Demetrescu, Matei; Kruse, Robinson

Conference Paper
Testing heteroskedastic time series for normality


Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association


This Version is available at:
http://hdl.handle.net/10419/113221

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Testing heteroskedastic time series for normality

Matei Demetrescu†
Christian-Albrechts-University of Kiel

Robinson Kruse
CREATEs and Leibniz University Hannover

Preliminary version: February 20, 2015
Please do not quote

Abstract

Normality testing is an evergreen topic in statistics and econometrics and other disciplines. The paper focuses on testing economic time series for normality in a robust way, taking specific data features such as serial dependence and time-varying volatility into account. Here, we suggest tests based on raw moments of probability integral transform of standardized time series. The use of raw moments is advantageous as they are quite sensitive to deviations from the null other than asymmetry and excess kurtosis. To standardize the series, nonparametric estimators of the (time-varying) variance may be used, but the mean as a function of time has to be estimated parametrically. Short-run dynamics is taken into account using the Heteroskedasticity and Autocorrelation Robust [HAR] approach of Kiefer and Vogelsang (2005, ET). The effect of estimation uncertainty arising from estimated standardization is accounted for by providing a necessary modification. In a simulation study, we compare the suggested tests to a benchmark test by Bai and Ng (2005, JBES). The results show that the new tests are performing well in terms of size (which is mainly due to the adopted fixed-b framework for long-run covariance estimation), but also in terms of power. An empirical application to G7 industrial production growth rates sheds further light on the empirical usefulness and limitations 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 Philipp Sibbertsen for helpful comments.

†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. Also, nonnormality may 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 complicated, 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 with in econometric practice. Consider for instance the situation where a series is marginally normal, but exhibits one break in the variance. The marginal distribution is a mixture of two normals, which is nonnormal (and typically leptokurtic), so a normality test ignoring the variance break will reject the null more often than the nominal level of the test indicates. The reasoning extends to more general patterns of variance changes. And indeed, economic data are often found to exhibit time-varying volatility. Examples can be found in the field of financial data (see among others Guidolin and Timmermann, 2006; Amado and Teräsvirta, 2014; Teräsvirta and Zhao, 2011; Amado and Teräsvirta, 2013) and also macroeconomic time series (see e.g. Stock and Watson, 2002; Sensier and van Dijk, 2004; Clark, 2009, 2011; Justiniano and Primiceri, 2008) such as asset returns, economic growth or price changes. 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 parameters. The main reason to do so is that PITs may be more sensitive against alternatives with non-zero skewness and excess kurtosis since they take higher-order moments into account. PITs have already been used successfully by Knüppel (2014), though without accounting for estimation effects. The marginal mean and variance as a function of time have to be estimated; one important contribution of the paper is to show that the variance function may be estimated in a nonparametric fashion. As a consequence the practitioner does not have to specify a model for the variance explicitly. This is not the case about the deterministic mean component, whose estimation has asymptotic effects. We show how to modify the naive moment-based test statistics such that pivotal inference is possible.

Regarding robustness against serial dependence, we adopt the approach by Kiefer and Vogelsang (2005), who propose a new asymptotic framework, known as fixed-\(b\) asymptotics. The main feature 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 new non-standard asymptotic limiting distributions of tests statistics (like \(t\), Wald and \(F\)).

¹Phillips 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).
Importantly, the critical values obtained from such distributions reflect the choice of bandwidth and kernel even as $T \to \infty$. Available simulation results convincingly demonstrate that the fixed-$b$ approach may provide much more accurate finite-sample inference.\(^2\) Another important contribution of the paper is to show how to account for the adjustment for deterministic mean components other than a constant mean.

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 estimation uncertainty is located in Section 3. It also contains some extensions to a more general mean function. 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, 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 Setup

The series of interest $x_t$ is marginally normal under the null. The series exhibits a constant mean $\mu$, but a time-varying variance $\sigma_t^2$ as given by the following component model

$$x_t = \mu + \sigma_t z_t, \quad t = 1, 2, \ldots, T,$$

where $z_t$ is unconditionally homoskedastic and otherwise short-range dependent, while the time-varying variance is induced by the deterministic triangular array $\sigma_t = \sigma_{tT}$. The following assumptions make the notions of short-run dependence and time-varying variance precise.

**Assumption 1** Let $z_t$ be a zero-mean strictly stationary series with unity long-run variance, $L_{2+\delta}$-bounded for some $\delta > 0$, and strong mixing with coefficients $\alpha(j)$ for which

$$\alpha(j) < A j^{-\beta} \quad \text{for some } \beta > \frac{2+2\delta}{\delta}.$$  

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 actually integrated of order zero. The condition also allows for mild form of conditional heteroskedasticity, so the observed series $x_t$ may exhibit both conditional and unconditional heteroskedasticity. Together with marginal normality (implying the finiteness of moments of any order) Assumption 1 ensures e.g. weak convergence of the suitably normalized partial sums of $z_t$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} z_t \Rightarrow W(s), \quad (1)$$

\(^2\)Recent contributions to this field include inter alia Yang and Vogelsang (2011), Vogelsang and Wagner (2013) or Sun (2014a,b).
where \( W \) is a standard Wiener process (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.

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 \( \sigma_t \) as marginal (long-run) standard deviation. The variance function itself is taken to satisfy typical conditions in the literature (cf. Cavaliere, 2004):

**Assumption 2** The triangular array \( \sigma_t \) is given as \( \sigma_t = \sigma(t/T) \) where \( \sigma(\cdot) \) is piecewise Lipschitz and bounded away from zero on \([0,1]\).

We base our test of the null hypothesis on moments of transformed series rather than the original series \( x_t \). With \( \Phi \) being the cdf (and \( \varphi \) 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. The test for normality will be based on the first \( K \) moments (including at least the first and second ones) of the uniformly distributed PITs. It holds under the null of uniformly distributed PITs that

\[
E(p^k_t) = \frac{1}{k+1} ; \quad k \in \mathbb{N} \tag{2}
\]

such that, under Assumption 1,

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

where \((B_1, \ldots B_K)^t\) is a \(K\)-variate Brownian motion with covariance matrix \( \Omega \). Because \( p_t \) is only marginally uniform, \( \Omega \) depends in general on the specific data generating process at hand. We shall resort to an estimate thereof (obtained by the standard spectral density based approach, 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 \( \Omega \). This follows the approach of Bai and Ng (2005) or Bontemps and Meddah (2005) to deal with serial dependence of unknown form.

Suppose that the test can be based directly on empirical moments of \( p_t \) (i.e. under known parameters \( \mu \) and \( \sigma_t \)). With \( m_k = \frac{1}{T} \sum_{t=1}^{T} p^K_t \), 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 \( \omega_k^2 \) being the long-run variance of \( p^K_t \). Let \( \hat{\omega}_k^2 = \sum_{j=-T+1}^{T-1} \kappa \left( \frac{j}{T} \right) \hat{\gamma}_j \) denote an estimator of \( \omega^2 \) with proportional bandwidth \( B = [bT], b > 0 \). The \( \hat{\gamma}_j \)'s denote the usual autocovariance estimator at lag \( j \):

\[
\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} \left( p^K_t - \bar{p}^k \right) \left( p^K_{t-j} - \bar{p}^k \right) .
\]
For $b \in (0, 1)$ we have (see Kiefer and Vogelsang, 2005)

$$t_k^2 \Rightarrow KV_{1,b,\kappa}.$$ 

For simplicity we work with smooth, “Andrews-type” kernels $\kappa(\cdot, \cdot)$ with real support, but the results extend in a straightforward manner to truncation kernels. The limiting distribution under fixed-$b$ can then be characterized in more detail as

$$KV_{1,b,\kappa} \equiv W(1) \left( -\int_0^1 \int_0^1 \frac{1}{b^2} \kappa'' \left( \frac{r-s}{b} \right) (W(r) - rW(1))(W(s) - sW(1))' \, dr \, ds \right)^{-1} W(1).$$

For $b \to 0$, the standard asymptotic framework applies and thus, $t_k^2 \Rightarrow \chi^2_1$. We shall work in the following with fixed-$b$ asymptotics only, with the understanding that usual $b \to 0$ asymptotics are encompassed since $KV_{K,b,\kappa}$ converges to $\chi^2_K$ for $b \to 0$ c.f. (Kiefer and Vogelsang, 2005).

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

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

with $\hat{\Omega}$ being the corresponding estimator of $\Omega$, $\hat{\Omega} = \sum_{j=-T}^{T} \kappa(j) \hat{\Gamma}_j$ and $\hat{\Gamma}_j$ is the usual autocovariance matrix estimator for $(p_t, \ldots, p^K_t)'$. Similarly,

$$T_K \Rightarrow KV_{K,b,\kappa},$$

where

$$KV_{K,b,\kappa} \equiv W'(1) \left( -\int_0^1 \int_0^1 \frac{1}{b^2} \kappa'' \left( \frac{r-s}{b} \right) (W(r) - rW(1))(W(s) - sW(1))' \, dr \, ds \right)^{-1} W(1)$$

with $K$ the dimension of the vector $W$ of independent standard Wiener processes.

The use of $\hat{p}_t$ instead of $p_t$ for 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. Since we also rely on PITs, a robustification will be required. Because we rely on sample moments to build our test, a simple adjustment of the covariance matrix estimator will suffice, unlike in Bai (2003). See Section 3 for the precise details. But we discuss the issue of time-varying volatility before proceeding to evaluating the effects of estimation uncertainty.

For the PIT-based tests, it turns out that estimating an unknown mean requires a correction which parallels that of Bai and Ng (2005). Yet our main finding is that estimating an unknown variance function does not affect the limiting properties of the PITs. In particular, nonparametric estimation of the unknown variance function can be allowed for under relatively weak conditions as specified in Assumption 3 below. The mean function could be specified to be more general as well, but requires setting up a parametric model since $\sqrt{T}$-consistency of the resulting parameter estimators is required, unlike the case of the variance. In Section 3, we also discuss
such extensions. The estimator $\hat{\sigma}_t$ should satisfy the following requirements.

**Assumption 3** The (in general nonparametric) estimator $\hat{\sigma}_t$ satisfies

1. $\hat{\sigma}_{[sT]}$ is bounded away from zero with probability one;
2. $\sup_{s \in [0,1]} |\hat{\sigma}_{[sT]} - \sigma_{[sT]}| = O_p (T^{-\alpha})$ for some $1/4 < \alpha \leq 1/2$;
3. $|\hat{\sigma}_{[sT]} - \hat{\sigma}_{[rT]}| \leq T^{\beta} Q_T |s - r|$ for all $0 \leq r, s \leq 1$, where $Q_T$ is uniformly bounded in probability and $0 \leq \beta < 2\alpha - 1/2$;
4. $\hat{\sigma}_{[sT]}$ has the same jump discontinuities as $\sigma(\cdot)$.

Condition 1 is quite plausible fulfilled considering that $\sigma^2(\cdot)$ is bounded away from zero and the estimator is converging uniformly (Condition 2). The case where $\alpha = 1/2$ simply takes the possibility into account, that a parametric estimator of $\sigma(\cdot)$ is available and $\sqrt{T}$-consistency of the corresponding estimators is given. E.g. when the variance is known to be constant, the natural choice for an estimator is simply the sample variance; this estimator is also clearly satisfying the Lipschitz-type condition 3. Condition 4 presupposes that the break dates, should there be breaks in the variance, are known; this is not a critical restriction since such break dates can be estimated superconsistently.

The following lemma shows that the standard Nadaraya-Watson estimator fulfils the conditions of Assumption 3 if suitably choosing the bandwidth parameter.

**Lemma 1** Choosing $\hat{\sigma}_t = \sqrt{\hat{\sigma}^2 (t)}$ with $\hat{\sigma}^2 (s)$ the Nadaraya-Watson estimator of the variance as a function of time,

$$\hat{\sigma}^2 (s) = \frac{\sum_{t=1}^T (x_t - \bar{x})^2 \kappa \left( \frac{s - t/T}{h} \right)}{\sum_{t=1}^T \kappa \left( \frac{s - t/T}{h} \right)}, \quad s \in [0,1],$$

satisfies Assumption 3 for a bandwidth $h = CT^{-\beta}$ when $\beta \in \left( \frac{1}{6}, \frac{1}{2} \right)$, and $\kappa$, as well as $\kappa'$, satisfy the usual regularity conditions.

**Proof:** see the Appendix.

Note that the optimal (at least for the iid case) bandwidth $h = O (T^{-0.2})$ can be chosen under the conditions of Assumption 3.

### 3 Estimation uncertainty

#### 3.1 The effects of estimation uncertainty

Let

$$\hat{p}_t = \Phi \left( \frac{x_t - \bar{x}}{\hat{\sigma}_t} \right)$$

with $\hat{\sigma}_t$ being an estimator of the standard deviation $\sigma_t$ obeying Assumption 3 and $\bar{x}$ denotes the sample mean of $x_t$. Let $\hat{m}_k = \frac{1}{T} \sum_{t=1}^T \hat{p}_t^k$ denote the sample average of $\hat{p}_t^k$. 

6
The limiting distribution of the "naive" test statistics which simply replace \( m_k \) with \( \hat{m}_k \) in the definition of \( t_k \) or \( T_K \) will not follow a \( KV \)-distribution asymptotically any longer. Importantly, the weak convergence in (3) is replaced by the following limiting behavior.

**Lemma 2** Under Assumptions 1, 2 and 3 it holds as \( T \to \infty \) that

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

with \( W \) from (1), \( B_k \) from (3) and \( \vartheta_{k-1} = E \left( p_t^{k-1} \phi(z_t) \right) \). Moreover, the convergence holds jointly for \( k = 1, \ldots, K \) for any fixed \( K \in \mathbb{N} \).

**Proof:** see the Appendix.

Note that \( \vartheta_0 = E (\phi_l) = \int_{-\infty}^{\infty} \varphi^2(x) dx = \frac{1}{\sqrt{2\pi}} \); the higher-order expectations (for \( \vartheta_k, k \geq 1 \)) can be tabulated (see Section 4) but a closed-form expression does not seem to exist.

By examining the proof of Lemma 2, it can be seen that the drift term \( s k \vartheta_{k-1} \) is due to demeaning of \( x_t \) and not due to scaling with an estimated variance function; this relates to the relative broad array of choices in modelling the variance of the series to be tested for normality.

Still, the limiting theory for \( p_t \) does not apply directly whenever the deterministic additive mean component of \( x_t \) is estimated, and a feasible correction is required.

### 3.2 Long-run covariance matrix estimation

By Lemma 2 we have that \( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \hat{p}_t^k - \frac{1}{k+1} \right) \Rightarrow B_k(1) - k \vartheta_{k-1} W(1) \) which makes the “numerator” of the naive test statistic normally distributed in the limit. 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

\[
\hat{T}_K = T \left( \hat{m}_1 - \frac{1}{2}, \ldots, \hat{m}_K - \frac{1}{K+1} \right) \hat{\Omega}^{-1} \left( \hat{m}_1 - \frac{1}{2}, \ldots, \hat{m}_K - \frac{1}{K+1} \right)^T
\]

with

\[
\hat{\Omega} = V \hat{\Xi} V'
\]

where \( \hat{\Xi} \) is a long-run covariance matrix estimator of \( (p_t, \ldots, p_t^K, z_t)^T \) (based on \( (\hat{p}_t, \ldots, \hat{p}_t^K, \hat{z}_t)^T \)), and \( V = \left( I_K: \ i_K \right) \) with \( i_K = \left( \vartheta_0, \ldots, K \vartheta_{K-1} \right)' \). Although the long-run variance of \( z_t \) is assumed to be unity (c.f. Assumption 1), its estimation is required for the limiting distribution to belong to the \( KV \) family of distributions; see the proof of Proposition 1 below. Moreover, the long-run covariances of \( z_t \) and powers of \( p_t \) are unknown so imposing the unity restriction would not lead to significant benefits, not even in relatively small samples.

**Proposition 1** Under Assumptions 1, 2 and 3, it holds as \( T \to \infty \) that

\[
\hat{T}_K \Rightarrow KV_{K,b,k}.
\]

**Proof:** see the Appendix.
Several remarks are in order. First, the approach involves the expectations \( \vartheta_k = \mathbb{E}(p_k \varphi(z_t)) \) which are specific to the normal distribution via \( \varphi(\cdot) \). Our framework clearly allows testing other null distributions in location-scale models. The values of \( \vartheta_k \) would have to be determined on a case-by-case basis, but the approach leads to test statistics with the same pivotal limiting distributions. Finally, 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 often the case when e.g. density forecasts are evaluated; see Knüppel (2014).

3.3 Extensions

Let us consider an extended model for the mean of the observed time series \( x_t \)

\[
x_t = m(t/T, \theta) + \sigma_t z_t ,
\]

where we set \( \sigma_t = 1 \) for ease of presentation. The main arguments are not affected by this simplification. Note also that normalizing the time is not restrictive, since one may redefine a classical linear trend model \( m = \theta_1 + \theta_2 t \) as \( m = \theta_1 + (T\theta_2) t/T \) without loss of generality. We take the mean component to satisfy the following requirements.

**Assumption 4** Let \( m(s, \theta) \) have uniformly continuous 2nd order partial derivatives. The first and second order partial derivatives w.r.t. \( \theta \) are weakly uniformly bounded in \( s \) in the sense that there exists a nondecreasing function \( f \) such that

\[
\max \left\{ \left\| \frac{\partial m(s, \theta)}{\partial \theta} \right\|, \left\| \frac{\partial^2 m(s, \theta)}{\partial \theta \partial \theta'} \right\| \right\} \leq f(\|\theta\|) \text{ for all } s.
\]

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

Based on this model, one obtains

\[
\hat{p}_t = \Phi(\hat{z}_t) = \Phi(x_t - m(t/T, \hat{\theta}))
\]

by plugging in an estimator \( \hat{\theta} \) 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 4, 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

\[
\hat{p}_t = \Phi(z_t - m(t/T, \hat{\theta}) + m(t/T, \theta))
\]

such that the estimation has an effect, just like in Lemma 2. The following Lemma provides the analogous result when \( m \) is not just an intercept.

**Lemma 3** Under Assumptions 1 and 4 with \( \sigma_t = 1 \) known, it holds as \( T \to \infty \) that

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[tT]} \left( \hat{p}_t^k - \frac{1}{k+1} \right) \Rightarrow B_k(s) - k \vartheta_{k-1} \mu'(s, \theta) \Theta
\]
where $\Theta (1) = \left( \int_0^1 \frac{\partial m(s, \theta)}{\partial \theta} \frac{\partial m(s, \theta)}{\partial \theta} \, ds \right)^{-1} \int_0^1 \frac{\partial m(s, \theta)}{\partial \theta} \, dW (s)$, $\mu (s, \theta) = \int_0^s \frac{\partial m(x, \theta)}{\partial \theta} \, dx$ and $\vartheta_{k-1} = E \left( p_t^{-1} \varphi (z_t) \right)$ as before.

Proof: see the Appendix.

Before moving on, note that 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 (8) 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 $\frac{\partial m(y_t, \theta)}{\partial \theta} = w_t$. We obtain for stationary regressors that $\frac{1}{T} \sum_{t=1}^{[T]} \varphi (z_t) w_t \Rightarrow s E (\varphi (z_t) w_t)$ now, Bai and Ng (2005) assume that an intercept is always present in the regression, which is equivalent to setting $E (w_t) = 0$; they also assume the regressors to be independent of $z_t$, hence $E (\varphi (z_t) w_t) = 0$ and correspondingly $\mu (s) = 0$. 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 (8) will affect the limiting fixed-$b$ distribution of a statistic based on an estimated standardization. The normalized sample moments of $\hat{p}_t$ behave namely as

$$B_k (1) - k \vartheta_{k-1} \mu' (1, \theta) \left( \int_0^1 \frac{\partial m(s, \theta)}{\partial \theta} \frac{\partial m(s, \theta)}{\partial \theta} \, ds \right)^{-1} \int_0^1 \frac{\partial m(s, \theta)}{\partial \theta} \, dW (s),$$

where the covariance matrix of $\left( B_1, \ldots, B_K, \int_0^1 \frac{\partial m(s, \theta)}{\partial \theta} \, dW (s) \right)'$ is given by

$$\begin{pmatrix} \Omega & \Gamma \\ \Gamma' & 1 \end{pmatrix} \int_0^1 \frac{\partial m(s, \theta)}{\partial \theta} \frac{\partial m(s, \theta)}{\partial \theta} \, ds$$

with $\begin{pmatrix} \Omega & \Gamma \\ \Gamma' & 1 \end{pmatrix}$ being the long run covariance matrix of $\left( p_t, \ldots, p_t^k, z_t \right)'$. Even with a correction like in the previous subsection, the overall effect is much more intricate than in the simple demeaning case, and no KV-type distribution can be recovered, not even when demeaning $\hat{p}_t^b$ additionally. The bottom line is that different deterministic components will lead to different distributions (with the exception of the small-$b$ case, where $\chi^2$ 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 (such as intercept and trend), the feasible 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 a valid method.

4 Monte Carlo study

In our Monte Carlo simulation study we compare the $T_K$ test to the procedure of Bai and Ng (2005). The newly proposed test is carried out by using either the first one ($\tilde{T}_1$), two ($\tilde{T}_2$), three

---

3Details on the test proposed by Bai and Ng (2005) can be found in the Appendix.
Table 1: Empirical size results.

\[ T = 50 \]

<table>
<thead>
<tr>
<th>( b )</th>
<th>( \mathcal{T}_1 )</th>
<th>( \mathcal{T}_2 )</th>
<th>( \mathcal{T}_3 )</th>
<th>( \mathcal{T}_4 )</th>
<th>( b )</th>
<th>( \mathcal{T}_1 )</th>
<th>( \mathcal{T}_2 )</th>
<th>( \mathcal{T}_3 )</th>
<th>( \mathcal{T}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.046</td>
<td>0.015</td>
<td>0.014</td>
<td>0.017</td>
<td>0.1</td>
<td>0.048</td>
<td>0.024</td>
<td>0.025</td>
<td>0.023</td>
</tr>
<tr>
<td>0.2</td>
<td>0.045</td>
<td>0.019</td>
<td>0.021</td>
<td>0.023</td>
<td>0.2</td>
<td>0.053</td>
<td>0.027</td>
<td>0.030</td>
<td>0.032</td>
</tr>
<tr>
<td>0.3</td>
<td>0.048</td>
<td>0.023</td>
<td>0.026</td>
<td>0.026</td>
<td>0.3</td>
<td>0.056</td>
<td>0.027</td>
<td>0.027</td>
<td>0.035</td>
</tr>
<tr>
<td>0.4</td>
<td>0.047</td>
<td>0.019</td>
<td>0.026</td>
<td>0.028</td>
<td>0.4</td>
<td>0.048</td>
<td>0.029</td>
<td>0.030</td>
<td>0.038</td>
</tr>
<tr>
<td>0.5</td>
<td>0.045</td>
<td>0.023</td>
<td>0.027</td>
<td>0.023</td>
<td>0.5</td>
<td>0.048</td>
<td>0.028</td>
<td>0.040</td>
<td>0.036</td>
</tr>
<tr>
<td>0.6</td>
<td>0.045</td>
<td>0.021</td>
<td>0.029</td>
<td>0.023</td>
<td>0.6</td>
<td>0.048</td>
<td>0.025</td>
<td>0.036</td>
<td>0.036</td>
</tr>
<tr>
<td>0.7</td>
<td>0.041</td>
<td>0.021</td>
<td>0.026</td>
<td>0.023</td>
<td>0.7</td>
<td>0.046</td>
<td>0.031</td>
<td>0.033</td>
<td>0.035</td>
</tr>
<tr>
<td>0.8</td>
<td>0.043</td>
<td>0.022</td>
<td>0.024</td>
<td>0.027</td>
<td>0.8</td>
<td>0.043</td>
<td>0.028</td>
<td>0.032</td>
<td>0.037</td>
</tr>
<tr>
<td>0.9</td>
<td>0.044</td>
<td>0.022</td>
<td>0.026</td>
<td>0.023</td>
<td>0.9</td>
<td>0.044</td>
<td>0.027</td>
<td>0.035</td>
<td>0.033</td>
</tr>
<tr>
<td>1</td>
<td>0.044</td>
<td>0.023</td>
<td>0.027</td>
<td>0.025</td>
<td>1</td>
<td>0.045</td>
<td>0.028</td>
<td>0.036</td>
<td>0.035</td>
</tr>
</tbody>
</table>

BN 0.097

<table>
<thead>
<tr>
<th>( b )</th>
<th>( \mathcal{T}_1 )</th>
<th>( \mathcal{T}_2 )</th>
<th>( \mathcal{T}_3 )</th>
<th>( \mathcal{T}_4 )</th>
<th>( b )</th>
<th>( \mathcal{T}_1 )</th>
<th>( \mathcal{T}_2 )</th>
<th>( \mathcal{T}_3 )</th>
<th>( \mathcal{T}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.054</td>
<td>0.018</td>
<td>0.027</td>
<td>0.024</td>
<td>0.1</td>
<td>0.063</td>
<td>0.032</td>
<td>0.042</td>
<td>0.036</td>
</tr>
<tr>
<td>0.2</td>
<td>0.051</td>
<td>0.022</td>
<td>0.031</td>
<td>0.019</td>
<td>0.2</td>
<td>0.064</td>
<td>0.030</td>
<td>0.046</td>
<td>0.044</td>
</tr>
<tr>
<td>0.3</td>
<td>0.053</td>
<td>0.021</td>
<td>0.035</td>
<td>0.030</td>
<td>0.3</td>
<td>0.059</td>
<td>0.035</td>
<td>0.047</td>
<td>0.040</td>
</tr>
<tr>
<td>0.4</td>
<td>0.058</td>
<td>0.021</td>
<td>0.031</td>
<td>0.024</td>
<td>0.4</td>
<td>0.060</td>
<td>0.033</td>
<td>0.047</td>
<td>0.042</td>
</tr>
<tr>
<td>0.5</td>
<td>0.057</td>
<td>0.020</td>
<td>0.035</td>
<td>0.029</td>
<td>0.5</td>
<td>0.065</td>
<td>0.034</td>
<td>0.050</td>
<td>0.040</td>
</tr>
<tr>
<td>0.6</td>
<td>0.054</td>
<td>0.022</td>
<td>0.031</td>
<td>0.031</td>
<td>0.6</td>
<td>0.061</td>
<td>0.033</td>
<td>0.050</td>
<td>0.042</td>
</tr>
<tr>
<td>0.7</td>
<td>0.052</td>
<td>0.018</td>
<td>0.027</td>
<td>0.023</td>
<td>0.7</td>
<td>0.059</td>
<td>0.033</td>
<td>0.045</td>
<td>0.038</td>
</tr>
<tr>
<td>0.8</td>
<td>0.056</td>
<td>0.019</td>
<td>0.030</td>
<td>0.031</td>
<td>0.8</td>
<td>0.059</td>
<td>0.032</td>
<td>0.048</td>
<td>0.041</td>
</tr>
<tr>
<td>0.9</td>
<td>0.055</td>
<td>0.020</td>
<td>0.026</td>
<td>0.027</td>
<td>0.9</td>
<td>0.056</td>
<td>0.032</td>
<td>0.049</td>
<td>0.039</td>
</tr>
<tr>
<td>1</td>
<td>0.056</td>
<td>0.020</td>
<td>0.029</td>
<td>0.028</td>
<td>1</td>
<td>0.057</td>
<td>0.032</td>
<td>0.052</td>
<td>0.044</td>
</tr>
</tbody>
</table>

BN 0.115

\( \left( \mathcal{T}_3 \right) \) or four moments \( \left( \mathcal{T}_4 \right) \). 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 given by

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

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

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

The expectations \( \vartheta_{k-1} = E(p_k^{k-1} \varphi(z_t)) \) are simulated for \( k = 1, 2, 3, 4 \) with 1,000,000 observations and 10,000 Monte Carlo replications.\(^4\)

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 a bit 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. 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 obviously due to its upward size-distortions. For \( c > 0.5 \), the raw moment-based tests perform much better even though the tests are undersized. These considerations apply for both non-normal distributions under study. For the mixture with a \( \chi^2(3) \)-distribution, \( \tilde{T}_1 \) and \( \tilde{T}_2 \) perform best, while the \( \tilde{T}_2 \)-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

\(^4\)The resulting simulated values are as follows: \( \vartheta = (0.2820948, 0.1410473, 0.0857805, 0.0581472) \). Clearly, the simulated value for \( k = 1 \) matches perfectly with its theoretical counterpart. We therefore expect that the higher-order terms are simulated reasonably well.

\(^5\)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), $T = 250$. 
Table 2: Normality testing for G7 industrial production growth rates at the 5% level.

<table>
<thead>
<tr>
<th></th>
<th>skew</th>
<th>kurt</th>
<th>$\tilde{T}_1$</th>
<th>$\tilde{T}_2$</th>
<th>$\tilde{T}_3$</th>
<th>$\tilde{T}_4$</th>
<th>BN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN</td>
<td>-0.183</td>
<td>3.484</td>
<td>2.311</td>
<td>5.408</td>
<td>8.744</td>
<td>23.946</td>
<td>4.722</td>
</tr>
<tr>
<td>FRA</td>
<td>-0.765</td>
<td>95.986</td>
<td>2.511</td>
<td>123.517</td>
<td>153.766</td>
<td>160.868</td>
<td>1.164</td>
</tr>
<tr>
<td>GER</td>
<td>0.022</td>
<td>9.990</td>
<td>0.706</td>
<td>31.967</td>
<td>36.719</td>
<td>37.300</td>
<td>3.274</td>
</tr>
<tr>
<td>ITA</td>
<td>0.255</td>
<td>11.173</td>
<td>3.310</td>
<td>11.707</td>
<td>12.362</td>
<td>46.081</td>
<td>3.209</td>
</tr>
<tr>
<td>JPN</td>
<td>-2.122</td>
<td>19.471</td>
<td>2.758</td>
<td>67.236</td>
<td>73.401</td>
<td>78.418</td>
<td>2.729</td>
</tr>
<tr>
<td>UK</td>
<td>0.048</td>
<td>13.369</td>
<td>2.548</td>
<td>12.708</td>
<td>12.761</td>
<td>15.120</td>
<td>4.025</td>
</tr>
<tr>
<td>US</td>
<td>0.263</td>
<td>9.046</td>
<td>3.687</td>
<td>11.387</td>
<td>24.200</td>
<td>68.659</td>
<td>6.211</td>
</tr>
</tbody>
</table>

| cv    | 5.016 | 8.872 | 13.200 | 18.258 | 5.991 |

It can be clearly seen that the newly proposed test statistics typically disagree with the Bai and Ng (2005) test, except for the US, where the BN test barely 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 ($\tilde{T}_1$) 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 ($\tilde{T}_4$) statistics rejects. Apparently, it is able to detect even a relatively small deviation in the kurtosis (3.484) from its theoretical value of three. From a cross-sectional perspective, it can be seen that $\tilde{T}_2$ and $\tilde{T}_4$ 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.

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 1965, Feb to 2014, Oct yielding $T = 597$ 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.

It can be clearly seen that the newly proposed test statistics typically disagree with the Bai and Ng (2005) test, except for the US, where the BN test barely 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 ($\tilde{T}_1$) 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 ($\tilde{T}_4$) statistics rejects. Apparently, it is able to detect even a relatively small deviation in the kurtosis (3.484) from its theoretical value of three. From a cross-sectional perspective, it can be seen that $\tilde{T}_2$ and $\tilde{T}_4$ 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.

---

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.
Figure 5: Monthly G7 industrial production growth rates from 1965:02-2014:11 (FRED database).
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 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 heteroscedasticity is 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

Proof of Lemma 1

Write first the Nadaraya-Watson estimator as
\[
\hat{\sigma}^2(s) = \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})^2 \kappa \left( \frac{s - t/T}{h} \right).
\]

Note that the numerator of the Nadaraya-Watson estimator is positive w.p.1 since the summands are positive with positive probability. Since the denominator easily shown to be positive, under standard conditions for the kernel \(\kappa\), the first condition follows.

For the second, recall that the uniform convergence rate of the Nadaraya-Watson estimator with bandwidth \(h\) is, in the one-regressor setup, given by \(\ln T / \sqrt{Th}\) under our bandwidth restrictions; see e.g. Hansen (2008, Theorem 2) for the case of mixing errors. With \(h = CT^{-\beta}\), we have \(\alpha = 1/2 - \beta/2\) such that, when \(\beta \in (1/6; 1/4)\), we have \(\alpha \in (3/8; 5/6)\).

To establish the third condition, note that the first-order derivative of \(\hat{\sigma}(s)\) is given by the sum of two components,
\[
\frac{1}{h} \hat{\sigma}^2_s(s) + \frac{1}{T} \sum_{t=1}^{T} \kappa \left( \frac{s - t/T}{h} \right)
\]
where \(\hat{\sigma}^2_s(s)\) is the Nadaraya-Watson estimator of \(\sigma^2(\cdot)\) based on the same bandwidth but the kernel \(\kappa'\) and
\[
-\frac{1}{h} \hat{\sigma}^2(s) + \frac{1}{T} \sum_{t=1}^{T} \kappa' \left( \frac{s - t/T}{h} \right),
\]
with both \(\kappa\) and \(\kappa'\) satisfying the regularity conditions for the Nadaraya-Watson estimator, both \(\hat{\sigma}^2_s\) and \(\hat{\sigma}^2\) converge uniformly in probability and are thus uniformly bounded in probability (in \(T\)) for all \(s\), and the result follows due to the boundedness of the ratio \(\frac{1}{T} \sum_{t=1}^{T} \kappa \left( \frac{s - t/T}{h} \right)\).
Proof of Lemma 2

Write with a Taylor expansion

\[ \hat{p}_t = \Phi \left( z_t + \frac{\bar{x} - \bar{x}}{\bar{\sigma}_t} - z_t \right) = p_t + \varphi (z_t) \left( \frac{\bar{x} - \bar{x}}{\bar{\sigma}_t} - z_t \right) + \varphi' (\xi_t) \left( \frac{\bar{x} - \bar{x}}{\bar{\sigma}_t} - z_t \right)^2 \]

where \( \xi_t \) lies between \( \frac{x_t - \mu}{\sigma_t} = z_t \) and \( \frac{x_t - \bar{x}}{\bar{\sigma}_t} = \frac{\sigma_t}{\bar{\sigma}_t} (z_t - \bar{z}) \); note that \( \varphi' (\cdot) \) is bounded on \( \mathbb{R} \). In other words,

\[ \hat{p}_t = p_t + z_t \varphi (z_t) \left( \frac{\sigma_t}{\bar{\sigma}_t} - 1 \right) - \varphi (z_t) \frac{\sigma_t}{\bar{\sigma}_t} \bar{z} + \varphi' (\xi_t) \left( \frac{\sigma_t}{\bar{\sigma}_t} (z_t - \bar{z}) - z_t \right)^2 . \]

Let us now examine the behavior of the partial sums of the three terms on the r.h.s.

Under the null, \( z_t \varphi (z_t) \) has zero expectation (it is an odd function of a symmetric random variable so it is symmetric itself) and has finite variance with \( \varphi (\cdot) \) being bounded. Then, split the sample in \( B \) disjoint blocks of length \( J \), \( B = T/J \), and exploit the assumed smoothness\(^7\) of \( \sigma \) and \( \bar{\sigma} \) to arrive at

\[
\begin{align*}
\sum_{t=1}^{[sT]} z_t \varphi (z_t) \left( \frac{\sigma_t}{\bar{\sigma}_t} - 1 \right) & = \sum_{b=1}^{[sB]} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) \left( \frac{\sigma_{J(b-1)} + O \left( \frac{J}{T} \right)}{\sigma_{J(b-1)} + O_p \left( \frac{JT^\beta}{T} \right)} - 1 \right) \\
& = \sum_{b=1}^{[sB]} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) \left( \frac{\sigma_{J(b-1)}}{\sigma_{J(b-1)}} - 1 \right) + R_T \left( s \right) .
\end{align*}
\]

Now, \( \sigma_t \) being away from zero implies that \( \left| \frac{\sigma_{J(b-1)} + AT}{\sigma_{J(b-1)} + BT} - \frac{\sigma_{J(b-1)}}{\sigma_{J(b-1)}} \right| \leq CQ_T \frac{JT^\beta}{T} \) when \( AT = O \left( \frac{J}{T} \right) \) and \( |B_T| \leq Q_T \frac{JT^\beta}{T} \). Hence,

\[
|R_T \left( s \right)| \leq CT^\beta Q_T \frac{JT^\beta}{T} \sum_{b=1}^{[sB]} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) = O_p \left( T^\beta J \right) ;
\]

note furthermore that

\[
\sum_{b=1}^{[sB]} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) \leq \sum_{b=1}^{B} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) ;
\]

such that the \( O_p \left( T^\beta J \right) \) order is uniform over \([0, 1]\). Thus,

\[
\begin{align*}
\sum_{t=1}^{[sT]} z_t \varphi (z_t) \left( \frac{\sigma_t}{\bar{\sigma}_t} - 1 \right) & = \sum_{b=1}^{[sB]} \left( \frac{\sigma_{J(b-1)}}{\sigma_{J(b-1)}} - 1 \right) \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) + O_p \left( T^\beta J \right)
\end{align*}
\]

\(^7\)When there are breaks at known time, one may choose block boundaries at the break dates such that the argument remains unchanged.
with
\[
\sum_{b=1}^{[sB]} \left( \frac{\sigma_{J(b-1)}}{\sigma_{J(b-1)}} - 1 \right) \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) \leq \sup_b \left| \frac{\sigma_{J(b-1)}}{\sigma_{J(b-1)}} - 1 \right| \sum_{b=1}^{[sB]} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right)
\]
where \( \sup_b \left| \frac{\sigma_{J(b-1)}}{\sigma_{J(b-1)}} - 1 \right| = O_p \left( T^{-\alpha} \right) \) due to Assumption 3, and
\[
\sum_{b=1}^{[sB]} \left| \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) \right| \leq \sum_{b=1}^{B} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) = O_p \left( B\sqrt{J} \right),
\]

since \( \frac{1}{\sqrt{J}} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) \) is easily shown to be uniformly \( L_2 \)-bounded in \( s \) and as such uniformly \( L_1 \)-bounded such that
\[
E \left( \sum_{b=1}^{B} \sum_{j=1}^{J} z_{J(b-1)+j} \varphi \left( z_{J(b-1)+j} \right) \right) \leq CB\sqrt{J}
\]
with the Markov’s inequality completing the argument. Summing up,
\[
\sum_{t=1}^{[sT]} z_t \varphi \left( z_t \right) = O_p \left( \max \left\{ JT^3, T^{-\alpha} B\sqrt{J} \right\} \right)
\]
uniformly in \( s \in [0,1] \); since one can pick \( J = T^\kappa \) for some \( 1/2 - \beta > \kappa > 1 - 2\alpha \), both \( JT^3 \) and \( T^{-\alpha} B\sqrt{J} \) are \( o \left( T^{1/2} \right) \), and the order turns out to actually be \( o_p \left( T^{1/2} \right) \) as required.

Moreover, \( \varphi' \left( \xi_t \right) \left( \frac{\sigma_t}{\sigma_t} - \bar{z} - z_t \right)^2 \) also stays negligible upon cumulating because
\[
\sum_{t=1}^{[sT]} \varphi' \left( \xi_t \right) \left( z_t \left( \frac{\sigma_t}{\sigma_t} - 1 \right) - \bar{z} \frac{\sigma_t}{\sigma_t} \right)^2 = \sum_{t=1}^{[sT]} \varphi' \left( \xi_t \right) z_t^2 \left( \frac{\sigma_t}{\sigma_t} - 1 \right)^2 + \sum_{t=1}^{[sT]} \varphi' \left( \xi_t \right) \bar{z}^2 \frac{\sigma_t^2}{\sigma_t} - 2 \sum_{t=1}^{[sT]} \varphi' \left( \xi_t \right) z_t \left( \frac{\sigma_t}{\sigma_t} - 1 \right) \bar{z} \frac{\sigma_t}{\sigma_t},
\]
for whose three terms on the r.h.s. we may write uniformly in \( s \)
\[
\sum_{t=1}^{[sT]} \varphi' \left( \xi_t \right) z_t^2 \left( \frac{\sigma_t}{\sigma_t} - 1 \right)^2 \leq \max_{1 \leq t \leq T} \left( \frac{\sigma_t}{\sigma_t} - 1 \right)^2 \max \varphi' \sum_{t=1}^{[sT]} z_t^2 = O_p \left( T^{1-2\alpha} \right)
\]
since \( \sum_{t=1}^{[sT]} z_t^2 / \sum_{t=1}^{T} z_t^2 \Rightarrow s \) and \( \bar{\sigma}_t \) is uniformly \( T^\alpha \)-consistent and bounded away from zero,
\[
\sum_{t=1}^{[sT]} \varphi' \left( \xi_t \right) \bar{z}^2 \frac{\sigma_t^2}{\sigma_t^2} \leq \bar{z}^2 \max_{1 \leq t \leq T} \frac{\sigma_t^2}{\sigma_t^2} \left[ sT \right] \max \varphi' = O_p \left( 1 \right)
\]
and
\[
\left| \sum_{t=1}^{[sT]} \varphi' \left( \xi_t \right) z_t \left( \frac{\sigma_t}{\sigma_t} - 1 \right) \bar{z} \frac{\sigma_t}{\sigma_t} \right| \leq \bar{z} \sup_t \left| \frac{\sigma_t}{\sigma_t} \right| \sup_t \left| \frac{\sigma_t}{\sigma_t} - 1 \right| \max \varphi' \sum_{t=1}^{[sT]} \left| z_t \right| = O_p \left( T^{1/2-\alpha} \right).
\]
This way, when studying e.g. the partial sums of \( \hat{p}_t - \frac{1}{2} \), we have

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

\[
= \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \left( p_t - \frac{1}{2} \right) - sE(\varphi(z_t)) \sqrt{T} \varepsilon
\]

\[-\sqrt{T} \varepsilon \sum_{t=1}^{[sT]} \varphi(z_t) \left( \frac{\sigma_t}{\sigma_t} - 1 \right) + o_p(1),
\]

and the result follows for \( k = 1 \) by noting that \( \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \varphi(z_t) \Rightarrow sE(\varphi_t) \) and recalling that the \( o_p(1) \) term is uniform in \( s \). Note also that \( E(\varphi_t) = \int_{-\infty}^\infty \varphi^2(x) dx = \frac{1}{2\sqrt{\pi}} \) is positive and

\[
\left| \frac{1}{T} \sum_{t=1}^{[sT]} \varphi(z_t) \left( \frac{\sigma_t}{\sigma_t} - 1 \right) \right| \leq \sup_t \left| \frac{\sigma_t}{\sigma_t} - 1 \right| \left| \frac{1}{T} \sum_{t=1}^{[sT]} \varphi(z_t) \right| \leq CsT^{-\alpha} = o_p(1)
\]

uniformly in \( s \in [0,1] \).

For the higher-order moments we have along the same lines

\[
\sum_{t=1}^{[sT]} \hat{p}_t^k = \sum_{t=1}^{[sT]} p_t^k - k \varepsilon \sum_{t=1}^{[sT]} p_t^{k-1} \varphi(z_t) + o_p(T^{1/2})
\]

with the \( o_p(T^{1/2}) \) term being uniform in \( s \). Hence, for all \( 1 \leq k \leq K \),

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

with the \( o_p \) term being uniform for \( s \in [0,1] \). A multivariate functional central limit theorem for mixing processes then completes the result (see e.g. Davidson, 1994, Chapter 29).

**Proof of Proposition 1**

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

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

uniformly in \( s \) thanks to the arguments used in the proof of Lemma 2. Then,

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

\[
\Rightarrow W(s) - sW(1).
\]
Let

\[ \mathcal{B}(t) = (B_1(t), \ldots, B_K(t), W(t))' \]

and

\[ \mathcal{B} = (B_1(t) - s \vartheta_0 W(t), \ldots, B_K(t) - s K \vartheta_{K-1} W(t), W(t) - sW(t))' ; \]

using the arguments of the proof of Theorem 2 in Kiefer and Vogelsang (2005) together with the Lemma 2, we obtain

\[ \hat{T}_K \Rightarrow (V \mathcal{B})'(1) \left( V - \int_0^1 \int_0^1 \frac{1}{b^2} \kappa'' \left( \frac{r - s}{b} \right) (\mathcal{B}(r) - r \mathcal{B}(1)) (\mathcal{B}(s) - s \mathcal{B}(1))' drds \right) V' \left( \frac{1}{V} \right)^{-1} \mathcal{B}(1). \]

Note further that

\[ V(\hat{\mathcal{B}}(s) - s \hat{\mathcal{B}}(1)) = V(\hat{\mathcal{B}}(s) - s \hat{\mathcal{B}}(1)), \]

and let \( Y = V \hat{\mathcal{B}} \) such that

\[ \hat{T}_K \Rightarrow Y'(1) \left( - \int_0^1 \int_0^1 \frac{1}{b^2} \kappa'' \left( \frac{r - s}{b} \right) (Y(r) - rY(1)) (Y(s) - sY(1))' drds \right) Y(1) \]

where \( Y \) is a multivariate Brownian motion; since its long-run covariance matrix cancels out, the r.h.s. is the required \( K \mathcal{N}_{K,b,\kappa} \) distribution.

**Proof of Lemma 3**

Begin by discussing the limiting behavior of the NLS estimators \( \hat{\theta} \). We have under Assumptions 1 and 4 that

\[ \sqrt{T} (\hat{\theta} - \theta) \Rightarrow \left( \int_0^1 \frac{\partial m(s, \theta) \partial m(s, \theta)'}{\partial \theta} ds \right)^{-1} \int_0^1 \frac{\partial m(s, \theta)}{\partial \theta} dW(s) ; \]

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

Begin with \( k = 1 \); with the application of the mean value theorem (or Taylor series expansion with rest term in differential form) we obtain

\[ \hat{p}_t = p_t + \varphi(z_t) \left( m \left( \frac{t}{T}, \theta \right) - m \left( \frac{t}{T}, \hat{\theta} \right) \right) + \varphi'(\xi_t) \left( m \left( \frac{t}{T}, \theta \right) - m \left( \frac{t}{T}, \hat{\theta} \right) \right)^2 \]

where \( \xi_t \) lies between \( z_t \) and \( z_t - m \left( \frac{t}{T}, \hat{\theta} \right) + m \left( \frac{t}{T}, \theta \right) \) for each \( t \). The exact values for \( \xi_t \) do not matter since \( \varphi' \) is bounded. A second expansion, here about \( \theta \), is required for the trend function \( m \):

\[ m \left( \frac{t}{T}, \theta \right) - m \left( \frac{t}{T}, \hat{\theta} \right) = - \frac{\partial m \left( \frac{t}{T}, \theta \right)'}{\partial \theta} (\hat{\theta} - \theta) - (\hat{\theta} - \theta)' \frac{\partial^2 m \left( \frac{t}{T}, \theta \right)}{\partial \theta \partial \theta'} \bigg|_{\theta = \hat{\theta}_t} (\hat{\theta} - \theta) \]

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

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} (\hat{p}_t - \frac{1}{2}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} (p_t - \frac{1}{2}) - \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \varphi (z_t) \frac{\partial m (t/T, \theta)}{\partial \theta} \right)' (\hat{\theta} - \theta) \\
- (\hat{\theta} - \theta)' \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \varphi (z_t) \frac{\partial^2 m (t/T, \theta)}{\partial \theta \partial \theta'} \bigg|_{\theta = \hat{\theta}} \right) (\hat{\theta} - \theta) + R_T
\]

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

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

For the fourth summand, \( R_T \), we have with a first-order Taylor expansion, \( m (t/T, \theta) - m (t/T, \hat{\theta}) = \frac{\partial m(t/T, \theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} (\hat{\theta} - \theta) \) with \( \varphi_t \) between \( \theta \) and \( \hat{\theta} \) for each \( t \), that

\[
R_T = \left( \hat{\theta} - \theta \right)' \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \varphi' (\xi_t) \frac{\partial m (t/T, \theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} \right) \left( \hat{\theta} - \theta \right).
\]

Similarly, \( \varphi' \) is bounded and \( \left| \frac{\partial m(t/T, \theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} \right| \leq f (\| \hat{\theta} \|) \leq f \left( \max \left\{ \| \theta \| : \| \hat{\theta} \| \right\} \right) \) for all \( t \), it follows that \( \sup_s R_T = O_p (T^{-1/2}) \).

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

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

the same arguments show that analogous relations hold for \( \hat{p}_t^k \). With \( \sqrt{T} (\hat{\theta} - \theta) = O (1) \) and \( \frac{1}{T} \sum_{t=1}^{[sT]} \hat{p}_t^{k-1} \varphi (z_t) \frac{\partial m(t/T, \theta)}{\partial \theta} = E \left( \hat{p}_t^{k-1} \varphi (z_t) \right) \int_0^s \frac{\partial m(r, \theta)}{\partial \theta} dr = \hat{\vartheta}_{k-1} \mu (s, \theta) \), the desired result follows.
The Bai and Ng (2005) test procedure

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

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

where

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

and

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

$\bar{y}$, $\hat{\sigma}$ and $\hat{\Phi}$ are consistent estimators. The theoretical long-run covariance matrix $\Phi$ is given by $\Phi = \lim_{T \to \infty} T E(\bar{Z} \bar{Z}')$ with $Z' = \left[ y_t - \mu, (y_t - \mu)^2 - \sigma^2, (y_t - \mu)^3, (y_t - \mu)^4 - 3\sigma^4 \right]$ and $\bar{Z}$ being the sample mean of $Z_t$. The limiting distribution of $\mu_{34}$ 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.
Critical values

Table 3: Critical values via response curves from the $K V_{K,b,κ}$-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.

<table>
<thead>
<tr>
<th>$K = 1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>$a_0$</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>$a_3$</td>
</tr>
<tr>
<td>0.9</td>
<td>2.7055</td>
<td>6.1598</td>
<td>8.6142</td>
<td>-3.3854</td>
</tr>
<tr>
<td>0.95</td>
<td>3.8415</td>
<td>10.2574</td>
<td>15.6231</td>
<td>-7.0320</td>
</tr>
<tr>
<td>0.975</td>
<td>5.0239</td>
<td>15.8489</td>
<td>24.5892</td>
<td>-12.5751</td>
</tr>
<tr>
<td>0.99</td>
<td>6.6349</td>
<td>26.3361</td>
<td>36.1330</td>
<td>-19.6341</td>
</tr>
<tr>
<td>0.995</td>
<td>7.8794</td>
<td>37.5823</td>
<td>41.2076</td>
<td>-21.6338</td>
</tr>
<tr>
<td>$K = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>$a_0$</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>$a_3$</td>
</tr>
<tr>
<td>0.9</td>
<td>4.6052</td>
<td>15.5300</td>
<td>33.0455</td>
<td>-18.0050</td>
</tr>
<tr>
<td>0.95</td>
<td>5.9915</td>
<td>24.2350</td>
<td>48.4528</td>
<td>-27.7431</td>
</tr>
<tr>
<td>0.975</td>
<td>7.3778</td>
<td>35.6889</td>
<td>62.8696</td>
<td>-36.8917</td>
</tr>
<tr>
<td>0.99</td>
<td>9.2103</td>
<td>53.2832</td>
<td>88.7896</td>
<td>-55.9722</td>
</tr>
<tr>
<td>0.995</td>
<td>10.5966</td>
<td>71.9545</td>
<td>96.5536</td>
<td>-60.2045</td>
</tr>
<tr>
<td>$K = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>$a_0$</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>$a_3$</td>
</tr>
<tr>
<td>0.9</td>
<td>6.2514</td>
<td>30.2793</td>
<td>67.5629</td>
<td>-42.2680</td>
</tr>
<tr>
<td>0.95</td>
<td>7.8147</td>
<td>45.5956</td>
<td>88.1783</td>
<td>-56.1070</td>
</tr>
<tr>
<td>0.975</td>
<td>9.3484</td>
<td>63.5918</td>
<td>109.2760</td>
<td>-70.7583</td>
</tr>
<tr>
<td>0.99</td>
<td>11.3449</td>
<td>94.2752</td>
<td>127.9765</td>
<td>-84.0108</td>
</tr>
<tr>
<td>0.995</td>
<td>12.8382</td>
<td>121.7357</td>
<td>137.7951</td>
<td>-91.2883</td>
</tr>
<tr>
<td>$K = 4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>$a_0$</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>$a_3$</td>
</tr>
<tr>
<td>0.9</td>
<td>7.7794</td>
<td>54.1072</td>
<td>94.7069</td>
<td>-61.0147</td>
</tr>
<tr>
<td>0.95</td>
<td>9.4877</td>
<td>76.3485</td>
<td>121.5104</td>
<td>-79.8180</td>
</tr>
<tr>
<td>0.975</td>
<td>11.1433</td>
<td>102.1803</td>
<td>145.6040</td>
<td>-97.0618</td>
</tr>
<tr>
<td>0.99</td>
<td>13.2767</td>
<td>142.5323</td>
<td>169.0490</td>
<td>-113.2457</td>
</tr>
<tr>
<td>0.995</td>
<td>14.8603</td>
<td>177.5045</td>
<td>183.2276</td>
<td>-123.6561</td>
</tr>
</tbody>
</table>
References


