Nonparametric volatility change detection

Maria Mohr and Natalie Neumeyer
Department of Mathematics, University of Hamburg

June 10, 2019

Abstract

We consider a nonparametric heteroscedastic time series regression model and suggest testing procedures to detect changes in the conditional variance function. The tests are based on a sequential marked empirical process and thus combine classical CUSUM tests with marked empirical process approaches known from goodness-of-fit testing. The tests are consistent against general alternatives of a change in the conditional variance function, a feature that classical CUSUM tests are lacking. We derive a simple limiting distribution and in the case of univariate covariates even obtain asymptotically distribution-free tests. We demonstrate the good performance of the tests in a simulation study and consider exchange rate data as a real data application.

Key words: change point, conditional variance function, CUSUM, heteroscedasticity, kernel estimation, Kolmogorov-Smirnov test, marked empirical process, structural change
AMS 2010 Classification: Primary 62M10, Secondary 62G08, 62G10

1 Introduction

The paper is concerned with the investigation of structural stability of the conditional variance function (volatility function) in nonparametric heteroscedastic time series regression models. Those models have gained much attention over the last decades and contain as special cases nonparametric AR-ARCH models, which are also called nonparametric CHARN (conditional heteroscedastic autoregressive nonlinear) models; see Fan and Yao (2003) or Gao (2007) for overviews. They have been successfully applied to model econometric time series such as foreign exchange rates or stock market indices, see e.g. Yang et al. (1999) and Zhao and Wu (2008). Here tests for structural changes in the volatility function are of special importance.

A lot of research has been devoted to the parametric case, notably for ARCH and GARCH models. Among others, Kokoszka and Leipus (1999) suggested a CUSUM type test for parameter stability in ARCH models, while Kulperger and Yu (2005) considered partial sums of higher powers of residuals to test for a parameter change in GARCH models. Berkes et al. (2004) considered tests for parameter stability in GARCH models based on likelihood ratios. Kengne's (2012) test, which is based on quasi likelihood estimators, is applicable to more general parametric causal time series models. Lee and Lee (2014) suggested a residual based CUSUM test for change points in parametric AR-GARCH models, while Lee and Song (2008) and Song and Kang (2018) considered ARMA-GARCH models. Very few results are available in the nonparametric framework. Chen et al. (2005) studied a nonparametric heteroscedastic time series model with a scale change in volatility. However, they assume a compact support of regressors, which is problematic when considering autoregression models. Tests for change points in the unconditional variance in time series models have been considered as well. Lee et al. (2003) considered parametric autoregression models, as well as fixed design nonparametric regression models with strongly mixing errors using a CUSUM testing procedure. Chen and Tian (2014) constructed a ratio test for change point detection in the variance in random design nonparametric regression models. However, their test does not allow for autoregressive effects, as a compact support of regressors is assumed. A related strand of the literature deals with change point detection in the error distribution of a time series regression model. In the parametric framework Koul (1996) considered non-linear regression models and Ling (1998) non-stationary AR models, to just mention a few, while Selk and Neumeyer (2013) considered nonparametric heteroscedastic autoregression models.

Recently, Mohr and Neumeyer (2019) suggested a test for change point in the regression function in nonparametric time series models. They combine traditional CUSUM tests as considered by Hidalgo (1995), Honda (1997) and Su and Xiao (2008) in the nonparametric context with the marked empirical process approach originally suggested by Stute (1997) and widely used in the goodness-of-fit literature. Compared with the CUSUM approach the new test shows better power properties, in theory as well as in finite sample simulations. In the paper at hand we will modify the CUSUM marked empirical process test in order to test for a change point in the conditional volatility function. We obtain tests with very simple limiting distributions, which are consistent against general fixed alternatives. In the case of univariate covariates one can even obtain tests that are asymptotically distribution-free.

The paper is organized as follows. In section 2 we define the process on which the test statistics are built. In section 3 we give the limiting distribution of the process under the null hypothesis of no change in the variance function. We further discuss consistency against fixed alternatives of one change point. In section 4 we describe a simulation study and discuss a real data example of currency exchange rates. Section 5 concludes the paper, whereas in the appendix we list the regularity assumptions and prove the asymptotic results.

2 The model and test statistic

Consider a strictly stationary and strongly mixing time series $(Y_t, \boldsymbol{X}_t)$, $t \in \mathbb{Z}$, following the nonparametric model

$$Y_t = m(\boldsymbol{X}_t) + U_t, \tag{2.1}$$

where $E[U_t|\mathcal{F}^t] = 0$ a.s. for the sigma-field $\mathcal{F}^t = \sigma(U_{j-1}, \boldsymbol{X}_j : j \le t)$, and $m : \mathbb{R}^d \to \mathbb{R}$ does not depend on t. Further, let the following representation for the innovations U_t hold,

$$U_t = \sigma_t(\boldsymbol{X}_t)\varepsilon_t,\ t \in \mathbb{Z}, \tag{2.2}$$

for some functions $\sigma_t : \mathbb{R}^d \to \mathbb{R}$ and an i.i.d. sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$, such that ε_t is independent of $\boldsymbol{X}_j$ for all $j \le t$ and fulfills $E[\varepsilon_1] = 0$, $E[\varepsilon_1^2] = 1$ and $E[\varepsilon_1^4] < \infty$. With these restrictions, σ_t^2 is the variance function of Y_t, conditioned on $\boldsymbol{X}_t$, as

$$\mathrm{Var}(Y_t|\boldsymbol{X}_t) = E[U_t^2|\boldsymbol{X}_t] = \sigma_t^2(\boldsymbol{X}_t) \text{ a.s.}$$

The d-dimensional absolutely continuous covariate $\boldsymbol{X}_t$ may include finitely many lagged values of Y_t, for instance $\boldsymbol{X}_t = (Y_{t-1}, \ldots, Y_{t-d})^T$, such that the model includes nonparametric AR-ARCH models.

Our aim is to test whether the function $\sigma_t^2(\cdot)$ is stable in time t. Given observations $(Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)$ the null hypothesis

$$H_0 : \sigma_t^2(\cdot) = \sigma^2(\cdot),\ t = 1, \ldots, n,$$

for some not further specified function $\sigma^2 : \mathbb{R}^d \to \mathbb{R}$ (not depending on time t) will be considered.

The idea is to base tests for H_0 on a sequential marked empirical process of residuals,

$$\hat{T}_n(s, \boldsymbol{z}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \left((Y_t - \hat{m}_n(\boldsymbol{X}_t))^2 - \hat{\sigma}_n^2(\boldsymbol{X}_t) \right) \omega_n(\boldsymbol{X}_t) I\{\boldsymbol{X}_t \le \boldsymbol{z}\} \tag{2.3}$$

indexed in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. Throughout $I\{\ldots\}$ denotes an indicator function. Further $\omega_n(\cdot) = I\{\cdot \in \boldsymbol{J}_n\}$ is a weight function with $\boldsymbol{J}_n$ specified in assumption **(J)** in appendix A. The regression and volatility functions are estimated as

$$\hat{m}_n(\boldsymbol{x}) = \frac{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right) Y_j}{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right)}$$

and

$$\hat{\sigma}_n^2(\boldsymbol{x}) = \frac{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right) (Y_j - \hat{m}_n(\boldsymbol{x}))^2}{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right)},$$

respectively, with kernel function K and bandwidth h_n as considered in the assumptions in appendix A. The null hypothesis H_0 of no change in the variance will be rejected for large values of, e.g., a Kolmogorov-Smirnov type test statistic

$$T_{n1} := \sup_{\boldsymbol{z}\in\mathbb{R}^d} \sup_{s\in[0,1]} \left|\hat{T}_n(s,\boldsymbol{z})\right|$$

due to the following motivation. Note that the volatility function σ_t^2 from (2.2) can be viewed as regression function in a regression model

$$U_t^2 = \sigma_t^2(\boldsymbol{X}_t) + \xi_t, \quad t \in \mathbb{Z},$$

with covariate $\boldsymbol{X}_t$, response variable U_t^2 and innovations $\xi_t = U_t^2 - \sigma_t^2(\boldsymbol{X}_t)$, that satisfy $E[\xi_t|\boldsymbol{X}_t] = 0$ and $E[\xi_t^2|\boldsymbol{X}_t] = \sigma_t^4(\boldsymbol{X}_t)E[(\varepsilon_t^2-1)^2]$ a.s. However, this is not a feasible model as $U_t = Y_t - m(\boldsymbol{X}_t)$ is unobservable and has to be estimated. The term

$$(Y_t - \hat{m}_n(\boldsymbol{X}_t))^2 - \hat{\sigma}_n^2(\boldsymbol{X}_t) =: \hat{\xi}_t$$

in the definition of the process $\hat{T}_n$ can be seen as estimator for the innovation ξ_t in the 'non-feasible' model above under the null hypothesis $\sigma_t^2(\cdot) = \sigma^2(\cdot)\forall t$. Thus $n^{-1/2}\hat{T}_n$ will vanish for $n \to \infty$ under the null hypothesis. The limiting process of $\hat{T}_n$ will be given in Corollary 3.2 below. From this result critical values for a test based on the Kolmogorov-Smirnov type test statistic T_{n1} can be approximated. The behavior of T_{n1} under fixed alternatives will be demonstrated in Remark 3.3 in order to motivate consistency of the test. The process $\hat{T}_n$ is a consistent improvement of CUSUM tests analogous to the procedure in Mohr and Neumeyer (2019) developed for changes in the regression function.

Remark 2.1. *In model (2.1) we assume a regression function m that is stable in time t. For testing of a change in the variance function this assumption makes sense if beforehand one can test for a change in the regression function applying a testing procedure which only reacts sensitive to changes in the regression function, not to changes in the variance function. Mohr and Neumeyer (2019) provide such a bootstrap test, which can be applied in cases of unstable variances, but as desired only reacts sensitive to changes in the regression function. Consecutively applying the bootstrap test in Mohr and Neumeyer (2019) and, if it does not reject, the test in the paper at hand, gives the knowledge of whether a change occurs in the mean or the variance function.*

3 Asymptotic results

Under the regularity assumptions in appendix A one can derive the following decomposition of the process $\hat{T}_n$ defined in (2.3) in terms of the process

$$T_n(s,\boldsymbol{z}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \xi_t I\{\boldsymbol{X}_t \leq \boldsymbol{z}\}, \quad s \in [0,1], z \in \mathbb{R}^d, \tag{3.1}$$

as well as the weak convergence of T_n.

Theorem 3.1. *Assume model (2.1), (2.2) under the null hypothesis H_0 and assumptions **(G)**, **(ξ)**, **(M)**, **(J)**, **(F1)**, **(F2)**, **(K)**, **(B1)** and **(B2)** from appendix A.*

(i) *Then, $\hat{T}_n(s,\boldsymbol{z}) = T_n(s,\boldsymbol{z}) - sT_n(1,\boldsymbol{z}) + o_P(1)$ uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$.*

(ii) *The process $T_n = \{T_n(s,\boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\}$ converges weakly in $\ell^\infty([0,1]\times\mathbb{R}^d)$ to a centered Gaussian process G with*

$$\mathrm{Cov}\left(G(s_1,\boldsymbol{z}_1), G(s_2,\boldsymbol{z}_2)\right) = (s_1 \wedge s_2)\Sigma(\boldsymbol{z}_1 \wedge \boldsymbol{z}_2),$$

where $\Sigma(\boldsymbol{z}) := E[(\varepsilon_1^2 - 1)^2]\int_{(-\boldsymbol{\infty},\boldsymbol{z}]} \sigma^4(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$.

Here and throughout we define $(-\boldsymbol{\infty},\boldsymbol{z}] = (-\infty, z_1]\times\cdots\times(-\infty,z_d]$ for $\boldsymbol{z} = (z_1,\ldots,z_d) \in \mathbb{R}^d$. The proof of Theorem 3.1 is given in appendix B. An application of the continuous mapping theorem and Slutsky's lemma give the following weak convergence result for the process $\hat{T}_n$.

Corollary 3.2. *Suppose that the assumptions of Theorem 3.1 and H_0 are satisfied. Then the process $\hat{T}_n$ converges weakly in $\ell^\infty([0,1]\times\mathbb{R}^d)$ to a centered Gaussian process G_0 with*

$$\mathrm{Cov}\left(G_0(s_1,\boldsymbol{z}_1), G_0(s_2,\boldsymbol{z}_2)\right) = (s_1 \wedge s_2 - s_1 s_2)\Sigma(\boldsymbol{z}_1 \wedge \boldsymbol{z}_2).$$

The continuous mapping theorem then implies convergence in distribution of the Kolmogorov-Smirnov test statistic,

$$T_{n1} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{\boldsymbol{z}\in\mathbb{R}^d} \sup_{s\in[0,1]} |G_0(s,\boldsymbol{z})|\,.$$

In particular in the case $d = 1$ using continuity of Σ and the scaling property of the Brownian motion, it holds that T_{n1} converges in distribution to $c^{1/2}T$, where

$$T = \sup_{s\in[0,1]} \sup_{t\in[0,1]} |K_0(s,t)|$$

and K_0 is a Kiefer-Müller process. The constant $c = E[((Y_1 - m(X_1))^2 - \sigma^2(X_1))^2]$ can be consistently estimated as

$$\hat{c}_n := \frac{1}{n}\sum_{i=1}^n \left((Y_i - \hat{m}_n(X_i))^2 - \hat{\sigma}_n^2(X_i)\right)^2 \omega_n(X_i),$$

and the test statistic $T_{n1}/\hat{c}_n^{1/2}$ is asymptotically distribution-free. We reject H_0 at asymptotic level α if $T_{n1}/\hat{c}_n^{1/2}$ is larger than the (known) $(1-\alpha)$-quantile of T.

Remark 3.3. *To see that the test is consistent against simple fixed alternatives of one change in the volatility function,*

$$H_1 : \exists s_0 \in (0,1) : \sigma^2_{n,t}(\cdot) = \begin{cases} \sigma^2_{(1)}(\cdot), & t = 1,\ldots,\lfloor ns_0\rfloor \\ \sigma^2_{(2)}(\cdot), & t = \lfloor ns_0\rfloor + 1,\ldots,n, \end{cases}$$

for some functions with $\sigma^2_{(1)} \not\equiv \sigma^2_{(2)}$, consider a triangular array

$$Y_{n,t} = m(\boldsymbol{X}_{n,t}) + U_{n,t},\ \ t = 1, \dots, n,$$

with regression function m stable in time and innovations such that $E[U_{n,t}|\mathcal{F}^t_n] = 0$ and $E[U^2_{n,t}|\boldsymbol{X}_{n,t}] = \sigma^2_{n,t}(\boldsymbol{X}_{n,t})$ a.s. Further assume that the covariate $\boldsymbol{X}_{n,t}$ is absolutely continuous with density function $f_{n,t}$. Then $\hat{\sigma}^2_n(\boldsymbol{x})$ will estimate the function

$$\bar{\sigma}^2_n(\boldsymbol{x}) = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})\sigma^2_{n,i}(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})} = \left(\sigma^2_{(1)}(\boldsymbol{x}) - \sigma^2_{(2)}(\boldsymbol{x})\right) \frac{\frac{1}{n}\sum\limits_{i=1}^{\lfloor ns_0 \rfloor} f_{n,i}(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})} + \sigma^2_{(2)}(\boldsymbol{x}).$$

Now assume that for each $s \in (0,1)$, the limit of $n^{-1}\sum_{i=1}^{\lfloor ns \rfloor} f_{n,i}$ exists and denote it by $\bar{f}^{(s)}$. Then $n^{-1/2}\hat{T}_n(s_0, \boldsymbol{z})$ will converge in probability to the integral

$$\int\limits_{(-\boldsymbol{\infty}, \boldsymbol{z}]} \left(\sigma^2_{(1)}(\boldsymbol{u}) - \sigma^2_{(2)}(\boldsymbol{u})\right) \bar{f}^{(s_0)}(\boldsymbol{u}) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) d\boldsymbol{u},$$

which, under H_1, does not vanish for at least one $\boldsymbol{z} = \boldsymbol{z}_0$ (provided that $\bar{f}^{(s_0)} \neq \bar{f}^{(1)}$). As $T_{n1} \geq |\hat{T}_n(s_0, \boldsymbol{z}_0)|$, the test statistic will converge to infinity in probability and the test is consistent.

Remark 3.4. *A traditional CUSUM test statistic in our context would be defined as $\sup_{s\in[0,1]} |\hat{T}_n(s,\infty)|$. With the same reasoning as in Remark 3.3, $n^{-1/2}\hat{T}_n(s_0,\infty)$ will converge in probability to*

$$\int \left(\sigma^2_{(1)}(\boldsymbol{u}) - \sigma^2_{(2)}(\boldsymbol{u})\right) \bar{f}^{(s_0)}(\boldsymbol{u}) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) d\boldsymbol{u},$$

which could be zero, even under the alternative H_1. In such a case the CUSUM test is not consistent.

4 Finite sample properties

4.1 Simulations

A Monte Carlo study is conducted in order to compare the results for T_{n1} from section 2 and a Cramér-von Mises type test $T_{n2} := \sup_{z\in\mathbb{R}} \int_0^1 |\hat{T}_n(s,z)|^2 ds$ with those of the traditional CUSUM versions denoted by $KS := \sup_{s\in[0,1]} |\hat{T}_n(s,\infty)|$ and $CM := \int |\hat{T}_n(s,\infty)|^2 ds$. All simulations are carried out with a level of 5%, 1000 replications and for sample sizes $n \in \{100, 300, 500\}$. For the nonparametric estimators $\hat{m}_n$ and $\hat{\sigma}^2_n$ we use an Epanechnikov kernel K and $h_n = n^{-1/3}$ as a simple ad hoc bandwidth. Furthermore, we set $c_n = \log(n)$ for the weighting function. The data is simulated from the following models.

$$\text{(model 1)} \qquad Y_t = m(X_t) + \sigma_t(X_t)\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1),$$
$$\sigma_t(x) = \begin{cases} 0.5\exp(-0.2x), & t = 1, \ldots, \lfloor ns_0 \rfloor \\ 0.5\exp(0.2x), & t = \lfloor ns_0 \rfloor + 1, \ldots, n, \end{cases}$$

where X_t is an exogenous variable following the AR(1) model $X_t = 0.4X_{t-1} + \xi_t$ with ξ_t being i.i.d. $\sim \mathcal{N}(0,1)$.

$$\text{(model 2)} \qquad Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1),$$
$$\sigma_t(x) = \begin{cases} \sqrt{0.1 + 0.1x^2}, & t = 1, \ldots, \lfloor ns_0 \rfloor \\ \sqrt{0.1 + 0.7x^2}, & t = \lfloor ns_0 \rfloor + 1, \ldots, n. \end{cases}$$

For both model 1 and 2 we consider $s_0 \in \{0, 0.25, 0.5, 0.75, 1\}$ and two different choices for the regression function, namely $m(x) = 0.5x$ (case (a)) and $m(x) = -0.5x$ (case (b)).

Model 1 is a heteroscedastic regression model with autoregressive covariables while model 2 is a heteroscedastic autoregression (AR-ARCH) model. In both cases H_0 is satisfied for $s_0 \in \{0, 1\}$ and H_1 is satisfied for $s_0 \in \{0.25, 0.5, 0.75\}$. Further, note that data generated from both models fulfill the stationarity and mixing assumptions when $s_0 \in \{0, 1\}$ (see Remark A.1 in appendix A).

Table 1 shows the rejection frequencies for model 1. To summarize the performance of the tests it is to mention that all level simulations ($s_0 \in \{0, 1\}$) show reasonably good results. The tests based on T_{n1} and T_{n2} show nice consistency properties ($s_0 \in \{0.25, 0.5, 0.75\}$), rejecting the null more frequently with increasing sample sizes, where T_{n2} has larger power. The classical CUSUM tests, however, clearly fail in detecting the change, having a power that does not exceed 10% for all cases (see Remark 3.4). All of the tests perform rather poorly when the sample size is small, i.e. for $n = 100$. Furthermore, we note that changes occurring at $s_0 = 0.5$ are easiest to detect.

The corresponding results in model 2 can be found in table 2. The level of 5% is approximately hold for all tests, even in the case where the variance has a relatively large influence ($s_0 = 0$). The power simulations suggest that our tests as well as the classical CUSUM tests result in reasonable rejection probabilities, detecting the change more often for increasing sample sizes. Again changes in $s_0 = 0.5$ are easiest to detect.

4.2 Data example

In this section we will apply our test to a financial data set that is concerned with exchange rates of currencies. Exchange rate regimes indicate how a country manages its currency with respect to other currencies, it can vary from "fixed", over "pegged" to "floating". In the case of a fixed regime, the currency is more or less fixed to some other currency.

Table 1: Rejection frequencies in model 1

		model 1 (a)				model 1 (b)			
s_0	n	T_{n1}	T_{n2}	KS	CM	T_{n1}	T_{n2}	KS	CM
0	100	0.035	0.058	0.046	0.042	0.052	0.064	0.059	0.047
	300	0.053	0.073	0.058	0.048	0.056	0.068	0.064	0.055
	500	0.057	0.071	0.062	0.055	0.062	0.064	0.059	0.043
0.25	100	0.041	0.080	0.052	0.048	0.053	0.081	0.060	0.054
	300	0.112	0.157	0.072	0.062	0.099	0.155	0.072	0.051
	500	0.187	0.266	0.069	0.050	0.216	0.294	0.097	0.080
0.50	100	0.063	0.122	0.053	0.057	0.068	0.120	0.073	0.066
	300	0.210	0.276	0.091	0.073	0.199	0.279	0.081	0.068
	500	0.413	0.521	0.097	0.067	0.428	0.510	0.096	0.074
0.75	100	0.055	0.092	0.074	0.067	0.051	0.084	0.061	0.066
	300	0.130	0.174	0.079	0.053	0.120	0.196	0.080	0.068
	500	0.222	0.291	0.086	0.074	0.239	0.304	0.096	0.074
1	100	0.045	0.072	0.062	0.057	0.046	0.076	0.055	0.055
	300	0.053	0.064	0.056	0.041	0.076	0.088	0.081	0.064
	500	0.063	0.071	0.072	0.050	0.064	0.070	0.064	0.051

Table 2: Rejection frequencies in model 2

		model 2 (a)				model 2 (b)			
s_0	n	T_{n1}	T_{n2}	KS	CM	T_{n1}	T_{n2}	KS	CM
0	100	0.036	0.066	0.041	0.045	0.039	0.070	0.046	0.045
	300	0.056	0.066	0.062	0.041	0.040	0.054	0.046	0.045
	500	0.059	0.064	0.070	0.056	0.059	0.074	0.066	0.064
0.25	100	0.054	0.094	0.086	0.079	0.068	0.096	0.080	0.081
	300	0.165	0.209	0.218	0.214	0.153	0.202	0.200	0.194
	500	0.317	0.365	0.405	0.376	0.274	0.338	0.364	0.350
0.50	100	0.086	0.134	0.123	0.137	0.100	0.141	0.143	0.142
	300	0.414	0.433	0.507	0.470	0.423	0.438	0.510	0.470
	500	0.743	0.746	0.829	0.780	0.748	0.746	0.809	0.782
0.75	100	0.076	0.110	0.109	0.115	0.082	0.128	0.115	0.119
	300	0.329	0.361	0.410	0.376	0.340	0.353	0.402	0.368
	500	0.655	0.636	0.724	0.667	0.631	0.614	0.705	0.651
1	100	0.049	0.065	0.054	0.050	0.044	0.082	0.053	0.048
	300	0.069	0.063	0.068	0.054	0.054	0.073	0.063	0.051
	500	0.064	0.075	0.081	0.052	0.056	0.065	0.061	0.045

Contrarily with a floating regime the currency is allowed to fluctuate freely by market forces. Pegged regimes are somehow in between, the currency then has limited flexibility when compared with other currencies. As Zeileis et al. (2010) point out, information on the exchange rate regime of a country is not always fully disclosed by the corresponding central bank. Hence, data driven methods such as linear regression became popular to classify the exchange rate regime in operation. Zeileis et al. (2010) suggest that a vanishing error variance can be interpreted as a fixed currency regime, while a small or large error variance can indicate a pegged or floating regime respectively. This is illustrating that the error variance is an important quantity when looking for changes in the exchange rate regime. As such changes are often caused by policy interventions, tests for sudden breaks (rather than smooth transitions) are of reasonable interest.

We consider the exchange rates of the Chinese Yuan Renminbi (CNY) regressed on the exchange rates of the US Dollar (USD). The reason to do so is that China decided to give up on a fixed exchange rate to the US dollar in 2005. More precisely, we consider 251 data points which are the daily log-difference returns from July 26nd, 2005 to July 25nd, 2006 of the CNY and USD each with respect to the Swiss franc (CHF) as numeraire currency. This is the first year of observations of a data set considered by Zeileis et al. (2010) as well as Kirch and Weber (2018). Both studies use a linear regression model and a basket of four currencies as regressors, namely the USD, Japanese yen (JPY), Euro (EUR) and the British Pound (GBP). However, the results of Zeileis et al. (2010) indicate nearly vanishing regression coefficients for the JPY, the EUR and the GBP over the whole investigated time period from July 26nd, 2005 to July 31st, 2009.

We first apply the bootstrap test by Mohr and Neumeyer (2019) to test for changes in the unknown regression function. With a p-value of 90% it suggests a stable regression function.

Secondly, we apply our test based on T_{n1} using the 95%-quantile of the limiting distribution T as critical values. The test clearly rejects the null with a p-value smaller than 0.001%, indicating a change in the conditional variance function. The possible change point can be estimated by $\operatorname{argmax}_{s\in[0,1]}(\sup_{z\in\mathbb{R}}|\hat{T}_n(s,z)|)$ and suggests a change of the exchange rate regime in March 3rd, 2006 which is consistent with the results of Zeileis et al. (2010). Figure 1 shows the cumulative sum, $\sup_{z\in\mathbb{R}}|\hat{T}_n(\cdot,z)|$ (top plot), as well as the exchange rates of the CNY plotted against the time (bottom plot). The green dashed line is indicating the critical value while the red dashed line corresponds to the estimated change point.

Note that applying the tests to the full data set, no change in the regression function is detected (p-value 16%), but a change in the variance is clearly detected (p-value smaller than 0.001%). However, as the data set is rather large and from the findings of Zeileis et al. (2010) we expect more than one change in the variance when looking at the full set of observations, which makes the estimation of possible changes more complicated (see also section 5).

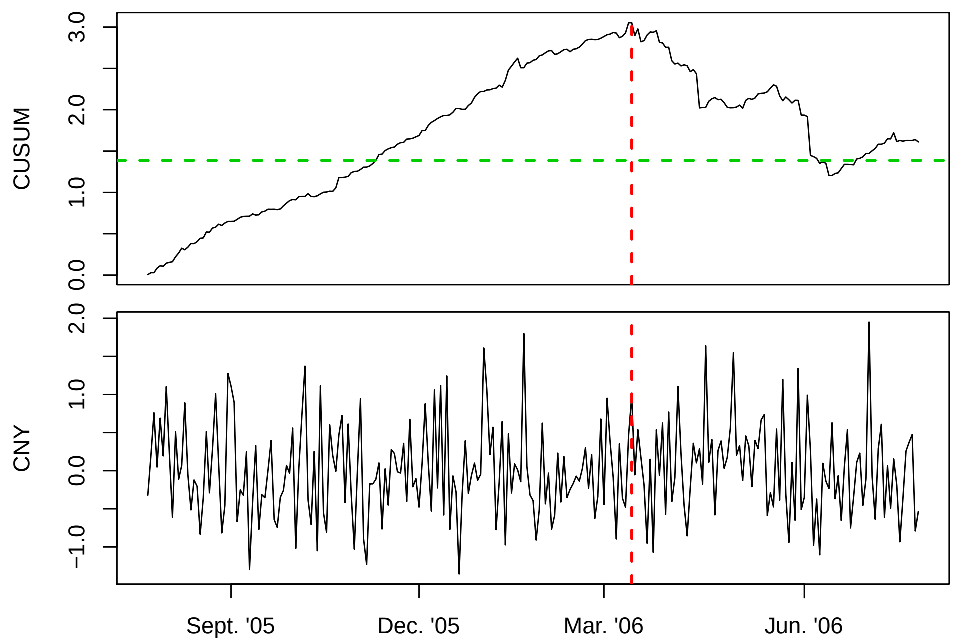

Figure 1: Exchange rate data: CUSUM and time series

5 Concluding remarks

This paper closes a gap in the change point testing theory for nonparametric time series models. Assume that one already has accepted that there is no change in the (nonparametric) regression function, but one suspects a change in the (nonparametric) volatility function. In such a case the new test gives a valid procedure. To the best knowledge of the authors the new test is the first that can be applied to (nonparametric) autoregressive models (no assumption of bounded support of the covariates) and is consistent against general alternatives of a change point in the variance function.

Under the assumption that only one change occurs, an estimator for the change point is given by $\operatorname{argmax}_{s\in[0,1]}(\sup_{\boldsymbol{z}\in\mathbb{R}^d}|\hat{T}_n(s,\boldsymbol{z})|)$. Asymptotic properties of this estimator will be considered in future research. If more than one change occurs it might be necessary to modify this estimator. For instance Fryzlewicz (2014) proposes a wild binary segmentation procedure for the estimation of multiple changes in a simple piecewise-constant signal model, which possibly can be adapted to our setting.

For our theoretical result Theorem 3.1 we need stationarity under the null. However, if there are no changes in both regression function m and variance function σ^2, there still could be a change in the error distribution of ε_t. In this case, a bootstrap test similar to the wild bootstrap proposal of Mohr and Neumeyer (2019) can be conducted that is sensible to changes in the variance function but not to changes in the error distribution. If both tests of Mohr and Neumeyer (2019) and the bootstrap version of the test at hand do not indicate a change in the regression and variance function respectively, the procedure of Selk and Neumeyer (2013) can be used to detect changes in the error distribution.

A Assumptions

(G) Let $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ be strictly stationary and α-mixing with mixing coefficient $\alpha(\cdot)$ such that $\alpha(t) = O(a^{-t})$ for some $a \in (1,\infty)$.

($\boldsymbol{\xi}$) For $\xi_t := U_t^2 - \sigma^2(\boldsymbol{X}_t)$ let there exist some $\gamma > 0$ and some even $Q > (d+1)(2+\gamma)$ such that $E[\xi_t|\mathcal{F}^t] = 0$, where $\mathcal{F}^t = \sigma(U_{j-1}, \boldsymbol{X}_j : j \le t)$, $E[\xi_t^2|\boldsymbol{X}_t] = \tau^2(\boldsymbol{X}_t)$ and $E[|\xi_t|^{Q\frac{2+\gamma}{2}}|\boldsymbol{X}_t] \le c(\boldsymbol{X}_t)^Q$ a.s. for all $t \in \mathbb{Z}$, for some functions $c, \tau^2 : \mathbb{R}^d \to \mathbb{R}$ with $\int \bar{c}(\boldsymbol{u}) f(\boldsymbol{u}) d(\boldsymbol{u}) \le M_1$ for some $M_1 < \infty$ and $\bar{c}(\boldsymbol{u}) = \max\left\{\tau^2(\boldsymbol{u}), c(\boldsymbol{u})^2, \ldots, c(\boldsymbol{u})^Q\right\}$.

($\boldsymbol{\sigma}$) For Q, γ from assumption **($\boldsymbol{\xi}$)** let $\int |\sigma^2(\boldsymbol{u})|^{Q\frac{2+\gamma}{2}} f(\boldsymbol{u}) d(\boldsymbol{u}) \le M_2$ for some $M_2 < \infty$.

(M) For some $b > 2$ let $E[|Y_1|^{2b}] < \infty$ and let $\boldsymbol{X}_1$ be absolutely continuous with density function $f : \mathbb{R}^d \to \mathbb{R}$ that satisfies $\sup_{\boldsymbol{x}\in\mathbb{R}^d} E[|Y_1|^{2b}|\boldsymbol{X}_0 = \boldsymbol{x}] f(\boldsymbol{x}) < \infty$ and $\sup_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x}) < \infty$. Let there exist some $j^* < \infty$ such that $\sup_{\boldsymbol{x}_1,\boldsymbol{x}_j} E[Y_1^2 Y_j^2|\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_j = \boldsymbol{x}_j] f_{1j}(\boldsymbol{x}_1, \boldsymbol{x}_j) < \infty$ for all $j \ge j^*$, where f_{1j} is the density function of $(\boldsymbol{X}_1, \boldsymbol{X}_j)$.

(J) Let $(c_n)_{n\in\mathbb{N}}$ be a positive sequence of real numbers satisfying $c_n \to \infty$ and $c_n = O((\log n)^{1/d})$ and let $\boldsymbol{J}_n = [-c_n, c_n]^d$.

(F1) For some $C < \infty$ and c_n from assumption **(J)** let $\boldsymbol{I}_n = [-c_n - Ch_n, c_n + Ch_n]^d$, where h_n is from assumption **(B1)** and **(B2)** and let $\delta_n^{-1} = \inf_{\boldsymbol{x}\in\boldsymbol{J}_n} f(\boldsymbol{x}) > 0$ for all $n \in \mathbb{N}$. Further, let for some $r, l \in \mathbb{N}$ and for all $n \in \mathbb{N}$

$$p_n = \max_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d \\ 1\le|\boldsymbol{k}|\le l+1+r}} \sup_{\boldsymbol{x}\in\boldsymbol{I}_n} |D^{\boldsymbol{k}} f(\boldsymbol{x})| < \infty$$

$$0 < q_n = \max\left\{ \max_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d \\ 0\le|\boldsymbol{k}|\le l+1+r}} \sup_{\boldsymbol{x}\in\boldsymbol{I}_n} |D^{\boldsymbol{k}} m(\boldsymbol{x})|, \max_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d \\ 0\le|\boldsymbol{k}|\le l+1+r}} \sup_{\boldsymbol{x}\in\boldsymbol{I}_n} |D^{\boldsymbol{k}} \sigma(\boldsymbol{x})| \right\} < \infty,$$

where $|\boldsymbol{i}| = \sum_{j=1}^d i_j$ and $D^{\boldsymbol{i}} = \frac{\partial^{|\boldsymbol{i}|}}{\partial x_1^{i_1} \ldots \partial x_d^{i_d}}$ for $\boldsymbol{i} = (i_1, \ldots, i_d) \in \mathbb{N}_0^d$.

(F2) For q_n from assumption **(F1)**, c_n from assumption **(J)** and C from assumption **(K)**, let for all $\boldsymbol{k} \in \mathbb{N}_0^d$ with $|\boldsymbol{k}| = 2$,

$$\max\left\{ \sup_{\boldsymbol{x}\in[-c_n-2h_nC, c_n+2h_nC]^d} \left|D^{\boldsymbol{k}} m(\boldsymbol{x})\right|, \sup_{\boldsymbol{x}\in[-c_n-2h_nC, c_n+2h_nC]^d} \left|D^{\boldsymbol{k}} \sigma(\boldsymbol{x})\right| \right\} = O(q_n).$$

(K) Let $K : \mathbb{R}^d \to \mathbb{R}$ be symmetric in each component, $l+1$ times differentiable with $\int_{\mathbb{R}^d} K(\boldsymbol{z}) d\boldsymbol{z} = 1$ and compact support $[-C, C]^d$. Additionally, let $r \ge 2$ and

$\int_{\mathbb{R}^d} K(\boldsymbol{z})\boldsymbol{z}^{\boldsymbol{k}} d\boldsymbol{z} = 0$ for all $\boldsymbol{k} \in \mathbb{N}_0^d$ with $1 \leq |\boldsymbol{k}| \leq r-1$, where $\boldsymbol{z}^{\boldsymbol{k}} = z_1^{k_1} \cdots z_d^{k_d}$. For all $L \in \{K\} \cup \{D^{\boldsymbol{k}}K : \boldsymbol{k} \in \mathbb{N}_0^d$ with $1 \leq |\boldsymbol{k}| \leq l+1\}$ let $|L(\boldsymbol{u})| < \infty$ for all $\boldsymbol{u} \in \mathbb{R}^d$ and $|L(\boldsymbol{u}) - L(\boldsymbol{u'})| \leq \Lambda \|\boldsymbol{u} - \boldsymbol{u'}\|$ for some $\Lambda < \infty$ and for all $\boldsymbol{u}, \boldsymbol{u'} \in \mathbb{R}^d$. (Here, r, l and C are from assumption **(F1)**.)

(B1) For δ_n, p_n, q_n and r, l from assumption **(F1)** let

$$\left(\sqrt{\frac{\log n}{n h_n^{d+2(l+1)}}} + h_n^r p_n\right) p_n^{l+1} \delta_n^{l+2} = O(1),$$

and for some $\eta \in (0,1)$ let

$$\left(\sqrt{\frac{\log n}{n h_n^{d+2(l+1)}}} + h_n^r p_n\right) p_n^{l+\eta} q_n^2 \delta_n^{l+1+\eta} = o(1).$$

(B2) For l, p_n, q_n, δ_n from assumption **(F1)** and η from assumption **(B1)** let

$$\frac{(\log n)^{3+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}} h_n^d}} q_n^3 \delta_n^2 = o(1), \ \frac{\log h_n}{\sqrt{n h_n^d}} = o(1), \ \sqrt{n} h_n^r p_n q_n^2 = o(1), \ (\log n)^3 h_n q_n^3 = o(1)$$

and $\dfrac{(\log n)^{2+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{1}{q}-\frac{d}{l+\eta}}}} q_n \delta_n = o(1)$ for $q = Q\frac{2+\gamma}{2}$ with Q and γ from assumption $(\boldsymbol{\xi})$.

Remark A.1. *Assumption **(G)** is fulfilled by data following causal and stationary ARMA models as they have an MA(∞) representation with coefficients that decay exponentially fast (see for instance Fan and Yao (2003) Subsection 2.6.1 (iii), p. 69). For more general nonlinear AR-ARCH processes both Lu (1998) and Liebscher (2005) give sufficient conditions on regression function, volatility function and the innovations under which the mixing condition in **(G)** holds. In the linear model*

$$Y_t = a_1 Y_{t-1} + \cdots + a_d Y_{t-d} + (b_0 + b_1 Y_{t-1}^2 + \cdots + b_d Y_{t-d}^2)^{1/2} \varepsilon_t, \ t \in \mathbb{Z},$$

where $(\varepsilon_t)_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, the condition in Lu (1998) simplifies to $(\sum_{i=1}^d |a_i|)^2 + \sum_{i=1}^d b_i < 1$.

Remark A.2. *In order to satisfy the first bandwidth assumption in **(B2)**, a necessary condition is $l + \eta > d$, hence for higher dimensional covariate $\boldsymbol{X}_t$, the existence of higher order partial derivatives of f and m is needed. In order to satisfy both the first and third bandwidth assumption in **(B2)** at the same time, depending on the dimension d and the smoothness parameters l and η, the order of the kernel r needs to be chosen such that $r > \frac{d}{2}\frac{l+\eta}{l+\eta-d}$ holds. As a rule of thumb, one can choose $h_n = O(n^{-k})$ for some $0 < k < \frac{1}{d} - \frac{1}{l+\eta}$ and a kernel, such that $r > \frac{1}{2k}$. That choice satisfies the assumptions given negligible rates for q_n and δ_n.*

*Further note that the last constraint in **(B2)** is merely a trade off between existence of moments of ξ_t, dimension d and smoothness parameters l and η. It is satisfied if $q > \frac{l+\eta}{l+\eta-d}$ (given negligible rates for q_n and δ_n).*

B Proofs

Lemma B.1. *Under the assumptions of Theorem 3.1 and under H_0 the following rates of convergence can be obtained for the kernel estimators $\hat{m}_n$ and $\hat{\sigma}_n^2$,*

(i) (a) $\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})| = O_P\left(\left(\sqrt{\frac{\log n}{nh_n^d}} + h_n^r p_n\right) q_n\delta_n\right)$,

(b) $\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right)\right| = O_P\left(\left(\sqrt{\frac{\log n}{nh_n^{d+2|\boldsymbol{k}|}}} + h_n^r p_n\right) p_n^{|\boldsymbol{k}|} q_n \delta_n^{|\boldsymbol{k}|+1}\right)$ *for all* $1 \leq |\boldsymbol{k}| \leq l+1$,

(c) $$\sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{J}_n\\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^\eta} = o_P(1) \text{ for all } |\boldsymbol{k}| = l,$$

(ii) (a) $\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |\hat{\sigma}_n^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x})| = O_P\left(\left(\sqrt{\frac{\log n}{nh_n^d}} + h_n^r p_n\right) q_n^2\delta_n\right)$,

(b) $\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left|D^{\boldsymbol{k}}\left(\hat{\sigma}_n^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x})\right)\right| = O_P\left(\left(\sqrt{\frac{\log n}{nh_n^{d+2|\boldsymbol{k}|}}} + h_n^r p_n\right) p_n^{|\boldsymbol{k}|} q_n^2 \delta_n^{|\boldsymbol{k}|+1}\right)$ *for all* $1 \leq |\boldsymbol{k}| \leq l+1$,

(c) $$\sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{J}_n\\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{\sigma}_n^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{\sigma}_n^2(\boldsymbol{y}) - \sigma^2(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^\eta} = o_P(1) \text{ for all } |\boldsymbol{k}| = l.$$

Note that the results for the Nadaraya-Watson estimator $\hat{m}_n$ in (i) are also stated in Lemma A.1 in Mohr and Neumeyer (2019). The proof of Lemma B.1 is similar to the proof of Theorem 8 in Hansen (2008) and omitted for the sake of brevity.

Lemma B.2. *Under the assumptions of Theorem 3.1 and under H_0 we have uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} U_i(m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} = o_P(1).$$

Proof. For some l-times differentiable function $h : \boldsymbol{J}_n \to \mathbb{R}$ define the norm

$$\|h\|_{l+\eta} := \max_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d\\ 1\leq|\boldsymbol{k}|\leq l}} \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left|D^{\boldsymbol{k}}h(\boldsymbol{x})\right| + \max_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d\\ |\boldsymbol{k}|=l}} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{J}_n\\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{\left|D^{\boldsymbol{k}}h(\boldsymbol{x}) - D^{\boldsymbol{k}}h(\boldsymbol{y})\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^\eta}$$

and the function class $\mathcal{H} := \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n) := \{h : \boldsymbol{J}_n \to \mathbb{R} : \|h\|_{l+\eta} \leq 1, \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|h(\boldsymbol{x})| \leq z_n(\log n)^{1/2}\}$ with $z_n := q_n\delta_n((\log n)/(nh_n^d))^{1/2}$. The third bandwidth condition in **(B2)** implies

$$\left(\sqrt{\frac{\log n}{nh_n^d}} + h_n^r p_n\right) q_n\delta_n = O\left(\sqrt{\frac{\log n}{nh_n^d}} q_n\delta_n\right)$$

and thus Lemma B.1 (i) implies that $P(\hat{h}_n \in \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n)) \to 1$ as $n \to \infty$ holds for $\hat{h}_n(\boldsymbol{x}) = (m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x})$. It is then sufficient to consider $n^{-1/2}\sum_{i=1}^{\lfloor ns\rfloor} h(\boldsymbol{X}_i)U_i I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}$ for $s \in [0,1]$, $\boldsymbol{z} \in \mathbb{R}^d$ and $h \in \mathcal{H}$. Furthermore, using **($\boldsymbol{\xi}$)** and **($\boldsymbol{\sigma}$)** it can be shown that for $q := Q\frac{2+\gamma}{2} > 2$

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} h(\boldsymbol{X}_i)U_i I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\big(h(\boldsymbol{X}_i)U_i I\{|U_i| \leq n^{1/q}\} I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} - E[h(\boldsymbol{X}_i)U_i I\{|U_i| \leq n^{1/q}\} I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}]\big) \\
&+ o_P(1)
\end{aligned}$$

holds uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. Defining the function class $\mathcal{F} := \{(u, \boldsymbol{x}) \mapsto uI\{|u| \leq n^{1/q}\} I\{\boldsymbol{x} \leq \boldsymbol{z}\} : \boldsymbol{z} \in \mathbb{R}^d\}$ and imposing $(U_1, \boldsymbol{X}_1) \sim P$, the assertion then follows if we show

$$\sup_{s\in[0,1]}\sup_{\varphi\in\mathcal{F}}\sup_{h\in\mathcal{H}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\Big(h(\boldsymbol{X}_i)\varphi(U_i, \boldsymbol{X}_i) - \int h\varphi dP\Big)\right| = o_P(1).$$

To this end let $\varepsilon_{n1} = n^{-1/2}n^{-1/q}$, $\varepsilon_{n2} = n^{-1/2}$ and $\varepsilon_{n3} = n^{-1/2}/(\log n)$ and let further $0 = s_1 < \cdots < s_{K_n} = 1$ partition $[0,1]$ in intervals of length $2\varepsilon_{n1}$ such that $K_n = O(\varepsilon_{n1}^{-1})$. Furthermore, we use the bracketing numbers $J_n := N_{[\,]}\big(\varepsilon_{n2}, \mathcal{F}, \|\cdot\|_{L_2(P)}\big)$ and $M_n := N_{[\,]}(\varepsilon_{n3}, \mathcal{H}, \|\cdot\|_\infty)$, where $\|\cdot\|_\infty$ is the supremum norm on $\boldsymbol{J}_n$. Let $[\varphi_1^l, \varphi_1^u], \ldots, [\varphi_{J_n}^l, \varphi_{J_n}^u]$ denote the brackets needed to cover $\mathcal{F}$. Let furthermore $[h_1^l, h_1^u], \ldots, [h_{M_n}^l, h_{M_n}^u]$ define the brackets needed to cover $\mathcal{H}$. It can be shown that $J_n = O\big(\varepsilon_{n2}^{-2d}\big)$ and $M_n = O(\exp(c_n^d \varepsilon_{n3}^{-d/(l+\eta)}))$ and further

$$\begin{aligned}
&\sup_{s\in[0,1]}\sup_{\varphi\in\mathcal{F}}\sup_{h\in\mathcal{F}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\Big(h(\boldsymbol{X}_i)\varphi(U_i, \boldsymbol{X}_i) - \int h\varphi dP\Big)\right| \\
&\leq \max_{\substack{1\leq k\leq K_n\\1\leq j\leq J_n\\1\leq m\leq M_n}}\sup_{\varphi\in[\varphi_j^l,\varphi_j^u]}\sup_{h\in[h_m^l,h_m^u]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\Big(h(\boldsymbol{X}_i)\varphi(U_i, \boldsymbol{X}_i) - \int h\varphi dP\Big)\right| + o_P(1) \\
&\leq \max_{\substack{1\leq k\leq K_n\\1\leq j\leq J_n\\1\leq m\leq M_n}}\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\Big(h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i, \boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i, \boldsymbol{X}_i) \geq 0\} - \int h_m^u\varphi_j^u I\{h_m^u\varphi_j^u \geq 0\} dP\Big)\right|,\right. \\
&\qquad\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\Big(h_m^l(\boldsymbol{X}_i)\varphi_j^l(U_i, \boldsymbol{X}_i) I\{h_m^l(\boldsymbol{X}_i)\varphi_j^l(U_i, \boldsymbol{X}_i) < 0\} - \int h_m^l\varphi_j^l I\{h_m^l\varphi_j^l < 0\} dP\Big)\right|, \\
&\qquad\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\Big(h_m^l(\boldsymbol{X}_i)\varphi_j^l(U_i, \boldsymbol{X}_i) I\{h_m^l(\boldsymbol{X}_i)\varphi_j^l(U_i, \boldsymbol{X}_i) \geq 0\} - \int h_m^l\varphi_j^l I\{h_m^l\varphi_j^l \geq 0\} dP\Big)\right|,
\end{aligned}$$

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left(h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)<0\}-\int h_m^u\varphi_j^u I\{h_m^u\varphi_j^u<0\}dP\right)\right|\Bigg\}$$
$$+\,o_P(1).$$

In what follows we only consider the first line on the right hand side, while the other ones can be treated similarly. We apply Theorem 2.1 of Liebscher (1996) to the random variable (for m, j, k fixed)

$$Z_i := \left(h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)\geq 0\}-\int h_m^u\varphi_j^u I\{h_m^u\varphi_j^u\geq 0\}dP\right) I\left\{\tfrac{i}{n}\leq s_k\right\}.$$

The mixing coefficient of $\{Z_t : 1 \leq t \leq n\}$ can be bounded by the mixing coefficient of $\{(U_t, \boldsymbol{X}_t) : t \in \mathbb{Z}\}$ due to Bradley (1985), Section 2, remark (iv). Further, the variables are centered and have a bound of order $O(z_n(\log n)^{1/2}n^{1/q})$. Applying Theorem 2.1 to $\sum_{i=1}^n Z_i$ yields for all $\epsilon > 0$ and $n \in \mathbb{N}$ large enough

$$\begin{aligned}
&P\Bigg(\max_{\substack{1\leq k\leq K_n\\1\leq j\leq J_n\\1\leq m\leq M_n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\Big(h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)\geq 0\}\right.\\
&\qquad\qquad\left.-\int h_m^u\varphi_j^u I\{h_m^u\varphi_j^u\geq 0\}dP\Big)\right|>\epsilon\Bigg)\\
&\leq\sum_{\substack{1\leq k\leq K_n\\1\leq j\leq J_n\\1\leq m\leq M_n}}P\Bigg(\left|\sum_{i=1}^{\lfloor ns_k\rfloor}\Big(h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i)\varphi_j^u(U_i,\boldsymbol{X}_i)\geq 0\}\right.\\
&\qquad\qquad\left.-\int h_m^u\varphi_j^u I\{h_m^u\varphi_j^u\geq 0\}dP\Big)\right|>\sqrt{n}\epsilon\Bigg)\\
&\leq K_nJ_nM_n4\exp\left(-\frac{n\epsilon^2}{64n\left\lfloor(nh_n^d)^{1/2}\right\rfloor z_n^2\log(n)+\frac{8}{3}n^{1/2}\epsilon\left\lfloor(nh_n^d)^{1/2}\right\rfloor z_n\log(n)^{1/2}n^{1/q}}\right)\\
&\quad+K_nJ_nM_n4\frac{n}{\left\lfloor(nh_n^d)^{1/2}\right\rfloor}\alpha\left(\left\lfloor(nh_n^d)^{1/2}\right\rfloor\right)\\
&=o(1),
\end{aligned}$$

where the first, second and last bandwidth constraint in **(B2)** were used in the last equality. Details are omitted for the sake of brevity.

□

Lemma B.3. *Under the assumptions of Theorem 3.1 and under H_0 we have uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(U_i^2-\hat{\sigma}_n^2(\boldsymbol{X}_i)\right)\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}=T_n(s,\boldsymbol{z})-sT_n(1,\boldsymbol{z})+o_P(1).$$

Note that the proof of Lemma B.3 is similar to the proof of Theorem 3.1 (i) in Mohr and Neumeyer (2019). It will only be sketched for the sake of brevity.

Proof. Using $\xi_t = U_t^2 - \sigma^2(\boldsymbol{X}_t)$ under H_0, it holds that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(U_i^2-\hat{\sigma}_n^2(\boldsymbol{X}_i)\right)\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\xi_i\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}+\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(\sigma^2(\boldsymbol{X}_i)-\hat{\sigma}_n^2(\boldsymbol{X}_i)\right)\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}.
\end{aligned}$$

By strict stationarity of $\{(\xi_t, \boldsymbol{X}_t) : t \in \mathbb{Z}\}$ and the moment constraints from **($\boldsymbol{\xi}$)** we deduce that uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\xi_i\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}=T_n(s,\boldsymbol{z})+o_P(1).$$

Making use of the uniform convergence rates of $\hat{\sigma}_n^2$ stated in Lemma B.1 (ii) we furthermore obtain

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(\sigma^2(\boldsymbol{X}_i)-\hat{\sigma}_n^2(\boldsymbol{X}_i)\right)\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}\\
&=s\sqrt{n}\int\left(\sigma^2(\boldsymbol{x})-\hat{\sigma}_n^2(\boldsymbol{x})\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x}+o_P(1),
\end{aligned}$$

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. Continuing by inserting the definition of $\hat{\sigma}_n^2$, using $Y_i = m(\boldsymbol{X}_i) + U_i$ and finally $\xi_i = U_i^2 - \sigma^2(\boldsymbol{X}_i)$ under H_0, it holds that

$$\begin{aligned}
&\sqrt{n}\int\left(\sigma^2(\boldsymbol{x})-\hat{\sigma}_n^2(\boldsymbol{x})\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x}\\
&=\sqrt{n}\int\left(\sigma^2(\boldsymbol{x})-\frac{1}{n}\sum_{i=1}^{n}K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)(Y_i-\hat{m}_n(\boldsymbol{x}))^2\frac{1}{\hat{f}_n(\boldsymbol{x})}\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int\left(\sigma^2(\boldsymbol{x})-(Y_i-\hat{m}_n(\boldsymbol{x}))^2\right)K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x}\\
&=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\int K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} &\text{(B.1)}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int\left(\sigma^2(\boldsymbol{x})-\sigma^2(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} &\text{(B.2)}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int(m(\boldsymbol{X}_i)-\hat{m}_n(\boldsymbol{x}))^2K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} &\text{(B.3)}\\
&+\frac{2}{\sqrt{n}}\sum_{i=1}^{n}U_i\int(m(\boldsymbol{X}_i)-\hat{m}_n(\boldsymbol{x}))K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le\boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x}. &\text{(B.4)}
\end{aligned}$$

Concerning (B.1) and (B.2), it can be shown that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}+o_P(1),$$
$$= T_n(1,\boldsymbol{z})+o_P(1),$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}\left(\sigma^2(\boldsymbol{x})-\sigma^2(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x}=o_P(1),$$

uniformly in $\boldsymbol{z}\in\mathbb{R}^d$ respectively. Using the uniform rates of convergences of $\hat{m}_n$ from Lemma B.1 (i) (a), which also hold on the slightly larger set $\boldsymbol{I}_n=[-c_n-Ch_n,c_n+Ch_n]^d$, it can be shown that the term (B.3) is negligible uniformly in $\boldsymbol{z}\in\mathbb{R}^d$. Finally, using similar methods as for the proof of Lemma B.2, it can be shown that the term (B.4) is as well negligible uniformly in $\boldsymbol{z}\in\mathbb{R}^d$. Putting the results together, the assertion of the lemma follows.

□

Proof of Theorem 3.1. The assertion (i) follows by Lemma B.2 and Lemma B.3 and by Lemma B.1 (i) (a) together with the bandwidth constraints as

$$\begin{aligned}\hat{T}_n(s,\boldsymbol{z}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}(m(\boldsymbol{X}_i)-\hat{m}_n(\boldsymbol{X}_i))^2\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}\\ &+\frac{2}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}U_i(m(\boldsymbol{X}_i)-\hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}\\ &+\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(U_i^2-\hat{\sigma}_n^2(\boldsymbol{X}_i)\right)\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}.\end{aligned}$$

For (ii) note that $\{(\xi_t,\boldsymbol{X}_t):t\in\mathbb{Z}\}$ is strictly stationary and strongly mixing under H_0 and assumption **(G)**. Denote by P the marginal distribution of $(\xi_1,\boldsymbol{X}_1)$. The assertion then follows by an application of Corollary 2.7 in Mohr (2019) to the sequential empirical process $\{n^{-1/2}\sum_{i=1}^{\lfloor ns\rfloor}(\varphi(\xi_i,\boldsymbol{X}_i)-\int\varphi dP):s\in[0,1],\varphi\in\mathcal{F}\}$ indexed in the function class $\mathcal{F}:=\{(\xi,\boldsymbol{x})\mapsto\xi I\{\boldsymbol{x}\le\boldsymbol{z}\}:\boldsymbol{z}\in\mathbb{R}^d\}$. The conditions that are needed for the asymptotic equicontinuity of the process are implied by assumptions **(G)** and **($\boldsymbol{\xi}$)**. The convergence of the finite dimensional distributions can be shown by applying Corollary 1 in Rio (1995), which is a central limit theorem for strongly mixing triangular arrays.

□

References

Berkes, I., Horváth, L., and Kokoszka, P. (2004). Testing for parameter constancy in GARCH(p,q) models. *Stat. Probab. Lett.*, 70:182–195.

Bradley, R. C. (1985). Basic properties of strong mixing conditions. In Eberlein, E. and Taqqu, M. S., editors, *Dependence in Probability and Statistics*, pages 165–192. Birkhäuser, Boston.

Chen, G., Choi, Y. K., and Zhou, Y. (2005). Nonparametric estimation of structural change points in volatility models for time series. *J. Econom.*, 126:79–114.

Chen, Z. and Tian, Z. (2014). Ratio tests for variance change in nonparametric regression. *Statistics*, 48:1–16.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42:2243–2281.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* Chapman & Hall/CRC, Boca Raton.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econom. Theory*, 24:726–748.

Hidalgo, J. (1995). A Nonparametric Conditional Moment Test for Structural Stability. *Econom. Theory*, 1:671–698.

Honda, T. (1997). The CUSUM tests with nonparametric regression residuals. *J. Japan Statist. Soc.*, 27:45–63.

Kengne, W. C. (2012). Testing for parameter constancy in general causal time-series models. *J. Time Ser. Anal.*, 33:503–518.

Kirch, C. and Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electron. J. Stat.*, 12:1579–1613.

Kokoszka, P. and Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lith. Math. J.*, 39:182–195.

Koul, H. L. (1996). Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Ann. Statist.*, 24:380–404.

Kulperger, R. and Yu, H. (2005). High Moment Partial Sum Processes of Residuals in GARCH Models and Their Applications. *Ann. Statist.*, 33:2395–2422.

Lee, S. and Lee, J. (2014). Residual Based Cusum Test for Parameter Change in AR-GARCH Models. In Huynh, V.-N., Kreinovich, V., and Sriboonchitta, S., editors, *Modeling Dependence in Econometrics*, pages 101–111. Springer, Cham.

Lee, S., Na, O., and Na, S. (2003). On the CUSUM of squares test for the variance change in nonstationary and nonparametric time series models. *Ann. Inst. Statist. Math.*, 55:467–485.

Lee, S. and Song, J. (2008). Test for parameter change in ARMA models with GARCH innovations. *Stat. Probab. Lett.*, 78:1990–1998.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Process. Appl.*, 65:69–80.

Liebscher, E. (2005). Towards a Unified Approach for Proving Geometric Ergodicity and Mixing Properties of Nonlinear Autoregressive Processes. *J. Time Ser. Anal.*, 26:669–689.

Ling, S. (1998). Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *Ann. Statist.*, 26:741–754.

Lu, Z. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Stat. Sinica*, 8:1205–1217.

Mohr, M. (2019). A weak convergence result for sequential empirical processes under weak dependence. *Stochastics*. `https://doi.org/10.1080/17442508.2019.1602132`.

Mohr, M. and Neumeyer, N. (2019). Consistent nonparametric change point detection combining CUSUM and marked empirical processes. Preprint available at `https://arxiv.org/abs/1901.08491`.

Rio, E. (1995). About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Statist.*, 1:35–61.

Selk, L. and Neumeyer, N. (2013). Testing for a change of the innovation distribution in nonparametric autoregression: The sequential empirical process approach. *Scand. J. Statist.*, 40:770–788.

Song, J. and Kang, J. (2018). Parameter change tests for ARMA-GARCH models. *Comput. Stat. Data Anal.*, 121:41–56.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.*, 25:613–641.

Su, L. and Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. *Stat. Interface*, 1:347–366.

Yang, L., Härdle, W., and Nielsen, J. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *J. Time Ser. Anal.*, 20:579–604.

Zeileis, A., Shah, A., and Patnaik, I. (2010). Testing, Monitoring and Dating Structural Changes in Exchange Rate Regimes. *Comput. Stat. Data Anal.*, 54:1696–1706.

Zhao, Z. and Wu, W. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.*, 36:1854–1879.