Specification and structural break tests for additive models with applications to realized variance data

M. R. Fengler* E. Mammen† M. Vogt‡

April 17, 2015

Abstract

We study two types of testing problems in a nonparametric additive model setting: We develop methods to test (i) whether an additive component function has a given parametric form and (ii) whether an additive component has a structural break. We apply the theory to a nonparametric extension of the linear heterogeneous autoregressive model which is widely employed to describe realized variance data. We find that the linearity assumption is often rejected, but actual deviations from linearity are mild.

AMS 1991 subject classifications. 62G08, 62G10

Journal of economic literature classification. C14, C58

Keywords and phrases. Additive models; Backfitting; Nonparametric time series analysis; Specification tests; Structural break tests; Realized variance; Heterogeneous autoregressive model

---

*University of St. Gallen, Institute of Mathematics and Statistics, Bodanstrasse 6, 9000 St. Gallen, Switzerland. Tel. +41 71 224 2457. Fax: +41 71 224 2894. E-mail: matthias.fengler@unisg.ch. Supported by Swiss National Science Foundation, Grant 144033: "Analysis and models of cross asset dependency structures in high-frequency data."

†Institute for Applied Mathematics, Heidelberg University, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany; Laboratory of Stochastic Analysis and its Applications, Higher School of Economics, 26, Ulitsa Shabolovka, Moscow, Russia. E-mail: mammen@math.uni-heidelberg.de. Support by Deutsche Forschungsgemeinschaft through the Research Training Group RTG 1953 is gratefully acknowledged. Research was prepared within the framework of a subsidy granted to the HSE by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

‡Department of Mathematics and Statistics, University of Konstanz, 78457 Konstanz, Germany. Tel. +49 7531 88 2757. E-mail: michael.vogt@uni-konstanz.de.

1 Introduction

Additive models are an important structured nonparametric regression framework. Compared to fully nonparametric models, they have the advantage that the regression function can be estimated without running into the curse of dimensionality problem. For this reason, they are particularly useful in applications where the dimensionality of the regressors is too high to fit a fully nonparametric model. Examples from economics and finance include, e.g., estimating a production function (Liu and Yang; 2010), studying the determinants of migration (Linton and Härdle; 1996), or modeling volatility (Yang et al.; 1999; Linton and Mammen; 2005). Remarkably, there is an abundant literature on estimation, but work on testing problems in additive models is scarce. The goal of this paper is to develop two diagnostic tests for additive models, and with the help of these, to explore the assumptions underlying a classical model for realized variance data.

We work with the following model setup. We observe a sample of time series data $\{(Y_t, \boldsymbol{X}_t) : t = 1, \ldots, T\}$, where $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,d})^\top$. The data are described by the additive model

$$Y_t = m_0 + \sum_{j=1}^{d} m_j(X_{t,j}) + \varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{1}$$

where $\mathbb{E}[\varepsilon_t | \boldsymbol{X}_t] = 0$ and m_j are unknown nonparametric functions. We tackle two testing problems within this framework: we test (i) whether an additive component function m_j has a given parametric form; and (ii) whether there is a structural break in an additive component m_j. To construct the tests, we build on the smooth backfitting estimator (SBE) of Mammen et al. (1999). The SBE avoids the drawbacks of the ordinary backfitting algorithm of Hastie and Tibshirani (1990), which may break down in strongly correlated designs (Nielsen and Sperlich; 2005). This will be important for the applications we have in mind. Moreover, the asymptotic properties of the SBE are better understood.[1]

Our test for parametric specification is introduced in Section 2. Roughly speaking, it compares a nonparametric and a parametric fit of the additive component function m_j under consideration. More specifically, it measures an L_2-distance between a smooth backfitting estimate and a parametric fit of m_j. There is a variety of tests on parametric specification in fully nonparametric models; see, among others, Härdle and Mammen (1993), González-Manteiga and Cao-Abad (1993), Hjellvik et al. (1998), Zheng (1996) and Kreiß et al. (2008). In contrast, the number of testing procedures for parametric specification in additive models is limited. Two notable exceptions are Fan and Jiang (2005) and Haag (2008). Fan and Jiang (2005) investigate the behavior of generalized

[1] Alternative estimators for additive models include kernel based marginal integration techniques (Newey; 1994; Tjøstheim and Auestad; 1994; Linton and Nielsen; 1995), sieve estimators (Chen; 2007), and penalized splines (Eilers and Marx; 2002).

likelihood ratio tests based on classical backfitting estimators and derive a variety of asymptotic results for them. Our approach differs from theirs in that we work with an L_2-type test statistic and base our test on smooth rather than classical backfitting. Haag's approach is more similar to ours, because he also considers an L_2-test statistic based on the SBE. His method, however, is only able to test whether the entire additive regression function $m(x) = m_0 + \sum_{j=1}^{d} m_j(x_j)$ belongs to a certain parametric family. In contrast, our test allows us to ask whether a specific component m_j has a given parametric form.

In Section 3, we tackle the problem of testing for structural breaks in the additive model (1). More specifically, we test whether an additive component function m_j has the same functional form before and after a pre-specified break point in time. Our method is based on a similar idea as our test on parametric specification. In particular, we compute two smooth backfitting estimates of m_j based on the data before and after the break point and compare them by means of an L_2-distance. The test allows us to check each component function separately for an unknown functional change, while the remaining functions may or may not undergo a structural break. Testing nonparametric functions for structural breaks or, more generally, for functional instability over time has been considered, e.g., in Hidalgo (1995) and Delgado and Hidalgo (2000). In additive models, the literature is again much more sparse. Indeed, to the best of our knowledge, testing for structural breaks in additive models by means of backfitting methods has not been considered so far.

In Section 4, we apply our methods to a major workhorse model for financial market volatility: the heterogeneous autoregressive (HAR) model for realized variance (RV) as suggested by Corsi (2009). As a core assumption, this model is linear in the regressors. But is there empirical support for this assumption? Based on the diagnostic tools that we provide in this paper, we can answer this question. In particular, we suggest a non-parametric version of the HAR model that belongs to the class of additive models (1) and use our tests to investigate the linearity assumption. In addition, we discuss the problem of measurement error of RV and provide size and power simulations for our tests.

In the empirical analysis, we study RV data for 17 global futures contracts and indices from 2003 to 2010. Using the structural break test, we decide whether to split up the RV time series into a pre-crisis and a crisis sample. We then test for linearity. Recently, Lahaye and Shaw (2014) have used conventional methods to test the linear HAR model against a fully nonparametric model without any additive structure. Interestingly, the authors cannot reject the linear model. With our procedure, in contrast, we can detect a number of violations of the linearity assumption, but the actual deviations from linearity are mild. We thus conclude that the linear specification of the HAR model is well taken in most cases. This evidence may explain why most refinements of the linear HAR model achieve only tiny improvements in terms of predictive power (Corsi et al.; 2012). A small forecasting exercise confirms this expectation.

2 Testing for a parametric specification

In what follows, we investigate the question whether one of the additive components $m_1, \dots, m_d$ in model (1) admits a certain parametric form. Without loss of generality, we restrict attention to m_1, i.e., we test whether m_1 belongs to a parametric family of functions $\{m_\theta : \theta \in \Theta\}$, where Θ denotes the parameter space. The null hypothesis is thus given by

$$H_0 : m_1 \in \big\{m_\theta : \theta \in \Theta\big\}.$$

To identify the additive component functions $m_1, \dots, m_d$ in model (1), we normalize them to satisfy $\int m_j(x_j)p_j(x_j)dx_j = 0$. Here, p_j is the marginal density of the j-th regressor $X_{t,j}$. To keep the notation as simple as possible, we assume throughout that the regressors $\boldsymbol{X}_t$ have bounded support. Without loss of generality, the support is supposed to equal the unit cube $[0,1]^d$. The case of unbounded support can be incorporated by slightly modifying the test statistic. We comment on this in Appendix A. In the next subsection, we introduce our test statistic. The asymptotic distribution of the statistic is derived in the second subsection. Finally, we describe a wild bootstrap procedure to improve the small sample behavior of the test. The technical assumptions and proofs of the main results can be found in Appendix A.

2.1 The test statistic

First suppose that the constant m_0 and the functions $m_2, \dots, m_d$ were known. In this situation, we could base our test on the one-dimensional model

$$Z_t = m_1(X_{t,1}) + \varepsilon_t,$$

where $Z_t = Y_t - m_0 - \sum_{j=2}^d m_j(X_{t,j})$, and use standard nonparametric procedures to test the hypothesis H_0. In particular, we could apply the kernel-based test of Härdle and Mammen (1993) which measures an L_2-distance between a parametric fit and a kernel smoother of the function m_1.

As we do not observe m_0 and the functions $m_2, \dots, m_d$, we replace them by a set of estimates, which are obtained by the smooth backfitting procedure introduced in Mammen et al. (1999). We focus attention on a version of the smooth backfitting algorithm which is based on Nadaraya-Watson smoothers and comment on a local linear version below. The smooth backfitting estimators $\tilde{m}_0, \dots, \tilde{m}_d$ of the functions $m_0, \dots, m_d$ are defined as the minimizers of the criterion

$$\sum_{t=1}^{T} \int_{[0,1]^d} \left\{ Y_t - f_0 - \sum_{j=1}^{d} f_j(x_j) \right\}^2 \boldsymbol{K}_g(x, \boldsymbol{X}_t) dx, \tag{2}$$

where the minimization runs over all additive functions $f(x) = f_0 + f_1(x_1) + \cdots + f_d(x_d)$ whose components satisfy $\int_0^1 f_j(x_j)\tilde{p}_j(x_j)dx_j = 0$ for $j = 1, \ldots, d$. Here, $\tilde{p}_j$ is a standard kernel density estimator of p_j given by $\tilde{p}_j(x_j) = \frac{1}{T}\sum_{t=1}^{T} K_g(x_j, X_{t,j})$. Moreover, g is the bandwidth and $\boldsymbol{K}_g(v, w) = \prod_{j=1}^{d} K_g(v_j, w_j)$ is a product kernel. The factors $K_g(v_j, w_j)$ are modified kernel weights of the form

$$K_g(v_j, w_j) = \frac{K_g(v_j - w_j)}{\int_0^1 K_g(s - w_j)ds},$$

where $K_g(s) = g^{-1}K(s/g)$ and the kernel function $K(\cdot)$ integrates to one. These modified kernel weights have the property that $\int_0^1 K_g(v_j, w_j)dv_j = 1$ for all w_j, which is needed to derive the asymptotic results for the backfitting estimators.

Given the estimates $\tilde{m}_0, \tilde{m}_2, \ldots, \tilde{m}_d$, the variables Z_t can be approximated by $\tilde{Z}_t = Y_t - \tilde{m}_0 - \sum_{j=2}^{d} \tilde{m}_j(X_{t,j})$. Based on the sample $\{\tilde{Z}_t, X_{t,1}\}_{t=1}^{T}$, we can construct a parametric and a nonparametric estimator of the function m_1. Denote by $m_{\hat{\theta}}$ the parametric estimator, which satisfies the high-level condition (A8) in Appendix A, and denote by $\hat{m}$ a Nadaraya-Watson smoother of m_1 with bandwidth h, i.e.,

$$\hat{m}(w) = \frac{\sum_{t=1}^{T} K_h(w - X_{t,1})\tilde{Z}_t}{\sum_{t=1}^{T} K_h(w - X_{t,1})}.$$

As we will see below, the bandwidth h differs from g. In particular, for the theory to work, we have to undersmooth the backfitting estimates and thus choose g to converge faster to zero than h.

The idea of our test is to measure the distance between the two estimates $m_{\hat{\theta}}$ and $\hat{m}$. More specifically, we set up a test statistic of the type introduced in Härdle and Mammen (1993) which measures an L_2-distance between the parametric and the nonparametric estimate. The statistic is defined as

$$S_T = Th^{1/2} \int \left(\hat{m}(w) - \mathcal{K}_{h,T} m_{\hat{\theta}}(w) \right)^2 \pi(w) dw \ ,$$

where

$$\mathcal{K}_{h,T} g(\cdot) = \frac{\sum_{t=1}^{T} K_h(\cdot - X_{t,1}) g(X_{t,1})}{\sum_{t=1}^{T} K_h(\cdot - X_{t,1})}$$

and π is a weight function with bounded support $\mathrm{supp}(\pi) \subseteq [0, 1]$ and $\int \pi(x)dx = 1$. As proposed in Härdle and Mammen (1993), we smooth the parametric estimates $m_{\hat{\theta}}$ by applying the operator $\mathcal{K}_{h,T}$ to it. This artificially creates a bias term which cancels with the bias part of the kernel smoother $\hat{m}$.

Our test statistic is based on Nadaraya-Watson type estimators. Alternatively, local linear estimators could be used. Specifically, we may estimate the functions $m_0, m_2, \dots, m_d$ by a local linear based version of the smooth backfitting approach; see Mammen et al. (1999) for a formal definition and the technical details. Let us denote the resulting estimates by $\tilde{m}_0^{LL}, \tilde{m}_2^{LL}, \dots, \tilde{m}_d^{LL}$ and write $\tilde{Z}_t^{LL} = Y_t - \tilde{m}_0^{LL} - \sum_{j=2}^d \tilde{m}_j^{LL}(X_{t,j})$. With this notation at hand, we can replace $\hat{m}$ by the local linear smoother

$$\hat{m}^{LL}(w) = \frac{\sum_{t=1}^T W_h(w, X_{t,1}) \tilde{Z}_t^{LL}}{\sum_{t=1}^T W_h(w, X_{t,1})},$$

where $W_h(w, X_{t,1}) = K_h(w - X_{t,1})[Q_{T,2} - (w - X_{t,1})Q_{T,1}]$ and $Q_{T,j} = \sum_{t=1}^T K_h(w - X_{t,1})(w - X_{t,1})^j$ for $j = 1, 2$. Analogously as in the Nadaraya-Watson-based case, we may now define our test statistic by

$$S_T^{LL} = Th^{1/2} \int \left(\hat{m}^{LL}(w) - \mathcal{K}_{h,T}^{LL} m_{\hat{\theta}}(w)\right)^2 \pi(w) dw,$$

where the operator $\mathcal{K}_{h,T}^{LL}$ is given by

$$\mathcal{K}_{h,T}^{LL} g(\cdot) = \frac{\sum_{t=1}^T W_h(\cdot, X_{t,1}) g(X_{t,1})}{\sum_{t=1}^T W_h(\cdot, X_{t,1})}.$$

As in the Nadaraya-Watson case, this operator helps to get rid of the bias part of the nonparametric estimate.

2.2 Asymptotic distribution

We now examine the asymptotic behavior of our test. For simplicity, we focus on the theoretical analysis of the Nadaraya-Watson based statistic S_T. The statistic S_T^{LL} can be handled by similar arguments. We derive the limit distribution of S_T under local alternatives of the form

$$m_1(w) = m_{1,T}(w) = m_{\theta_0}(w) + c_T \Delta(w), \tag{3}$$

where m_{θ_0} is a parametric function with $\theta_0 \in \Theta$, Δ is a bounded function of w and $c_T = T^{-1/2} h^{-1/4}$. This nests the null hypothesis with $\Delta \equiv 0$.

Theorem 1. *Assume that the conditions (A1)–(A8) of Appendix A are satisfied. Then*

$$S_T - B_T \xrightarrow{d} N\left(\int (\mathcal{K}_h \Delta(w))^2 \pi(w) dw, V\right),$$

where $\mathcal{K}_h g(\cdot) = \int K_h(\cdot - u) g(u) du$ *and*

$$B_T = h^{-1/2} \kappa_0 \int \frac{\sigma^2(w)\pi(w)}{p_1(w)} dw$$
$$V = 2\kappa_1 \int \frac{[\sigma^2(w)]^2 \pi^2(w)}{p_1^2(w)} dw.$$

Here, p_1 is the marginal density of $X_{t,1}$, $\sigma^2(w) = \mathbb{E}[\varepsilon_t^2|X_{t,1} = w]$, $\kappa_0 = \int K^2(u)du$ and $\kappa_1 = \int(\int K(u)K(u+v)du)^2dv$.

Importantly, our test statistic has the same limit distribution as the test which is based on the one-dimensional model $Z_t = m_1(X_{t,1}) + \varepsilon_t$ with $Z_t = Y_t - m_0 - \sum_{j=2}^d m_j(X_{t,j})$. Thus, the uncertainty stemming from estimating the additive components $m_0, m_2, \ldots, m_d$ does not show up in the asymptotic distribution. Put differently, the test has the following oracle property: It has the same limit distribution as in the case where the components $m_0, m_2, \ldots, m_d$ are known.

2.3 Bootstrap

To improve the small sample behavior of our test, we set up a wild bootstrap procedure. The bootstrap sample is given by $\{Z_t^*, X_{t,1}\}_{t=1}^T$ with

$$Z_t^* = m_{\hat{\theta}}(X_{t,1}) + \varepsilon_t^*. \tag{4}$$

The bootstrap residuals are constructed as $\varepsilon_t^* = \hat{\varepsilon}_t \cdot \eta_t$, where $\hat{\varepsilon}_t = \tilde{Z}_t - \hat{m}(X_{t,1})$ are the estimated residuals and $\{\eta_t\}_{t=1}^T$ is some sequence of i.i.d. variables with zero mean and unit variance that is independent of the sample $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^T$. Denote by $m_{\hat{\theta}^*}$ and $\hat{m}^*$ the parametric and nonparametric estimates of m_1 calculated from the bootstrap sample $\{Z_t^*, X_{t,1}\}_{t=1}^T$. Replacing the estimates $m_{\hat{\theta}}$ and $\hat{m}$ in S_T by the bootstrap analogues $m_{\hat{\theta}^*}$ and $\hat{m}^*$ yields the bootstrap statistic

$$S_T^* = Th^{1/2} \int \big(\hat{m}^*(w) - \mathcal{K}_{h,T} m_{\hat{\theta}^*}(w)\big)^2 \pi(w)dw.$$

The next theorem shows that the bootstrap is consistent.

THEOREM 2. *Assume that the conditions (A1)–(A7) and (A8*) of Appendix A are satisfied. Then*

$$S_T^* - B_T \xrightarrow{d} N\big(0, V\big)$$

conditional on the sample $\big\{(Y_t, \boldsymbol{X}_t)\big\}_{t=1}^T$ with probability tending to one.

3 Testing for breaks

In this section, we discuss how to test for structural breaks in the additive model (1). In the presence of a structural break, the model can be written as

$$Y_t = \begin{cases} m_0^{ante} + \sum_{j=1}^d m_j^{ante}(X_{t,j}) + \varepsilon_t & \text{for } t < t^* \\ m_0^{post} + \sum_{j=1}^d m_j^{post}(X_{t,j}) + \varepsilon_t & \text{for } t \geq t^*, \end{cases} \tag{5}$$

where t^* is the break point and the functions m_j^{ante} and m_j^{post} denote the additive components before and after the break. Given the break t^*, we are interested in testing whether the various component functions have the same form before and after the break. More precisely, for each $j \in \{1, \ldots, d\}$, we want to test the hypothesis

$$H_0 : \; m_j^{ante}(x_j) = m_j^{post}(x_j) \quad \text{for almost all } x_j.$$

In the sequel, we assume that t^* is known. This is motivated by our application where we have a natural candidate for the break date. Our theory carries over to the case when t^* is unknown and can be estimated by using additional data. It changes if the break point is estimated by using only the observations from the sample $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^T$ because in this case, the break point is not defined under the null hypothesis where $m_j^{ante} \equiv m_j^{post}$. Moreover, the estimators of the additive functions will suffer from an additional bias because the break point is estimated such that the curves fitted before and after the break differ as strongly as possible.

Testing for a structural break is particularly difficult to handle when (5) is an autoregression, i.e., when we observe the time series $\{Y_t\}_{t=1}^T$ and set $X_{t,j} = Y_{t-j}$ for $j = 1, \ldots, d$ in (5). The reason is that the autoregressive process $\{Y_t\}_{t=1}^T$ is nonstationary in the presence of structural breaks. Specifically, the variables Y_t and Y_s will have different distributions at time points $s \neq t$ with $s, t \geq t^*$. To incorporate the autoregressive case, we thus cannot simply assume our data $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^T$ to be stationary. We rather have to take into account potential nonstationarities caused by structural breaks in the additive component functions. Appendix B provides the technical details on the nonstationary behavior we allow for. Moreover, it contains the proofs and theoretical arguments related to our structural break test.

3.1 The test statistic

Without loss of generality, we give an explicit definition of our test statistic only for the case $j = 1$. The statistic is based on the comparison of smooth backfitting estimators of m_1^{ante} and m_1^{post}. To introduce these estimators, we modify the discussion following equation (2) in the previous section. The Nadaraya-Watson smooth backfitting estimators $\tilde{m}_0^\ell, \ldots, \tilde{m}_d^\ell$ ($\ell = ante, post$) are defined as the minimizers of the criterion

$$\sum_{t \in \mathcal{T}_\ell} \int_{[0,1]^d} \Big\{ Y_t - f_0 - \sum_{j=1}^d f_j(x_j) \Big\}^2 \boldsymbol{K}_g(x, \boldsymbol{X}_t) dx, \tag{6}$$

where $\mathcal{T}_{ante} = \{t : 1 \leq t \leq t^* - 1\}$ and $\mathcal{T}_{post} = \{t : t^* \leq t \leq T\}$. The minimization runs over all additive functions $f(x) = f_0 + f_1(x_1) + \cdots + f_d(x_d)$ whose components satisfy $\int_0^1 f_j(x_j) \tilde{p}_j^\ell(x_j) dx_j = 0$ for $j = 1, \ldots, d$. Here, $\tilde{p}_j^\ell(x_j)$ is equal to $\frac{1}{T} \sum_{t \in \mathcal{T}_\ell} K_g(x_j, X_{t,j})$, where the kernel $K_g(v_j, w_j)$ is defined as in the previous section. Up to a factor, $\tilde{p}_j^\ell(x_j)$ can be interpreted as a kernel estimator of the average density of $X_{t,j}$ for $t \in \mathcal{T}_\ell$.

To construct our test statistic, we proceed similarly as in Section 2. To start with, we consider the variables $Z_t^\ell = Y_t - m_0^\ell - \sum_{j=2}^d m_j^\ell(X_{t,j})$ for $t \in \mathcal{T}_\ell$ and $\ell = ante, post$. These can be approximated by $\tilde{Z}_t^\ell = Y_t - \tilde{m}_0^\ell - \sum_{j=2}^d \tilde{m}_j^\ell(X_{t,j})$. Based on the sample $\{\tilde{Z}_t^\ell, X_{t,1}\}_{t \in \mathcal{T}_\ell}$, we can construct the Nadaraya-Watson smoother of m_1^ℓ with bandwidth h,

$$\hat{m}_1^\ell(w) = \frac{\sum_{t \in \mathcal{T}_\ell} K_h(w - X_{t,1}) \tilde{Z}_t^\ell}{\sum_{t \in \mathcal{T}_\ell} K_h(w - X_{t,1})}.$$

Our test statistic is now defined as

$$S_T = Th^{1/2} \int \big(\mathcal{K}_{h,T}^{1,post} \hat{m}_1^{ante}(x) - \mathcal{K}_{h,T}^{1,ante} \hat{m}_1^{post}(x) - \hat{\delta} \big)^2 \pi(x) dx, \tag{7}$$

where

$$\begin{aligned}
\mathcal{K}_{h,T}^{1,ante} g(\cdot) &= \frac{\sum_{t=1}^{t^*-1} K_h(\cdot - X_{t,1}) g(X_{t,1})}{\sum_{t=1}^{t^*-1} K_h(\cdot - X_{t,1})}, \\
\mathcal{K}_{h,T}^{1,post} g(\cdot) &= \frac{\sum_{t=t^*}^{T} K_h(\cdot - X_{t,1}) g(X_{t,1})}{\sum_{t=t^*}^{T} K_h(\cdot - X_{t,1})}, \\
\hat{\delta} &= \int \big(\mathcal{K}_{h,T}^{1,post} \hat{m}_1^{ante}(x) - \mathcal{K}_{h,T}^{1,ante} \hat{m}_1^{post}(x) \big) \pi(x) dx
\end{aligned}$$

and π is a weight function with bounded support $\mathrm{supp}(\pi) \subseteq [0,1]$ and $\int \pi(x) dx = 1$. Note that $\hat{\delta}$ is chosen such that

$$S_T = \min_{\delta \in \mathbb{R}} Th^{1/2} \int \big(\mathcal{K}_{h,T}^{1,post} \hat{m}_1^{ante}(x) - \mathcal{K}_{h,T}^{1,ante} \hat{m}_1^{post}(x) - \delta \big)^2 \pi(x) dx.$$

The construction of this test statistic can be motivated as follows: In a first attempt, one could consider a test based on the statistic $\min_{\delta \in \mathbb{R}} Th^{1/2} \int \big(\hat{m}_1^{ante}(x) - \hat{m}_1^{post}(x) - \delta \big)^2 \pi(x) dx$. The estimates $\hat{m}_1^{ante}(x)$ and $\hat{m}_1^{post}(x)$ in this statistic have different asymptotic bias terms. For this reason, the test behaves like a linear test and not like an overall goodness-of-fit test; see Härdle and Mammen (1993) for a related discussion. Our test statistic corrects for this disadvantage because, as one can show, $\mathcal{K}_{h,T}^{1,post} \hat{m}_1^{ante}(x)$ and $\mathcal{K}_{h,T}^{1,ante} \hat{m}_1^{post}(x)$ have the same asymptotic bias and thus the bias terms cancel when we take the difference of the two smoothed estimates.

As an alternative, we could consider the test statistic $\min_{\delta \in \mathbb{R}} Th^{1/2} \int \big(\hat{m}_1^{LL,ante}(x) - \hat{m}_1^{LL,post}(x) - \delta \big)^2 \pi(x) dx$, where $\hat{m}_1^{LL,ante}$ and $\hat{m}_1^{LL,post}$ are local linear smoothers based on a local linear version of the backfitting algorithm. Now, no additional smoothing of the estimates is required because the asymptotic bias terms of the two estimates do not differ. The reason is that the bias of a local linear estimator does not depend on the design density of the covariates. This holds true both for local linear smoothers in classical regression models and for the local linear smooth backfitting estimators in additive models (Mammen et al.; 1999).

3.2 Asymptotic distribution

We now derive the asymptotic distribution of S_T under local alternatives of the following form: The function m_1^{ante} is fixed and

$$m_1^{post}(x) = m_1^{ante}(x) + c_T \Delta(x),$$

where $c_T = T^{-1/2}h^{-1/4}$ and Δ is some bounded function. The other component functions, i.e., the functions m_j^{ℓ} for $j > 1$ and $\ell = ante, post$ are assumed to be fixed. Importantly, we allow for structural breaks in the other components, i.e., we allow for the possibility that $m_j^{ante} - m_j^{post} \neq 0$ for some $j > 1$. One can show that the asymptotics of S_T do not change if m_j^{ante} and m_j^{post} are not fixed and additional uniform smoothness conditions are imposed on them. For $\Delta \equiv 0$, we obtain a specification that lies on our null hypothesis; for $\Delta \not\equiv 0$ we get a neighbored point in the alternative. The limit distribution of S_T is given by the following theorem.

THEOREM 3. *Suppose that assumptions (B1)–(B5) of Appendix B are satisfied. Then*

$$S_T - B_T \xrightarrow{d} N(\mu, V),$$

where $\mu = \int \Delta^2(x)\pi(x)dx - \left[\int \Delta(x)\pi(x)dx\right]^2$ *and*

$$B_T = h^{-1/2} K^{(2)}(0) \int \left[c^{-1}\sigma^{ante}(x)^2 + (1-c)^{-1}\sigma^{post}(x)^2\right] \pi(x)dx$$

$$V = 2K^{(4)}(0) \int \left[\frac{1}{c^2}\frac{\sigma^{ante}(x)^4}{p_1^{ante}(x)^2} + \frac{2}{c(1-c)}\frac{\sigma^{ante}(x)^2\sigma^{post}(x)^2}{p_1^{ante}(x)p_1^{post}(x)} + \frac{1}{(1-c)^2}\frac{\sigma^{post}(x)^4}{p_1^{post}(x)^2}\right] \pi(x)^2\, dx.$$

Here, $\sigma^{ante}(x)^2$ *is the conditional variance of* ε_t *given* $X_{t,1} = x$ *for* $t < t^*$ *and* $\sigma^{post}(x)^2$ *is the conditional variance of* ε_t *given* $X_{t,1} = x$ *for* $t \geq t^*$. *Furthermore,* $K^{(r)}$ *denotes the* r*-times convolution product of* K *(for* $r \geq 1$*) and* c *is the limit of* t^*/T *for* $T \to \infty$.

When deriving the above result, we have to take care of the following two points: (i) As already discussed above, we cannot simply assume that the process $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^T$ is stationary but have to take into account potential nonstationarities caused by structural breaks in the component functions. (ii) We have to show that by the additional smoothing operations $\mathcal{K}_{h,T}^{1,ante}$ and $\mathcal{K}_{h,T}^{1,post}$ the bias terms cancel in the test statistics. Appendix B provides the details on how to deal with these two issues. As for (i), we will assume that there exist stationary processes X_t^{ante} and X_t^{post} such that X_t is approximated by X_t^{ante} for $t < t^*$ and X_t is approximated by X_t^{post} for $t > t^*$. Appendix B gives a rigorous definition of these approximating processes. In addition, it provides conditions under which such approximating processes exist in the autoregressive case, in particular when considering the nonparametric HAR model of our empirical analysis in Section 4.

3.3 Bootstrap

To improve the small sample behavior, we suggest bootstrapping the test statistic. Denote the bootstrap sample by $\{(Z_t^*, X_{t,1})\}_{t=1}^T$, where

$$Z_t^* = \bar{m}_1(X_{t,1}) + \varepsilon_t^*$$

and $\bar{m}_1$ is an average of $\tilde{m}_1^{ante}$ and $\tilde{m}_1^{post}$. The bootstrap residuals are constructed as $\varepsilon_t^* = \hat{\varepsilon}_t \cdot \eta_t$, where $\hat{\varepsilon}_t = \tilde{Z}_t^{ante} - \hat{m}_1^{ante}(X_{t,1})$ for $t < t^*$ and $\hat{\varepsilon}_t = \tilde{Z}_t^{post} - \hat{m}_1^{post}(X_{t,1})$ for $t \geq t^*$ are the estimated residuals and $\{\eta_t\}_{t=1}^T$ is some sequence of i.i.d. variables with zero mean and unit variance that is independent of the sample $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^T$. Denote the bootstrap analogue of $\hat{m}_1^\ell(x)$ by $\hat{m}_1^{*,\ell}(x)$ for $\ell = ante, post$. The bootstrap statistic is then defined as

$$S_T^* = Th^{1/2} \int \big(\mathcal{K}_{h,T}^{1,post} \hat{m}_1^{*,ante}(x) - \mathcal{K}_{h,T}^{1,ante} \hat{m}_1^{*,post}(x) - \hat{\delta}_1^* \big)^2 \pi(x) dx$$

where

$$\hat{\delta}_1^* = \int \big(\mathcal{K}_{h,T}^{1,post} \hat{m}_1^{*,ante}(x) - \mathcal{K}_{h,T}^{1,ante} \hat{m}_1^{*,post}(x) \big) \pi(x) dx.$$

The following theorem states that the bootstrap works.

THEOREM 4. *Suppose that assumptions (B1)–(B5) of Appendix B are satisfied. Then*

$$S_T^* - B_T \xrightarrow{d} N\big(0, V\big)$$

conditional on the sample $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^T$ *with probability tending to one.*

4 Additive modeling of realized variance

In the following, we present a nonparametric extension of the heterogeneous autoregressive (HAR) model of Corsi (2009). Section 4.1 introduces the model. Section 4.2 discusses whether and how our estimation methods are affected by measurement errors in realized variance (RV) data. In Section 4.3, we simulate data from various HAR models and use these to investigate the size and power properties of our test procedures. The empirical applications follow in Section 4.4.

4.1 A nonparametric HAR model

To introduce the nonparametric HAR model, let V_t denote RV or a transformation of it such as realized volatility ($\sqrt{\text{RV}}$) or logarithmic RV ($\log \text{RV}$). Moreover, define $V_t^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} V_{t-j}$, $n \in \mathbb{N}_+$, to be an average of V_t over the past n trading days. Finally, denote by $\iota = (\iota_1, \ldots, \iota_d)^\top \in \mathbb{N}_+^d$ an index vector where $\iota_1 < \iota_2 \ldots < \iota_d$. The nonparametric HAR model is given by

$$V_t^{(\iota_1)} = m_0 + \sum_{j=1}^{d} m_j\big(V_{t-1}^{(\iota_j)}\big) + \varepsilon_t \quad \text{for } t = 1, \ldots T, \tag{8}$$

where m_0 is a constant, m_j $(j = 1, \ldots, d)$ are smooth functions of unknown shape, also called the variance component functions, and $\mathbb{E}[\varepsilon_t | V_{t-1}^{(\iota_1)}, \ldots, V_{t-1}^{(\iota_d)}] = 0$. The model is a special case of (1) and obtained by setting $Y_t = V_t^{(\iota_1)}$ and $X_{t,j} = V_{t-1}^{(\iota_j)}$ for $j = 1, \ldots, d$. The most commonly used index vector is $\iota = (1, 5, 22)^\top$, which corresponds to a daily lag and averages of the daily variances over the last week and the last month, respectively.[2] This choice is motivated by the idea that market participants that have different investment horizons, such as daily, weekly and monthly time scales, provoke different types of variance components by their trading activities (Corsi; 2009). As a standard assumption, the variance component functions in (8) are assumed to be linear, i.e., $m_j(x) = \theta_j x$, $\theta_j \in \mathbb{R}$, in which case the model reduces to a restricted AR(ι_d) model.

Despite its simplicity, the linear HAR model is a major benchmark for describing RV data. It well captures the principal stylized fact, i.e., the slowly decaying sample autocorrelation function of RV, and has a forecasting power that is hard to beat. Few studies have addressed the topic of potential nonlinearities in the HAR model for RV. McAleer and Medeiros (2008) consider a multiple regime smooth transition HAR model and Corsi et al. (2012) propose a tree-structured HAR model. Chen et al. (2013) suggest a linear HAR model whose coefficients are allowed to vary slowly in time. With the exception of Lahaye and Shaw (2014), who consider the very general model $V_t^{(1)} = m\big(V_{t-1}^{(1)}, V_{t-1}^{(5)}, V_{t-1}^{(22)}\big) + \varepsilon_t$, all these studies assume a specification that is linear conditionally on the regime or locally in time. In spite of these efforts, the actual improvements that are achieved in terms of predictive ability vis-à-vis the linear model have proved to be comparably tiny. See Corsi et al. (2012) for a survey on the HAR model.[3]

4.2 Measurement error

Under appropriate conditions, RV that is constructed from high-frequency intra-day data can be considered as an estimator of the latent variance of the daily return process. There are situations, however, in which one may be interested in studying the latent variable

[2] This choice is owed to Corsi (2009) and has frequently been adopted. Testing the index itself is beyond the scope of this text; for linear models, this is investigated in Audrino and Knaus (2014).

[3] Aside from modeling RV data by means of neglected nonlinearities and structural breaks, an alternative strand of the literature uses long memory processes (Andersen et al.; 2001, 2003). Very recent research of Hillebrand and Medeiros (2014) suggests that both features – long memory and nonlinearities/structural breaks – may also jointly contribute to the observational patterns of RV data.

rather than its estimate. In what follows, we show that this is possible with the help of our methods as long as the estimation error is not too large. For simplicity, we restrict attention to the case where $V_t^{(1)}$ is realized volatility, i.e. $\sqrt{\mathrm{RV}}$, at day t and $V_t^{*(1)}$ is the underlying latent daily volatility estimated by $V_t^{(1)}$. Our arguments carry over to the case where $V_t^{(1)}$ is RV or the logarithm thereof.

Following Asai et al. (2012), the measurement error can be modeled by

$$V_t^{(1)} = V_t^{*(1)} + \frac{u_t}{n_t^{\beta}}, \tag{9}$$

where n_t is the number of high-frequency observations at day t and β satisfies $1/6 \leq \beta \leq 1/2$, the exact value of β depending on the assumptions that are adopted for additional microstructure noise in the observed logarithmic price process. Instead of $V_t^{(1)}$, we may want to describe the latent volatility $V_t^{*(1)}$ by a HAR model, that is,

$$V_t^{*(\iota_1)} = m_0 + \sum_{j=1}^{d} m_j\big(V_{t-1}^{*(\iota_j)}\big) + \varepsilon_t. \tag{10}$$

This model is more complicated than (8) because $V_t^{*(\iota_j)}$ is not observed. To estimate the functions m_j, we replace the latent variables $V_t^{*(\iota_j)}$ by the RV estimates $V_t^{(\iota_j)}$ and apply the smooth backfitting algorithm to them.

We now investigate how the SBE gets affected by the estimation error in RV. Without loss of generality, we set $d = 2$ in (10) and consider the model

$$V_t^{*(1)} = m_0 + m_1(V_{t-1}^{*(1)}) + m_2(V_{t-1}^{*(5)}) + \varepsilon_t. \tag{11}$$

Moreover, we impose the following assumptions:

- (C1) The errors ε_t have the form $\varepsilon_t = \sigma(V_{t-1}^{*(1)}, V_{t-1}^{*(5)})\xi_t$ with i.i.d. residuals ξ_t and some volatility function $\sigma(\cdot)$. For any t, ξ_t is independent of $\{(V_{t-j}^{*(1)}, V_{t-j}^{(1)}) : j \geq 1\}$.
- (C2) $\mathbb{E}[|u_{t-j}||V_{t-1}^{(1)}, V_{t-1}^{(5)}] \leq C < \infty$ for some constant C, $j = 0, \ldots, 5$, and all t.
- (C3) The derivatives m_1' and m_2' are absolutely bounded.

(C1) is a very common assumption in the literature on nonlinear AR models; see, e.g., Tjøstheim and Auestad (1994). (C2) and (C3) are technical conditions required to control the measurement error.

Model (11) can be rewritten in terms of realized volatility as follows. Using (9) together with a Taylor expansion, we obtain that

$$V_t^{(1)} = m_0 + m_1(V_{t-1}^{(1)}) + m_2(V_{t-1}^{(5)}) + W_t + \varepsilon_t,$$

where

$$W_t = -\frac{u_t}{n_t^{\beta}} + m_1'\left(\tilde{V}_{t-1}^{(1)}\right)\frac{u_{t-1}}{n_{t-1}^{\beta}} + m_2'\left(\tilde{V}_{t-1}^{(5)}\right)\frac{1}{5}\sum_{j=1}^{5}\frac{u_{t-j}}{n_{t-j}^{\beta}}$$

and $\tilde{V}_{t-j}^{(\iota)}$ denotes an intermediate point between $V_{t-j}^{(\iota)}$ and $V_{t-j}^{*(\iota)}$ for $\iota = 1, 5$. Hence,

$$\mathbb{E}\big[V_t^{(1)}\big|V_{t-1}^{(1)}, V_{t-1}^{(5)}\big] = m_0 + m_1(V_{t-1}^{(1)}) + m_2(V_{t-1}^{(5)}) + \rho_n(V_{t-1}^{(1)}, V_{t-1}^{(5)}), \tag{12}$$

where $\rho_n(V_{t-1}^{(1)}, V_{t-1}^{(5)}) = \mathbb{E}[W_t|V_{t-1}^{(1)}, V_{t-1}^{(5)}]$. According to (12), the additive structure of model (11) gets lost when it is expressed in terms of realized volatility, i.e., the regression function $m(x) = \mathbb{E}[V_t^{(1)}|V_{t-1}^{(1)} = x_1, V_{t-1}^{(5)} = x_2]$ does not have an additive form.

The SBE has the following property: If the true regression function m is not additive, it estimates the L_2-projection of m onto the space of additive functions. This projection converges to the additive function $m_{\text{add}}(x) = m_0 + m_1(x_1) + m_2(x_2)$ as the sample size increases: Under (C2) and (C3), the term $\rho_n(V_{t-1}^{(1)}, V_{t-1}^{(5)})$ can be bounded by

$$\big|\rho_n(V_{t-1}^{(1)}, V_{t-1}^{(5)})\big| \le \sum_{j=0}^{5}\frac{C'}{n_{t-j}^{\beta}}\mathbb{E}\big[|u_{t-j}|\big|V_{t-1}^{(1)}, V_{t-1}^{(5)}\big] \le \frac{C''}{n^{\beta}}, \tag{13}$$

where C' and C'' are sufficiently large constants and for simplicity, we set $n_t = n$ for all t, i.e., the number of intraday observations is the same at each day. Equations (12)–(13) imply that m approaches the additive function m_{add} as $n \to \infty$. In particular, the squared L_2-distance between m and m_{add} is bounded by

$$\mathbb{E}\Big[\big\{m(V_{t-1}^{(1)}, V_{t-1}^{(5)}) - m_{\text{add}}(V_{t-1}^{(1)}, V_{t-1}^{(5)})\big\}^2\Big] = \mathbb{E}\big[\rho_n^2(V_{t-1}^{(1)}, V_{t-1}^{(5)})\big] \le \frac{C}{n^{2\beta}} \tag{14}$$

with some sufficiently large constant C. Hence, m_{add} is asymptotically identical to the L_2-projection of m. As a result, the SBE converges to the additive function m_{add}, i.e., it consistently estimates the variance component functions m_1 and m_2 of model (11) despite the presence of measurement error.

The term $\rho_n(V_{t-1}^{(1)}, V_{t-1}^{(5)})$ can be interpreted as a bias induced by the measurement error. The convergence rate of the SBE heavily depends on how quickly this bias converges to zero, or put differently, on how fast n tends to infinity compared to T. If n grows so quickly that $((Tg)^{-1/2} + g^2)n^{\beta} \to \infty$, where g is the bandwidth of the SBE, there is no loss in terms of the rate. If n grows more slowly, the rate slows down. As a consequence, our methods will asymptotically not be affected by the measurement error as long as the bias $\rho_n(V_{t-1}^{(1)}, V_{t-1}^{(5)})$ converges to zero sufficiently fast. This finding parallels the results in Corradi et al. (2009) who estimate the predictive density of the true daily volatility from noisy RV measures. In small samples, the term $\rho_n(V_{t-1}^{(1)}, V_{t-1}^{(5)})$ may of course strongly bias the SBE estimator even though it washes out asymptotically. Ideally, we would thus like to refine our estimation methods and use techniques that explicitly correct for this bias. Asai et al. (2012) devise such a bias correction for the linear HAR model. In our nonparametric setting, it is however much more difficult and not entirely clear how to construct such a correction. For the time being, we thus advocate to proceed as described above.

4.3 Simulations

In order to get some insights into the size and power properties of our testing procedures in finite samples, we run simulations both under the null and alternative hypotheses. We define the following size and power functions: $\alpha_T(h) = P(S_T > s^*_\alpha | H_0)$ and $\beta_T(h) = P(S_T > s^*_\alpha | H_A)$, where H_0 and H_A denote the null and the alternative, respectively, and s^*_α is the stochastic approximate critical value at the α-level obtained from the bootstrap. The set-up we choose is motivated from our applications to RV data.

4.3.1 Specification tests

As a first case, we test for linearity of the function m_2 and consider the following models:

$$V_t^{(1)} = m_0 + m_1(V_{t-1}^{(1)}) + m_{2,i}(V_{t-1}^{(5)}) + m_3(V_{t-1}^{(22)}) + \varepsilon_t,$$

where $V_t^{(\iota_j)}$ is defined in Section 4.1; moreover, $m_0 = 0$, $m_1(x) = 0.3x$, $m_3(x) = 0.3x$, and $m_{2,i}(x) = (b_i x - c_i) \cdot I\{x < 0.25\} + a_i x \cdot I\{-0.25 \leq x \leq 0.25\} + (b_i x + c_i) \cdot I\{x > 0.25\}$, where $a_0 = 0.3$, $a_1 = 0.4$, $a_2 = 0.5$, $a_3 = 0.6$, $b_0 = 0.3$, $b_1 = 0.2$, $b_2 = 0.1$, $b_3 = 0.0$, and c_i are constants chosen such that the functions are continuous. The case $m_{2,0}$ corresponds to the null, whereas $m_{2,i}$ $(i = 1, 2, 3)$ are alternative hypotheses that diverge from the null by getting progressively steeper at the center and flatter in the tails. Since the kinked functions $m_{2,i}$ under the alternative do not satisfy the smoothness assumptions of the SBE, they are mollified by applying the smoothing operator $\mathcal{K}_b g(\cdot) = \int K_b(\cdot - u) g(u) du$ to them, where $K_b(u) = K^{qua}(u/b)/b$ with $K^{qua}(v) = 15/16(1 - v^2)^2 I(|v| \leq 1)$ and we set $b = 0.15$. For the disturbances ε_t, we assume (a) i.i.d. normal errors with mean zero and variance $\sigma_\varepsilon^2 = 0.35^2$; (b) heteroscedastic errors generated by $\varepsilon_t = \sigma(V_{t-1}^{(5)})\xi_t$, where ξ_t is i.i.d. standard normal, and $\sigma(V_{t-1}^{(5)}) = 0.15 \left\{ \arctan\left(4V_{t-1}^{(5)}\right) + \frac{\pi}{2} \right\} + 0.1$. This parametrization gives about the same unconditional variance as the homoscedastic case.[4]

The sample size is $T = 1000$, which corresponds to about 60% of the sample size for the specification tests carried out on the whole sample, the size of the bootstrap samples is $M = 1000$, and the tests are computed $N = 1000$ times to approximate the size and power functions. We construct the bootstrap error using residuals that are perturbed

[4]Both parametrizations are motivated by our empirical applications. In an earlier draft of the paper, we also simulated with GARCH(1,1) errors. Although GARCH errors do not conform with Assumptions (A4) and (B1) in the appendix, we have obtained very similar results.

with standard normal random variates, and for all tests, we use the uniform density as the weight function of the test statistic.

As regards the bandwidth selection, we simulate 50 samples under the null. On each of these samples, we run the plug-in bandwidth selector of Mammen and Park (2005) and take the average of these bandwidths. The plug-in rule selects the bandwidth that approximately minimizes the integrated mean squared error of the smooth backfitting estimates. Since the asymptotic bias expression of the Nadaraya-Watson SBE is very involved, we replace it by the bias expression of the local linear backfitting (Mammen et al.; 1999) in the plug-in rule. We obtain the undersmoothed bandwidths of the pilot estimates by dividing the plug-in bandwidths by the factor 1.4.[5] In the simulations, the bandwidths g_1 and g_3 (for m_1 and m_3) are kept fixed, but we vary the bandwidth of m_2 in a neighborhood of the bandwidth selected by the plug-in rule. Of course, for each pilot estimation step, this bandwidth is divided by the factor 1.4 as well. Because the plug-in rule picks very similar bandwidths under both error scenarios, we use the same ones in either case. In what follows, we do not only report the estimated bandwidths themselves, but in parentheses also the values they amount to when the data are normalized to the unit interval. This gives an idea of the effective size of the bandwidths.

In Table 2, we report the results of the simulations for three common significance levels; in addition, Fig. 1 provides size discrepancy plots and power curves on [0, 0.25], i.e., for the usual domain of interest. As can be seen in the upper left block of Table 2, but also in the upper left frame of Fig. 1, for homoscedastic errors, the actual size is held optimally for a bandwidth of 0.5 (about 0.25 on the unit interval), which is a bit smaller than the plug-in rule would have suggested (0.6, about 0.3 on the unit interval). For smaller test bandwidths, the test tends to reject overly, whereas it is undersized for larger ones. These findings are as reported, e.g., in Härdle and Mammen (1993, their Table 1). For heteroscedastic errors (upper right block of Table 2, upper right frame of Fig. 1), the optimal size is for bandwidths between 0.5 and 0.6. Moreover, the test is more sensitive to smaller bandwidths, but it holds the size much better for larger bandwidths. This is an important observation, for in our empirical applications, we will operate in a heteroscedastic environment. The power results given in the lower panels of Table 2 and Fig. 1 are acceptable as well. The power seems to be weak for the case of $m_{2,1}$, but this is only a tiny deviation from the null hypothesis. For the remaining cases, the power looks good, with the additional qualification that in the heteroscedastic error scenario

[5] This factor is obtained by the following heuristic argument. By our theory, $h \approx c_1 T^{-1/5}$ and $g \approx c_2 T^{-1/4}$. Unless the two constants c_1 and c_2 differ drastically, we have $h/g \approx 1.4$ for the sample sizes relevant for our simulations and empirical applications. The simulations that we provide here suggest that this heuristic works fine.

the power is a bit weaker than in the homoscedastic one. This appears natural because heteroscedastic errors may create artificial noisy structures on the estimates which make it more difficult for the test to detect the deviations from the null hypothesis.

4.3.2 Structural break tests

For the structural break test, we consider two cases. In setting (a), we consider the fully linear model

$$V_t^{(1)} = \begin{cases} m_0^{ante} + m_1^{ante}(V_{t-1}^{(1)}) + m_2^{ante}(V_{t-1}^{(5)}) + m_3^{ante}(V_{t-1}^{(22)}) + \varepsilon_t \ , & \text{for } t < t^*, \\ m_0^{post} + m_1^{post}(V_{t-1}^{(1)}) + m_{2,i}^{post}(V_{t-1}^{(5)}) + m_3^{post}(V_{t-1}^{(22)}) + \varepsilon_t \ , & \text{for } t \geq t^*, \end{cases}$$

where $m_0^\ell = 0$, $m_1^\ell(x) = 0.3x$, $m_3^\ell(x) = 0.3x$ for $\ell = \{ante, post\}$, $m_2^{ante}(x) = 0.3x$ and $m_{2,i}^{post}(x) = a_i x$ with $a_0 = 0.3$, $a_1 = 0.2$, $a_2 = 0.1$, $a_3 = 0.0$. For setting (b), we set $m_2^{ante}(x) = 0.15x \cdot I(x \leq 0) + 0.3xI(x > 0)$ and $m_{2,i}^{post}(x) = 0.15x \cdot I(x \leq 0) + a_i xI(x > 0)$ with $a_0 = 0.3$, $a_1 = 0.15$, $a_2 = 0.0$, $a_3 = -0.15$, with the kinks being mollified as described above. Thus, we consider a nonlinear function that only is changed in some parts. For both settings, a_0 corresponds to the null, and a_i $(i = 1, 2, 3)$ are alternative hypotheses that increasingly diverge from the null. We use the two error specifications as above, but we now have $T = 1800$, so that the ante and post samples are $T = 900$. This corresponds to the sample sizes of the empirical application. For the bandwidth selection, we proceed as discussed above, and the weight function of the test statistic is the uniform density. The break point is treated as known.

Tables 3 and 4 give the detailed results for three levels of significance; Fig. 2 and Fig. 3 provide the corresponding plots. For both settings, the test holds the size best at a bandwidth of about 0.6 (about 0.3 on the unit interval) which corresponds to about what the plug-in rule suggests for estimating these models. In comparison to the specification test, the best size accuracy is now achieved for a larger bandwidth. We suspect that the reason for this is as follows. The design density of the regressors may differ on the ante and the post samples; in particular, there may be regions where one density has considerably less mass than the other. In order to avoid poor and instable function fits in these regions, larger bandwidths are needed. Considering the simulation results under the alternative, we find good power properties, even for the less drastic deviations from the null hypothesis. In summary, the simulations suggest that the tests have good size and power properties.

4.4 Empirical applications

4.4.1 Data, realized variance measurement, and break date

The high-frequency data we use are intra-day equity index calculations and intra-day trades of futures on equity indices, fixed income instruments, currencies, and commodities; see Table 1 for an overview.[6] The data range is 2003–2010 for the tests and 2011–2013 for the forecasting exercise in Section 4.4.3. For the futures contracts, we use the most active front-contract as roll-over convention. The raw price data are cleaned as suggested in Barndorff-Nielsen et al. (2009). To compute intra-day log-returns, we construct an equidistant 5-minutes tick data series from the observed prices by means of the previous tick method (Andersen et al.; 2001).

For this work, we use a robust measure of realized variance. This is because for our nonparametric estimation technique, we would like to avoid outliers that might influence our inference. For this reason, we also exclude the overnight return from the analysis. Our measure of choice is the median RV (MedRV) estimator of Andersen et al. (2012). Let $\{r_{i,t}\}_{i=1}^{M}$, be a sample of intra-day returns observed at day t; then

$$\text{MedRV} = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{M}{M-2} \right) \sum_{i=2}^{M-1} \text{med}(r_{i-1,t}, r_{i,t}, r_{i+1,t})^2 \ .$$

As recommended in Andersen et al. (2012), we use the MedRV estimator with subsampling in order to reduce the effect of market microstructure noise.

Our sample covers the financial crisis which may have triggered a structural break. We therefore split the data set into a pre-crisis and a crisis sample and use our methods to test for a structural break. In Fig. 4, we plot the S&P 500 closing prices between 2003 and 2011 along with the spread of the London interbank offered rate (3 months, USD) over the overnight indexed swap (LOIS) which is a recognized measure of credit risk within the banking sector (Thornton; 2009). The spread is close to zero up to July/August 2007, after which it spikes up. In view of this graph, we set the break date on July 25, 2007, which is one of the last days on which the LOIS is reported below 10 basis points.

4.4.2 Empirical results

We model log RV[7] by means of the HAR model (8) using daily, weekly and monthly

[6] They are provided by *Tick Data*; see `http:\\www.tickdata.com`.

[7] This is a typical choice because it makes RV data approximately normal. One may, however, perceive this case as a particular Box-Cox transform with transformation parameter $\lambda = 0$. In an unreported analysis, we study the results of the specification tests for $\lambda \in [-0.25, 0.25]$. This interval cannot be made larger, because the SBE breaks down when the data are too unequally distributed. We find the results stable for variations in this neighborhood, except in some cases where the evidence is already weak for log RV.

variance components. We do not account for measurement error. The tests are implemented exactly as in the simulations except that we define the uniform density on the 2%-98% inner quantile range of the data to mitigate the influence of boundary effects. The bandwidths selected by the plug-in rule are reported in Table 5.

We start our discussion with the structural breaks in Table 6. Breaks are identified in all three component functions: in CF, TY, and US in the daily component functions; in CF, KM, XX, TY, and NG in the weekly component; and in XX, TY, US, NG, and SY in the monthly component. Thus, with the exception of natural gas and soybeans, most of the breaks are found for the equity and the fixed income instruments. This is a plausible finding because assets whose prices are predominantly determined by long-term global consumption perspectives may be less affected by a financial crisis.

In Table 7, we present the specification tests for the linear HAR model. The tests are carried out either on the pre-crisis and the crisis samples separately or on the full sample, depending on the outcome of the structural break test. The figures show that with the exception of three series (FT, GC, JY), there is at least one component function for which linearity is rejected. In most cases, the linearity assumption can be questioned for the daily variance function.

To give a better impression of the functional forms that we can document, we display the estimates for which linearity is rejected in Fig. 5. The estimates are normalized to a common support on the unit interval in order to make them visually better comparable. Most of the nonlinear estimates of m_1 exhibit mild convex forms (see top panel in Fig. 5). Thus, the marginal impact of lagged daily log RV on tomorrow's log RV is smaller in low volatility regimes and increases as volatility rises. According to the motivation of the HAR model, this means that at higher variance levels, daily trading activities drive future log RV more predominantly than in calm markets. Considering that crises times require more frequent hedging activities also by long-term investors, this appears to be a reasonable finding. See Table 15.3 in Corsi et al. (2012) for similar evidence. For the m_1 estimates of the two US fixed income futures (TY, US), this interpretation, however, does not hold.

For the estimates of the weekly and monthly variance functions, the picture is less coherent. As shown in the lower left panel of Fig. 5, three of the m_2-estimates (CF, SP, XX) become flatter for higher volatility levels; if anything, they display a gently concave shape. In contrast, the m_2 function of NG exhibits two plateaus. As regards m_3 in the lower right panel, two estimates show a convex shape for low variance levels (HG, CN), yet the estimates of KM, HG, and EC again flatten with higher variance levels. In terms of the HAR model, the flattening of the weekly and monthly variance functions at the highest variance levels implies that the marginal impact of weekly and monthly trading activities on future log RV diminishes. In some senses, this observation complements our interpretation of the convex daily variance function.

For a few cases, we finally present all function estimates of the nonparametric HAR model. In Fig. 6 and Fig. 7, we contrast the nonparametric fits against the linear model, which we display along with the pointwise asymptotic 95% confidence intervals of the nonparametric fits. The top panel of Fig. 6 shows the estimates of the S&P500 (SP) data, the lower panel those of the Euro-USD currency future (EC). In Fig. 7, two model fits are presented after accounting for structural breaks. We display the KOSPI 200 index (KM, top panel), for which a structural break is detected in the weekly variance function, and the 30yrs US-TBond (US, lower panel), which has a break in the daily and the monthly variance functions.

Summarizing, on the one hand, we find compelling statistical evidence for nonlinear variance component functions. This suggests that the linear HAR model is misspecified. On the other hand, visual inspection reveals that the actual deviations from linearity are moderate. This finding may explain why nonlinear extensions of the HAR model typically attain only slight improvements over the baseline model in terms of predictive ability.

4.4.3 Does the additive HAR model have any predictive value?

It could well be that the nonlinear forms we detect in our diagnostic analysis do not provide any out-of-sample value, because the deviation from linearity is too small or tied to the given sample period. To investigate the predictive ability of our model compared to the linear HAR model, we proceed as follows. We increase all bandwidths, which we obtain from the plug-in rule, by 15% and estimate the models with a local linear SBE. The larger bandwidths vis-à-vis the in-sample results are chosen to avoid overfitting, which the in-sample fits could suffer from. We do not use the Nadaraya-Watson SBE, because its estimates tend toward a constant the larger the bandwidths. This is an undesirable feature for prediction purposes. In contrast, the local linear SBE tends toward the linear model, which is our benchmark here. Taken to the extreme, for infinitely large bandwidths, we would even recover the linear HAR model and would obtain the same predictive ability.

Using the entire sample from January 2003 to December 2010, we estimate the linear and the nonlinear HAR model for the series for which no structural breaks are found.[8] The predictions are evaluated on data dating from 2011 to 2013 (about 730 sample days). For the weekly and monthly predictions, we use a direct modeling approach, i.e., in (8), we set $Y_t = V_{t+4}^{(5)}$ and $Y_t = V_{t+21}^{(22)}$, respectively. These aggregates are computed from log RV.

[8] The series with a structural break drop out, because the ranges of the functions estimated on the post-samples are too small for the predictions to be computed; the same applies to the series CL and HG. For some dates, this problem still occurs for the series under investigation; we then linearly extrapolate the estimated nonparametric functions.

We determine new bandwidths and re-estimate the models.

In Table 8, we display the root mean squared error (RMSE, top panel) and the mean absolute error (MAE, middle panel) of the exercise. In about two-thirds of all tests, the forecast error of the nonlinear model is smaller than that of the linear model. The lowest panel of Table 8 shows the p-values of the test of superior predictive ability of Hansen (2005). We employ the studentized version of the test and block-bootstrap its distribution using 1000 draws and a block size of twelve. About one third of the 30 tests are statistically significant. Given the difficulties to beat the linear HAR with nonlinear approaches, see, e.g., McAleer and Medeiros (2008), we read these results as encouraging evidence for our nonlinear modeling approach. Clearly, more ample investigations and comparisons with other nonlinear modeling approaches are necessary to fully ascertain the benefits.

A Appendix

In this appendix, we prove the results concerning the test on parametric specification from Section 2. Throughout, the symbol C is used to denote a universal real constant that may take a different value on each occurrence. Without loss of generality, we consider the case $d = 2$, i.e., we work with the model

$$Y_t = m_0 + m_1(X_{t,1}) + m_2(X_{t,2}) + \varepsilon_t.$$

We make the following assumptions:

(A1) The process $\{(\boldsymbol{X}_t, \varepsilon_t)\}$ is strictly stationary and strongly mixing with mixing coefficients α satisfying $\alpha(k) \leq a^k$ for some $0 < a < 1$.

(A2) The variables $\boldsymbol{X}_t = (X_{t,1}, X_{t,2})$ have compact support, w.l.o.g. $[0,1]^2$. The density p of $\boldsymbol{X}_t$ and the densities $p_{(0,l)}$ of $(\boldsymbol{X}_t, \boldsymbol{X}_{t+l})$, $l = 1, 2, \ldots$, are uniformly bounded. Furthermore, p is bounded away from zero on $[0,1]^2$.

(A3) The functions m_1 and m_2 are twice continuously differentiable. The second derivatives are Lipschitz continuous of order β for some small $\beta > 0$, i.e. $|m_i''(u) - m_i''(v)| \leq C|u - v|^\beta$ for $i = 1, 2$. Moreover, p is twice continuously differentiable.

(A4) The residuals are of the form $\varepsilon_t = \sigma(\boldsymbol{X}_t)\xi_t$. Here, σ is a Lipschitz continuous function and $\{\xi_t\}$ is an i.i.d. process having the property that ξ_t is independent of $\boldsymbol{X}_s$ for $s \leq t$. The variables ξ_t satisfy $\mathbb{E}[\xi_t^{4+\delta}] < \infty$ for some small $\delta > 0$ and are normalized such that $\mathbb{E}[\xi_t^2] = 1$.

(A5) There exists a real constant C and a natural number l^* such that $\mathbb{E}[|\xi_t||\boldsymbol{X}_t, \boldsymbol{X}_{t+l}] \leq C$ for all $l \geq l^*$.

(A6) The kernel K is bounded, symmetric about zero and has compact support $[-C_1, C_1]$ for some $C_1 > 0$. Moreover, it fulfills the Lipschitz condition that $|K(u) - K(v)| \leq L|u - v|$ for some $L > 0$.

(A7) The bandwidth g is of the order $T^{-(1/4+\delta)}$ for some small $\delta > 0$ and h is such that $g \ll h \ll T^{-2/11}$, where $a_T \ll b_T$ means that $a_T/b_T \to 0$.

(A8) It holds that

$$m_{\theta_0}(w) - m_{\hat{\theta}}(w) = \frac{1}{T}\sum_{t=1}^{T}\langle q(w), r(X_{t,1})\rangle\tilde{\varepsilon}_t + o_p\Big(\sqrt{\frac{g}{T(\log T)h}}\Big)$$

uniformly in w, where θ_0 is defined in (3), $\tilde{\varepsilon}_t = \varepsilon_t + (m_2(X_{t,2}) - \tilde{m}_2(X_{t,2}))$ and q and r are bounded functions taking values in $\mathbb{R}^k$ for some k. Here, $\langle\cdot,\cdot\rangle$ denotes the usual Euclidean inner product for vectors.

For the results on the wild bootstrap procedure, we replace (A8) by an analogous assumption in the bootstrap world.

(A8*) Let $\hat{\theta}^*$ be the parameter estimate based on the bootstrap sample $\{(Y_t^*, \boldsymbol{X}_t)\}$. It holds that uniformly in w,

$$m_{\hat{\theta}}(w) - m_{\hat{\theta}^*}(w) = \frac{1}{T}\sum_{t=1}^{T}\langle q^*(w), r^*(X_{t,1})\rangle\varepsilon_t^* + o_p\Big(\sqrt{\frac{g}{T(\log T)h}}\Big),$$

where q^* and r^* are bounded functions taking values in $\mathbb{R}^k$ for some k.

Note that we do not necessarily require exponentially decaying mixing rates as assumed in (A1). These could alternatively be replaced by sufficiently high polynomial rates at the cost of a more involved notation. It is also possible to relax (A2) and to allow for unbounded support of $\boldsymbol{X}_t$. In this case, however, we have to restrict our test to a compact subset of the potentially unbounded support. In particular, let $A = A_1 \times A_2$ be a compact subset of $\mathbb{R}^2$ contained in the support of $\boldsymbol{X}_t$ and suppose we want to test m_1 for parametric specification on the compact set A_1, i.e., we want to test the hypothesis $H_0^{(A_1)}$ that $m_1 : A_1 \to \mathbb{R}$ has a given parametric form. To do so, we have to modify the smoother $\hat{m}$ and the pilot estimators of the backfitting algorithm. Specifically, we replace $\hat{m}$ by

$$\frac{\sum_{t=1}^{T} I(\boldsymbol{X}_t \in A)K_h(w - X_{t,1})\tilde{Z}_t}{\sum_{t=1}^{T} I(\boldsymbol{X}_t \in A)K_h(w - X_{t,1})}$$

and modify the pilot estimates of the backfitting procedure in an analogous way; see Section 5 in Mammen et al. (1999) who work with the same modification. Rewriting the test statistic in terms of these modified estimators allows one to test $H_0^{(A_1)}$. (A1)–(A3) are standard conditions in the smooth backfitting literature (Mammen et al.; 1999). (A4) is a quite common assumption in the literature on kernel-based nonparametric tests; see, e.g., Fan and Li (1999) or Li (1999). It imposes a martingale difference structure on the residuals, which is needed to cope with the time series dependence of the model variables when deriving the limit distribution of the test statistic. (A5) is required to derive the uniform convergence rates of the Nadaraya-Watson estimators that enter the smooth backfitting procedure as pilot smoothers. Finally, condition (A8) is fulfilled, e.g., for weighted least squares estimators in linear models and under appropriate smoothness conditions for weighted least squares estimators in nonlinear settings; see Härdle and Mammen (1993) for details.

Before we come to the proof of Theorems 1 and 2, we list some properties of the backfitting estimators $\tilde{m}_1$ and $\tilde{m}_2$. For technical reasons, we undersmooth them by choosing the bandwidth g to be of the order $O(T^{-(1/4+\delta)})$ for some small $\delta > 0$. Moreover, we decompose $\tilde{m}_i$ $(i = 1, 2)$ into a stochastic part $\tilde{m}_i^A$ and a bias part $\tilde{m}_i^B$ according to $\tilde{m}_i(x_i) = \tilde{m}_i^A(x_i) + \tilde{m}_i^B(x_i)$. The two components are defined by

$$\tilde{m}_i^S(x_i) = \tilde{m}_i^{S,NW}(x_i) - \sum_{k \neq i} \int_0^1 \tilde{m}_k^S(x_k) \frac{\tilde{p}(x_k, x_i)}{\tilde{p}_i(x_i)} \, dx_k - \tilde{m}_0^S \tag{15}$$

for $S = A, B$. Here, $\tilde{p}$ is a kernel density estimator of the joint density of $\boldsymbol{X}_t = (X_{t,1}, X_{t,2})$ and $\tilde{p}_i$ is a kernel estimator of the marginal density p_i of $X_{t,i}$. Moreover, $\tilde{m}_i^{A,NW}$ and $\tilde{m}_i^{B,NW}$ denote the stochastic and the bias part of a Nadaraya-Watson estimator,

$$\tilde{m}_i^{A,NW}(x_i) = \frac{\sum_{t=1}^T K_g(w, X_{t,i}) \varepsilon_t}{\sum_{t=1}^T K_g(w, X_{t,i})} \tag{16}$$

$$\tilde{m}_i^{B,NW}(x_i) = \frac{\sum_{t=1}^T K_g(w, X_{t,i})[m_0 + m_1(X_{t,1}) + m_2(X_{t,2})]}{\sum_{t=1}^T K_g(w, X_{t,i})}. \tag{17}$$

Finally we let $\tilde{m}_0^A = \frac{1}{T} \sum_{t=1}^T \varepsilon_t$ and $\tilde{m}_0^B = \frac{1}{T} \sum_{t=1}^T \{m_0 + m_1(X_{t,1}) + m_2(X_{t,2})\}$. Under the assumptions from above, the stochastic part $\tilde{m}_i^A$ has the expansion

$$\tilde{m}_i^A(w) = \tilde{m}_i^{A,NW}(w) + \frac{1}{T} \sum_{t=1}^T r_{t,i}(w) \varepsilon_t + o_p(T^{-1/2}) \tag{18}$$

uniformly for $w \in [0, 1]$. Here, $r_{t,i}(\cdot) = r_i(\boldsymbol{X}_t, \cdot)$ are random functions that are absolutely uniformly bounded and fulfill the Lipschitz condition $|r_{t,i}(w) - r_{t,i}(w')| \leq C|w - w'|$. The expansion (18) has been derived in Mammen and Park (2005) in an i.i.d. setup. The proving strategy can however be easily extended to our stationary mixing framework.

We provide the details at the end of this appendix after the proof of Theorems 1 and 2. For the bias part $\tilde{m}_i^B$, we have the following uniform convergence result: Let $I_h = [2C_1 g, 1-2C_1 g]$ and $I_h^c = [0,1] \setminus I_h$ be the interior and the boundary region of the support of $X_{t,i}$, respectively. Then

$$\sup_{w \in I_h} |m_i(w) - \tilde{m}_i^B(w)| = O_p(g^2) \tag{19}$$

$$\sup_{w \in I_h^c} |m_i(w) - \tilde{m}_i^B(w)| = O_p(g). \tag{20}$$

This can be shown following the lines of the proof for Theorem 4 in Mammen et al. (1999).

Proof of Theorem 1

Let $m_1(\cdot) = m_{\theta_0}(\cdot) + c_T \Delta(\cdot)$ with $c_T = T^{-1/2} h^{-1/4}$ and denote by p_1 the marginal density of $X_{t,1}$. To shorten notation, we set $m_0 = 0$, i.e., we drop the model constant. Moreover, without loss of generality, we let $\pi(w) = I(w \in [0,1])$ and write $\int = \int_0^1$ for short. Straightforward calculations yield that

$$S_T = T h^{1/2} \int \big(U_{T,1}(w) + \ldots + U_{T,5}(w) \big)^2 dw + o_p(1)$$

with

$$U_{T,1}(w) = \frac{1}{T} \sum_{t=1}^T K_h(w - X_{t,1}) c_T \Delta(X_{t,1}) / p_1(w)$$

$$U_{T,2}(w) = \frac{1}{T} \sum_{t=1}^T K_h(w - X_{t,1}) \varepsilon_t / p_1(w)$$

$$U_{T,3}(w) = \frac{1}{T} \sum_{t=1}^T K_h(w - X_{t,1}) \big(m_2(X_{t,2}) - \tilde{m}_2(X_{t,2}) \big) / p_1(w)$$

$$U_{T,4}(w) = \frac{1}{T} \sum_{t=1}^T K_h(w - X_{t,1}) \Big(\frac{1}{T} \sum_{s=1}^T \langle q(X_{t,1}), r(X_{s,1}) \rangle \varepsilon_s \Big) / p_1(w)$$

$$U_{T,5}(w) = \frac{1}{T} \sum_{t=1}^T K_h(w - X_{t,1}) \Big(\frac{1}{T} \sum_{s=1}^T \langle q(X_{t,1}), r(X_{s,1}) \rangle \big(m_2(X_{s,2}) - \tilde{m}_2(X_{s,2}) \big) \Big) / p_1(w).$$

The two terms $U_{T,3}(w)$ and $U_{T,5}(w)$ capture the estimation error resulting from approximating the function m_2 by $\tilde{m}_2$. They can thus be regarded as measuring the difference between our test statistic and the statistic of the oracle case where the function m_2 is known. In what follows, we show that $U_{T,3}(w)$ and $U_{T,5}(w)$ are asymptotically negligible in the sense that

$$T h^{1/2} \int U_{T,j}(w) U_{T,3}(w) dw = o_p(1) \tag{21}$$

$$T h^{1/2} \int U_{T,j}(w) U_{T,5}(w) dw = o_p(1) \tag{22}$$

for all $j = 1, \dots, 5$. We thus arrive at

$$S_T = Th^{1/2} \int \big(U_{T,1}(w) + U_{T,2}(w) + U_{T,4}(w)\big)^2 dw + o_p(1) =: S_T' + o_p(1) \tag{23}$$

with S_T' basically being the statistic of the oracle case. (23) shows that our statistic S_T has the same limit distribution as that of the oracle case.

To complete the proof, we need to derive the asymptotic distribution of S_T'. The latter has exactly the same structure as the statistic from Proposition 1 in Härdle and Mammen (1993). Even though Härdle and Mammen derive their results in an i.i.d. setting, their proving strategy easily carries over to our mixing setup. We need only make some minor adjustments. Most importantly, we cannot apply a central limit theorem for quadratic forms of i.i.d. variables as they do. Nevertheless, assumption (A4) on the error terms allows us to work with a central limit theorem for martingale differences instead (e.g. with Theorem 1 in Chapter 8 of Pollard (1984)). On this basis we can proceed along the lines of their arguments to complete the proof. The details are omitted. □

Proof of (21) and (22). We limit our attention to the proof of (21), the arguments for (22) being fully analogous. Using the uniform expansion (18) for the stochastic part of the backfitting estimator $\tilde{m}_2$, we can write $U_{T,3}(w) = U_{T,3}^B(w) - U_{T,3}^{A,NW}(w) - U_{T,3}^{A,SBE}(w) + U_{T,3}^R(w)$, where the remainder term $U_{T,3}^R(w)$ is of the order $o_p(T^{-1/2})$ and

$$\begin{aligned}
U_{T,3}^{A,NW}(w) &= \frac{1}{T}\sum_{t=1}^T K_h(w - X_{t,1})\Big(\frac{1}{T}\sum_{s=1}^T \frac{K_g(X_{t,2}, X_{s,2})}{\frac{1}{T}\sum_{v=1}^T K_g(X_{t,2}, X_{v,2})}\varepsilon_s\Big)/p_1(w) \\
U_{T,3}^{A,SBE}(w) &= \frac{1}{T}\sum_{t=1}^T K_h(w - X_{t,1})\Big(\frac{1}{T}\sum_{s=1}^T r_{s,2}(X_{t,2})\varepsilon_s\Big)/p_1(w) \\
U_{T,3}^B(w) &= \frac{1}{T}\sum_{t=1}^T K_h(w - X_{t,1})\big(m_2(X_{t,2}) - \tilde{m}_2^B(X_{t,2})\big)/p_1(w).
\end{aligned}$$

We now show that for $j = 1, \dots, 5$,

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{A,NW}(w) dw = o_p(1) \tag{24}$$

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{A,SBE}(w) dw = o_p(1) \tag{25}$$

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{B}(w) dw = o_p(1). \tag{26}$$

The arguments for these three claims also imply that $Th^{1/2}\int U_{T,j}(w)U_{T,3}^R(w)dw = o_p(1)$, thus completing the proof of (21).

We start with the proof of (24) which consists of several steps. In the first step, we show that

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{A,NW}(w) dw = W_{T,j} + o_p(1) \tag{27}$$

with

$$W_{T,j} = Th^{1/2} \int U_{T,j}(w) \frac{1}{T} \sum_{t=1}^{T} \frac{K_h(w - X_{t,1})}{p_1(w)} \Big(\frac{1}{T} \sum_{s=1}^{T} \frac{K_g(X_{t,2}, X_{s,2})}{\kappa(X_{t,2})} \varepsilon_s \Big) dw$$

and $\kappa(u) = \mathbb{E}[K_g(u, X_{0,2})]$. We thus replace the sum $\frac{1}{T}\sum_{v=1}^{T} K_g(X_{t,2}, X_{v,2})$ in $U_{T,3}^{A,NW}$ by the moment $\kappa(X_{t,2})$ and show that the resulting error is asymptotically negligible. To do so, write $\frac{1}{T}\sum_{v=1}^{T} K_g(u, X_{v,2}) = \kappa(u) + R(u)$ with $R(u) = \frac{1}{T}\sum_{v=1}^{T}(K_g(u, X_{v,2}) - \mathbb{E}[K_g(u, X_{v,2})])$. As $\sup_{u\in[0,1]} |R(u)| = O_p(\sqrt{\log T/Tg})$, it holds that

$$\Big(\frac{1}{T} \sum_{v=1}^{T} K_g(u, X_{v,2}) \Big)^{-1} = \frac{1}{\kappa(u)} \Big(1 + \frac{R(u)}{\kappa(u)} \Big)^{-1} = \frac{1}{\kappa(u)} \Big(1 - \frac{R(u)}{\kappa(u)} + O_p\Big(\frac{\log T}{Tg} \Big) \Big)$$

uniformly in u. Plugging this into the term $U_{T,3}^{A,NW}(w)$, we easily arrive at (27).

In the next step, we split up $W_{T,j}$ into a leading term and a remainder which is asymptotically negligible. In particular, letting $\mathbb{E}_t[\,\cdot\,]$ denote the expectation with respect to the variables indexed by t, we show that

$$W_{T,j} = Th^{1/2} \int \frac{U_{T,j}(w)}{p_1(w)} \Big(\frac{1}{T} \sum_{s=1}^{T} \mathbb{E}_0 \Big[\frac{K_h(w - X_{0,1}) K_g(X_{0,2}, X_{s,2})}{\kappa(X_{0,2})} \Big] \varepsilon_s \Big) dw + R_{T,j}, \quad (28)$$

where the remainder term $R_{T,j}$ is given by

$$R_{T,j} = Th^{1/2} \int \frac{U_{T,j}(w)}{p_1(w)} \Big\{ \frac{1}{T^2} \sum_{s,t=1}^{T} \psi_{t,s}(w) \varepsilon_s \Big\} dw$$

with

$$\psi_{t,s}(w) = \frac{K_h(w - X_{t,1}) K_g(X_{t,2}, X_{s,2})}{\kappa(X_{t,2})} - \mathbb{E}_t \Big[\frac{K_h(w - X_{t,1}) K_g(X_{t,2}, X_{s,2})}{\kappa(X_{t,2})} \Big]$$

and satisfies $R_{T,j} = o_p(1)$. (28) can be seen as follows: To start with, apply the Cauchy-Schwarz inequality to obtain that $|R_{T,j}| \le C(\int U_{T,j}(w)^2 dw)^{1/2} \cdot Q_T^{1/2}$, where

$$Q_T = \int \Big\{ \frac{h^{1/2}}{T} \sum_{s,t=1}^{T} \psi_{t,s}(w) \varepsilon_s \Big\}^2 dw.$$

Below, we show that $Q_T^{1/2} = O_p(a_T)$ with $a_T = \kappa_T (\log T) g^{-3/4}$, where κ_T slowly diverges to infinity, e.g., $\kappa_T = \log\log T$. As $(\int U_{T,j}(w)^2 dw)^{1/2} = O_p(g)$ for all $j = 1, \dots, 5$, this immediately implies that $R_{T,j} = o_p(1)$.

Our strategy to verify that $Q_T^{1/2} = O_p(a_T)$ is to exploit the second moment structure of the term $Q_T^{1/2}$. More specifically, let M be a positive constant. Then by Chebychev's inequality,

$$\mathbb{P}\big(|Q_T^{1/2}| > Ma_T\big) \le \frac{\mathbb{E}[Q_T]}{(Ma_T)^2} = \frac{h}{(MTa_T)^2} \sum_{s,s',t,t'=1}^{T} \int \mathbb{E}\big[\psi_{t,s}(w)\psi_{t',s'}(w)\varepsilon_s\varepsilon_{s'}\big] dw.$$

We now write

$$
\begin{aligned}
&\frac{h}{(Ta_T)^2} \sum_{s,s',t,t'=1}^{T} \int \mathbb{E}\big[\psi_{t,s}(w)\psi_{t',s'}(w)\varepsilon_s\varepsilon_{s'}\big] dw \\
&= \frac{h}{(Ta_T)^2} \sum_{(s,s',t,t')\in\Gamma} \int \mathbb{E}\big[\psi_{t,s}(w)\psi_{t',s'}(w)\varepsilon_s\varepsilon_{s'}\big] dw \\
&\quad + \frac{h}{(Ta_T)^2} \sum_{(s,s',t,t')\in\Gamma^c} \int \mathbb{E}\big[\psi_{t,s}(w)\psi_{t',s'}(w)\varepsilon_s\varepsilon_{s'}\big] dw \\
&=: E_\Gamma + E_{\Gamma^c}.
\end{aligned}
$$

Here, Γ is the set of tuples (s,s',t,t') with $1 \le s,s',t,t' \le T$ such that (at least) one index is separated from the others and Γ^c is its complement. We say that an index, for instance t, is separated from the others if $\min\{|t-t'|,|t-s|,|t-s'|\} > C_2 \log T$, i.e., if it is further away from the other indices than $C_2 \log T$ for a constant C_2 to be specified later.

We now analyze E_Γ and E_{Γ^c} separately. By definition, the set Γ^c contains all index tuples (s,s',t,t') such that no index is separated. Hence, the number of tuples contained in Γ^c is smaller than $C(T\log T)^2$ for some sufficiently large constant C. This together with some straightforward calculations yields that $E_{\Gamma^c} \le C/\kappa_T^2 \to 0$. We next turn to E_Γ. Writing Γ as the union of the disjoint sets

$$
\begin{aligned}
\Gamma_1 &= \{(s,s',t,t') \in \Gamma \mid \text{the index } t \text{ is separated}\} \\
\Gamma_2 &= \{(s,s',t,t') \in \Gamma \mid (s,s',t,t') \notin \Gamma_1 \text{ and the index } s \text{ is separated}\} \\
\Gamma_3 &= \{(s,s',t,t') \in \Gamma \mid (s,s',t,t') \notin \Gamma_1 \cup \Gamma_2 \text{ and the index } t' \text{ is separated}\} \\
\Gamma_4 &= \{(s,s',t,t') \in \Gamma \mid (s,s',t,t') \notin \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \text{ and the index } s' \text{ is separated}\},
\end{aligned}
$$

we get that $E_\Gamma = E_{\Gamma_1} + E_{\Gamma_2} + E_{\Gamma_3} + E_{\Gamma_4}$ with

$$
E_{\Gamma_r} = \frac{h}{(Ta_T)^2} \sum_{(s,s',t,t')\in\Gamma_r} \int \mathbb{E}\big[\psi_{t,s}(w)\psi_{t',s'}(w)\varepsilon_s\varepsilon_{s'}\big] dw
$$

for $r = 1,\dots,4$. In what follows, we show that $E_{\Gamma_r} = o(1)$ for $r = 1,\dots,4$. Since the proof is exactly the same for $r = 1,\dots,4$, we focus attention on the term E_{Γ_1}. Let $\{I_n\}_{n=1}^{N_T}$ be a cover of the compact support $[0,1]$ of $X_{t,2}$. The elements I_n are intervals of length $1/N_T$ given by $I_n = [\frac{n-1}{N_T}, \frac{n}{N_T})$ for $n = 1,\dots,N_T-1$ and $I_{N_T} = [1-\frac{1}{N_T}, 1]$. The midpoint of the interval I_n is denoted by u_n. With this, we can write

$$
K_g(X_{t,2}, X_{s,2}) = \sum_{n=1}^{N_T} I(X_{s,2} \in I_n)\big[K_g(X_{t,2}, u_n) + (K_g(X_{t,2}, X_{s,2}) - K_g(X_{t,2}, u_n))\big]
$$

and thus $\psi_{t,s}(w) = \psi^A_{t,s}(w) + \psi^B_{t,s}(w)$ with

$$\begin{aligned}
\psi^A_{t,s}(w) &= \sum_{n=1}^{N_T} \Big\{ \frac{K_h(w - X_{t,1}) K_g(X_{t,2}, u_n)}{\kappa(X_{t,2})} - \mathbb{E}_t\Big[\frac{K_h(w - X_{t,1}) K_g(X_{t,2}, u_n)}{\kappa(X_{t,2})}\Big]\Big\} I(X_{s,2} \in I_n) \\
\psi^B_{t,s}(w) &= \sum_{n=1}^{N_T} \Big\{ \frac{K_h(w - X_{t,1})(K_g(X_{t,2}, X_{s,2}) - K_g(X_{t,2}, u_n))}{\kappa(X_{t,2})} \\
&\qquad - \mathbb{E}_t\Big[\frac{K_h(w - X_{t,1})(K_g(X_{t,2}, X_{s,2}) - K_g(X_{t,2}, u_n))}{\kappa(X_{t,2})}\Big]\Big\} I(X_{s,2} \in I_n).
\end{aligned}$$

Inserting this into the expression for E_{Γ_1}, we obtain $E_{\Gamma_1} = E^A_{\Gamma_1} + E^B_{\Gamma_1}$ with

$$\begin{aligned}
E^A_{\Gamma_1} &= \frac{h}{(Ta_T)^2} \sum_{(s,s',t,t') \in \Gamma_1} \int \mathbb{E}\big[\psi^A_{t,s}(w)\varepsilon_s \psi_{t',s'}(w)\varepsilon_{s'}\big] dw \\
E^B_{\Gamma_1} &= \frac{h}{(Ta_T)^2} \sum_{(s,s',t,t') \in \Gamma_1} \int \mathbb{E}\big[\psi^B_{t,s}(w)\varepsilon_s \psi_{t',s'}(w)\varepsilon_{s'}\big] dw.
\end{aligned}$$

We first consider $E^B_{\Gamma_1}$: The Lipschitz continuity of the kernel K yields that $|K_g(X_{t,2}, X_{s,2}) - K_g(X_{t,2}, u_n)| \le \frac{C}{g^2}|X_{s,2} - u_n|$, which in turn gives that $|\psi^B_{t,s}(w)| \le \frac{C}{hg^2 N_T}$. Plugging this into the expression for $E^B_{\Gamma_1}$ and letting N_T grow at a sufficiently fast rate, we obtain that $|E^B_{\Gamma_1}| = o(1)$. To deal with $E^A_{\Gamma_1}$, we write

$$E^A_{\Gamma_1} = \frac{h}{(Ta_T)^2} \sum_{(s,s',t,t') \in \Gamma_1} \Big(\sum_{n=1}^{N_T} \int \gamma_n(w) dw \Big)$$

with

$$\begin{aligned}
\gamma_n(w) = \mathbb{E}\Big[\Big\{ \frac{K_h(w - X_{t,1}) K_g(X_{t,2}, u_n)}{\kappa(X_{t,2})} - \mathbb{E}_t\Big[\frac{K_h(w - X_{t,1}) K_g(X_{t,2}, u_n)}{\kappa(X_{t,2})}\Big]\Big\} \\
\times I(X_{s,2} \in I_n)\varepsilon_s \psi_{t',s'}(w)\varepsilon_{s'}\Big].
\end{aligned}$$

By Davydov's inequality,

$$\begin{aligned}
\gamma_n(w) &= \mathrm{Cov}\Big(\frac{K_h(w - X_{t,1}) K_g(X_{t,2}, u_n)}{\kappa(X_{t,2})} - \mathbb{E}_t\Big[\frac{K_h(w - X_{t,1}) K_g(X_{t,2}, u_n)}{\kappa(X_{t,2})}\Big], \\
&\qquad I(X_{s,2} \in I_n)\varepsilon_s \psi_{t',s'}(w)\varepsilon_{s'}\Big) \\
&\le \frac{C}{(gh)^2}\big(\alpha(C_2 \log T)\big)^{1-\frac{2}{r}} \le \frac{C}{(gh)^2}\big(a^{C_2 \log T}\big)^{1-\frac{2}{r}} \le \frac{C}{(gh)^2} T^{-C_3}
\end{aligned}$$

with some $C_3 > 0$, where r is chosen slightly larger than 2. Note that we can make C_3 arbitrarily large by choosing C_2 sufficiently large. From this, it easily follows that $E^A_{\Gamma_1} = o(1)$. Putting everything together yields that $Q_T = O_p(a_T)$, which in turn shows that $R_{T,j} = o_p(1)$.

Thus far, we have shown that equation (28) holds with $R_{T,j} = o_p(1)$, i.e.,

$$\Big| Th^{1/2} \int U_{T,j}(w) U_{T,3}^{A,NW}(w) dw \Big|$$

$$= Th^{1/2} \int \frac{U_{T,j}(w)}{p_1(w)} \Big(\frac{1}{T} \sum_{s=1}^{T} \mathbb{E}_0 \Big[\frac{K_h(w - X_{0,1}) K_g(X_{0,2}, X_{s,2})}{\kappa(X_{0,2})} \Big] \varepsilon_s \Big) dw + o_p(1). \tag{29}$$

It is now straightforward to obtain (24) for $j = 1, 2, 4$. To get (24) for $j = 3, 5$, we repeat the arguments from above to simplify the expressions $U_{T,3}(w)$ and $U_{T,5}(w)$ which show up in (29) for $j = 3, 5$. Once this has been done, (24) easily follows for $j = 3, 5$ as well. $\square$

Proof of Theorem 2

The proof has the same structure as the proof of Theorem 1. By arguments analogous to those above, we can replace the estimator $\tilde{m}_2$ by the true function m_2 in the bootstrap statistic and show that the resulting error is asymptotically negligible. Once this has been done, the proof follows the line of the arguments in Härdle and Mammen (1993). $\square$

Proof of (18)

For the proof, we outline the arguments needed to extend Theorem 6.1 of Mammen and Park (2005) which is in the context of i.i.d. data. For an additive function $g(x) = g_1(x_1) + g_2(x_2)$, let

$$\tilde{\psi}_1 g(x) = g_1^*(x_1) + g_2(x_2)$$

with

$$g_1^*(x_1) = -\int_0^1 g_2(x_2) \frac{\tilde{p}(x_1, x_2)}{\tilde{p}_1(x_1)} dx_2 + \sum_{k=1}^{2} \int_0^1 g_k(x_k) \tilde{p}_k(x_k) dx_k$$

and define $\hat{\psi}_2 g(x)$ analogously. Using standard uniform convergence results for kernel estimators and exploiting our model assumptions, we can show that Lemma 3 in Mammen et al. (1999) applies in our case. For $\tilde{m}^A(x) = \tilde{m}_1^A(x_1) + \tilde{m}_2^A(x_2)$, we therefore have the expansion

$$\tilde{m}^A(x) = \sum_{r=0}^{\infty} \tilde{S}^r \tilde{\tau}(x),$$

where $\tilde{S} = \tilde{\psi}_2 \tilde{\psi}_1$ and $\tilde{\tau}(x) = \tilde{\psi}_2 [\tilde{m}_1^{A,NW}(x_1) - \tilde{m}_{0,1}^{A,NW}] + [\tilde{m}_2^{A,NW}(x_2) - \tilde{m}_{0,2}^{A,NW}]$ with $\tilde{m}_{0,i}^{A,NW} = \int_0^1 \tilde{m}_i^{A,NW}(x_i) \tilde{p}_i(x_i) dx_i$. Now decompose $\tilde{m}^A(x)$ according to

$$\tilde{m}^A(x) = \tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW} + \sum_{r=0}^{\infty} \tilde{S}^r \big(\tilde{\tau}(x) - (\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW}) \big)$$
$$+ \sum_{r=1}^{\infty} \tilde{S}^r \big(\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW} \big)$$

with $\tilde{m}^{A,NW}(x) = \tilde{m}_1^{A,NW}(x_1) + \tilde{m}_2^{A,NW}(x_2)$ and $\tilde{m}_0^{A,NW} = \tilde{m}_{0,1}^{A,NW} + \tilde{m}_{0,2}^{A,NW}$. We show that there exist absolutely bounded functions $a_t(x)$ with $|a_t(x) - a_t(y)| \leq C\|x - y\|$ for a constant C s.t.

$$\sum_{r=1}^{\infty} \tilde{S}^r(\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW}) = \frac{1}{T}\sum_{t=1}^{T} a_t(x)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{30}$$

uniformly in x. A similar claim holds for the term $\sum_{r=0}^{\infty} \tilde{S}^r(\tilde{\tau}(x) - (\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW}))$. From this, (18) easily follows.

The idea behind the proof of (30) is as follows: From the definition of the operators $\tilde{\psi}_i$, it can be seen that

$$\tilde{S}(\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW}) = \tilde{\psi}_2 S_{1,2}(x_1) \tag{31}$$

with

$$S_{1,2}(x_1) = -\int_0^1 \frac{\tilde{p}(x_1, x_2)}{\tilde{p}_1(x_1)}(\tilde{m}_2^{A,NW}(x_2) - \tilde{m}_{0,2}^{A,NW})dx_2.$$

In what follows, we show that $S_{1,2}(x_1)$ has the representation

$$S_{1,2}(x_1) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{p(x_1, X_{t,2})}{p_1(x_1)p_2(X_{t,2})} - 1\Big)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{32}$$

uniformly in x_1. Thus, it essentially has the desired form $\frac{1}{T}\sum_t w_{t,2}(x_1)\varepsilon_t$ with some weights $w_{t,2}$. This allows us to infer that

$$\tilde{S}(\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW}) = \frac{1}{T}\sum_{t=1}^{T} b_t(x)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{33}$$

uniformly in x with some absolutely bounded functions b_t satisfying $|b_t(x) - b_t(y)| \leq C\|x - y\|$ for some $C > 0$. Moreover, using standard uniform convergence results for kernel estimators, it can be shown that

$$\sum_{r=1}^{\infty} \tilde{S}^r(\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW}) = \sum_{r=1}^{\infty} S^{r-1}\tilde{S}(\tilde{m}^{A,NW}(x) - \tilde{m}_0^{A,NW}) + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{34}$$

uniformly in x, where S is defined analogously to $\tilde{S}$ with the density estimators replaced by the true densities. Combining (33) and (34) completes the proof.

To show (32), we exploit the mixing behavior of the variables $\boldsymbol{X}_t$. Plugging the definition of $\tilde{m}_2^{A,NW}$ into the term $S_{1,2}$, we can write

$$S_{1,2}(x_1) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\int_0^1 \frac{\tilde{p}(x_1, x_2)}{\tilde{p}_1(x_1)\tilde{p}_2(x_2)} K_g(x_2, X_{t,2})dx_2 - 1\Big)\varepsilon_t.$$

Again applying standard uniform convergence results for kernel estimators, we can further replace the density estimates in the above expression by the true densities. This yields

$$\begin{aligned} S_{1,2}(x_1) &= -\frac{1}{T}\sum_{t=1}^{T}\Big(\int_0^1 \frac{p(x_1,x_2)}{p_1(x_1)p_2(x_2)}K_g(x_2,X_{t,2})dx_2 - 1\Big)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \\ &=: S^*_{1,2}(x_1) + o_p\Big(\frac{1}{\sqrt{T}}\Big) \end{aligned}$$

uniformly for $x_1 \in [0,1]$. In the final step, we show that

$$S^*_{1,2}(x_1) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{p(x_1,X_{t,2})}{p_1(x_1)p_2(X_{t,2})} - 1\Big)\varepsilon_t + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

uniformly in x_1. This is done by applying a covering argument together with an exponential inequality for mixing variables. $\square$

B Appendix

We now provide the proofs for the results on the structural break test from Section 3. As in Appendix A, we assume for simplicity that $d = 2$. We make the following assumptions.

(B1) The residuals ε_t are of the form $\varepsilon_t = \sigma^{ante}(\boldsymbol{X}_t)\xi_t$ for $t < t^*$ and $\varepsilon_t = \sigma^{post}(\boldsymbol{X}_t)\xi_t$ for $t \geq t^*$. Here, σ^ℓ are Lipschitz continuous functions and $\{\xi_t\}$ is an i.i.d. process with the same properties as in (A4) from Appendix A.

(B2) For $\ell \in \{ante, post\}$, there exist strictly stationary and strongly mixing processes $\{(Y_t^\ell, \boldsymbol{X}_t^\ell) : t \in \mathcal{T}_\ell\}$ which satisfy the equation

$$Y_t^\ell = m_0^\ell + m_1^\ell(X_{t,1}^\ell) + m_2^\ell(X_{t,2}^\ell) + \varepsilon_t^\ell, \tag{35}$$

where $\mathbb{E}[\varepsilon_t^\ell | \boldsymbol{X}_t^\ell] = 0$ with $\varepsilon_t^\ell = \sigma^\ell(\boldsymbol{X}_t^\ell)\xi_t$ and $\boldsymbol{X}_t^\ell = (X_{t,1}^\ell, X_{t,2}^\ell)$. The processes $\{(Y_t^{ante}, \boldsymbol{X}_t^{ante}) : t \in \mathcal{T}_{ante}\}$ and $\{(Y_t^{post}, \boldsymbol{X}_t^{post}) : t \in \mathcal{T}_{post}\}$ are independent and have mixing coefficients α_ℓ with the property that $\alpha_\ell(k) \leq a^k$ for some $0 < a < 1$ and $\ell \in \{ante, post\}$.

(B3) There exist constants $\gamma \in (1/2, 1)$ and $C_\gamma > 0$ such that for $j = 1, 2$,

$$\begin{aligned} \sup\{|X_{t,j}^{post} - X_{t,j}| : t^* + C_\gamma \log T \leq t \leq T\} &= O_p(hT^{-\gamma}) \\ \sup\{|Y_t^{post} - Y_t| : t^* + C_\gamma \log T \leq t \leq T\} &= O_p(T^{-\gamma}) \\ \sup\{|X_{t,j}^{ante} - X_{t,j}| : C_\gamma \log T \leq t \leq t^*\} &= O_p(hT^{-\gamma}) \\ \sup\{|Y_t^{ante} - Y_t| : C_\gamma \log T \leq t \leq t^*\} &= O_p(T^{-\gamma}) \\ \sup\{|Y_t| : 0 \leq t \leq C_\gamma \log T \text{ or } t^* \leq t \leq t^* + C_\gamma \log T\} &= O_p(\log T). \end{aligned}$$

(B4) Assumptions (A2)–(A7) from Appendix A apply with $(\boldsymbol{X}_t, \varepsilon_t)$, $(p, p_{(0,l)})$ and (m_0, m_1, m_2, σ) replaced by $(\boldsymbol{X}_t^\ell, \varepsilon_t^\ell)$, $(p^\ell, p_{(0,l)}^\ell)$ and $(m_0^\ell, m_1^\ell, m_2^\ell, \sigma^\ell)$ for $\ell \in \{ante, post\}$. Here, p^ℓ is the density of $\boldsymbol{X}_t^\ell$ and $p_{(0,l)}^\ell$ is the joint density of $(\boldsymbol{X}_t^\ell, \boldsymbol{X}_{t+l}^\ell)$. The marginal density of $X_{t,j}^\ell$ is denoted by p_j^ℓ.

(B5) It holds that $t^*/T \to c$ with $c \in (0,1)$ for $T \to \infty$.

Taken together, assumptions (B1)–(B3) essentially say that the potentially nonstationary process $\{(Y_t, \boldsymbol{X}_t)\}$ can be approximated by the stationary process $\{(Y_t^{ante}, \boldsymbol{X}_t^{ante})\}$ on the ante sample, i.e., at time points $t < t^*$, and by $\{(Y_t^{post}, \boldsymbol{X}_t^{post})\}$ on the post sample, i.e., at time point $t \geq t^*$. At the end of this appendix, we give a more detailed discussion of (B2) and (B3). In particular, we provide conditions under which (B2) and (B3) hold in an autoregressive setup which nests the nonparametric HAR model from Section 4 as a special case. Note that the distribution of $\{(Y_t^{post}, \boldsymbol{X}_t^{post})\}$ depends on n if we consider local alternatives, since the regression function depends on n in this case. To keep the proofs readable, we however do not reflect this in the notation.

Some technical lemmas

We first introduce some notation. By $\tilde{m}_1^{ante}$ and $\tilde{m}_2^{ante}$, we denote the backfitting estimators of m_1^{ante} and m_2^{ante}, respectively, which are based on the observations $\{(Y_t, \boldsymbol{X}_t) : 1 \leq t \leq t^* - 1\}$. Analogously, we let $\tilde{m}_1^{post}$ and $\tilde{m}_2^{post}$ be the backfitting estimators of m_1^{post} and m_2^{post} which are based on the observations $\{(Y_t, \boldsymbol{X}_t) : t^* \leq t \leq T\}$. In our asymptotic analysis, we compare these estimators with the corresponding infeasible backfitting estimators of m_1^{ante}, m_2^{ante} and m_1^{post}, m_2^{post} that are based on the samples $\{(Y_t^{ante}, \boldsymbol{X}_t^{ante}) : 1 \leq t \leq t^* - 1\}$ and $\{(Y_t^{post}, \boldsymbol{X}_t^{post}) : t^* \leq t \leq T\}$, respectively. These are denoted by $\tilde{m}_1^{\dagger,ante}, \tilde{m}_2^{\dagger,ante}$ and $\tilde{m}_1^{\dagger,post}, \tilde{m}_2^{\dagger,post}$. In our next lemma, we argue that $\tilde{m}_j^\ell - \tilde{m}_j^{\dagger,\ell}$ is small for $\ell = ante, post$ and $j = 1, 2$.

Lemma 1. *Suppose that (B1)–(B5) are satisfied. Then for $\ell = ante, post$ and $j = 1, 2$,*

$$\sup_{x \in [0,1]} |\tilde{m}_j^{\dagger,\ell}(x) - \tilde{m}_j^\ell(x)| = O_p(T^{-\gamma}).$$

Proof of Lemma 1. We argue that for $\ell = ante, post$ and $j = 1, 2$,

$$\sup_{x \in [0,1]} |\bar{m}_j^{\dagger,\ell}(x) - \bar{m}_j^\ell(x)| = O_p(T^{-\gamma}), \tag{36}$$

where we compare the 'marginal estimators'

$$\bar{m}_j^{\dagger,ante}(x) = \frac{\sum_{t=1}^{t^*-1} K_g(x, X_{t,l}^{ante}) Y_t^{ante}}{\sum_{t=1}^{t^*-1} K_g(x, X_{t,l}^{ante})}, \qquad \bar{m}_j^{ante}(x) = \frac{\sum_{t=1}^{t^*-1} K_g(x, X_{t,l}) Y_t}{\sum_{t=1}^{t^*-1} K_g(x, X_{t,l})},$$

$$\bar{m}_j^{\dagger,post}(x) = \frac{\sum_{t=t^*}^{T} K_g(x, X_{t,l}^{post}) Y_t^{post}}{\sum_{t=t^*}^{T} K_g(x, X_{t,l}^{post})}, \qquad \bar{m}_j^{post}(x) = \frac{\sum_{t=t^*}^{T} K_g(x, X_{t,l}) Y_t}{\sum_{t=t^*}^{T} K_g(x, X_{t,l})}.$$

We prove (36) for $\ell = post$: An application of (B3) yields that

$$\begin{aligned}
&\left|\frac{1}{T}\sum_{t=t^*}^{T} K_g(x, X_{t,j}^{post})Y_t^{post} - \frac{1}{T}\sum_{t=t^*}^{T} K_g(x, X_{t,j})Y_t\right| \\
&\quad \leq \left|\frac{1}{T}\sum_{t\in\mathcal{T}_-} K_g(x, X_{t,j}^{post})Y_t^{post}\right| + \left|\frac{1}{T}\sum_{t\in\mathcal{T}_-} K_g(x, X_{t,j})Y_t\right| \\
&\qquad + \left|\frac{1}{T}\sum_{t\in\mathcal{T}_+} \left\{K_g(x, X_{t,j}^{post})Y_t^{post} - K_g(x, X_{t,j})Y_t\right\}\right| \\
&\quad = O_p((Tg)^{-1}(\log T)^2 + T^{-\gamma}) = O_p(T^{-\gamma})
\end{aligned}$$

uniformly in x, where $\mathcal{T}_- = \{t : t^* \leq t \leq t^* + C_\gamma \log T\}$ and $\mathcal{T}_+ = \{t : t^* + C_\gamma \log T < t \leq T\}$. This shows (36) for $\ell = post$.

The statement of the lemma now follows from (36) together with the theory developed in Mammen et al. (1999) for the smooth backfitting estimators. There it is explained that the smooth backfitting estimators result from the 'marginal estimators' by the application of an operator with the following property: a bounded function is mapped onto a bounded function. This can be seen from arguments given in Mammen et al. (1999); see, e.g., the proof of their equation (88). □

We now define

$$\begin{aligned}
\mathcal{K}_{h,T}^{\dagger,j,ante} g(\cdot) &= \frac{\sum_{t=1}^{t^*-1} K_h(\cdot - X_{t,j}^{ante})g(X_{t,j}^{ante})}{\sum_{t=1}^{t^*-1} K_h(\cdot - X_{t,j}^{ante})}, \\
\mathcal{K}_{h,T}^{\dagger,j,post} g(\cdot) &= \frac{\sum_{t=t^*}^{T} K_h(\cdot - X_{t,j}^{post})g(X_{t,j}^{post})}{\sum_{t=t^*}^{T} K_h(\cdot - X_{t,j}^{post})}, \\
\hat{\delta}^\dagger &= \int \big(\mathcal{K}_{h,T}^{\dagger,j,post}\hat{m}_j^{\dagger,ante}(x) - \mathcal{K}_{h,T}^{\dagger,j,ante}\hat{m}_j^{\dagger,post}(x)\big)\pi(x)dx,
\end{aligned}$$

where $\hat{m}_j^{\dagger,ante}$ and $\hat{m}_j^{\dagger,post}$ are the infeasible Nadaraya-Watson estimators based on the samples $\{(Y_t^{ante}, \boldsymbol{X}_t^{ante}) : 1 \leq t \leq t^* - 1\}$ and $\{(Y_t^{post}, \boldsymbol{X}_t^{post}) : t^* \leq t \leq T\}$, respectively. By using similar arguments as for the proof of (36), we can show the following lemma.

Lemma 2. *Let (B1)–(B5) be satisfied. Then for $\ell = ante, post$ and $j = 1, 2$,*

$$\begin{aligned}
\sup_{x\in[0,1]} &\Big|\big\{\mathcal{K}_{h,T}^{j,post}\hat{m}_j^{ante}(x) - \mathcal{K}_{h,T}^{j,ante}\hat{m}_j^{post}(x) - \hat{\delta}\big\} \\
&- \big\{\mathcal{K}_{h,T}^{\dagger,j,post}\hat{m}_j^{\dagger,ante}(x) - \mathcal{K}_{h,T}^{\dagger,j,ante}\hat{m}_j^{\dagger,post}(x) - \hat{\delta}^\dagger\big\}\Big| = O_p(T^{-\gamma}).
\end{aligned}$$

From Lemma 2, we get that $S_T^\dagger = S_T + o_p(1)$, where

$$S_T^\dagger = Th^{1/2}\int \big(\mathcal{K}_{h,T}^{\dagger,1,post}\hat{m}_1^{\dagger,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\hat{m}_1^{\dagger,post}(x) - \hat{\delta}^\dagger\big)^2\pi(x)dx.$$

Thus, for the statement of Theorem 3, it suffices to show that $S_T^\dagger - B_T$ has a limiting normal distribution with mean μ and variance V.

Proof of Theorems 3 and 4

We restrict attention to the proof of Theorem 3. The proof of Theorem 4 follows by analogous arguments. As discussed above, we have to show that

$$S_T^\dagger - B_T \xrightarrow{d} N\big(\mu, V\big).$$

To do so, we derive the expansion

$$S_T^\dagger = S_{\Delta,T} + S_{\varepsilon,T} + o_p(1), \tag{37}$$

where

$$\begin{aligned}
S_{\Delta,T} &= \int \big(\mathcal{K}_{h,T}^{\dagger,1,ante}\mathcal{K}_{h,T}^{\dagger,1,post}\Delta(x) - \overline{\delta}_\Delta\big)^2 \pi(x)dx \\
S_{\varepsilon,T} &= Th^{1/2} \int \big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{\varepsilon,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{\varepsilon,post}(x) - \overline{\delta}_\varepsilon\big)^2 \pi(x)dx
\end{aligned}$$

with

$$\begin{aligned}
\overline{m}_1^{\varepsilon,ante}(x) &= \frac{\sum_{t=1}^{t^*-1} K_h(x - X_{t,1}^{ante})\varepsilon_t^{ante}}{\sum_{t=1}^{t^*-1} K_h(x - X_{t,1}^{ante})} \\
\overline{m}_1^{\varepsilon,post}(x) &= \frac{\sum_{t=t^*}^{T} K_h(x - X_{t,1}^{post})\varepsilon_t^{post}}{\sum_{t=t^*}^{T} K_h(x - X_{t,1}^{post})} \\
\overline{\delta}_\Delta &= \int \mathcal{K}_{h,T}^{\dagger,1,ante}\mathcal{K}_{h,T}^{\dagger,1,post}\Delta(x)\pi(x)dx \\
\overline{\delta}_\varepsilon &= \int \Big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{\varepsilon,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{\varepsilon,post}(x)\Big) \pi(x)dx.
\end{aligned}$$

The theorem then follows from the two asymptotic statements

$$\begin{aligned}
S_{\Delta,T} &\xrightarrow{p} \mu \\
S_{\varepsilon,T} - B_T &\xrightarrow{d} N\big(0, V\big),
\end{aligned}$$

which can be shown by analogous arguments as in the proof of Theorem 1.

Proof of (37). It can be shown that

$$S_T^\dagger = Th^{1/2} \int \big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{+,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{+,post}(x) - \overline{\delta}^\dagger\big)^2 \pi(x)dx + o_p(1), \tag{38}$$

where $\overline{m}_1^{+,ante}(x) = \overline{m}_1^{m,ante}(x) + \overline{m}_1^{\Delta,ante}(x) + \overline{m}_1^{\varepsilon,ante}(x)$ and $\overline{m}_1^{+,post}(x) = \overline{m}_1^{m,post}(x) +$

$\overline{m}_1^{\Delta,post}(x) + \overline{m}_1^{\varepsilon,post}(x)$ with $\overline{m}_1^{\Delta,ante}(x) = 0$,

$$\overline{m}_1^{\Delta,post}(x) = T^{-1/2}h^{-1/4}\frac{\sum_{t=t^*}^{T} K_h(x - X_{t,1}^{post})\Delta(X_{t,1}^{post})}{\sum_{t=t^*}^{T} K_h(x - X_{t,1}^{post})}$$
$$\overline{m}_1^{m,ante}(x) = \frac{\sum_{t=1}^{t^*-1} K_h(x - X_{t,1}^{ante})m_1^{ante}(X_{t,1}^{ante})}{\sum_{t=1}^{t^*-1} K_h(x - X_{t,1}^{ante})}$$
$$\overline{m}_1^{m,post}(x) = \frac{\sum_{t=t^*}^{T} K_h(x - X_{t,1}^{post})m_1^{post}(X_{t,1}^{post})}{\sum_{t=t^*}^{T} K_h(x - X_{t,1}^{post})}$$
$$\overline{\delta}^{\dagger} = \int \big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{+,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{+,post}(x)\big)\pi(x)dx.$$

We omit the proof of (38). The basic argument is that the estimation error coming from the pilot smooth backfitting estimation can be asymptotically neglected. This can be seen as in the proofs of Theorems 1 and 2. Next, we show that

$$Th^{1/2}\int \big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{l,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{l,post}(x) - \overline{\delta}_l\big) \times \big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{k,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{k,post}(x) - \overline{\delta}_k\big)\pi(x)dx = o_p(1) \tag{39}$$

for $(l,k) \in \{(\varepsilon,\Delta),(\varepsilon,m),(m,\Delta),(m,m)\}$ with

$$\overline{\delta}_l = \int \big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{l,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{l,post}(x)\big)\pi(x)dx.$$

For $(l,k) = (\varepsilon,\Delta)$, claim (39) follows by direct calculations and standard kernel smoothing theory. For the other cases, it is implied by the two statements

$$\sup_{x\in[2C_1h,1-2C_1h]}\Big|\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{m,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{m,post}(x)\Big| = o_p(T^{-1/2}) \tag{40}$$
$$\sup_{x\in[0,1]}\Big|\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{m,ante}(x) - \mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{m,post}(x)\Big| = o_p(h^{-3/4}T^{-1/2}), \tag{41}$$

which we verify below. (37) now follows from (38) and (39).

It remains to show (40) and (41). To simplify notation, we write $\mu(x) = m_1^{ante}(x)$ as well as $\hat{p}_1^{ante}(x) = t_{ante}^{-1}\sum_{t=1}^{t^*-1} K_h(x - X_{t,1}^{ante})$ and $\hat{p}_1^{post}(x) = t_{post}^{-1}\sum_{t=t^*}^{T} K_h(x - X_{t,1}^{post})$ with $t_{ante} = t^* - 1$ and $t_{post} = T - t^* + 1$. In addition, we let $p_{1,h}^{ante}(x) = \mathbb{E}[\hat{p}_1^{ante}(x)]$ and $p_{1,h}^{post}(x) = \mathbb{E}[\hat{p}_1^{post}(x)]$. We first give a proof of (40): It holds that uniformly for $x \in [2C_1h, 1-2C_1h]$,

$$\begin{aligned}
&\Big(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{m,ante}\Big)(x) - m_1^{ante}(x) \\
&= t_{ante}^{-1}t_{post}^{-1}\sum_{t=t^*}^{T}\sum_{s=1}^{t^*-1}\frac{K_h(x - X_{t,1}^{post})K_h(X_{s,1}^{ante} - X_{t,1}^{post})}{\hat{p}_1^{post}(x)\hat{p}_1^{ante}(X_{t,1}^{post})}\big(\mu(X_{s,1}^{ante}) - \mu(x)\big) \\
&= t_{ante}^{-1}t_{post}^{-1}\sum_{t=t^*}^{T}\sum_{s=1}^{t^*-1}\frac{K_h(x - X_{t,1}^{post})K_h(X_{s,1}^{ante} - X_{t,1}^{post})}{\hat{p}_1^{post}(x)\hat{p}_1^{ante}(X_{t,1}^{post})}
\end{aligned}$$

$$
\begin{aligned}
&\quad\times\left(\mu'(x)\left(X_{s,1}^{ante}-x\right)+\int_x^{X_{s,1}^{ante}}\mu''(u)(X_{s,1}^{ante}-u)\,du\right)\\
&= t_{ante}^{-1}t_{post}^{-1}\sum_{t=t^*}^{T}\sum_{s=1}^{t^*-1}\frac{K_h(x-X_{t,1}^{post})K_h(X_{s,1}^{ante}-X_{t,1}^{post})}{p_{1,h}^{post}(X_{t,1}^{post})p_{1,h}^{ante}(X_{s,1}^{ante})}\\
&\quad\times\left(\mu'(x)\left(X_{s,1}^{ante}-x\right)+\frac{(p_{1,h}^{ante})'(x)}{p_{1,h}^{ante}(x)}(X_{s,1}^{ante}-X_{t,1}^{post})\mu'(x)\left(X_{s,1}^{ante}-x\right)\right.\\
&\qquad\left.+\frac{(p_{1,h}^{post})'(x)}{p_{1,h}^{post}(x)}(X_{t,1}^{post}-x)\mu'(x)\left(X_{s,1}^{ante}-x\right)+\int_x^{X_{s,1}^{ante}}\mu''(u)(X_{s,1}^{ante}-u)\,du\right)\\
&\quad+o_p(T^{-1/2})\\
&=\mu'(x)A_1(x)+\mu'(x)\frac{(p_1^{ante})'(x)}{p_1^{ante}(x)}A_2(x)+\mu'(x)\frac{(p_1^{post})'(x)}{p_1^{post}(x)}A_3(x)+A_4(x)+o_p(T^{-1/2}),
\end{aligned}
$$

where

$$
\begin{aligned}
A_1(x)&=\int_0^1 K_h(x-u)K_h(v-u)(v-x)\,du\,dv,\\
A_2(x)&=\int_0^1 K_h(x-u)K_h(v-u)(v-u)(v-x)\,du\,dv,\\
A_3(x)&=\int_0^1 K_h(x-u)K_h(v-u)(v-x)(u-x)\,du\,dv,\\
A_4(x)&=\int_0^1 K_h(x-u)K_h(v-u)\int_x^v \mu''(w)(v-w)\,dw\,du\,dv.
\end{aligned}
$$

Note that $A_1(x)=0$ and $A_2(x)=A_3(x)$. By the same type of arguments, one gets that

$$
\begin{aligned}
&\left(\mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{m,post}\right)(x)-m_1^{ante}(x)\\
&\quad=\mu'(x)A_1(x)+\mu'(x)\frac{(p_1^{post})'(x)}{p_1^{post}(x)}A_2(x)+\mu'(x)\frac{(p_1^{ante})'(x)}{p_1^{ante}(x)}A_3(x)+A_4(x)+o_p(T^{-1/2}).
\end{aligned}
$$

This shows (40). We finally turn to the proof of (41): Uniformly for $x\in[0,2C_1h]$,

$$
\begin{aligned}
&\left(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{m,ante}\right)(x)-m_1^{ante}(x)\\
&\quad=\mu'(x)\int_0^1 K_h(x-u)K_h(v-u)(v-x)\frac{p_1^{ante}(v)p_1^{post}(u)}{p_{1,h}^{ante}(u)p_{1,h}^{post}(x)}\,du\,dv+o_p(h^{-3/4}T^{-1/2})\\
&\quad=\mu'(x)\int_0^1\frac{K_h(x-u)K_h(v-u)(v-x)}{\int_0^1 K_h(u-w)\,dw\int_0^1 K_h(x-z)\,dz}\,du\,dv+o_p(h^{-3/4}T^{-1/2}).
\end{aligned}
$$

This can be shown similarly as in the proof of (40). By analogous arguments, we further get that uniformly for $x\in[0,2C_1h]$,

$$
\begin{aligned}
&\left(\mathcal{K}_{h,T}^{\dagger,1,ante}\overline{m}_1^{m,post}\right)(x)-m_1^{ante}(x)\\
&\quad=\mu'(x)\int_0^1\frac{K_h(x-u)K_h(v-u)(v-x)}{\int_0^1 K_h(u-w)\,dw\int_0^1 K_h(x-z)\,dz}\,du\,dv+o_p(h^{-3/4}T^{-1/2})\\
&\quad=\left(\mathcal{K}_{h,T}^{\dagger,1,post}\overline{m}_1^{m,ante}\right)(x)-m_1^{ante}(x)+o_p(h^{-3/4}T^{-1/2}).
\end{aligned}
$$

A similar asymptotic equality also holds for $x \in [1 - 2C_1 h, 1]$. This completes the proof of (41). $\square$

Verification of (B2) and (B3) in the HAR model

To start with, consider the autoregressive model

$$Y_t = m(Y_{t-1}, \dots, Y_{t-d}) + \varepsilon_t, \tag{42}$$

where $\mathbb{E}[\varepsilon_t | Y_{t-1}, \dots, Y_{t-d}] = 0$ and the errors ε_t have the form $\varepsilon_t = \sigma(Y_{t-1}, \dots, Y_{t-d})\xi_t$ with i.i.d. residuals ξ_t. Standard results to be found e.g. in Chen and Chen (2000) show that the process $\{Y_t\}$ defined in (42) has a stationary solution and is geometrically α-mixing under the following conditions:

(i) The variables ξ_t are i.i.d. with $\mathbb{E}[\xi_t] = 0$ and $\mathbb{E}|\xi_t| < \infty$, they have an everywhere positive and continuous density function, and ξ_t is independent of all Y_s with $s < t$.

(ii) The function m is bounded on every bounded set, that is, for any constant $C \geq 0$, $\sup_{\|x\| \leq C} |m(x)| < \infty$.

(iii) The function σ is such that $0 < \underline{\sigma} \leq \inf_{\|x\| \leq C} \sigma(x) \leq \sup_{\|x\| \leq C} \sigma(x) < \infty$ for any $C \geq 0$ and some constant $\underline{\sigma} > 0$.

(iv) There exist constants $a_j, b_j \geq 0$ $(j = 1, \dots, d)$, $c_1, c_2 \geq 0$ and $C_0 > 0$ such that

$$|m(x)| \leq \sum_{j=1}^{d} a_j |x_j| + c_1 \quad \text{for } \|x\| \geq C_0$$

$$|\sigma(x)| \leq \sum_{j=1}^{d} b_j |x_j| + c_2 \quad \text{for } \|x\| \geq C_0$$

and $\sum_{j=1}^{d} (a_j + b_j \mathbb{E}|\xi_t|) < 1$.

Our nonparametric HAR model (8) is a special case of (42). To see this, consider the model with a daily, a weekly and a monthly component function and suppose that $\varepsilon_t = \sigma(V_{t-1}^{(1)}, V_{t-1}^{(5)}, , V_{t-1}^{(22)})\xi_t$. In this case, the HAR model can be rewritten as

$$V_t^{(1)} = m(V_{t-1}^{(1)}, V_{t-2}^{(1)}, \dots, V_{t-22}^{(1)}) + \varepsilon_t, \tag{43}$$

where $m(x_1, \dots, x_{22}) = m_0 + m_1(x_1) + m_2(\frac{1}{5}\sum_{j=1}^{5} x_j) + m_3(\frac{1}{22}\sum_{j=1}^{22} x_j)$. Assuming that the components of (43) fulfill conditions (i)–(iv), we can infer that the HAR process $\{V_t^{(1)}\}$ (as well as the average processes $\{V_t^{(5)}\}$ and $\{V_t^{(22)}\}$) has a stationary solution which is geometrically mixing. By the same token, if the residuals ξ_t and the functions $(m_0^\ell, m_1^\ell, m_2^\ell, m_3^\ell, \sigma^\ell)$ satisfy (i)–(iv) for $\ell \in \{ante, post\}$, there exist strictly stationary and strongly mixing HAR processes $\{V_t^{ante,(1)} : t \in \mathcal{T}_{ante}\}$ and $\{V_t^{post,(1)} : t \in \mathcal{T}_{post}\}$. Since the innovations ξ_t are i.i.d., we can assume w.l.o.g. that these two processes are independent. As a result, assumption (B2) is satisfied for $Y_t^\ell = V_t^{\ell,(1)}$ and $\boldsymbol{X}_t^\ell = (V_{t-1}^{\ell,(1)}, V_{t-1}^{\ell,(5)}, V_{t-1}^{\ell,(22)})$ with $\ell = ante, post$.

We finally turn to the discussion of (B3). To keep the exposition as simple as possible, we suppose that $\sigma^{ante}(\cdot) = \sigma^{post}(\cdot) \equiv \bar{\sigma}$ for some constant $\bar{\sigma} > 0$. Now assume that the processes $\{V_t^{ante,(1)}\}$ and $\{V_t^{post,(1)}\}$ are stationary and suppose that

$$\Big|\frac{\partial}{\partial x} m_1^\ell(x)\Big| + \Big|\frac{\partial}{\partial y} m_2^\ell(y)\Big| + \Big|\frac{\partial}{\partial z} m_3^\ell(z)\Big| \leq \rho$$

for some $0 < \rho < 1$. Then (B3) is satisfied for $Y_t^\ell = V_t^{\ell,(1)}$ and $\boldsymbol{X}_t^\ell = (V_{t-1}^{\ell,(1)}, V_{t-1}^{\ell,(5)}, V_{t-1}^{\ell,(22)})$. To derive the first two inequalities of (B3), note that

$$\begin{aligned}
|V_t^{post,(1)} - V_t^{(1)}| &\leq \Big| m_1^{post}(V_{t-1}^{post,(1)}) - m_1^{post}(V_{t-1}^{(1)}) + m_2^{post}(V_{t-1}^{post,(5)}) - m_2^{post}(V_{t-1}^{(5)}) \\
&\qquad + m_3^{post}(V_{t-1}^{post,(22)}) - m_3^{post}(V_{t-1}^{(22)}) \Big| \\
&\leq \rho \max\Big\{ \Big|V_{t-1}^{post,(1)} - V_{t-1}^{(1)}\Big|, \Big|V_{t-1}^{post,(5)} - V_{t-1}^{(5)}\Big|, \Big|V_{t-1}^{post,(22)} - V_{t-1}^{(22)}\Big| \Big\} \\
&\leq \rho \max_{1 \leq \ell \leq 22} |V_{t-\ell}^{post,(1)} - V_{t-\ell}^{(1)}| \qquad (44)
\end{aligned}$$

for $t^* + C_\gamma \log T \leq t \leq T$. Choosing C_γ such that $\rho^{C_\gamma \log T} \ll hT^{-\gamma}$ and making iterative use of (44), we immediately obtain the first two inequalities of (B3). The third and fourth claim follow similarly. The last claim of (B3) is a simple consequence of standard moment conditions.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics* **61**: 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica* **71**(2): 579–625.

Andersen, T. G., Dobrev, D. and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation, *Journal of Econometrics* **169**: 75–93.

Asai, M., McAleer, M. and Medeiros, M. C. (2012). Modelling and forecasting noisy realized volatility, *Computational Statistics and Data Analysis* **56**: 217–230.

Audrino, F. and Knaus, S. (2014). Lassoing the HAR model: A model selection perspective on realized volatility dynamics, *Econometric Reviews* . In press.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shepard, N. (2009). Realised kernels in practice: trades and quotes, *Econometrics Journal* **12**(3): C1–C32.

Chen, M. and Chen, G. (2000). Geometric ergodicity of nonlinear autoregressive models with changing conditional variances, *Canadian Journal of Statistics* **28**(3): 605–613.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models, *in* J. J. Heckman and E. E. Leamer (eds), *Handbook of Econometrics*, Vol. 6B, Elsevier, chapter 76, pp. 5549–5632.

Chen, X. B., Gao, J., Li, D. and Silvapulle, P. (2013). Nonparametric estimation and parametric calibration of time-varying coefficient realized volatility models, *Working paper*, Monash University, Department of Econometrics and Business Statistics.

Corradi, V., Distaso, W. and Swanson, N. R. (2009). Predictive density estimators for daily volatility based on the use of realized measures, *Journal of Econometrics* **150**: 119–138.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* **7**(2): 174–196.

Corsi, F., Audrino, F. and Renò, R. (2012). HAR Modeling for realized volatility forecasting, *in* L. Bauwens, C. Hafner and S. Laurent (eds), *Handbook of Volatility Models and their Applications*, John Wiley & Sons, New York, chapter 15, pp. 363–382.

Delgado, M. A. and Hidalgo, J. (2000). Nonparametric inference on structural breaks, *Journal of Econometrics* **96**: 113–144.

Eilers, P. H. C. and Marx, B. D. (2002). Generalized linear additive smooth structures, *Journal of Computational and Graphical Statistics* **11**: 758–783.

Fan, J. and Jiang, J. (2005). Nonparametric inference for additive models, *Journal of the American Statistical Association* **100**(471): 890–907.

Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate u-statistics of absolutely regular processes with applications to model specification testing, *Journal of Nonparametric Statistics* **10**: 245–271.

González-Manteiga, W. and Cao-Abad, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation, *TEST* **2**: 161–188.

Haag, B. R. (2008). Non-parametric regression tests using dimension reduction techniques, *Scandinavian Journal of Statistics* **35**: 719–738.

Hansen, P. R. (2005). A test for superior predictive ability, *Journal of Business & Economic Statistics* **23**(4): 365–380.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *Annals of Statistics* **21**(4): 1926–1947.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*, Chapman and Hall, London.

Hidalgo, J. (1995). A nonparametric conditional moment test for structural stability, *Econometric Theory* **11**(4): 671–698.

Hillebrand, E. and Medeiros, M. C. (2014). Asymmetries, breaks, and long-range dependence: An estimation framework for time series of daily realized volatility, *Journal of Business and Economic Statistics* . In press.

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation, *Journal of Statistical Planning and Inference* **68**: 295–321.

Kreiß, J.-P., Neumann, M. H. and Yao, Q. (2008). Bootstrap tests for simple structures in nonparametric time series regression, *Statistics and Its Interface* **1**: 367–380.

Lahaye, J. and Shaw, P. (2014). Can we reject linearity in an HAR-RV model for the S&P 500? Insights from a nonparametric HAR-RV, *Economics Letters* . In press.

Li, Q. (1999). Consistent model specification tests for time series econometric models, *Journal of Econometrics* **92**: 101–147.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika* **82**: 93–100.

Linton, O. and Härdle, W. (1996). Estimation of additive regression models with known links, *Biometrika* **83**(3): 529–540.

Linton, O. and Mammen, E. (2005). Estimating semiparametric ARCH(∞) models by kernel smoothing methods, *Econometrica* **73**(3): 771–836.

Liu, R. and Yang, L. (2010). Spline-backfitted kernel smoothing of additive coefficient model, *Econometric Theory* **26**(1): 29–59.

Mammen, E., Linton, O. and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions, *Annals of Statistics* **27**(5): 1443–1490.

Mammen, E. and Park, B. (2005). Bandwidth selection for smooth backfitting in additive models, *Annals of Statistics* **33**(3): 1260–1294.

McAleer, M. and Medeiros, M. C. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries, *Journal of Econometrics* **147**(1): 104–119.

Newey, W. K. (1994). Kernel estimation of partial means, *Econometric Theory* **10**: 233–253.

Nielsen, J. P. and Sperlich, S. (2005). Smooth backfitting in practice, *Journal of the Royal Statistical Society B* **67**(1): 43–61.

Pollard, D. (1984). *Convergence of stochastic processes*, Springer-Verlag, New York.

Thornton, D. L. (2009). What the Libor-OIS spread says, *Economic Synopses 24*, Federal Reserve Bank of St. Louis.

Tjøstheim, D. and Auestad, B. (1994). Nonparametric identification of nonlinear time series: projections, *Journal of the American Statistical Association* **89**: 1398–1409.

Yang, L., Härdle, W. and Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean, *Journal of Time Series Analysis* **20**(5): 579–604.

Zheng, J. X. (1996). A consistent test of a functional form via nonparametric estimation techniques, *Journal of Econometrics* **75**: 263–289.

Symbol	Name	Type	Exchange	Data: tests		Data: forecast eval.	
CF	CAC 40 Index	Equity index	NYSE Liffe	2. Jan., 2004	31. Dec., 2010		
FT	FTSE 100 Index	Equity index	NYSE Liffe	2. Jan., 2004	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
KM	KOSPI 200 Index	Equity index	Korea Exchange	2. Jan., 2004	31. Dec., 2010		
NE	Nikkei 225 Index	Equity index	Osaka Securities Exchange	5. Jan., 2004	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
SP	S&P 500	Equity index future	CME Group	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
XX	EURO STOXX 50 Index	Equity index	Eurex	1. Jul., 2003	31. Dec., 2010		
BN	Euro-Bund 10 yr	Interest rate future	Eurex	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
TY	TNote 10 yr	Interest rate future	CBOT/CME Group	1. Jul., 2003	31. Dec., 2010		
US	TBond 30 yr	Interest rate future	CBOT/CME Group	1. Jul., 2003	31. Dec., 2010		
CL	Light Crude NYMEX	Energy future	NYMEX/CME Group	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
NG	Natural Gas NYMEX	Energy future	NYMEX/CME Group	1. Jul., 2003	31. Dec., 2010		
GC	Gold COMEX	Metal future	COMEX/CME Group	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
HG	Copper High Grade COMEX	Metal future	COMEX/CME Group	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
EC	Euro FX	Currency future	CME Group	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
JY	Japanese Yen	Currency future	CME Group	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
CN	Corn	Grain future	CBOT/CME Group	1. Jul., 2003	31. Dec., 2010	3. Jan., 2011	31. Dec., 2013
SY	Soybeans	Grain future	CBOT/CME Group	1. Jul., 2003	31. Dec., 2010		

Table 1: Overview of futures and indices used for the empirical part.

		Homosc. errors			Heterosc. errors		
		nominal size			nominal size		
	h	0.01	0.05	0.10	0.01	0.05	0.10
$m_{2,0}$	0.4	0.010	0.062	0.115	0.014	0.074	0.136
	0.5	0.007	0.041	0.095	0.013	0.060	0.118
	0.6	0.006	0.026	0.069	0.008	0.048	0.097
	0.7	0.004	0.026	0.057	0.008	0.042	0.086
$m_{2,1}$	0.4	0.065	0.209	0.316	0.079	0.205	0.316
	0.5	0.044	0.178	0.275	0.043	0.151	0.261
	0.6	0.019	0.107	0.212	0.026	0.111	0.194
	0.7	0.006	0.058	0.130	0.016	0.071	0.151
$m_{2,2}$	0.4	0.304	0.569	0.700	0.193	0.438	0.598
	0.5	0.229	0.501	0.646	0.114	0.311	0.489
	0.6	0.125	0.396	0.549	0.079	0.218	0.366
	0.7	0.053	0.234	0.416	0.045	0.138	0.262
$m_{2,3}$	0.4	0.687	0.880	0.948	0.325	0.620	0.781
	0.5	0.584	0.837	0.909	0.193	0.465	0.658
	0.6	0.425	0.749	0.848	0.113	0.307	0.485
	0.7	0.227	0.574	0.740	0.071	0.212	0.341

Table 2: Size and power simulations of the specification test under the two error scenarios. The first block labeled with $m_{2,0}$ gives the actual size $\alpha_T(h)$ of the test. The blocks below give the power $\beta_T(h)$ for the alternatives $m_{2,i}$ $(i = 1, 2, 3)$. h is the test bandwidth for m_2; the variation corresponds to 0.20-0.37 on the unit interval. For the pilot estimates, $g_1 = 0.64$ and $g_3 = 0.28$ (about 0.22 and 0.20 on the unit interval) and $g_2 = h/1.4$ are used.

		Homosc. errors			Heterosc. errors		
		nominal size			nominal size		
	h	0.01	0.05	0.1	0.01	0.05	0.1
$m_{2,0}$	0.4	0.016	0.077	0.136	0.019	0.072	0.124
	0.5	0.016	0.061	0.112	0.015	0.058	0.116
	0.6	0.012	0.049	0.096	0.011	0.048	0.104
	0.7	0.009	0.044	0.083	0.009	0.040	0.087
$m_{2,1}$	0.4	0.104	0.242	0.351	0.085	0.215	0.301
	0.5	0.073	0.192	0.287	0.057	0.152	0.242
	0.6	0.049	0.134	0.227	0.021	0.100	0.173
	0.7	0.027	0.090	0.169	0.010	0.058	0.126
$m_{2,2}$	0.4	0.388	0.621	0.735	0.362	0.581	0.695
	0.5	0.286	0.532	0.669	0.268	0.505	0.622
	0.6	0.181	0.408	0.555	0.161	0.373	0.514
	0.7	0.092	0.290	0.411	0.087	0.248	0.395
$m_{2,3}$	0.4	0.775	0.912	0.950	0.736	0.892	0.942
	0.5	0.697	0.872	0.935	0.665	0.850	0.915
	0.6	0.570	0.815	0.891	0.545	0.771	0.864
	0.7	0.393	0.698	0.820	0.380	0.677	0.785

Table 3: Size and power simulations of the structural breaks test, setting (a). The first block labeled $m_{2,0}$ gives the actual size $\alpha_T(h)$ of the test. The blocks below give the power $\beta_T(h)$ for the alternatives $m_{2,i}$ ($i = 1, 2, 3$). h is the test bandwidth for m_2; the variation corresponds to 0.20-0.37 on the unit interval. For the pilot estimates, $g_1 = 0.64$ and $g_3 = 0.28$ (about 0.23 and 0.21 on the unit interval) and $g_2 = h/1.4$ are used.

		Homosc. errors			Heterosc. errors		
		nominal size			nominal size		
	h	0.01	0.05	0.1	0.01	0.05	0.1
$m_{2,0}$	0.4	0.022	0.079	0.145	0.017	0.084	0.138
	0.5	0.011	0.061	0.120	0.008	0.054	0.118
	0.6	0.010	0.044	0.098	0.008	0.037	0.091
	0.7	0.009	0.037	0.073	0.006	0.029	0.073
$m_{2,1}$	0.4	0.072	0.178	0.275	0.053	0.146	0.226
	0.5	0.043	0.136	0.230	0.031	0.116	0.189
	0.6	0.023	0.106	0.177	0.019	0.072	0.145
	0.7	0.011	0.075	0.131	0.006	0.050	0.109
$m_{2,2}$	0.4	0.222	0.422	0.540	0.177	0.352	0.474
	0.5	0.154	0.356	0.482	0.120	0.303	0.421
	0.6	0.100	0.276	0.411	0.067	0.224	0.350
	0.7	0.057	0.190	0.321	0.041	0.151	0.261
$m_{2,3}$	0.4	0.423	0.631	0.746	0.338	0.559	0.664
	0.5	0.336	0.577	0.690	0.253	0.489	0.624
	0.6	0.237	0.494	0.630	0.164	0.405	0.555
	0.7	0.141	0.390	0.538	0.085	0.284	0.440

Table 4: Size and power simulations of the structural breaks test, setting (b). The first block labeled with $m_{2,0}$ gives the actual size $\alpha_T(h)$ of the test. The blocks below give the power $\beta_T(h)$ for the alternatives $m_{2,i}$ ($i = 1, 2, 3$). h is the test bandwidth for m_2; the variation corresponds to 0.23-0.40 on the unit interval. For the pilot estimates, $g_1 = 0.61$ and $g_3 = 0.26$ (both about 0.23 on the unit interval) and $g_2 = h/1.4$ are used.

Bandwidths			
Symbol	Daily	Weekly	Monthly
CF	0.233	0.189	0.231
FT	0.212	0.201	0.216
KM	0.222	0.233	0.270
NE	0.282	0.268	0.273
SP	0.202	0.190	0.190
XX	0.236	0.206	0.219
BN	0.252	0.191	0.192
TY	0.292	0.254	0.224
US	0.387	0.240	0.229
CL	0.327	0.216	0.217
NG	0.269	0.203	0.255
GC	0.322	0.307	0.255
HG	0.255	0.287	0.211
EC	0.325	0.231	0.229
JY	0.295	0.233	0.287
CN	0.253	0.217	0.282
SY	0.269	0.251	0.234

Table 5: Bandwidths obtained by means of the plug-in rule of Mammen and Park (2005) as described in Section 4.3. The bandwidths are reported relative to the unit interval; see Table 1 for the list of acronyms.

Test for structural breaks			
Symbol	Daily	Weekly	Monthly
CF	**0.097**	**0.025**	0.483
FT	0.102	0.182	0.929
KM	0.453	**0.032**	0.108
NE	0.367	0.247	0.378
SP	0.385	0.362	0.825
XX	0.360	**0.018**	**0.037**
BN	0.211	0.450	0.287
TY	**0.000**	**0.075**	**0.000**
US	**0.000**	0.690	**0.000**
CL	0.597	0.529	0.432
NG	0.577	**0.000**	**0.006**
GC	0.157	0.788	0.331
HG	0.206	0.308	0.504
EC	0.115	0.363	0.465
JY	0.158	0.322	0.988
CN	0.244	0.224	0.206
SY	0.671	0.114	**0.001**

Table 6: Structural break tests based on Nadaraya-Watson smooth backfitting as suggested in Section 3. Null hypothesis is equality of the functions on the ante and the post sample. The p-values are obtained from 1000 bootstrap replications. p-values are in bold when below 10%. Weighting function in the test statistic is a uniform density; see Table 1 for the list of acronyms.

Test for linear specification						
Symbol	Daily		Weekly		Monthly	
	Ante	Post	Ante	Post	Ante	Post
CF	**0.087**	0.111	0.212	**0.019**	0.476	0.179
FT	0.701		0.942		0.781	
KM	0.677	0.221	0.984	0.124	0.144	**0.011**
NE	**0.020**		0.575		0.244	
SP	**0.041**		**0.043**		0.217	
XX	**0.090**	**0.058**	0.514	**0.015**	0.809	0.164
BN	**0.013**		0.122		0.337	
TY	**0.004**	0.149	0.497	0.149	0.471	0.302
US	**0.001**	0.637	0.929	0.175	0.385	0.730
CL	0.238		0.289		0.724	
NG	**0.097**	0.888	**0.002**	0.785	0.279	**0.076**
GC	0.589		0.940		0.594	
HG	0.326		0.303		**0.016**	
EC	0.284		**0.007**		**0.014**	
JY	0.120		0.143		0.448	
CN	**0.033**		**0.002**		**0.076**	
SY	0.498	**0.021**	0.800	0.978	0.390	0.395

Table 7: Specification tests (Nadaraya-Watson smooth backfitting) on the full sample or on the subsamples in presence of a structural break according to Table 6. Null hypothesis is the linear specification in the respective component function. p-values are obtained from 1000 bootstrap replications. p-values are in bold when below 10%. Weighting function is the uniform density; see Table 1 for the list of acronyms.

RMSE

	1 day		1 week		1 month	
Symbol	HAR	npHAR	HAR	npHAR	HAR	npHAR
FT	0.4244	0.4220	0.3739	0.3711	0.4094	0.4060
NE	0.5262	0.5229	0.4121	0.4101	0.4316	0.4283
SP	0.5269	0.5230	0.4185	0.4118	0.4708	0.4591
BN	0.3635	0.3624	0.2664	0.2645	0.2618	0.2594
CL	0.4335	0.4521	0.2906	0.3222	0.2988	0.4307
GC	0.6157	0.6147	0.3572	0.3559	0.3175	0.3142
HG	0.5090	0.5454	0.3471	0.4300	0.3661	0.3613
EC	0.5031	0.5014	0.3049	0.3007	0.2759	0.2715
JY	0.5846	0.5848	0.3561	0.3532	0.3282	0.3347
CN	0.5021	0.5077	0.3289	0.3291	0.2901	0.2985

MAE

	1 day		1 week		1 month	
Symbol	HAR	npHAR	HAR	npHAR	HAR	npHAR
FT	0.3319	0.3298	0.2811	0.2800	0.3069	0.3061
NE	0.3821	0.3798	0.3086	0.3077	0.3265	0.3227
SP	0.4193	0.4181	0.3157	0.3118	0.3478	0.3434
BN	0.2777	0.2772	0.2053	0.2043	0.2025	0.2011
CL	0.3365	0.3559	0.2237	0.2480	0.2389	0.3196
GC	0.4846	0.4846	0.2747	0.2740	0.2482	0.2478
HG	0.3811	0.4124	0.2664	0.3216	0.2827	0.2831
EC	0.3891	0.3873	0.2410	0.2380	0.2151	0.2138
JY	0.4607	0.4616	0.2738	0.2721	0.2508	0.2563
CN	0.3831	0.3858	0.2620	0.2631	0.2410	0.2468

***p*-values of Hansen's SPA test**

	1 day		1 week		1 month	
Symbol	RMSE	MAE	RMSE	MAE	RMSE	MAE
FT	**0.081**	**0.035**	**0.012**	0.116	0.101	0.344
NE	**0.033**	**0.023**	**0.086**	0.106	0.112	**0.043**
SP	**0.032**	0.282	**0.019**	**0.039**	**0.020**	0.202
BN	0.252	0.342	0.169	0.334	0.244	0.328
CL	1.000	1.000	1.000	1.000	1.000	1.000
GC	**0.064**	1.000	0.170	0.261	0.115	0.385
HG	1.000	1.000	1.000	1.000	0.208	1.000
EC	**0.039**	**0.082**	**0.022**	**0.033**	**0.058**	0.306
JY	1.000	1.000	**0.057**	0.115	1.000	1.000
CN	1.000	1.000	1.000	1.000	1.000	1.000

Table 8: Forecast evaluations of the linear and the nonlinear HAR model. Top panel: root mean squared error (RMSE). Middle panel: mean absolute error (MAE). Lower panel: p-values of Hansen's test of superior predictive ability (SPA). Null hypothesis: the linear HAR model is not inferior to the nonlinear model.

Homosc. errors **Heterosc. errors**

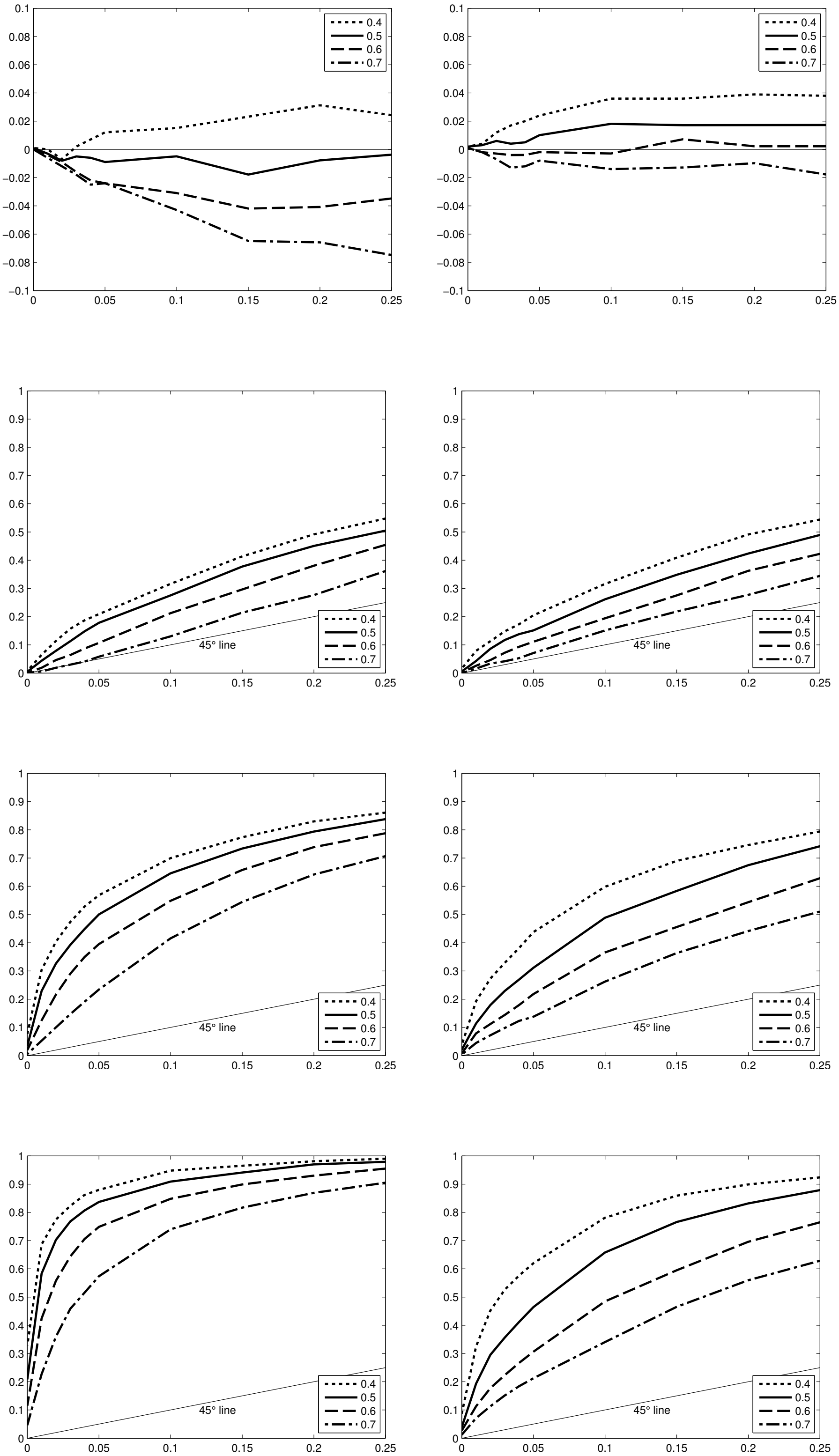

Figure 1: Size discrepancy and power curves for the specification test. Top row shows the size discrepancy curves for the four bandwidths $h = 0.4$, 0.5, 0.6, 0.7 (0.20-0.37 on the unit interval). The second, third and fourth row, from top to bottom, provide the power curves for $m_{2,i}$ $(i = 1, 2, 3)$, see Section 4.3.1. The left column features the homoscedastic error case, the right column the heteroscedastic error case.

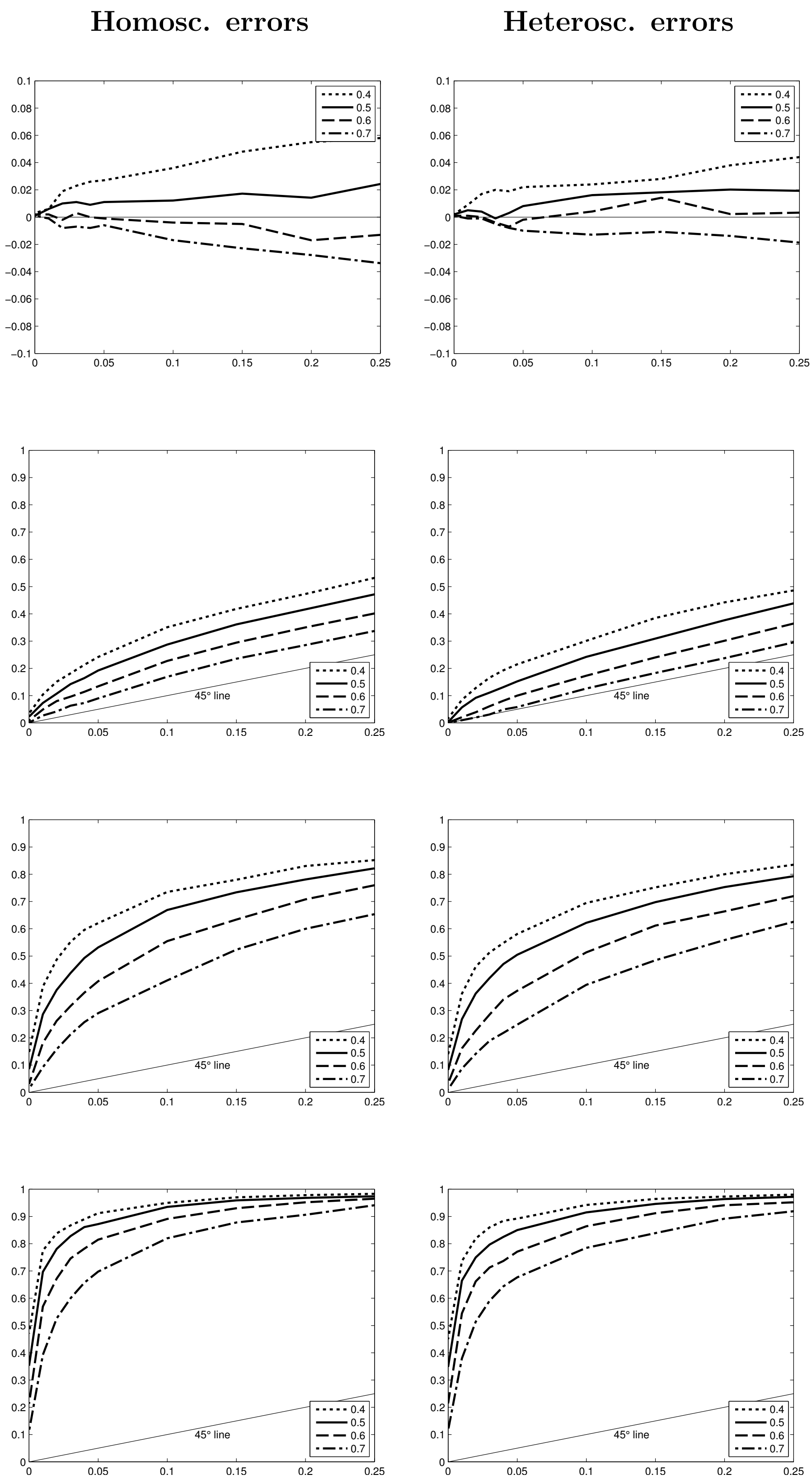

Figure 2: Size discrepancy and power curves for the structural break test, setting (a). Top row shows the size discrepancy curves for the four bandwidths $h = 0.4$, 0.5, 0.6, 0.7 (0.20-0.37 on the unit interval). The second, third and fourth row, from top to bottom, provide the power curves for $m_{2,i}$ $(i = 1, 2, 3)$, see Section 4.3.2. The left column features the homoscedastic error case, the right column the heteroscedastic error case.

Homosc. errors **Heterosc. errors**

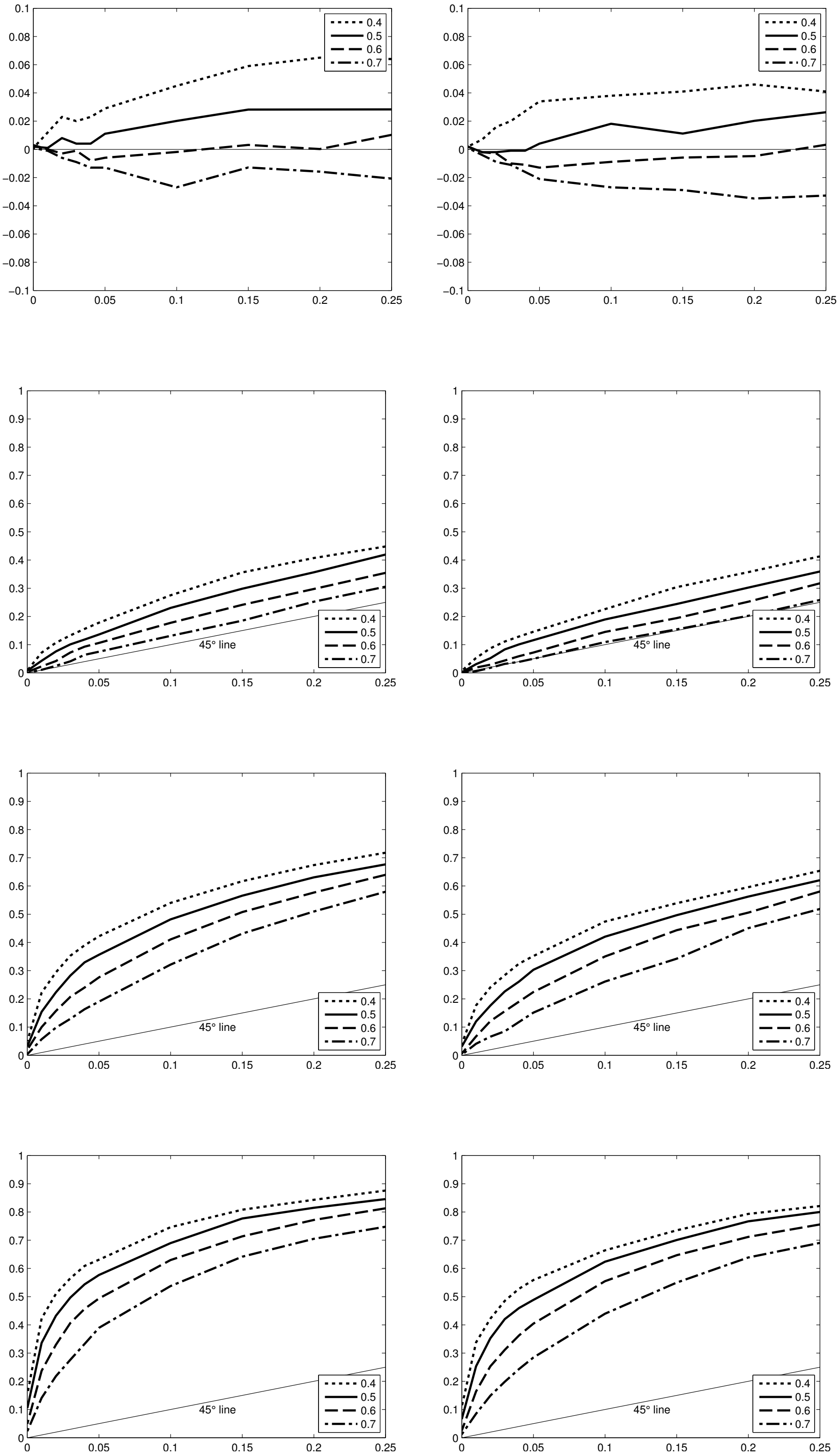

Figure 3: Size discrepancy and power curves for the structural break test, setting (b). Top row shows the size discrepancy curves for the four bandwidths $h = 0.4$, 0.5, 0.6, 0.7 (0.23-0.40 on the unit interval). The second, third and fourth row, from top to bottom, provide the power curves for $m_{2,i}$ $(i = 1, 2, 3)$, see Section 4.3.2. The left column features the homoscedastic error case, the right column the heteroscedastic error case.

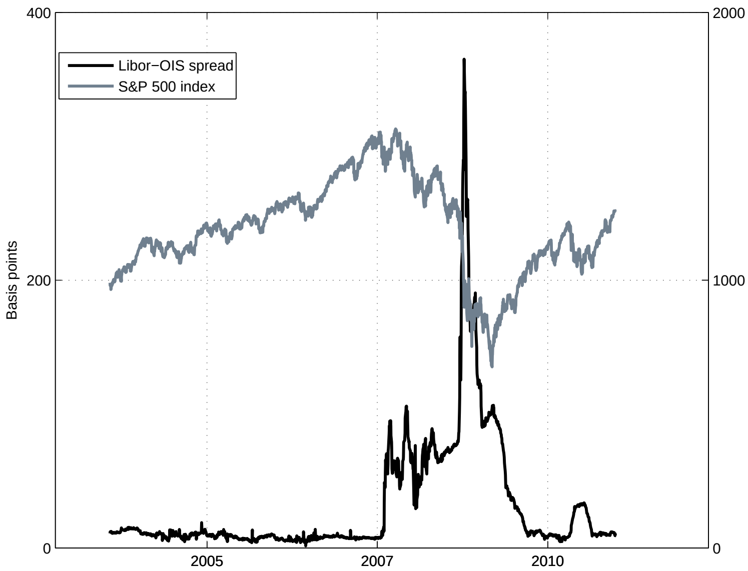

Figure 4: Difference of the 3-months USD Libor over the 3-months overnight indexed swap (left ordinate axis), S&P 500 index closing prices (right ordinate axis) from July 1, 2003 to Dec. 31, 2010. Source: Bloomberg.

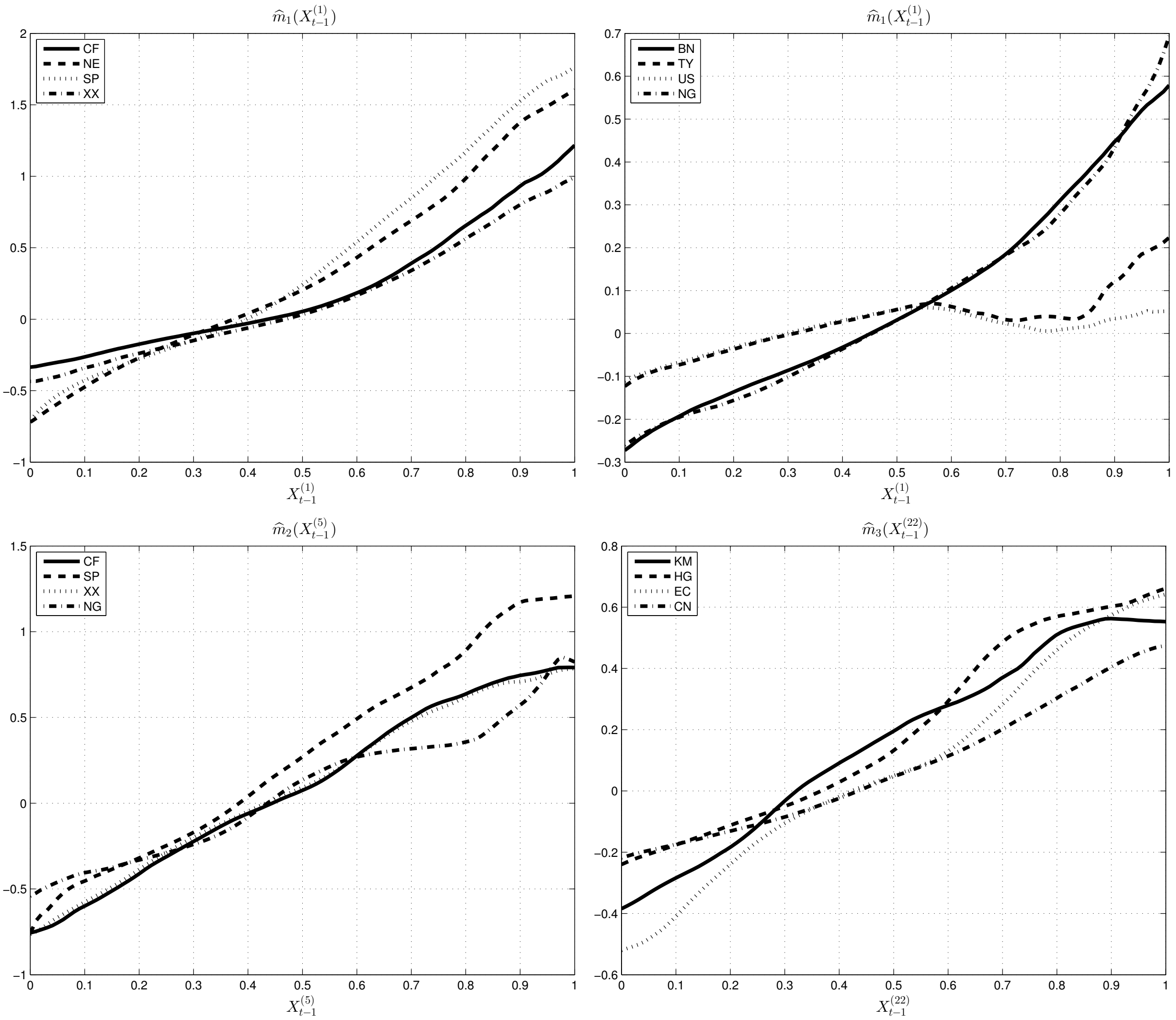

Figure 5: Nadaraya-Watson smooth backfitting estimates of the variance component functions, for which the linearity test rejects. All estimates are normalized to the unit interval and computed on the full, the ante or the post sample depending on the outcome of the structural break test given in Table 6. The top panel shows the estimates $\widehat{m}_1$ of CF (ante), NE (full), SP (full), XX (ante), BN (full), TY (ante), US (ante), NG (ante). The lower panel left shows $\widehat{m}_2$ of CF (post), SP (full), XX (post), NG (ante). The lower panel right shows $\widehat{m}_3$ of KM (post), HG (full), EC (full), CN (full); see Table 1 for the list of acronyms.

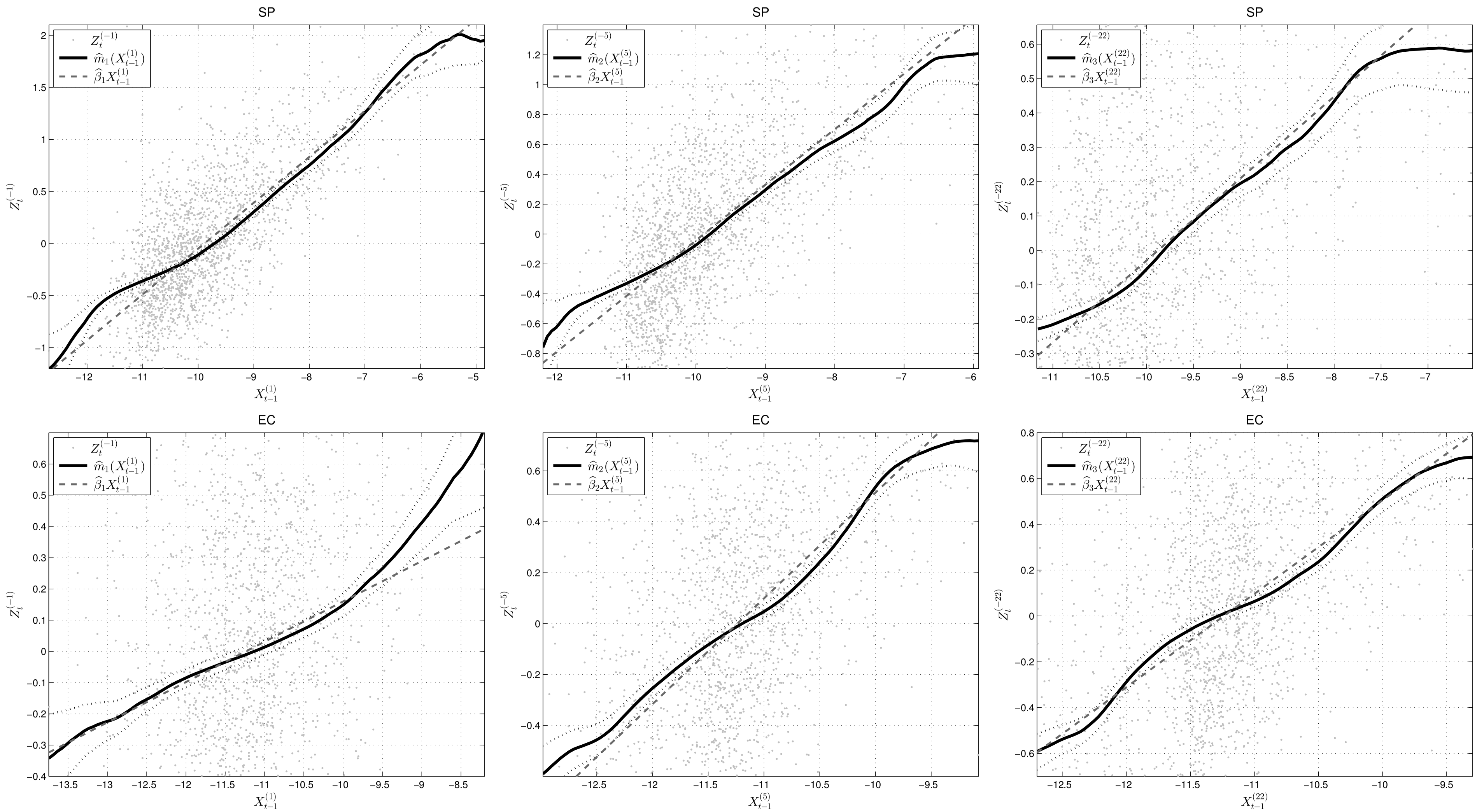

Figure 6: Nonparametric versus linear HAR model for the S&P500 (SP, top panel) and the Euro-USD currency future (EC, lower panel) with pointwise asymptotic 95% confidence intervals for the smooth backfitting estimates. Like the nonparametric component functions, the parametric fits are normed such that their expectation equals zero. $Z_t^{(-\iota_j)}$ is the residual response variable after subtracting all component functions m_i, $i \neq j$.

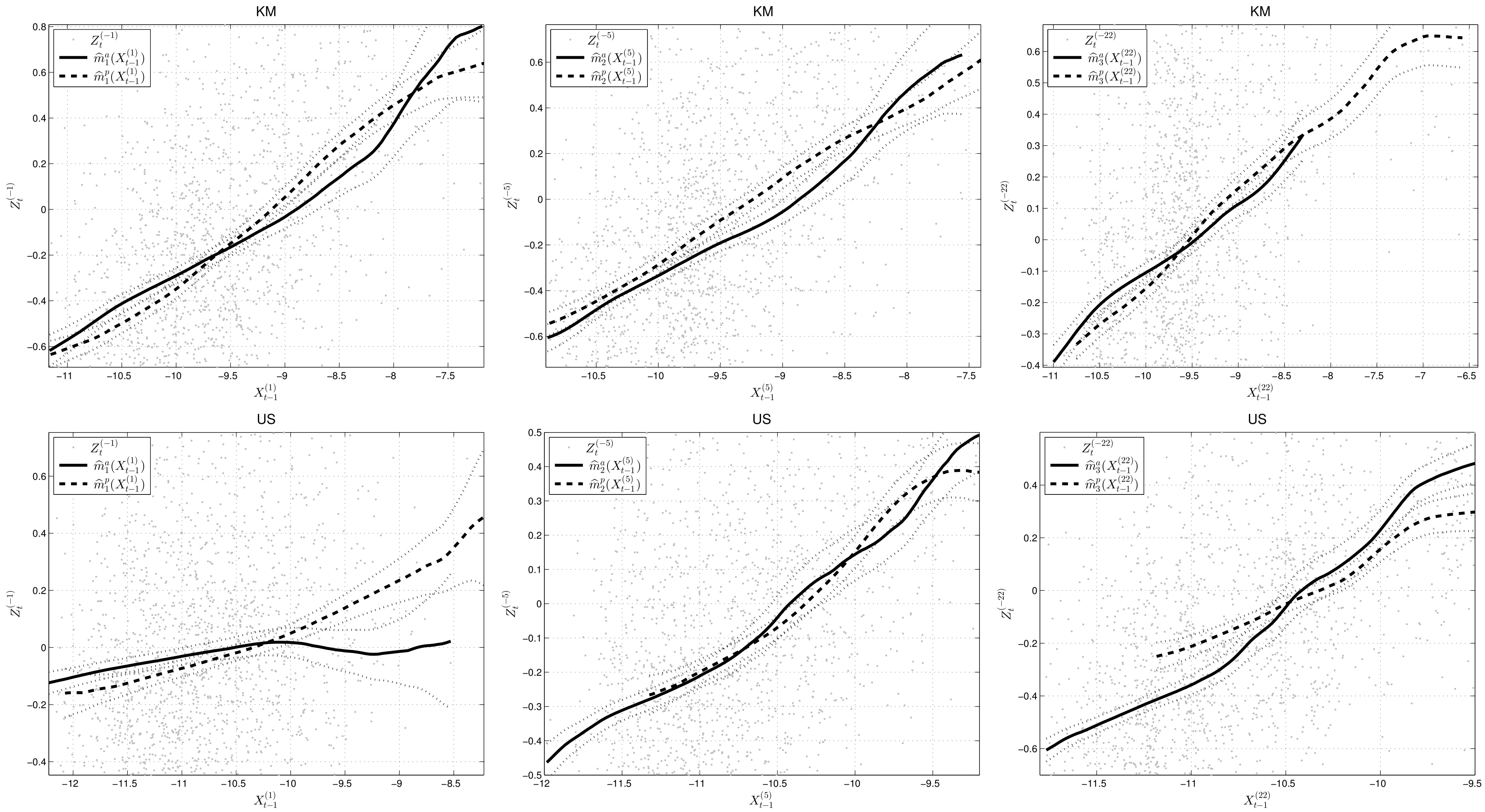

Figure 7: Nonparametric HAR models with pointwise asymptotic 95% confidence intervals for the smooth backfitting estimates for KOSPI 200 Index (KM, top panel) and the TBond 30yr (US, lower panel) for the ante and post sample. $Z_t^{(-\iota_j)}$ is the residual response variable after subtracting all component functions m_i, $i \neq j$.