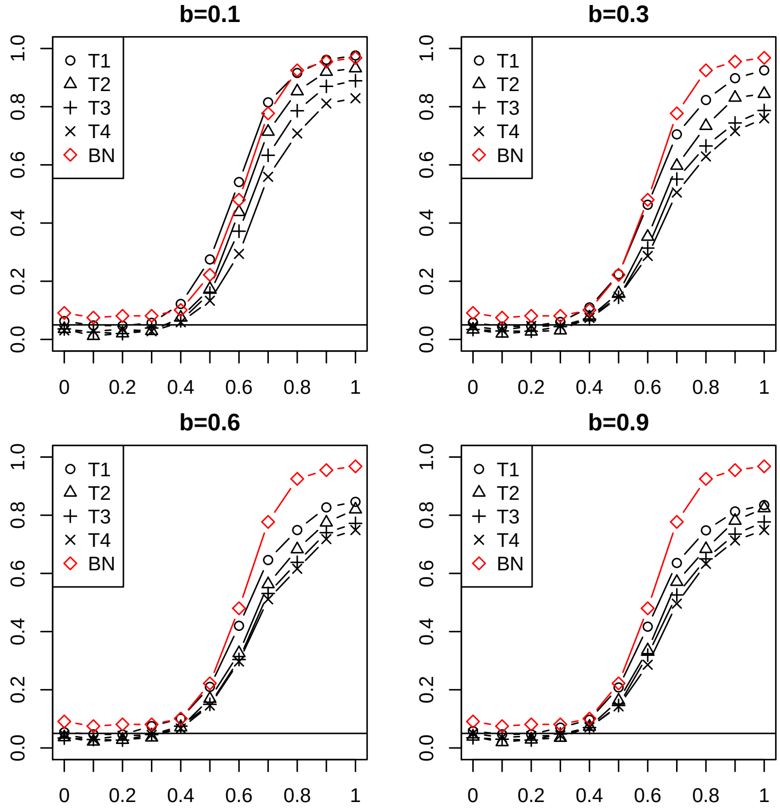

Figure 3: Rejection frequencies for mixed normal and $\chi^2(3)$ with weight $c \in [0, 1]$ (on x-axis), $T = 250$.

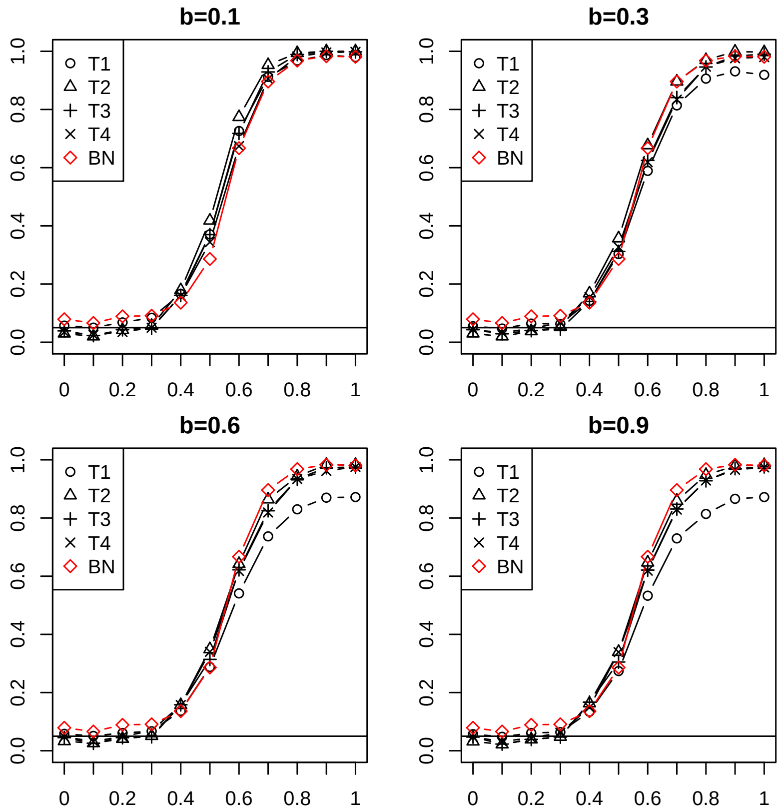

Figure 4: Rejection frequencies for mixed normal and log-normal with weight $c \in [0, 1]$ (on x-axis), $T = 250$.

size-distortions. For $c > 0.5$, the raw moment-based tests perform much better even though the tests are somewhat undersized. These considerations apply for both non-normal distributions under study. For the mixture with a $\chi^2(3)$-distribution, $\hat{\mathcal{T}}_1$ and $\hat{\mathcal{T}}_2$ perform best, while the $\hat{\mathcal{T}}_2$-statistic is most powerful against the mixture with a log-normal distribution. Overall, the particular bandwidth choice does not influence the results too much. But, for the larger sample size of $T = 250$ the choice gets much more important. As clearly seen from Figures 3 and 4, $b = 0.1$ appears to be the recommended choice.[6] For $b = 0.1$, the BN test can still be dominated in terms of power for both non-normal mixture distributions.

5 G7 industrial production growth rates

As an empirical application we consider monthly G7 industrial production growth rates obtained from the FRED database. The sample period covers the time from 1970, Jan to 2014, Oct yielding $T = 538$ observations. In Figure 5 we show the data together with QQ-plots against the normal distribution. It can be seen that for most countries, some discrepancies from normality are present, while Canada seems to be a counterexample. In Table 2 we report the outcome of different tests together with sample skewness and kurtosis. The nominal significance level is 5%. The fixed-bandwidth parameter is set equal to $b = 0.1$ as the Monte Carlo simulation results suggest. Rejections are indicated by bold faced values. We distinguish the case of constant (upper panel) and time-varying volatility (lower panel).

When assuming constant volatility, it can be clearly seen that the newly proposed test statistics typically disagree with the Bai and Ng (2005) test, except for France (FRA), where the BN test rejects the null hypothesis of normality. The moment-based tests lead to clear rejections in most cases. Interestingly, the test based on the first moment ($\hat{\mathcal{T}}_1$) is not significant in any case which reflects the fact that it is only sensitive towards skewness, but not to kurtosis. The sample statistics for the series indicate that actually kurtosis plays a much more important role in this application than skewness. An interesting result is obtained for Canada, where only the ($\hat{\mathcal{T}}_4$) statistics rejects. Apparently, it is able to detect even a relatively small deviation in the kurtosis (3.528) from its theoretical value of three under normality. From a cross-sectional perspective, it can be seen that $\hat{\mathcal{T}}_2$ and $\hat{\mathcal{T}}_4$ statistics lead to clearest test decisions in favor of non-normality. As a conclusion, the excess kurtosis in the distribution of G7 industrial production growth rates seems to be significant and shall be included in forecast models yielding predictive densities.

Turning to the results for time-varying volatility, we find strikingly different results for most countries. The US and UK seem to have rather fat tails than asymmetry

[6] It shall be noted that the power properties typically also depend on the kernel choice. So far, we have only considered the Bartlett kernel (which is known to deliver competitive power), but we are currently exploring further kernels as well.

Table 2: Normality testing for G7 industrial production growth rates at the 5% level.

Constant volatility

	skew	kurt	$\hat{\mathcal{T}}_1$	$\hat{\mathcal{T}}_2$	$\hat{\mathcal{T}}_3$	$\hat{\mathcal{T}}_4$	BN
CAN	-0.199	3.528	2.787	5.389	6.562	**21.390**	4.984
FRA	0.063	4.491	0.103	2.266	3.457	**23.543**	**8.401**
GER	0.061	10.395	1.365	**30.765**	**34.421**	**34.522**	2.725
ITA	0.333	11.800	3.721	**13.110**	**13.276**	**47.594**	3.356
JPN	-2.185	20.007	2.662	**61.277**	**61.525**	**65.050**	2.457
UK	0.059	13.375	2.208	**12.553**	12.671	15.752	4.068
US	0.264	10.110	2.033	**14.246**	**23.605**	**59.012**	5.556

Time-varying volatility

	skew	kurt	$\hat{\mathcal{T}}_1$	$\hat{\mathcal{T}}_2$	$\hat{\mathcal{T}}_3$	$\hat{\mathcal{T}}_4$	
CAN	-0.118	2.808	1.319	1.659	5.539	5.547	
FRA	0.150	3.222	1.414	4.753	6.282	10.277	
GER	-0.028	4.367	0.646	6.158	6.373	9.052	
ITA	0.183	4.023	0.942	3.980	4.052	4.063	
JPN	-0.364	4.061	3.284	3.644	4.334	4.765	
UK	-0.328	4.729	3.506	**14.540**	**16.641**	18.035	
US	0.287	4.791	1.170	**20.169**	**20.135**	**34.493**	
cv			5.016	8.872	13.200	18.258	5.991

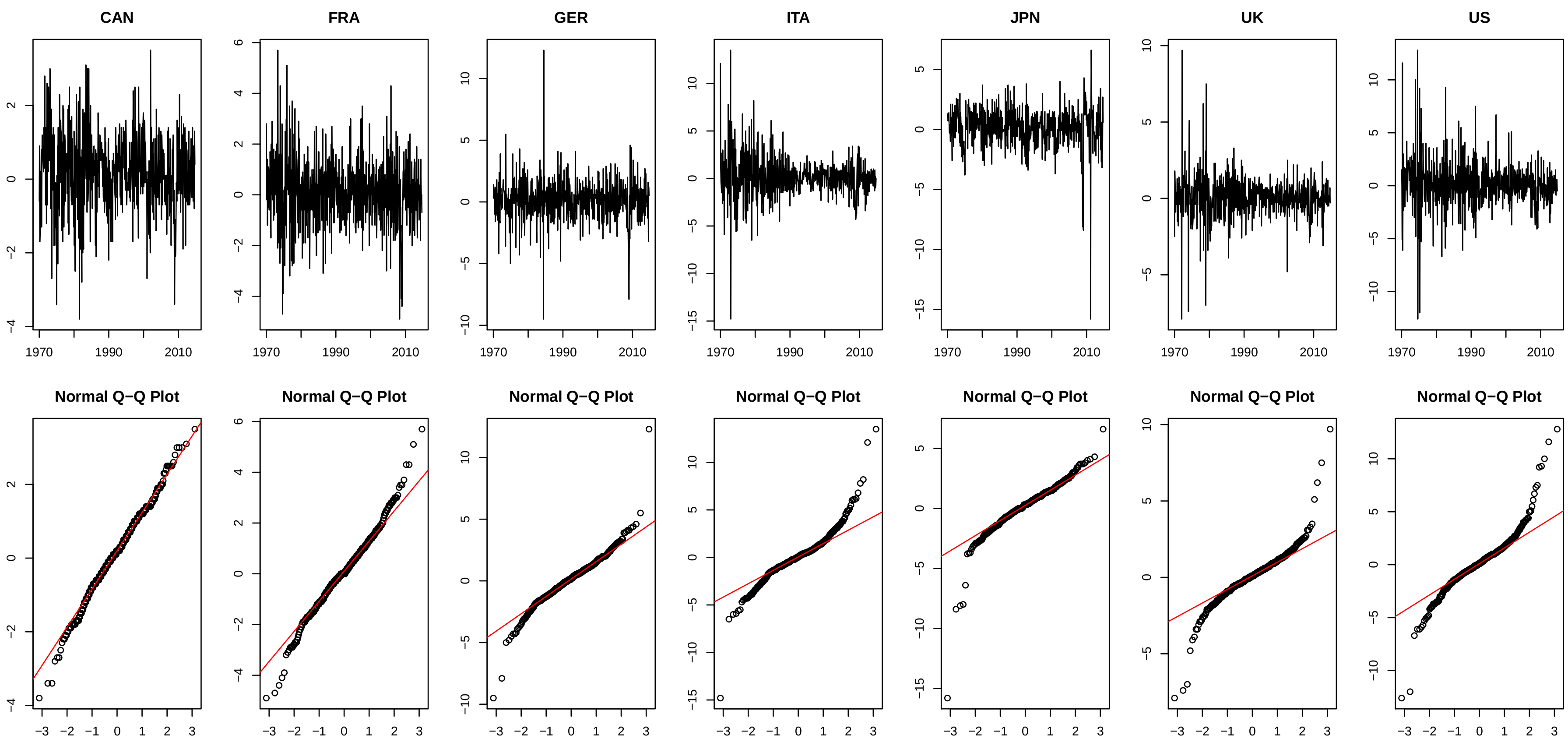

Figure 5: Monthly G7 industrial production growth rates from 1970:01-2014:11 (FRED database).

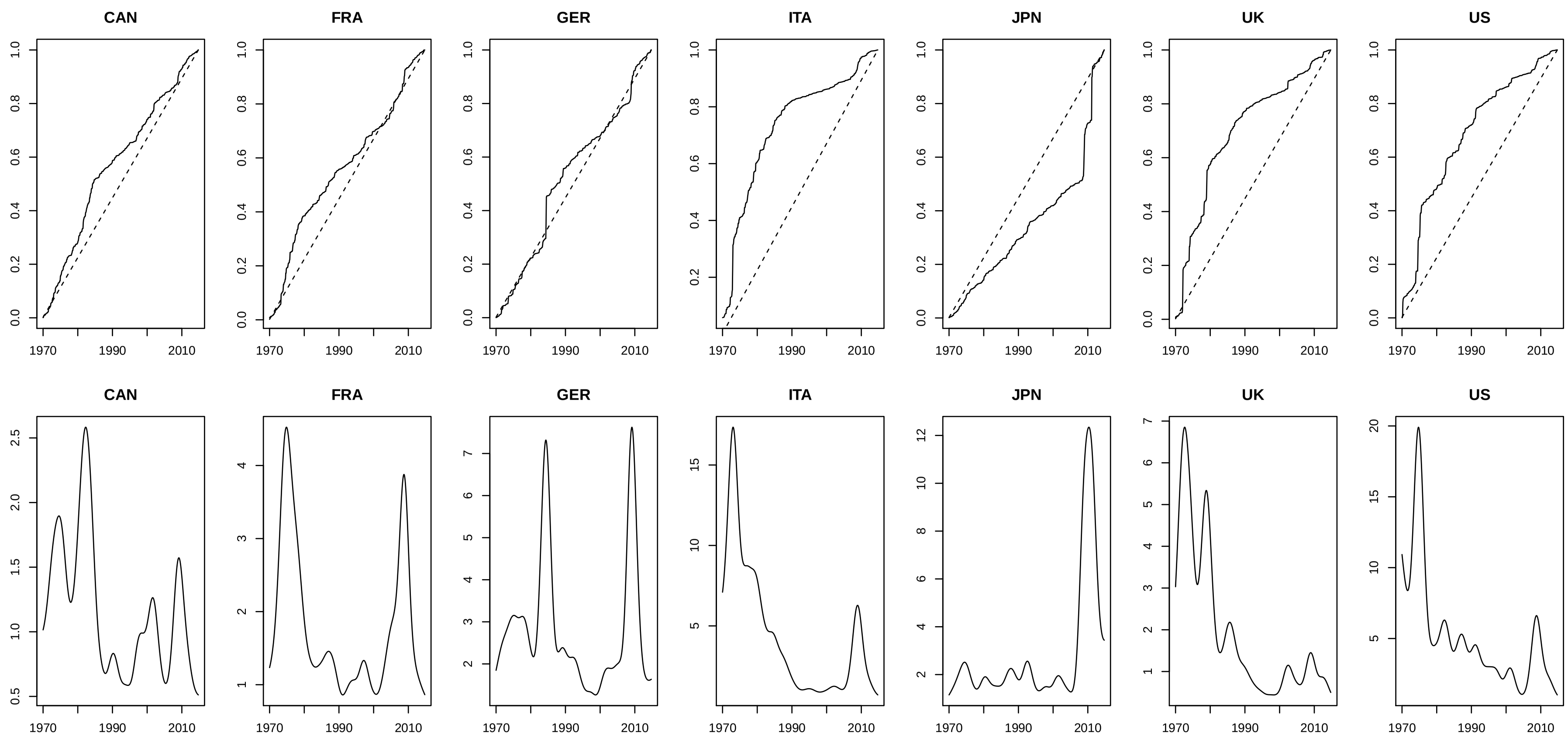

Figure 6: Variance profiles and nonparametric volatility estimation ($\hat{\sigma}_t^2$) for monthly G7 industrial production growth rates .

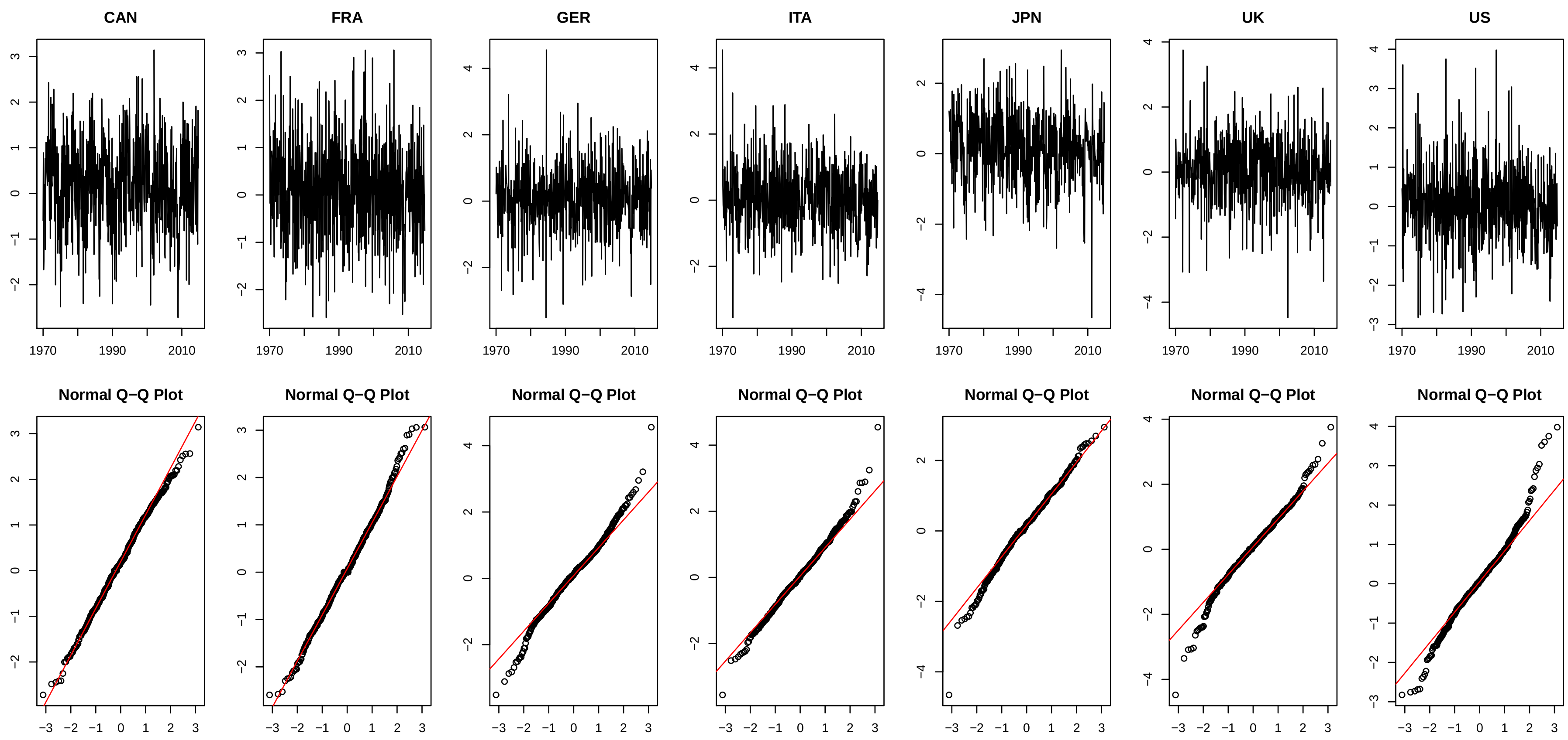

Figure 7: Monthly G7 industrial production growth rates standardized by $(\hat{\mu}, \hat{\sigma}_t)$.

6 Concluding remarks

This work considers the long-standing issue of testing for normality. The newly proposed tests are based on raw moment conditions of probability integral transformations. By doing so, we are able to construct tests which are more sensitive towards deviations from normality exhibiting zero skewness and zero excess kurtosis. The framework which we provide makes use of the so-called fixed-bandwidth approach for the estimation of long-run covariance matrices of different raw moments. As a result, the empirical size is well controlled for even in small samples under different types of autocorrelation. Time-varying unconditional mean and variance are found in many economic series. In order to cope with this typical empirical feature, our framework also allows for non-parametric time-varying variance estimation. As both, the mean and variance function of the time series are estimated, we provide a necessary correction which amounts to a modified long-run variance estimation. Our simulation study demonstrates that the suggested tests perform very well in finite samples. In an empirical application to G7 industrial production growth rates, we study the merits and limitations of the robust raw moment-based statistics.

Appendix

A Preliminary results

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

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

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

B Proofs

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 $t=1,\dots,T$. 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

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

where the absolute summability of the 8th order cumulants of z_t leads with standard arguments to the finiteness of this upper bound.

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

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

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 unform 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

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

Now,

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

where

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

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 $[sT]=M\left[sB\right]$ for the sake of the exposition. Then

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

The first summand on the r.h.s. is easily shown to be $O_p\left(\frac{\sqrt{TM}}{\tau}\sup_{t\in1,\ldots,T}\left|h\left(z_j\right)\right|\right)$. For the second, note that

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

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

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

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

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

so choosing B appropriately 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

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

For B_{1T}, we have

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

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 $\sigma^2\left(\cdot\right)$ about t/T to obtain

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

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 don't affect the result, $C_{2T}=0$ and

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

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 $g\left(z_t\right)$, we have $\sup_t|g\left(z_t\right)|=O_p\left(T^\gamma\right)=\sup_{t=1,\dots,T}z_t^2$ for any $\gamma>0$, and γ can then be chosen arbitrarily close to 0 to make the r.h.s. $o_p\left(1\right)$.

For B_{2T}, we have

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

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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mu_t-\hat{\mu}_t\right)^2=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{2\tau+1}\sum_{j=t-\tau}^{t+\tau}\left(\left(\mu_t-\mu_j\right)-\sigma_jz_j\right)\right)^2=O_p\left(T^{-\delta}\right)\qquad(7)$$

for some $0<\delta<\min\left\{\kappa_1-1/2;3/4-\kappa_2\right\}$, 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

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

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

and

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

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

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

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 $O(\tau/T)$ 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.

Summing up, $\sup_{s\in[0,1]}|A_{1T}| \xrightarrow{p} 0$; to complete the result, 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 $\sup_{s\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\hat{\sigma}_t^2}{\sigma_t^2}-1\right)^2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_t^2}{\sigma_t^2}-1\right)^2 = O_p\left(T^{-\delta}\right)$, analogously to Equation (7), so the result follows after choosing $\gamma<\delta$.

Proof of item 5

We have that

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

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 when 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 $\sup_{s\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{1}{\hat{\sigma}_t^2}\left(\mu_t-\hat{\mu}_t\right)^2$ vanish in probability. To show this, we have like in the proof of item 4 that

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

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

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

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

Proof of item 6

Note that $r_t=\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t}-z_t\right)\left(\varphi\left(z_t\right)+\varphi'\left(\xi_t\right)\left(\frac{x_t-\hat{\mu}_t}{\hat{\sigma}_t}-z_t\right)\right)$ 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

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

We now show the first summand 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

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

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 dealt with by writing

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

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

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

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

with $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_t^2}{\sigma_j^2}-1\right)^2$ vanishing like in the proof of item 4 and $\sup_t\left|z_t\right|=O_p\left(T^{\gamma}\right)$ for positive γ arbitrarily close to zero.

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 after 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 analog of B_{1T} from the proof of item 4 we write

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

where the sumands 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

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

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

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

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

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

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

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

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

Write with a Taylor expansion

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

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 that

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

uniformly in s, thanks to Lemma 2 item 6.

We may then focus on

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

where 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 2 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 $\sup_t\hat{\sigma}_t$ bounded in probability and nonnegative.

Using Lemma 2 again, items 7 and 8, we obtain uniformly in $s \in [0,1]$ that

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

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

The arguments of the proof of Theorem 2 in Kiefer and Vogelsang (2005), together with the result in Lemma 1, lead directly to the desired result.

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 may treat σ_t as known and set it to unity; similar effects are expected if σ_t is to be modeled as well. Let us hence consider 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 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 3 *Let $\mu(s, \boldsymbol{\theta})$ have uniformly continuous 2nd 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

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

by plugging in an estimator $\hat{\boldsymbol{\theta}}$ which is taken to be $\sqrt{T}$-consistent. The straightforward choice is the NLS estimator, which we employ in the following; some of the requirements of Assumption 3, e.g. referring to the Hessian of m, help establish the limiting behavior of the NLS estimator. Irrespective of what estimator is used, we note that

$$\hat{p}_t = \Phi\left(z_t - \mu\left({}^{t}\!/\!{}_{T}, \hat{\boldsymbol{\theta}}\right) + \mu\left({}^{t}\!/\!{}_{T}, \boldsymbol{\theta}\right)\right) \tag{8}$$

such that the estimation has an effect. The following Lemma provides the precise result when $\mathrm{E}(x_t)$ is parametrically adjusted.

Lemma 3 *Under Assumptions 1 through 3, it holds as $T \to \infty$ that*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left(\hat{p}_t^k - \frac{1}{k+1}\right) \Rightarrow B_k(s) - k\vartheta_{k-1}\boldsymbol{\delta}'(s, \boldsymbol{\theta})\,\boldsymbol{\Theta}(1) \tag{9}$$

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 3

Begin by discussing the limiting behavior of the NLS estimators $\hat{\boldsymbol{\theta}}$. We have under Assumptions 1 and 3 that

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

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 rest term in differential form) we obtain

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

where ξ_t lies between z_t and $z_t-\mu\left(t/T,\hat{\boldsymbol{\theta}}\right)+\mu\left(t/T,\boldsymbol{\theta}\right)$ for each 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:

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

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

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

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

Examining the third summand on the r.h.s., we note that the boundedness of φ' and the fact that $\left|\left.\frac{\partial m(t/T,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\right|\leq f\left(\|\boldsymbol{\vartheta}_t\|\right)\leq f\left(\max\left\{\|\boldsymbol{\theta}\|;\left\|\hat{\boldsymbol{\theta}}\right\|\right\}\right)$ make the partial sums of order $O_p\left(T\right)$, but $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}=O_p\left(T^{-0.5}\right)$ and the normalization with $\sqrt{T}$ make the entire summand vanish.

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

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

Similarly, φ' is bounded and $\left|\left.\frac{\partial\mu(t/T,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\vartheta}_t}\right|\leq f\left(\|\boldsymbol{\vartheta}_t\|\right)\leq f\left(\max\left\{\|\boldsymbol{\theta}\|;\left\|\hat{\boldsymbol{\theta}}\right\|\right\}\right)$ 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,

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

the same arguments show that analogous relations hold for $\hat{p}_t^k$. 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 6 *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 however between their result and our Lemma 3, since the result in* (9) *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 (9) will affect the limiting fixed-b distribution of a statistic based on a parametric estimated standardization. The effect is different from that deriven in Lemma 1, since the presence of $\boldsymbol{\Theta}\left(1\right)$ (as opposed to $W_1\left(s\right)$) indicates some sort bridge behavior of the limit process. Moreover, the components $\boldsymbol{\Theta}$ and $\boldsymbol{\delta}$ depend on the specific model μ chosen. The statistics can be made pivotal like in Section 3, but the limiting distributions are not the usual fixed-b ones, 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). This implies the need to simulate the distributions for each specific type of deterministic component accounted for in the data. 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.

For the case of a linear trend, we have the following procedure simplified by the linearity of the mean function. Detrendx_t using OLS regression and standardize the detrended series with $\hat{\sigma}_t$ to obtain $\hat{z}_t$. With $\left(\hat{p}_t,\ldots,\hat{p}_t^K,\hat{z}_t\right)'$, compute like in the mean case an fixed-b estimate of the long-run covariance matrix Ξ and, based on it, the scaling matrix $\hat{\hat{\Omega}}=V\hat{\Xi}V$ with V like beforeand then $\hat{\mathcal{T}}$ from (5). Then,

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

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

with

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

for smooth kernels and

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

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

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

and, for detrending, the second-level Brownian bridge

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

Proof of Proposition 2

To deal with detrending, let $\mu = \theta_1$ in Lemma 3 to obtain

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

We then need to examine the limiting behavior of the suitably normalized partial sums of $\hat{z}_t$. To this end, note that

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

uniformly in s thanks to the arguments used in the proof of Lemma 1. Then,

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\hat{z}_t &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\frac{\sigma_t}{\hat{\sigma}_t}(z_t - \bar{z}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}(z_t - \bar{z}) + \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\left(\frac{\sigma_t}{\hat{\sigma}_t} - 1\right)(z_t - \bar{z}) \\ &\Rightarrow W_1(s) - sW_1(1).\end{aligned}$$

Let now

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

and

$$\tilde{\boldsymbol{B}} = (B_1(s) - s\,\vartheta_0 W_1(1), \ldots, B_K(s) - s\,K\vartheta_{K-1}W_1(1), W_1(s) - sW_1(1))';$$

using the arguments of the proof of Theorem 2 in Kiefer and Vogelsang (2005) together with the Lemma 1, we obtain e.g. for smooth kernels

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

Note further that

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

and let $\boldsymbol{Y} = V\bar{\boldsymbol{B}}$ such that

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

where $\boldsymbol{Y}$ is a multivariate Brownian motion; since its long-run covariance matrix cancels out, the r.h.s. is the required distribution. The result for the Bartlett kernel follows analogously.

To deal with detrending, let $\mu = \theta_1 + \theta_2 s$ in Lemma 3 to obtain

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

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

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

where

$$Y_T = \left[\begin{array}{c} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t-\bar{y})^3 \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(y_t-\bar{y})^4-3\hat{\sigma}^4] \end{array} \right]$$

and

$$\hat{\gamma} = \left[\begin{array}{cccc} -3\hat{\sigma}^2 & 0 & 1 & 0 \\ 0 & -6\hat{\sigma}^2 & 0 & 1 \end{array} \right]$$

$\bar{y}$, $\hat{\sigma}$ and $\hat{\Phi}$ are consistent estimators. The theoretical long-run covariance matrix Φ is given by $\Phi = \lim_{T\to\infty} T\,\mathrm{E}(\bar{Z}\bar{Z}')$ with $Z' = \left[\begin{array}{cccc} y_t-\mu, & (y_t-\mu)^2-\sigma^2, & (y_t-\mu)^3, & (y_t-\mu)^4-3\sigma^4 \end{array} \right]$ and $\bar{Z}$ being the sample mean of Z_t. The limiting distribution of μ_{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.

E Critical values

Table 3: Critical values via response curves from the $\mathcal{KV}_{K,b,\kappa}$-distribution. κ is the Bartlett kernel. The regression is given by $cv(b) = a_0 + a_1 b + a_2 b^2 + a_3 b^3 + error$ with corresponding R^2. Nominal significance levels are 0.9, 0.95, 0.975, 0.99 and 0.995.

	a_0	a_1	a_2	a_3	R^2
$K=1$					
0.9	2.7055	6.1598	8.6142	-3.3854	0.9998
0.95	3.8415	10.2574	15.6231	-7.0320	0.9997
0.975	5.0239	15.8489	24.5892	-12.5751	0.9995
0.99	6.6349	26.3361	36.1330	-19.6341	0.9994
0.995	7.8794	37.5823	41.2076	-21.6338	0.9991
$K=2$					
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
0.995	10.5966	71.9545	96.5536	-60.2045	0.9994
$K=3$					
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
0.995	12.8382	121.7357	137.7951	-91.2883	0.9994
$K=4$					
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
0.995	14.8603	177.5045	183.2276	-123.6561	0.9996

References

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

Amado, C. and T. Teräsvirta (2014). Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance 25*(1), 15–35.

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

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

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *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.

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.

Cavaliere, G. (2004). Unit root tests under time-varying variances. *Econometric Reviews 23*(3), 259–292.

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

Cavaliere, G. and A. M. R. Taylor (2009). Heteroskedastic time series with a unit root. *Econometric Theory 25*(5), 1228–1276.

Clark, T. E. (2009). Is the Great Moderation over? An empirical analysis. *Economic Review 4*, 5–42.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford university press.

Demetrescu, M. and C. Hanck (2012). Unit root testing in heteroskedastic panels using the Cauchy estimator. *Journal of Business & Economic Statistics 30*(2), 256–264.

Durbin, J. (1973). *Distribution theory for tests based on the sample distribution function*, Volume 9. Siam.

Guidolin, M. and A. Timmermann (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics 21*, 1–22.

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.

Knüppel, M. (2014). Evaluating the calibration of multi-step-ahead density forecasts using raw moments. *Journal of Business & Economic Statistics forthcoming*.

Lanne, M., J. Luoto, and P. Saikkonen (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting 28*(3), 623–631.

Lanne, M. and P. Saikkonen (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics 3*(3), article 2.

Lanne, M. and P. Saikkonen (2013). Noncausal vector autoregression. *Econometric Theory 29*(3), 447–481.

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

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.

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

Phillips, P. C. B. and K. L. Xu (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis 27*(2), 289–308.

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.

Teräsvirta, T. and Z. Zhao (2011). Stylized facts of return series, robust estimates and three popular models of volatility. *Applied Financial Economics 21*(1-2), 67–94.

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

Westerlund, J. (2014). Heteroscedasticity robust panel unit root tests. *Journal of Business & Economic Statistics 32*(1), 112–135.

Xu, K.-L. (2008). Bootstrapping autoregression under non-stationary volatility. *The Econometrics Journal 11*(1), 1–26.

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.