Universität St.Gallen

Additive modeling of realized variance: tests for parametric specifications and structural breaks

Matthias R. Fengler, Enno Mammen, Michael Vogt

November 2013 Discussion Paper no. 2013-32

School of Economics and Political Science, Department of Economics

University of St. Gallen

Editor: Martina Flockerzi
University of St.Gallen
School of Economics and Political Science
Department of Economics
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35

Electronic Publication: http://www.seps.unisg.ch

Additive modeling of realized variance: tests for parametric specifications and structural breaks[1]

Matthias R. Fengler, Enno Mammen, Michael Vogt

Authors' address:

Matthias R. Fengler
Institute for Mathematics and Statistics
University of St.Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 224 2457
Fax +41 71 224 2894
Email matthias.fengler@unisg.ch

Enno Mammen
Department of Economics
University of Mannheim
L7, 3-5
D-68131 Mannheim
Tel. +49 621 181 1927
Email: emammen@rumms.uni-mannheim.de

Michael Vogt
Department of Mathematics and Statistics
University of Konstanz
D-78457 Konstanz
Tel. + 49 7531 88 2757
E-mail michael.vogt@uni-konstanz.de

[1] Supported by Swiss National Science Foundation, Grant 144033: "Analysis and models of cross asset dependency structures in high-frequency data" and by the Deutsche Forschungsgemeinschaft Project MA1026/11-1.

Abstract

For an additive autoregression model, we study two types of testing problems. First, a parametric specification of a component function is compared against a nonparametric fit. Second, two nonparametric fits of two different time periods are tested for equality. We apply the theory to a nonparametric extension of the linear heterogeneous autoregressive (HAR) model. The linear HAR model is widely employed to describe realized variance data. We find that the linearity assumption is often rejected, in particular on equity, fixed income, and currency futures data; in the presence of a structural break, nonlinearity appears to prevail on the sample before the outbreak of the financial crisis in mid-2007.

Keywords

Additive models; Backfitting; Nonparametric time series analysis; Specification tests; Realized variance; Heterogeneous autoregressive model.

JEL Classification

C14, C58.

1 Introduction

A major workhorse used in the research on financial market volatility over the past decade is the heterogeneous autoregressive model (HAR). When applied to time series of realized variance (RV) data estimated from intra-day tick prices, it captures the stylized facts of volatility data in an accurate manner and shows a forecasting power that is hard to beat. Although much effort has been invested into further refinements of the model, the actual improvements achieved in terms of forecasting ability against the simplest baseline specification have proved to be comparably tiny.[1]

The popularity of this model, first suggested in Corsi (2009), may be attributed to the following characteristics: (1) it builds on the intuitive heuristic of the heterogeneous market hypothesis, according to which the observed variance process emerges as a cascade of variance components induced by market participants with daily, weekly and monthly investment horizons; (2) the model can approximate long range dependence of volatility data in a simple manner; and (3) it is parsimonious and straightforward to estimate, as it is a restricted AR(22) model.

As a core assumption in the HAR model, the component functions are linear in the regressors. But is this assumption supported by the data? In this work, we provide the diagnostic tools to answer this question. We first generalize the baseline HAR model by allowing its component functions to have an arbitrary nonparametric form. This nonparametric HAR model belongs to the class of additive models. We then suggest a procedure to test whether the nonparametric components of the model have a parametric structure. Importantly, our method allows us to test each component separately for parametric specification. It can thus be used to verify the linearity assumption of the HAR model for each of the component functions. To accommodate potential time variation of the component functions, we additionally propose a test for a structural break. This test allows for checking each component function for an unknown functional change, while the remaining functions may or may not undergo a structural break. For both tests, the wild bootstrap procedure to obtain critical values is shown to be consistent. In the empirical part, we apply the theory to a large data set of high-frequency ticks of futures and indices derived from a variety of underlying assets and test the linearity assumption of the baseline HAR model.

Few studies have addressed the topic of potential nonlinearities in the HAR model for RV. McAleer and Medeiros (2008a) consider a multiple regime smooth transition model where regimes are triggered by large negative returns. In a similar vein, Corsi et al. (2012) propose a tree-structured HAR model, in which regimes depend both on returns

[1] See Corsi et al. (2012) for a recent survey.

and variance levels. Chen et al. (2013) suggest a linear HAR model whose coefficients are allowed to vary slowly in time. All these studies, however, assume a specification that is still linear conditionally on the regime or locally in time. We differ from these studies in allowing the variance component functions to be fully nonparametric.

Our model is related to regression and autoregression models, where the (auto-)regression function is modeled as a sum of nonparametric functions. Nonparametric additive models were first studied in Stone (1985, 1986) where it was pointed out that they are an efficient compromise between modeling approaches that are too complex on the one hand or too inaccurate on the other. Backfitting is one of the state of the art techniques for estimating additive models; see Hastie and Tibshirani (1990). In the testing procedures that we will discuss in this paper, we will make use of the smooth backfitting algorithm introduced in Mammen et al. (1999). Smooth backfitting avoids problems of the ordinary backfitting algorithm where covariates are too strongly correlated, and its asymptotic statistical properties are also better understood. Other estimation approaches for additive models include sieve estimators (Chen; 2007), penalized splines (Eilers and Marx; 2002), and marginal integration (Newey; 1994; Tjøstheim and Auestad; 1994).

Our test statistics are L_2-distances between a nonparametric and a parametric fit and between two nonparametric fits for two different time periods. Tests of this form have been proposed in a series of papers for a large range of testing problems in non- and semiparametric regression. Early research includes Härdle and Mammen (1993), González-Manteiga and Cao-Abad (1993), Hjellvik et al. (1998), Zheng (1996), and Fan et al. (2001). Whereas these papers treat i.i.d. data, more recent work discusses testing problems in time series models; see, e.g., Dette and Sprekelsen (2004), Kreiß et al. (2008), Aït-Sahalia et al. (2009), and Leucht (2012). Test problems with additive alternatives are considered in Fan and Jiang (2005) and Haag (2008). In this paper, we will develop a complete asymptotic theory for two types of testing problems. First, we will treat the testing problem where the parametric specification of an additive component is tested against a nonparametric fit. In a second step, we consider the case where the nonparametric fits for different time periods are compared. The theory applies to the case of a nonparametric HAR model.

In an ample empirical analysis, we estimate the nonparametric HAR model on 17 RV time series of global futures and indices on equity, fixed income instruments, currencies, metals, and energy and agricultural commodities. Across the entire sample period from July 2003 to December 2010, at a significance level at 10%, our tests uncover nonlinear features in about 27 of the 51 (i.e., three component functions times 17) additive functions to be estimated, most often in the daily component. In many cases, the nonlinear functional patterns continue to exist after accounting for a structural break. Hence nonlinearity is not an artifact of missing a structural break in an otherwise linear model. Across the different asset classes, nonlinearity in our sample is most prevalent for equity instruments and currency instruments.

While the number of positive test outcomes seems large, the actual degree of nonlinearity in the regressors is in fact moderate. For instance, we typically find that daily component functions are mildly convex, implying that the marginal impact of today's RV has a larger impact on tomorrow's RV in turbulent times than in calm market conditions. Conversely, if nonlinear, weekly and monthly component functions tend to be concavely shaped. This implies that the predictive power of weekly and monthly components is largest in calm market conditions. Finally, during times of market distress, we observe a preference for the simple model, as in most cases the linear specification is not rejected on the post sample that comprises the financial crisis.

In summary, our empirical study sheds light on RV modeling in various ways. Corroborating the forecasting literature, we conclude that the linear specification of the HAR model is well taken, particularly for non-equity instruments and generally in turbulent market phases. The moderate deviations from the linear model explain why nonlinear extensions may attain only incremental improvements of the forecasting ability of the baseline model. In addition to these insights, our results provide intuition about how to extend existing parametric modeling approaches for RV data. One obvious possibility would be a regime-switching model with parsimoniously parametrized nonlinear component functions conditional on the regimes.

The paper is structured as follows. In Section 2, we introduce the nonparametric HAR model. The specification test is introduced in Section 3, the structural break test in Section 4. Section 5 presents the data, after which follows the empirical analysis in Section 6. An appendix provides proofs and additional information on realized variance estimation.

2 The nonparametric HAR model

We investigate a nonparametric extension of the heterogeneous autoregressive (HAR) model introduced by Corsi (2009) for modeling realized variance. Let V_t be a daily observation of the RV time series and define $V_t^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} V_{t-j}$, $n \in \mathbb{N}_+$, i.e., as an average of V_t over the past n trading days. Furthermore, denote by $\iota = (\iota_1, \dots, \iota_d)^\top \in \mathbb{N}_+^d$ an index set with $\iota_1 < \iota_2 \dots < \iota_d$. The model is given by

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

where m_0 is a constant, $m_j(\cdot)$, $j = 1, \ldots, d$, are smooth nonparametric link functions of unknown shape, also called the (variance) component functions, and $\mathbb{E}[\varepsilon_t | V_{t-1}^{(\iota_1)}, \ldots, V_{t-1}^{(\iota_d)}] = 0$. Formally, this extends the notion of an additive model as introduced by Hastie and Tibshirani (1990) to the time series setting; see Chen and Tsay (1993).

Usually, the link functions 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 standard, yet parsimoniously parametrized AR(ι_d) model, since the autoregressive coefficients are restricted to be equal for each of the variance components. A very common index set 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 conceptual idea that market participants with different investment horizons, such as daily, weekly, and monthly time-scales, are active in the market. In this framework, one assumes that short-term variance does not impact long-term investment behavior, whereas long-term variance affects short-term trading decisions through expectations of future risk. By recursive substitutions of each long-term variance forecast into the dynamic equation of the next level shorter dated variance, an additive cascade of variance components emerges, where short-term variance is a function of past variance at the same time-scale and the expectations of the longer-term period variance components; see Corsi (2009) for details.

This motivates the model structure as given in (1). The linear specification, however, is ad hoc and at best justified by the enormous body of literature employing the linear HAR model for prediction purposes. In the following, we will formally test this assumption.

3 Testing for a parametric specification

In what follows, we investigate the question of whether one of the additive components in model (1), say m_j, admits a certain parametric form. Put differently, we want to test whether m_j 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

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

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] This choice is suggested in Corsi (2009) and has frequently been adopted in the literature. Testing the index $\iota = (1, 5, 22)^\top$ itself is beyond the scope of this text. For linear models, this is done in Craioveanu and Hillebrand (2010) and Audrino and Knaus (2012).

3.1 The test statistic

We develop our testing procedure in a general additive time series regression setup which nests the nonparametric HAR model (1) as a special case. The setup is given by the equation

$$Y_t = m_0 + \sum_{j=1}^{d} m_j(X_{t,j}) + \varepsilon_t \quad \text{for } t = 1, \dots, T \tag{2}$$

with $\mathbb{E}[\varepsilon_t | \boldsymbol{X}_t] = 0$ and $\boldsymbol{X}_t = (X_{t,1}, \dots, X_{t,d})^\top$. The nonparametric HAR model is obtained by setting $Y_t = V_t^{(\iota_1)}$ and $\boldsymbol{X}_t = \left(V_{t-1}^{(\iota_1)}, \dots, V_{t-1}^{(\iota_d)}\right)^\top$. To identify the additive component functions $m_1, \dots, m_d$ in (2), 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, for example $[0,1]^d$. The case of unbounded support can be incorporated by slightly modifying the test statistic. We comment on this in Appendix A.

We are interested in the question of whether one of the component functions $m_1, \dots, m_d$ has a parametric form. Restricting attention to the first component m_1, the null hypothesis is given by

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

If the functions $m_2, \dots, m_d$ were known, we could base our test on the one-dimensional model

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

with $Z_t = Y_t - \sum_{j=2}^{d} m_j(X_{t,j})$. In this case, standard nonparametric procedures could be used to test the hypothesis H_0. In particular, one 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 the functions $m_2, \dots, m_d$, we replace them by a set of estimators. In particular, we estimate them 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

$$\sum_{t=1}^{T} \int_0^1 \Big\{ Y_t - f_0 - \sum_{j=1}^{d} f_j(x_j) \Big\}^2 K_g(x_j, X_{t,j}) dx_j, \tag{3}$$

where the minimization runs over all additive functions $f(x) = f_0 + f_1(x_1) + \dots + f_d(x_d)$ whose components satisfy $\int_0^1 f_j(x_j) \tilde{p}_j(x_j) dx_j = 0$ for $j = 1, \dots, d$. Here, $\tilde{p}_j$ is a standard kernel density estimator of p_j given by $\tilde{p}_j(x) = \frac{1}{T} \sum_{t=1}^{T} K_g(x_j, X_{t,j})$. Moreover, g is the bandwidth and

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

is a modified kernel weight, where $K_g(v) = \frac{1}{g}K(\frac{v}{g})$ and the kernel function $K(\cdot)$ integrates to one. These weights have the property that $\int_0^1 K_g(v,w)dv = 1$ for all w, which is needed to derive the asymptotic results for the backfitting estimates.

Given the estimators $\tilde{m}_2, \dots, \tilde{m}_d$, the variables Z_t can be approximated by $\tilde{Z}_t = Y_t - \sum_{j=2}^d \tilde{m}_j(X_{t,j})$. 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 (A7) 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 estimators $m_{\hat{\theta}}$ 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

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

where π is a weight function with bounded support and $\int \pi(x)dx = 1$ and

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

As proposed in Härdle and Mammen (1993), we smooth the parametric estimator $m_{\hat{\theta}}$ 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 $m_2, \dots, m_d$ 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 estimators by $\tilde{m}_2^{LL}, \dots, \tilde{m}_d^{LL}$ and write $\tilde{Z}_t^{LL} = Y_t - \sum_{j=2}^d \tilde{m}_j^{LL}(X_{t,j})$. 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

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

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.

3.2 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. To start with, we derive the limiting distribution of S_T under the null hypothesis, i.e., for a parametric function $m_1 = m_{\theta_0}$ with $\theta_0 \in \Theta$. To get an idea of the power of the test, we additionally compute the asymptotic distribution under local alternatives of the form $m_1(w) = m_{1,T}(w) = m_{\theta_0}(w) + c_T \Delta(w)$ where 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)–(A7) of Appendix A are satisfied and let $h = O(T^{-1/5})$ as well as $g = O(T^{-1/4-\delta})$ for some small $\delta > 0$. Then*

$$S_T - B_T - \textstyle\int (\mathcal{K}_h \Delta)^2 \pi(w) dw \xrightarrow{d} N\big(0, V\big)$$

with $\mathcal{K}_h g(\cdot) = \int K_h(\cdot - u) g(u) du$,

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

$$V = 2\kappa_1 \int \frac{[\sigma^2(w)]^2 \pi^2(w)}{p_1^2(w)} dw,$$

where 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 $\kappa_1 = \int (\int K(u) K(u+v) du)^2 dv$.

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

3.3 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 $\{\eta_t\}$ is some sequence of i.i.d. variables with zero mean and unit variance that is independent of the sample $\{\tilde{Z}_t, X_{t,1}\}_{t=1}^T$. Here, we have used the unrestricted estimator $\hat{m}$ to construct the residuals $\hat{\varepsilon}_t$. Alternatively, it is possible to work with residuals which are based on the fit $m_{\hat{\theta}}$ under the null. Denote by $m_{\hat{\theta}^*}$ and $\hat{m}^*$ the parametric and nonparametric estimator of m_1 calculated from the bootstrap sample $\{Z_t^*, X_{t,1}\}$. 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 above-defined bootstrap is consistent.

THEOREM 2. *Assume that the conditions (A1)–(A6) and (A7*) of Appendix A are satisfied and let $h = O(T^{-1/5})$ as well as $g = O(T^{-1/4-\delta})$ for some small $\delta > 0$. Then*

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

conditional on the sample $\{\tilde{Z}_t, X_{t,1}\}_{t=1}^T$ with probability tending to one.

4 Testing for breaks

4.1 The test statistic

In this section, we will discuss tests for breaks in the autoregression functions. Our main motivation comes from the concern that the nonlinear structure detected by the specification test may be spurious due to a neglected break in the functional components. In the presence of a structural break, our data follow the nonparametric autoregression model

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

We will discuss tests of break points based on the comparison of estimators of m_j^{ante} and m_j^{post}. Here, t^* is the break point. In this paper, we assume that t^* is known. Our discussion carries over to the case when t^* is unknown and when it can be estimated by using additional data. Our theory changes if the break point is estimated by using only the observations $X_1, ..., X_T$. This is because for the hypothesis where $m_j^{ante} \equiv m_j^{post}$, the break point is not defined and thus t^* cannot be replaced by a consistent estimator.

Given the break point, we are interested in tests that separately check if there are changes in m_j^{ante} and m_j^{post}, $j = 1, \ldots, d$, i.e., we test the hypothesis

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

Without loss of generality we now give an explicit definition of the test statistic only for the case $j = 1$. Our test is based on smooth backfitting fits of the functions m_j^{ante} and m_j^{post}. We therefore modify the discussion following Equation (3) in the last section to this new setting. The Nadaraya-Watson smooth backfitting estimators $\tilde{m}_0^l, \ldots, \tilde{m}_d^l$ ($l = ante, post$) are defined as the minimizers of the criterion

$$\sum_{t \in \mathfrak{T}_l} \int_{I_j} \Big\{ Y_t - f_0 - \sum_{j=1}^{d} f_j(x_j) \Big\}^2 K_g(x_j, X_{t,j}) dx_j, \tag{5}$$

where $\mathfrak{T}_{ante} = \{t : 1 \le t \le t^*\}$ and $\mathfrak{T}_{post} = \{t : t^* + 1 \le t \le T\}$. The minimization runs over all additive functions $f(x) = f_0 + f_1(x_1) + \cdots + f_d(x_d)$ whose components satisfy $\int_0^1 f_j(x_j) \tilde{p}_j^l(x_j) dx_j = 0$ for $j = 1, \ldots, d$. Here, $\tilde{p}_j^l(x)$ is equal to $\frac{1}{T} \sum_{t \in \mathfrak{T}_l} K_g(x_j, X_{t,j})$. Up to a factor, this can be interpreted as a kernel estimator of the average density of $X_{t,j}$ for $t \in \mathfrak{T}_l$ with j and l fixed. Moreover, the kernel $K_g(v, w)$ is defined as in the last section. Note that in case of a breakpoint, it makes no sense to assume that $X_{t,j}$ have the same distribution for $t \in \mathfrak{T}_l$ with j and l fixed.

Given the estimators $\tilde{m}_2^l, \ldots, \tilde{m}_d^l$, for $l = ante, post$, the variables Z_t^l can be approximated by $\tilde{Z}_t^l = Y_t - \sum_{j=2}^d \tilde{m}_j^l(X_{t,j})$ for $t \in \mathfrak{T}_l$. Based on the samples $\{\tilde{Z}_t^l, X_{t,1}\}_{t \in \mathfrak{T}_l}$, we can construct the Nadaraya-Watson smoother of m_1^l with bandwidth h, i.e.,

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

For $j \ne 1$ the estimators $\hat{m}_j^l$ are defined by analogue constructions. We are now in the position to define our test statistic as

$$S_{j,T} = T h^{1/2} \int \left(\mathcal{K}_{h,T}^{j,post} \hat{m}_j^{ante}(x) - \mathcal{K}_{h,T}^{j,ante} \hat{m}_j^{post}(x) - \tilde{\delta} \right)^2 \pi(x) dx \tag{6}$$

where π is a weight function with $\int \pi(x) \, dx = 1$ and

$$\begin{aligned}
\mathcal{K}_{h,T}^{j,post} g(\cdot) &= \frac{\sum_{t=t^*+1}^{T} K_h(\cdot, X_{t-1,j}) g(X_{t-1,j})}{\sum_{t=t^*+1}^{T} K_h(\cdot, X_{t-1,j})}, \\
\mathcal{K}_{h,T}^{j,ante} g(\cdot) &= \frac{\sum_{t=2}^{t^*} K_h(\cdot, X_{t-1,j}) g(X_{t-1,j})}{\sum_{t=2}^{t^*} K_h(\cdot, X_{t-1,j})}, \\
\tilde{\delta} &= \int \left(\mathcal{K}_{h,T}^{j,post} \hat{m}_j^{ante}(x) - \mathcal{K}_{h,T}^{j,ante} \hat{m}_j^{post}(x) \right) \pi(x) dx.
\end{aligned}$$

Note that $\tilde{\delta}$ is chosen such that

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

We now motivate the construction of this test statistic. In a first attempt, one could consider the test statistic $\min_{\delta \in \mathbb{R}} Th^{1/2} \int \left(\hat{m}_j^{ante}(x) - \hat{m}_j^{post}(x) - \delta \right)^2 \pi(x) dx$. This test statistic has the problem that $\hat{m}_j^{ante}(x)$ and $\hat{m}_j^{post}(x)$ have different asymptotic bias terms. One can show that therefore this 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, $\mathcal{K}_{h,T}^{j,post} \hat{m}_j^{ante}(x)$ and $\mathcal{K}_{h,T}^{j,ante} \hat{m}_j^{post}(x)$ have the same asymptotic bias and thus the bias terms cancel when we take the difference of the two smoothed estimators.

We now discuss the asymptotic distribution of $S_{j,T}$ in the following setting: the functions m_k^ℓ are fixed for $(k, \ell) \neq (j, post)$. For $(k, \ell) = (j, post)$ we assume that

$$m_j^{post}(x) = m_j^{ante}(x) + T^{-1/2} h^{-1/4} \Delta(x).$$

One can show that the asymptotics of $S_{j,T}$ does not change if $m_\ell^{post}(x) - m_\ell^{ante}(x)$ also converges to 0 for $\ell \neq j$. For $\Delta \equiv 0$ we get a specification that lies on our hypothesis; for $\Delta \not\equiv 0$ we get a neighbored point of the alternative.

Bearing in mind the asymptotic discussion in the last section, we have to take care of the following two points: (i) Because of the break point, the process $(X_{t,1}, \ldots, X_{t,d}, Y_t)$ is no longer stationary. (ii) We have to show that by the additional smoothing operations $\mathcal{K}_{h,T}^{j,ante}$ and $\mathcal{K}_{h,T}^{j,post}$ the bias terms cancel in the test statistics. Details are given in Appendix B. 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 $t \ll t^*$ and that X_t is approximated by X_t^{post} for $t \gg t^*$; see Appendix B. In Appendix B, we also provide conditions under which these approximations apply to the nonparametric HAR model of Section 2.

The limit distribution of $S_{j,T}$ is given by the following theorem:

THEOREM 3. *Under assumptions (B1)–(B5) of Appendix B, the statistic*

$$S_{j,T} - h^{-1/2} K^{(2)}(0) \int_{[0,1]} \left[c^{-1} \sigma_{ante}^2(x) + (1-c)^{-1} \sigma_{post}^2(x) \right] \pi(x)\ dx$$

has a limiting normal distribution with mean $M_S = \int \Delta^2(x) \pi(x) dx - \left[\int \Delta(x) \pi(x) dx \right]^2$ *and variance*

$$V_S = 2K^{(4)}(0) \int_{[0,1]} \left[\frac{1}{c^2} \frac{\sigma_{ante}(x)^4}{p_j^{ante}(x)^2} + \frac{2}{c(1-c)} \frac{\sigma_{ante}(x)^2 \sigma_{post}(x)^2}{p_j^{ante}(x) p_j^{post}(x)} + \frac{1}{(1-c)^2} \frac{\sigma_{post}(x)^4}{p_j^{post}(x)^2} \right] \pi(x)^2\ dx.$$

Here, $\sigma_{ante}^2(x)$ *is the conditional variance of* ε_t^{ante} *given* $X_{t-1,j} = x$ *and* $\sigma_{post}^2(x)$ *is the conditional variance of* ε_t^{post} *given* $X_{t-1,j} = x$. *Furthermore,* $K^{(r)}$ *denotes the* r*-times convolution product of* K *(for* $r \geq 1$*).*

In the statement, c denotes the limit of t^*/T for $T \to \infty$. As an alternative one could consider the test statistic $\min_{\delta \in \mathbb{R}} Th^{1/2} \int \left(\tilde{m}_j^{LL,ante}(x) - \tilde{m}_j^{LL,post}(x) - \delta\right)^2 \pi(x)dx$, where $\tilde{m}_j^{LL,ante}$ and $\tilde{m}_j^{LL,post}$ are local linear smooth backfitting estimators. Now no additional smoothing of the estimators is required because the asymptotic bias terms of the two estimators do not differ. The reason is that bias terms of the local linear estimator do not depend on the density of the covariables $(X_{t,1}, \ldots, X_{t,d})$. This is the case for local linear estimation in classical regression models and it also holds for the smooth backfitting estimator in additive models; see Mammen et al. (1999). In our applications of both test statistics, the statistic based on Nadaraya-Watson smoothing and the statistic based on local linear smoothing, it turned out that the local linear estimator is highly variable in the boundary region and that this leads to instabilities of the local linear test. As we will point out, our conclusion is that Nadaraya-Watson smoothing is preferable in testing whereas local linear smoothing leads to more reliable estimation results.

4.2 Bootstrap

To test against the parametric alternative, we suggest bootstrapping the test statistic to improve on the small sample behavior. For simplicity of the exposition, set $j = 1$, for example, and denote the bootstrap sample as $\{Y_t^*, X_{t,1}, \ldots, X_{t,d}\}_{t=1}^T$ with

$$Y_t^* = \begin{cases} \tilde{m}_0^{ante} + \bar{m}_1\left(X_{t,1}\right) + \sum_{j=2}^d \tilde{m}_j^{ante}\left(X_{t,j}\right) + \varepsilon_t^{*,ante} & \text{for } t \leq t^*, \\ \tilde{m}_0^{post} + \bar{m}_1\left(X_{t,1}\right) + \sum_{j=2}^d \tilde{m}_j^{post}\left(X_{t,j}\right) + \varepsilon_t^{*,post} & \text{for } t > t^*. \end{cases}$$

where $\bar{m}_1$ is an average of $\tilde{m}_1^{ante}$ and $\tilde{m}_1^{post}$. The bootstrap residuals are constructed as $\varepsilon_t^{*,k} = \hat{\varepsilon}_t^k \cdot \eta_t$, where $\hat{\varepsilon}_t^k = Y_t - \tilde{m}_0^k - \sum_{j=1}^d \tilde{m}_j^{ante}\left(X_{t,j}\right)$ with $k = ante, post$, are the estimated residuals and $\{\eta_t\}$ is some sequence of i.i.d. variables with zero mean and unit variance that is independent of the sample $\{Y_t, X_{t,1}, \ldots, X_{t,d}\}_{t=1}^T$. Here, we define bootstrap residuals under the alternative. As also discussed in the last section, one can also use the residuals of the fit under the null hypothesis. Asymptotically, for neighbored alternatives both boostrap tests will have the same power. Thus differences in their performance must be checked by finite sample simulations. In our simulations, we did not find major differences.

Denote by $\tilde{m}_1^{*,k}(x)$, $k = ante, post$, the bootstrap analogue of $\tilde{m}_1^k(x)$. The bootstrap statistic is then defined as

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

where

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

The following theorem states that the bootstrap works.

Theorem 4. *Assume that the conditions (B1)–(B5) of Appendix B are satisfied. Then*

$$S^*_{j,T} - h^{-1/2}K^{(2)}(0)\int_{[0,1]}\left[c^{-1}\sigma^2_{ante}(x) + (1-c)^{-1}\sigma^2_{post}(x)\right]\pi(x)\; dx \;\xrightarrow{d}\; N\big(0, V_S\big)$$

conditional on the sample $\{Y_t, X_{t,1}, \ldots, X_{t,d}\}_{t=1}^T$ *with probability tending to one.*

5 Data description

The high-frequency data we use ranges from 2003 to 2010 and is provided by *Tick Data*[3] offering validated tick by tick data on a large number of global equity, currencies, commodities and interest rate futures and indices. The symbols we use for this work are summarized in Table 1. Prior to estimation of the models, the raw price data are cleaned as suggested in Barndorff-Nielsen et al. (2009). We then construct an equidistant 5-minutes tick data series from observed prices by means of the previous tick method. This follows suggestions in Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001).[4] Futures are rolled based on the day tick count for the various contracts traded to ensure that only price data from the most liquidly traded contract is employed. Finally, we compute intra-day log-returns.

For this work, we are only interested in modeling jump-free measures of realized variance, i.e., estimates of integrated variance. This is because including jumps as predictors is questionable in nonparametric modeling given the sparsity of jump data. Furthermore, the relevance of jumps in RV regressions is still subject to research (Andersen et al.; 2007; Corsi et al.; 2010, among others). We therefore subject the return series to a testing procedure (detailed in Appendix C) by which we separate jumps from non-jump returns. Throughout this study, we only consider intraday returns, i.e., we ex ante qualify overnight returns as jumps. Our estimate of daily RV is then estimated by summing the intraday squared 5-minutes returns, but correcting for the contribution of jumps; see Appendix C. Finally, we compute the (log) weekly and monthly RV series using the index set $\iota = (1, 5, 22)$, i.e. we set $V_t^{(1)} = \log RV_t$ to compute $V_t^{(5)}$ and $V_t^{(22)}$ as explained in Section 2. The log-transformation is commonly applied to RV data.

For illustration, we present three representative daily log-RV series of the S&P500 index (SP), the 10yr TNote (TY) and natural gas (NG) in Fig. 1. The equity index hovered at about $11\% \approx \sqrt{250\exp(-10)}$ annual volatility till mid 2007, while the fixed income series even declined from 8% to about 3% annual volatility. Both series soared through 2007-2008, where equity futures witnessed levels of RV of almost 80% annualized

[3]See `www.tickdata.com`.

[4]The previous tick method assigns the last traded price to the sampling time; see Dacorogna et al. (2001) and Hansen and Lunde (2006) for detailed discussions of alternative sampling schemes.

volatility at the climax of the financial crisis. For both instruments, RV declined through 2009 but picked up again during 2010 with the Euro-zone sovereign bond crisis looming. This is in stark contrast to the natural gas future (NG) which is traded at high RV levels of about 40% annualized volatility in the entire sample. This, however, does not imply that NG was unaffected by the crisis. In fact, and in anticipation of our empirical results, we find a structural break in NG as well as in TY, but not in SP, which demonstrates the limits of visual inspection of nonlinear time series.

6 Empirical analysis

For the empirical analysis we proceed as follows. We first make some preliminary comments on bandwidth choice and model estimation in Section 6.1. In Section 6.2, we estimate the additive model on the whole sample and test each of the variance components, i.e., the daily, weekly, and monthly functions, against linearity. In Section 6.3, we ask whether these results are driven by a structural break that could have occurred at the outbreak of the financial crisis in mid 2007. If our test identifies a structural break, we repeat the tests for nonlinearity on each subsample. This is to check whether the identified nonlinear shapes are artifacts of ignoring a structural break. In Section 6.4, we report further results on alternative testing procedures and robustness checks.

6.1 Bandwidth choice

For bandwidth selection of the smooth backfitting estimators $\tilde{m}_j$, we use a plug-in rule which works as follows. We calculate the integrated mean squared error of $\tilde{m}_j$; see Mammen et al. (1999) for the relevant expressions. Since the asymptotic bias expression of the Nadaraya-Watson based backfitting estimators is very involved, we work with the bias expression of the local linear backfitting instead. We then iteratively choose the bandwidths which minimize the integrated mean squared error. This follows bandwidth selection rules for additive models proposed in Mammen and Park (2005). The bandwidths resulting from this procedure are collected in Table 2.

As is apparent from Theorems 1 and 3, the initial smooth backfitting estimates need to be undersmoothed. To achieve this undersmoothing, these initial fits are obtained using bandwidths that are 30% smaller than those displayed in Table 2.

6.2 Tests against the linear specification

We now test the linearity hypothesis of the simple HAR model for each variance component function. The tests are implemented as prescribed by Theorems 1 and 2 for the Nadaraya-Watson based smooth backfitting estimator. The bootstrap sample is constructed as described in Equation (4). We experimented by drawing residuals from both the alternative and the null hypothesis. Both results are very similar, with the latter slightly more conservative. We hence only report results obtained for residuals drawn from the null hypothesis. The weight function is the empirical density of $X_{t,j}$, which mitigates potential boundary effects. All p-values are computed from 10,000 bootstrap replications.

From Table 3, it is evident that there is a case for nonlinearity. For the daily component, in 11 out of the 17 series, we find evidence of there being a deviation from the simple linear specification at p-values below 10% (CF, FT, XX, BN, TY, US, NG, HG, EC, JY, SY; see Table 1 for the acronyms). For the weekly component, the evidence is weaker. Here, we find nonlinearity in about half the cases (CF, XX, BN, CL, EC, JY, CN). For the monthly component, 9 tests reject the linear specification at a 10% significance level (FT, KM, NE, SP, TY, US, HG, EC, CN).

Across asset classes, evidence against linearity appears to be strongest among assets that can be expected to be most strongly affected during an economic downturn, such as assets under distress such as equity (CF, FT, KM, NE, SP, XX), for all of which at least one component is nonlinear, and safe haven assets such as fixed income (BN, TY, US) and currency futures (EC, JY), which even have two nonlinear components. For assets whose prices are predominantly determined by long-term global consumption perspectives, such as energy (CL, NG), metal (HG) and food commodities (SY), the evidence is less compelling, as typically only one component is affected at comparatively large p-values. Interestingly, despite its role as a primary safe haven asset during the financial and the sovereign bond crisis, gold (GC) appears to be best fitted with the linear HAR model.

In Fig. 3, we display a number of nonlinear fits of the first variance component function for both the equity-based assets and natural gas (CF, FT, XX, NG in top left panel) and the fixed income and currency futures (BN, US, EC, JY in top right panel). All in all, the degree of nonlinearity is moderate. Discounting boundary effects, we find upward curving, mildly convex shapes as a common stylized facts among these daily component functions. Hence, at increased levels of RV, the marginal impact of short-term trading on future daily RV is larger. According to the usual interpretation of the HAR model, this means that at higher variance levels, daily trading activities drive RV more predominantly than in calm markets. This is in contrast to the weekly and monthly component functions shown in the lower panels of Fig. 3. They appear to be concavely shaped, which implies that the marginal predictive relevance of the weekly and monthly variance components on future RV diminishes at higher variance levels. Thus, there appears to be a shift in the relative predictive importance of the different variance functions depending on the level of daily realized variance. Allowing for nonlinear variance components captures these patterns nicely.

In Fig. 4, we display the two examples of nonparametric HAR model for the French index CAC40 (CF, top panel) and USD-Yen currency future (JY, lower panel). The estimated functions are contrasted with the linear fits. According to our tests in Table 3 for CF, there is nonlinearity in the components $\tilde{m}_1$ and $\tilde{m}_2$, but not in $\tilde{m}_3$. This is confirmed by the plots. Because of convexity, the linear model overfits $\tilde{m}_1$ in the middle of the domain and underfits in the boundary regions. Converse observations apply to $\tilde{m}_2$. Clearly, $\tilde{m}_3$ for CF is not different from the linear fit. For $\tilde{m}_1$ of JY, similar observations as for the daily variance function of CF apply. For $\tilde{m}_2$ the evidence is less compelling, as the p-value is at about 7%. Therefore, this function could also be assumed to be linear – as well as obviously $\tilde{m}_3$.

6.3 Structural break tests and specification tests on subsamples

6.3.1 Choice of break date

The data set includes the global financial crisis following the collapse of the U.S. real estate markets towards the end of 2007 as well as the initial phase of the subsequent Euro-zone sovereign bond crisis beginning early in 2010. The severity of the market distress during the financial crisis can be inferred from Fig. 2, which shows daily S&P500 index closing prices along with the spread of the London interbank offered rate over the overnight indexed swap (Libor-OIS, 3-months, USD). The Libor-OIS spread is a widely recognized measure of credit risk within the banking sector (Thornton; 2009). As is visible, the spread is close to zero up to July/August 2007, after which it spikes up, reaching unprecedented levels of more than 350 basis points during the climax of the crisis in 2008/2009.

In structural break analysis, we will use the sharp increase of the spread in July/August 2007 as an indicator for a potential structural break in the RV series. More precisely, the break date is assumed to be July 25, 2007, where the sample is split. This was one of the last days on which the Libor-OIS spread was below 10 basis points. By the end of August the Libor-OIS spread was already larger than 45 basis points; see Fig. 2. In Section 6.4, we subject this choice to a sensitivity analysis.

6.3.2 Discussion of results

Structural break tests are based on the test statistic presented in Section 4. As a weight function in the statistic (6), we use $\pi = c_\pi \, p_j^{ante} p_j^{post} / (p_j^{ante} + p_j^{post})$, where the empirical density function on the subsample is denoted by p_j^i, $j = 1, 2, 3$, $i = \{ante, post\}$, and c_π is a constant such that $\int \pi(u) du = 1$. This weight function puts emphasis on the joint overlap of the ante sample and post sample data. All p-values are computed from 10,000 bootstrap replications. The specification tests against the linear HAR model on the subsamples are carried out exactly as described above.

Table 4 presents the results of the structural break tests. Breaks are identified in all three component functions but are most frequent in the daily component functions. In the daily component functions, breaks are found for the equity and fixed income-based instruments (CF, FT, KM, XX, TY, US), and corn (CN). TY and US exhibit breaks in all three components. Additional breaks in the weekly component are found for NG and SY and for the monthly component in KM, XX and EC. With few exceptions, the component functions with breaks display evidence of nonlinearity when estimated on the whole sample. It will thus be of interest to test for nonlinearity on the subsamples.

In Table 5, the subsamples are submitted to the specification test against the linear model. We first consider the daily component function. Here, for three out of the four remaining equity instruments (CF, FT, XX), the linear daily variance function is rejected in the ante sample but not in the post sample. The same observation applies to corn (CN). In contrast, in the 30 yrs TBond future data (US), the test strongly rejects linearity on the post sample. In all these cases (except CN), the tests in Table 3 already show evidence of nonlinearity on the whole sample.

For the weekly component functions, it is again the estimate on the ante sample that is rejected in most cases (US, NG, and EC). For EC, the post sample estimate is nonlinear as well. For the monthly function, evidence is more evenly distributed. For KM and EC, the tests on the ante and the post sample reject; for US and XX, the tests reject only on the ante and the post sample, respectively. In three cases (US, NG, XX), this nonlinearity was undetected when tested on the whole sample.

In Fig. 5, we contrast the component functions of interest estimated on the ante versus the post sample. As found for the estimates on the whole sample in Section 6.2, the daily component functions of the equity instruments CF, FT, XX, and corn (CN) estimated on the ante sample retain mildly convex shapes. The weekly component function of NG (middle plot in lower panel of Fig. 5) seems to consist of two approximately linear parts, more strongly sloped for low levels of RV and flatter for high levels. A similar observation applies to the monthly ante component function of EC. The 10 yrs treasury note future (TY) is the only series for which no test rejects. The nonlinearity previously detected on the whole sample is likely due to ignoring a structural break in an otherwise linear model. A linear HAR models on each subsample is a perfect choice; see Fig. 6.

In summary, the picture that emerges is twofold. First, RV series of equity, fixed income and currency futures exhibit structural breaks in at least one component function at the outbreak of the crisis period; second, the nonlinearity in the component function remains robust in at least one of the subsamples, usually in the ante, i.e., the pre-crisis sample. Interestingly, we find that the linear model is hardly rejected on the post sample: in crisis times, it is the simple model that fits best.

6.4 Additional statistical procedures and robustness checks

In order to complete the picture of our empirical findings, we compare our results with alternative testing procedures. As remarked in the main text, our tests could be carried out using the local linear smooth backfitting procedure. We therefore repeat the analysis of Section 6.2.[5] For most series, the local linear-based tests in Table 6 reject where the Nadaraya-Watson-based test rejects (Table 3), thus confirming our previous results. Yet, there appears a tendency of the local linear smooth backfitting-based test to find more rejections of the linearity assumption.[6] This could have been expected as local polynomials are instable when the design density is thin. In our situation, this may occur at the boundaries of the support and might therefore result in spurious evidence of nonlinearity; see Seifert and Gasser (1996) for a discussion of this issue. The Nadaraya-Watson based test is more robust in this respect.

Due to the aforementioned instability of the local linear-based testing, we do not consider this approach for the structural breaks. When the mutual support of the functions in the ante and the post sample does not fully overlap, boundary effects are overemphasized, which leads to unreasonable test outcomes. We rather consider alternative break dates. From Fig. 1 and Fig. 2, one sees that the increase in the Libor-OIS spread also marks the beginning of soaring volatility levels. Nevertheless, setting the structural break date in this way could invite criticism of being arbitrary. As a safeguard, we shifted this date two weeks before the July 25, 2007, and two and fours weeks after this date. In all cases, the conclusion about the structural break was the same as in Table 4.[7] This interpretation can also be supported by asymptotic considerations that show that the test statistic smoothly depends on the chosen break point; see the fifth bound in Assumptions (B2).

On the other hand, one could argue that by mid-2009 the crisis had calmed down, implying that one should test on a smaller post sample, for example from July 25, 2007,

[5]We again use the bandwidth from Table 2. In the plug-in procedure we used for bandwidth determination the bias of the Nadaraya-Watson-based backfitting was approximated with the bias from local linear-based backfitting; therefore the implementation is actually accurate for the local linear estimator.

[6]There are also a few cases in which Nadaraya-Watson-based tests reject the null hypothesis while the local linear-based tests do not. It should be noted, however, that with the exception of CN (weekly) and SP (monthly), p-values were about 7% or larger; thus evidence in favor of nonlinearity is already weak in the Nadaraya-Watson-based test.

[7]Table available from authors upon request.

to May 20, 2009. This date corresponds to the time when the level of the Libor-OIS spread reached about 50 basis points. We therefore ran the structural break tests on this smaller post sample. In comparison to Table 4, we find that the weekly components in NG and US as well as the monthly component function in KM do not reject the null of a structural break, whereas NE in the monthly component does.[8] Thus, on the smaller sample, which ranges over the dramatic period of the financial crisis only, we find fewer structural breaks rather than more. This may be caused by the smaller sample size. Given all these results, we conclude that our model estimated from the broader post sample and choice of the break date are justified and that the significant differences between ante and post functions are not driven by temporary changes only.

We finish this section by reporting the subsample tests against the linear HAR model based on local linear smooth backfitting; see Table 7. As observed on the entire sample, test results using the local linear-based estimator encompass the findings from the main section but also provide once more further evidence of nonlinearity. This may again be due to the sensitivity of the local linear estimator.

7 Concluding remarks

We provide the diagnostic tool to detect nonlinearity and to test for structural breaks in additive models that are estimated from dependent data by means of smooth backfitting. We apply the theory to a nonparametric extension of the linear heterogeneous autoregressive model (HAR) suggested by Corsi (2009) to model realized variance data. We find the linearity assumption is widely rejected, in particular on equity, fixed income and currency futures data; in the presence of a structural break, nonlinearity prevails on the sample ending before the outbreak of the financial crisis.

In this work, we concentrated on comparing the linear HAR specification against the additive model. This basic nonparametric HAR model could readily be extended in several ways, e.g., by adding additional parametric components that capture jumps, as in Andersen et al. (2007) or Corsi et al. (2010), and by including a nonparametric leverage function as in Corsi and Renò (2012). Furthermore, one could build on our results by replacing the nonparametric estimates by approximate parametric components and conducting a forecasting exercise. A potential avenue would be to add ideas from regime-switching models that have parsimoniously parametrized nonlinear variance component functions conditionally on the regimes. We leave this for further research.

[8]Table available from authors upon request.

A Parametric specification tests

In this appendix, we prove the results concerning the test on parametric specification from Section 3. 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 on the model components:

(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, e.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, the first partial derivatives of p exist and are continuous.

(A4) The kernel K is bounded, symmetric about zero and has compact support ($[-C_1, C_1]$, for instance). Moreover, it fulfills the Lipschitz condition that there exist a positive constant L with $|K(u) - K(v)| \leq L|u - v|$.

(A5) 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 $\mathbb{E}[\xi_t^{6+\delta}] < \infty$ for some small $\delta > 0$ and are normalized such that $\mathbb{E}[\xi_t^2] = 1$.

(A6) 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 $l \geq l^*$.

(A7) It holds that

$$m_{\hat{\theta}}(w) - m_{\theta_0}(w) = \frac{1}{T}\sum_{t=1}^{T}\langle q(w), r(X_{t,1})\rangle\tilde{\varepsilon}_t + o_p((T\log T)^{-1/2})$$

uniformly in w, where $\tilde{\varepsilon}_t = \varepsilon_t + (m_2(X_{t,2}) - \tilde{m}_2(X_{t,2}))$ 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 (A7) by an analogous assumption in the bootstrap world.

(A7*) Let $\hat{\theta}^*$ be the parameter estimate based on the bootstrap sample $\{Y_t^*, \boldsymbol{X}_t\}$. It holds that

$$m_{\hat{\theta}^*}(w) - m_{\theta_0}(w) = \frac{1}{T}\sum_{t=1}^{T}\langle q(w), r(X_{t,1})\rangle \varepsilon_t^* + o_p((T\log T)^{-1/2})$$

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

We make some brief remarks on the above assumptions. First, 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 drop the boundedness assumption in (A2) and to allow for unbounded support of $\boldsymbol{X}_t$. To do so, we have to modify the smoother $\hat{m}$ and the pilot estimators of the backfitting algorithm. Specifically, let $A = A_1 \times A_2$ be a compact subset of $\mathbb{R}^2$ and suppose that the support of the weight function π is contained in A_1, i.e., $\mathrm{supp}(\pi) \subseteq A_1$ with $\mathrm{supp}(\pi)$ being the support of π. With this notation at hand, we can 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 handle regressors with unbounded support. (A2)–(A4) are standard conditions in the smooth backfitting literature (Mammen et al.; 1999). (A6) is required to derive the uniform convergence rates of the Nadaraya-Watson estimators that enter the smooth backfitting procedure as pilot smoothers. (A5) 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 limiting distribution of the test statistic. Finally, condition (A7) is fulfilled, e.g., for weighted least squares estimators in linear models and under appropriate smoothness conditions for weighted least squares estimates 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 $\delta > 0$. Under the assumptions from above, $\tilde{m}_i$, for $i = 1, 2$, can be written as $\tilde{m}_i = \tilde{m}_i^A + \tilde{m}_i^B$ with $\tilde{m}_i^A$ having the expansion

$$\tilde{m}_i^A(w) = \tilde{m}_i^{A,NW}(w) + \frac{1}{T}\sum_{t=1}^{T} r_t(w)\varepsilon_t + o_p(T^{-1/2}) \tag{7}$$

uniformly for $w \in [0,1]$. Here, $\tilde{m}_i^{A,NW}$ is the stochastic part of a one-dimensional Nadaraya-Watson estimator given by

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

and $r_t(\cdot) = r(\boldsymbol{X}_t, \cdot)$ are random functions that are absolutely uniformly bounded and fulfill the Lipschitz condition $|r_t(w) - r_t(w')| \le C|w - w'|$. The expansion (7) 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 omit the details. 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

$$\sup_{w \in I_h} |m_i(w) - \tilde{m}_i^B(w)| = O_p(g^2) \tag{8}$$

$$\sup_{w \in I_h^c} |m_i(w) - \tilde{m}_i^B(w)| = O_p(g). \tag{9}$$

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 $X_{t,1}$. Moreover, without loss of generality set $\pi(w) = I(w \in [0,1])$ and write $\int = \int_0^1$ for short. Some 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

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

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

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

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

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

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

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}(w) dw = o_p(1) \tag{10}$$

$$Th^{1/2} \int U_{T,j}(w) U_{T,5}(w) dw = o_p(1) \tag{11}$$

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

$$S_T = Th^{1/2} \int \big(U_{T,1}(w) + U_{T,2}(w) + U_{T,4}(w)\big)^2 dw + o_p(1) =: S_T' + o_p(1) \tag{12}$$

with S_T' basically being the statistic of the oracle case. (12) thus 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 (A5) 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. $\square$

Proof of (10) and (11). We limit our attention to the proof of (10), the arguments for (11) being exactly the same. Using the uniform expansion (7) for the stochastic part of the backfitting estimator $\tilde{m}_2$, we can write $U_{T,3}(w) = U_{T,3}^{A,NW}(w) + U_{T,3}^{A,SBF}(w) + U_{T,3}^{B}(w)$ with

$$U_{T,3}^{A,NW}(w) = -\frac{1}{T}\sum_{t=1}^{T} K_h(w - X_{t,1})\Big(\frac{1}{T}\sum_{s=1}^{T} \frac{K_g(X_{t,2}, X_{s,2})}{\frac{1}{T}\sum_{v=1}^{T} K_g(X_{t,2}, X_{v,2})}\varepsilon_s\Big)\Big/p_1(w)$$

$$U_{T,3}^{A,SBF}(w) = -\frac{1}{T}\sum_{t=1}^{T} K_h(w - X_{t,1})\Big(\frac{1}{T}\sum_{s=1}^{T} r_s(X_{t,2})\varepsilon_s\Big)\Big/p_1(w)$$

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

To complete the proof of (10), it thus suffices to show that

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{A,NW}(w) dw = o_p(1) \tag{13}$$

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{A,SBF}(w) dw = o_p(1) \tag{14}$$

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{B}(w) dw = o_p(1) \tag{15}$$

for $j = 1, \ldots, 5$.

We start with the proof of (13) which consists of several steps. In the first step, we show that

$$Th^{1/2} \int U_{T,j}(w) U_{T,3}^{A,NW}(w) dw = W_{T,j} + o_p(1) \tag{16}$$

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 $\kappa(u) = \mathbb{E}[K_g(u, X_{v,2})]$ and $R(u) = \frac{1}{T}\sum_{v=1}^{T}(K_g(u, X_{v,2}) - \mathbb{E}[K_g(u, X_{v,2})])$. As $\sup_{u \in [0,1]} |R(u)| = O_p(\sqrt{\log T / Tg})$, it further 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 (16).

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

$$W_{T,j} = Th^{1/2} \int U_{T,j}(w) \Big(\frac{1}{T^2} \sum_{s,t=1}^{T} \frac{K_h(w - X_{t,1})}{p_1(w)} \mathbb{E}_t \Big[\frac{K_g(X_{t,2}, X_{s,2})}{\kappa(X_{t,2})} \Big] \varepsilon_s \Big) dw + R_{T,j}, \tag{17}$$

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

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

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

$$Q_T = \int \Big\{ \frac{h^{1/2}}{T} \sum_{s,t=1}^{T} K_h(w - X_{t,1}) \Big(\frac{K_g(X_{t,2}, X_{s,2})}{\kappa(X_{t,2})} - \mathbb{E}_t \Big[\frac{K_g(X_{t,2}, X_{s,2})}{\kappa(X_{t,2})} \Big] \Big) \varepsilon_s \Big\}^2 dw.$$

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 $j = 1, \ldots, 5$, 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,

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

with

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

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 $1 \leq s, s', t, t' \leq T$ 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. With this in mind, it is easily seen that the number of tuples contained in Γ^c is smaller than $C(T \log T)^2$ for some sufficiently large constant C. Moreover, with the help of the Cauchy-Schwarz inequality it is straightforward to compute that $\int \mathbb{E}[\psi_{t,s}(w)\psi_{t',s'}(w)\varepsilon_s\varepsilon_{s'}]dw \leq Ch^{-1}g^{-3/2}$. As a consequence,

$$E_{\Gamma^c} \leq \frac{h}{(Ta_T)^2} \sum_{(s,s',t,t')\in\Gamma} \int \mathbb{E}\big[\psi_{t,s}(w)\psi_{t',s'}(w)\varepsilon_s\varepsilon_{s'}\big] dw \leq \frac{C}{\kappa_T^2} \to 0.$$

We next turn to E_Γ. First note that Γ can be written as the union of the disjoint sets

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

Hence, $E_\Gamma = E_{\Gamma_1} + E_{\Gamma_2} + E_{\Gamma_3} + E_{\Gamma_4}$ 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 $\psi_{t,s}(w) = \psi_{t,s}^A(w) + \psi_{t,s}^B(w)$ with

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

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

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

We first consider $E_{\Gamma_1}^A$: 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 $|\psi_{t,s}^B(w)| \le \frac{C}{hg^2 N_T}$. Plugging this into the expression for $E_{\Gamma_1}^B$ and letting N_T grow at a sufficiently fast rate, we arrive at $|E_{\Gamma_1}^B| = o(1)$. To deal with $E_{\Gamma_1}^A$, we write

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

with

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

By Davydov's inequality, it holds that

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

with some $C_3 > 0$, where q and r are chosen slightly larger than $\frac{4}{3}$ and 4, respectively. Note that we can make C_3 arbitrarily large by choosing C_2 to be large enough. 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 (17) holds with $R_{T,j} = o_p(1)$. Slightly rearranging the terms in (17), we arrive at

$$\Big|Th^{1/2} \int U_{T,j}(w) U^{A,NW}_{T,3}(w) dw\Big|$$
$$= \Big|Th^{1/2} \int U_{T,j}(w) \underbrace{\Big(\frac{1}{T}\sum_{t=1}^{T} \frac{K_h(w - X_{t,1})}{p_1(w)}\Big)}_{O_p(1) \text{ uniformly in } w} \underbrace{\Big(\frac{1}{T}\sum_{s=1}^{T} \mathbb{E}_0\Big[\frac{K_g(X_{0,2}, X_{s,2})}{\kappa(X_{0,2})}\Big]\varepsilon_s\Big)}_{=O_p(T^{-1/2}) \text{ uniformly in } w} dw\Big| + o_p(1)$$
$$\leq Th^{1/2}\ O_p(T^{-1/2})\ \Big(\int U_{T,j}(w)^2 dw\Big)^{1/2} + o_p(1),$$

where the last line is by the Cauchy-Schwarz inequality. From this, (13) immediately follows for $j = 1, 2, 4$. Using the arguments from Steps 1 and 2 above and noting that

$$\Big(\int U^B_{T,j}(w)^2 dw\Big)^{1/2} = O_p(g^2) \tag{18}$$

for $j = 3, 5$, some straightforward additional considerations yield (13) for $j = 3, 5$ as well. The results (14) and (15) follow by similar arguments. This completes the proof. □

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 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). □

B Structural breaks tests

We make the following assumptions. As in Appendix A, we assume for simplicity that $d = 2$.

(B1) For $l = ante, post$, there exist stationary processes $(X_{t,1}^l, X_{t,2}^l, Y_t^l)$ with:

$$Y_t^l = m_0^l + m_1^l \left(X_{t,1}^l\right) + m_2^l \left(X_{t,2}^l\right) + \varepsilon_t^l \,. \tag{19}$$

The process $(X_{t,1}^{ante}, X_{t,2}^{ante}, \varepsilon_t^{ante}, X_{t,1}^{post}, X_{t,2}^{post}, \varepsilon_t^{post})$ is stationary and α-mixing with

$$\alpha(t) \leq C_\alpha \eta^t$$

for some $0 < C_\alpha < \infty$ and $0 < \eta < 1$.

(B2) There exist $1/2 < \gamma < 1$ and a constant C_γ such that for $j = 1$ and $j = 2$

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

(B3) Assumptions (A2)–(A6) from Appendix A apply with $X_{t,1}, X_{t,2}, \varepsilon_t, \sigma, m_0, m_1$ and m_2 replaced by $X_{t,1}^{ante}$, $X_{t,2}^{ante}$, ε_t^{ante}, σ^{ante}, m_0^{ante}, m_1^{ante} and m_2^{ante} or by $X_{t,1}^{post}$, $X_{t,2}^{post}$, ε_t^{post}, m_0^{post}, σ^{post}, m_1^{post} and m_2^{post}. In Assumption (A6), we put $\varepsilon_t^l = \sigma^l(X_{t,1}^l, X_{t,2}^l)\xi_t^l$ *for* $l = ante, post$ and we assume that ξ_t^{ante} and ξ_t^{post} are independent of $(X_s^{ante}, X_s^{post} : s \leq t)$.

(B4) It holds that $t^*/T \to c$ for $T \to \infty$.

(B5) It holds that $h = O(T^{-1/5})$ and $g = O(T^{-1/4-\delta})$ for some small $\delta > 0$.

Before we come to the proofs of Theorems 3 and 4, we briefly discuss how Assumptions (B1)–(B4) hold under reasonable conditions for the nonparametric HAR process of Section 2:

$$V_t^{(\iota_1)} = \begin{cases} m_0^{ante} + m_1^{ante}\left(V_{t-1}^{(\iota_1)}\right) + m_2^{ante}\left(V_{t-1}^{(\iota_2)}\right) + \varepsilon_t^{ante} & \text{for } t \leq t^*,\\ m_0^{post} + m_1^{post}\left(V_{t-1}^{(\iota_1)}\right) + m_2^{post}\left(V_{t-1}^{(\iota_2)}\right) + \varepsilon_t^{post} & \text{for } t > t^*. \end{cases}$$

For a discussion of the existence of stationary solutions of (19), it needs to be determined if there exist stationary solutions of the equations

$$V_t^{j,(\iota_1)} = m_0^j + m_1^j\left(V_{t-1}^{j,(\iota_1)}\right) + m_2^{ante}\left(V_{t-1}^{j,(\iota_2)}\right) + \varepsilon_t^j \qquad \text{for } 1 \leq t \leq T, j = ante, post$$

that fulfill the mixing conditions of (B1). For such a discussion, see, e.g., Chen and Chen (2001). The following lemma states a condition under which (B2) holds for the nonparametric HAR process.

LEMMA 1. *Assume that stationary solutions of (19) exist and that for $j = ante, post$*

$$\left|\frac{\partial}{\partial x} m_1^j(x)\right| + \left|\frac{\partial}{\partial x} m_2^j(x)\right| < \rho$$

for some $0 < \rho < 1$. Then Assumption (B2) holds for $(Y_t, X_{t,1}, X_{t,2}) = (V_t^{(\iota_1)}, V_{t-1}^{(\iota_1)}, V_{t-1}^{(\iota_2)})$.

Proof of Lemma 1. Choose $\gamma > 1/2$. Suppose that $\iota_1 < \iota_2$. For the proof of the first two inequalities of (B2), one makes iterative use of the following inequality

$$\begin{aligned} |V_t^{post,(\iota_1)} - V_t| &\leq \left| m_1^{post}\left(V_{t-1}^{post,(\iota_1)}\right) - m_1^{post}\left(V_{t-1}^{(\iota_1)}\right) + m_2^{post}\left(V_{t-1}^{post,(\iota_2)}\right) - m_2^{post}\left(V_{t-1}^{(\iota_2)}\right)\right| \\ &\leq \rho \max\left\{\left|V_{t-1}^{post,(\iota_1)} - V_{t-1}^{(\iota_1)}\right|, \left|V_{t-1}^{post,(\iota_2)} - V_{t-1}^{(\iota_2)}\right|\right\} \\ &\leq \rho \max_{1\leq l\leq \iota_2} |V_{t-l}^{post,(\iota_1)} - V_{t-l}| \end{aligned}$$

for $t^* + C_\gamma \log T \leq t \leq T$. The third and fourth claims of (B2) follow similarly.

For the proof of the last claim of (B2), note first that our moment condition on the errors imply that for $j = ante, post$

$$\sup\{|\varepsilon_t^j| : 0 \leq t \leq C_\gamma \log T \text{ or } t^* \leq t \leq t^* + C_\gamma \log T\} = O_p((\log T)^{1/4}) = O_p((\log T)).$$

□

In Section 2, we assume that the autoregression function of the HAR process consists of three additive components m_1^j, m_2^j and m_3^j for $j = ante, post$. For this specification of the HAR process, we have to replace the assumption of Lemma 1 by:

$$\left|\frac{\partial}{\partial x} m_1^j(x)\right| + \left|\frac{\partial}{\partial x} m_2^j(x)\right| + \left|\frac{\partial}{\partial x} m_3^j(x)\right| < \rho$$

for $j = ante, post$ for some $0 < \rho < 1$.

Some technical lemmas

Define now the backfitting estimators $\tilde{m}_1^{ante}$ and $\tilde{m}_2^{ante}$ of m_1^{ante} and m_2^{ante}, respectively, based on the observations $Y_1, X_{1,1}, X_{1,2}, \ldots, Y_{t^*-1}, X_{t^*-1,1}, X_{t^*-1,2}$, and the backfitting estimators $\tilde{m}_1^{post}$ and $\tilde{m}_2^{post}$ of m_1^{post} and m_2^{post}, respectively, based on the observations Y_{t^*}, $X_{t^*,1}$, $X_{t^*,2}, \ldots$, $Y_T, X_{T,1}, X_{T,2}$. In our asymptotic analysis, we compare these estimators with the corresponding infeasable backfitting estimators of m_1^{ante}, m_2^{ante}, m_1^{post} and m_2^{post}, respectively, based on the observations $Y_1^{ante}, X_{1,1}^{ante}, X_{1,2}^{ante}, ..., Y_{t^*-1}^{ante}, X_{t^*-1,1}^{ante}, X_{t^*-1,2}^{ante}$ or $Y_{t^*}^{post}, X_{t^*,1}^{post}, X_{t^*,2}^{post}, ..., Y_T^{post}, X_{T,1}^{post}, X_{T,2}^{post}$, respectively. The latter estimators are denoted by $\tilde{m}_1^{\dagger,ante}, \tilde{m}_2^{\dagger,ante}$, $\tilde{m}_1^{\dagger,post}$ and $\tilde{m}_2^{\dagger,post}$, respectively. In our next lemma, we argue that $\tilde{m}_l^j - \tilde{m}_l^{\dagger,j}$ is small for $j = ante, post$ and $l = 1$ and $l = 2$.

LEMMA 2. *Under the assumptions (B1)–(B4) we have that for $j = ante, post$ and $l = 1$ and $l = 2$*

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

Proof of Lemma 2. We argue that we have for $j = ante, post$ and $l = 1$ and $l = 2$

$$\sup_{x\in[0,1]} |\bar{m}_l^{\dagger,j}(x) - \bar{m}_l^j(x)| \;=\; O_p(T^{-\gamma}), \tag{20}$$

where we compare the following 'marginal estimators'

$$\begin{aligned}
\bar{m}_l^{\dagger,ante}(x) &= \frac{\sum_{t=1}^{t^*-1} K_h(x, X_{t,l}^{ante}) Y_t^{ante}}{\sum_{t=1}^{t^*-1} K_h(x, X_{t,l}^{ante})}, \\
\bar{m}_l^{\dagger,post}(x) &= \frac{\sum_{t=t^*}^{T} K_h(x, X_{t,l}^{post}) Y_t^{post}}{\sum_{t=t^*}^{T} K_h(x, X_{t,l}^{post})}, \\
\bar{m}_l^{ante}(x) &= \frac{\sum_{t=1}^{t^*-1} K_h(x, X_{t,l}) Y_t}{\sum_{t=1}^{t^*-1} K_h(x, X_{t,l})}, \\
\bar{m}_l^{post}(x) &= \frac{\sum_{t=t^*}^{T} K_h(x, X_{t,l}) Y_t}{\sum_{t=t^*}^{T} K_h(x, X_{t,l})}.
\end{aligned}$$

We prove (20) for $j = post$. Application of Assumption (B2) yields that

$$\begin{aligned}
&\left| \frac{1}{T} \sum_{t=t^*}^{T} K_h(x, X_{t,l}^{post}) Y_t^{post} - \frac{1}{T} \sum_{t=t^*}^{T} K_h(x, X_{t,l}) Y_t \right| \\
&\quad \le \left| \frac{1}{T} \sum_{t\in\mathfrak{T}_-} K_h(x, X_{t,l}^{post}) Y_t^{post} \right| + \left| \frac{1}{T} \sum_{t\in\mathfrak{T}_-} K_h(x, X_{t,l}) Y_t \right| \\
&\qquad + \left| \frac{1}{T} \sum_{t\in\mathfrak{T}_+} \left\{ K_h(x, X_{t,l}^{post}) Y_t^{post} - K_h(x, X_{t,l}) Y_t \right\} \right| \\
&\quad = O_P(T^{-1}(\log T)^2 + T^{-\gamma}) \\
&\quad = O_P(T^{-\gamma}),
\end{aligned}$$

where $\mathfrak{T}_- = \{t : t^* \le t \le t^* + C_\gamma \log T\}$ and $\mathfrak{T}_+ = \{t : t^* + C_\gamma \log T < t \le T\}$. This shows (20) for $j = post$. The statement of the lemma follows from the theory developed in Mammen et al. (1999) for the smooth backfitting estimator. There it is explained that the smooth backfitting estimator results from the 'marginal estimators' by the application of an operator that has 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). $\square$

We now define

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

By using similar arguments as for the proof of (20), one gets the following lemma:

LEMMA 3. *Under the assumptions (B1)–(B4) we have that for $j = ante, post$ and $l = 1$ and $l = 2$*

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

From Lemma 3, we get that $S_{j,T}^{\dagger} = S_{j,T} + o_P(1)$, where

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

Thus, for the statement of Theorem 3, it suffices to show that

$$S_{j,T}^{\dagger} - h^{-1/2} K^{(2)}(0) \int_{[0,1]} \left[c^{-1}\sigma_{ante}^2(x) + (1-c)^{-1}\sigma_{post}^2(x)\right] \pi(x)\, dx$$

has a limiting normal distribution with mean M_S and variance V_S.

Proofs of Theorems 3 and 4

The proofs follow the lines of arguments given in the proofs of Theorems 1 and 2. The discussion of the last subsection has shown that the observations can be approximated by stationary random variables. Thus we can show asymptotic normality with the same tools as in Appendix A. Our assumptions on the bandwidths show that the bias terms of the pilot smooth backfitting estimation step are of lower order. Furthermore, one can show by similar arguments that the bias terms of Nadaraya-Watson smoothing of the second step cancel.

C Measurement of realized variance

For reasons explained in the main text, we are only interested in modeling the continuous component in RV. Here, we sketch the estimation procedure, which follows ideas outlined, e.g., in Andersen et al. (2007), Andersen, Bollerslev, Frederiksen and Nielsen (2010), and Audrino and Hu (2011).

Suppose the log-price of a financial asset y_t follows the process

$$dy(t) = \mu(t)dt + \sigma(t)dW(t) + a(t)dJ(t) \ , \tag{21}$$

where $\mu(t)$ is the drift, $\sigma(t)$ the instantaneous volatility function, $W(t)$ a Wiener and $J(t)$ a counting process with discrete jump sizes $a(t) = y(t) - y(t^-)$. It is well known[9] that in this case the 'classical' realized variance estimator that sums squared intra-day returns captures the contribution to quadratic variation from both the continuous and the discontinuous part of the process. More precisely, in recording $y_{i,t}, i = 1, \ldots, M$, at M equally spaced time points during day t, define the intra-day high-frequency returns as $r_{i,t} = y\Big(t - 1 + \frac{i}{M}\Big) - y\Big(t - 1 + \frac{i-1}{M}\Big)$, $i = 1, \ldots, M$. Then

$$RV_t = \sum_{i=1}^{M} r_{i,t}^2 \xrightarrow{P} \int_{t-1}^{t} \sigma^2(s)ds + \sum_{t-1 \leq s \leq t} a^2(s) \ , \tag{22}$$

where the term $\int_{t-1}^{t} \sigma^2(s)ds$ constitutes the contribution stemming from the continuous part only (also called integrated variance).

We therefore test whether each return $r_{i,t}$ is a jump based on the jump test due to Lee and Mykland (2008). Whenever the Lee-Mykland test statistic $T_{t,i}^J$ exceeds the critical value β^* at a significance level of 1%, the return is classified as a jump by $a_{i,t} = r_{i,t}\mathbb{1}_{\{T_{t,i}^J > \beta^*\}}$, where $\mathbb{1}_{\mathcal{A}}$ is the indicator function of the event $\mathcal{A}$. The measure of daily variation that only comprises the continuous part of RV is estimated by

$$cRV_t = RV_t - JV_t \ , \tag{23}$$

where

$$JV_t = \sum_{i=1}^{M} a_{i,t}^2 - \frac{M_t^J}{M - M_t^J} IV_t \ , \tag{24}$$

$$IV_t = \sum_{i=1}^{M} r_{i,t}^2 \mathbb{1}_{\{T_{t,i}^J \leq \beta^*\}} \tag{25}$$

with $M_t^J = \sum_{i=1}^{M} \mathbb{1}_{\{T_{t,i}^J > \beta^*\}}$ denoting the number of jumps in day t. The correction term in (24) is designed to correct the total RV due to jumps by the average level of integrated variance measured on the non-jump returns.

[9] See McAleer and Medeiros (2008b) and Andersen, Bollerslev and Diebold (2010) for surveys, including discussions of the related literature.

References

Aït-Sahalia, Y., Fan, J. and Peng, H. (2009). Nonparametric transition-based tests for diffusions, *Journal of the American Statistical Association* **104**: 1102–1116.

Andersen, T. G., Bollerslev, T. and Diebold, F. (2010). Parametric and nonparametric measurement of volatility, *in* Y. Aït-Sahalia and L. Hansen (eds), *Handbook of Financial Econometrics*, Vol. 1, North Holland, Amsterdam, pp. 67–137.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility, *Review of Economics and Statistics* **89**(4): 701–720.

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. (2001). The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* **96**: 42–55.

Andersen, T. G., Bollerslev, T., Frederiksen, P. H. and Nielsen, M. Ø. (2010). Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns, *Journal of Applied Econometrics* **25**(2): 233–261.

Audrino, F. and Hu, Y. (2011). Volatility Forecasting: Downside Risk, Jumps and Leverage Effect. Working Paper.

Audrino, F. and Knaus, S. (2012). Lassoing the HAR model: A model selection perspective on realized volatility dynamics, *Technical report*, St. Gallen University.

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, G. and Chen, M. (2001). On a class of nonlinear AR(p) models with nonlinear ARCH errors, *Australian & New Zealand Journal of Statistics* **43**(4): 445–454.

Chen, R. and Tsay, R. S. (1993). Nonlinear additive ARX models, *Journal of the American Statistical Association* **88**: 955–967.

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.

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.

Corsi, F., Pirino, D. and Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting, *Journal of Econometrics* **159**: 276–288.

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling, *Journal of Business and Economic Statistics* **30**(3): 368–380.

Craioveanu, M. and Hillebrand, E. (2010). Why it is OK to use the HAR-RV (1,5,21) model, *Working paper*, Louisiana State University.

Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B. and Pictet, O. V. (2001). *An Introduction to High-Frequency Finance*, Academic Press, London.

Dette, H. and Sprekelsen, I. (2004). Some comments on specification tests in nonparametric absolutely regular processes, *Journal of Time Series Analysis* **25**: 159–172.

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, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon, *Annals of Statistics* **29**: 153–193.

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. and Lunde, A. (2006). Realised variance and market microstructure noise, *Journal of Business and Economic Statistics* **24**: 127–218.

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.

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.

Lee, S. and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics, *Review of Financial Studies* **21**(6): 2535–2563.

Leucht, A. (2012). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency, *Bernoulli* **18**: 552–585.

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

McAleer, M. and Medeiros, M. C. (2008b). Realized volatility: A review, *Econometric Reviews* **27**(1–3): 10–45.

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

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

Seifert, B. and Gasser, T. (1996). Finite sample variance of local polynomials: Analysis and solutions, *Journal of the American Statistical Association* **91**: 267–275.

Stone, C. J. (1985). Additive regression and other nonparametric models, *Annals of Statistics* **13**(2): 689–705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models, *Annals of Statistics* **14**(2): 592–606.

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.

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

Tables and Figures

Symbol	**Name**	**Type**	**Exchange**	**data range**	
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
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
SP	S&P 500	Equity index future	CME Group	1. Jul., 2003	31. Dec., 2010
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
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
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
HG	Copper High Grade COMEX	Metal future	COMEX/CME Group	1. Jul., 2003	31. Dec., 2010
EC	Euro FX	Currency future	CME Group	1. Jul., 2003	31. Dec., 2010
JY	Japanese Yen	Currency future	CME Group	1. Jul., 2003	31. Dec., 2010
CN	Corn	Grain future	CBOT/CME Group	1. Jul., 2003	31. Dec., 2010
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.

Bandwidths			
Symbol	**Daily**	**Weekly**	**Monthly**
CF	0.213	0.206	0.209
FT	0.204	0.197	0.194
KM	0.251	0.283	0.301
NE	0.214	0.210	0.238
SP	0.206	0.222	0.195
XX	0.217	0.246	0.207
BN	0.188	0.197	0.271
TY	0.234	0.253	0.217
US	0.240	0.256	0.200
CL	0.353	0.214	0.185
NG	0.175	0.182	0.227
GC	0.245	0.260	0.259
HG	0.314	0.255	0.224
EC	0.169	0.198	0.199
JY	0.171	0.202	0.358
CN	0.216	0.218	0.255
SY	0.208	0.254	0.239

Table 2: Optimal bandwidths obtained by minimizing the IMSE; see Section 6.1 for details. Integrated squared bias is approximated from the local linear smooth backfitting. Estimation is carried out on [0,1]; see Table 1 for the list of acronyms.

Symbol	**Daily**	**Weekly**	**Monthly**
CF	**0.007**	**0.023**	0.467
FT	**0.013**	0.665	**0.073**
KM	0.293	0.362	**0.002**
NE	0.106	0.868	**0.000**
SP	0.364	0.201	**0.010**
XX	**0.020**	**0.049**	0.321
BN	**0.003**	**0.017**	0.572
TY	**0.062**	0.111	**0.001**
US	**0.003**	0.519	**0.012**
CL	0.133	**0.060**	0.445
NG	**0.053**	0.195	0.160
GC	0.344	0.436	0.322
HG	**0.094**	0.122	**0.000**
EC	**0.000**	**0.006**	**0.000**
JY	**0.000**	**0.006**	0.146
CN	0.180	**0.014**	**0.072**
SY	**0.098**	0.803	0.198

Table 3: Specification test based on Nadaraya-Watson smooth backfitting as suggested in Section 3. Null hypothesis is the linear specification in the respective component function. The p-values are obtained from 10,000 bootstrap replications of the estimate with residuals obtained from the null hypothesis. p-values are in bold when below 10%. Weighting function in the test statistic is the empirical density function; see Table 1 for the list of acronyms.

Symbol	**Daily**	**Weekly**	**Monthly**
CF	**0.014**	0.688	0.416
FT	**0.019**	0.909	0.841
KM	**0.070**	0.833	**0.022**
NE	0.133	0.648	0.104
SP	0.591	0.735	0.537
XX	**0.028**	0.947	**0.080**
BN	0.570	0.656	0.157
TY	**0.000**	**0.042**	**0.001**
US	**0.000**	**0.069**	**0.005**
CL	0.194	0.142	0.690
NG	0.109	**0.021**	0.696
GC	0.765	0.815	0.288
HG	0.595	0.106	0.455
EC	0.102	**0.081**	**0.014**
JY	0.168	0.396	0.185
CN	**0.007**	0.270	0.514
SY	0.200	**0.015**	0.062

Table 4: Structural breaks test based on Nadaraya-Watson smooth backfitting as suggested in Section 4. Null hypothesis is equality of the function on the ante and the post sample. The p-values are obtained from 10,000 bootstrap replications of the estimate with residuals obtained from the null hypothesis. p-values are in bold when below 10%. Assumed break date is July 25, 2007. Weighting function in the test statistic is a weighted product of the empirical density function on the subsamples; see Table 1 for the list of acronyms.

	Daily		Weekly		Monthly	
	Ante	Post	Ante	Post	Ante	Post
CF	**0.002**	0.114				
FT	**0.000**	0.435				
KM	0.174	0.378			**0.016**	**0.075**
XX	**0.002**	0.266			0.596	**0.030**
TY	0.315	0.228	0.474	0.420	0.410	0.127
US	0.865	**0.001**	**0.001**	0.124	**0.002**	0.131
NG			**0.012**	0.573		
EC			**0.005**	**0.021**	**0.000**	**0.002**
CN	**0.003**	0.847				
SY			0.761	0.634		

Table 5: Specification tests (Nadaraya-Watson smooth backfitting) on subsamples where a structural break is detected according to Table 4. Null hypothesis is the linear specification in the respective component function. p-values are obtained from 10,000 bootstrap replications with residuals obtained from the null hypothesis. p-values are in bold when below 10%. Weighting function in the test statistic is the empirical density function; see Table 1 for the list of acronyms.

Symbol	**Daily**	**Weekly**	**Monthly**
CF	**0.000**	**0.000**	0.587
FT	**0.000**	**0.073**	0.194
KM	0.167	0.840	**0.022**
NE	**0.000**	**0.013**	**0.009**
SP	**0.001**	0.389	0.375
XX	**0.000**	**0.001**	0.504
BN	**0.005**	**0.011**	0.560
TY	0.216	**0.019**	**0.000**
US	**0.005**	0.278	**0.001**
CL	0.381	0.173	**0.059**
NG	**0.004**	**0.005**	0.583
GC	**0.035**	0.215	0.989
HG	0.821	**0.046**	**0.014**
EC	**0.054**	**0.003**	**0.000**
JY	**0.000**	**0.010**	0.869
CN	0.104	0.155	**0.048**
SY	0.449	0.520	0.228

Table 6: Specification tests with LL smooth backfitting. Null hypothesis is the linear specification in the respective component function. The p-values are obtained from 10,000 bootstrap replications of the estimate with residuals obtained from the null hypothesis. p-values are in bold when below 10%. Weighting function in the test statistic is the empirical density function; see Table 1 for the list of acronyms.

	Daily		**Weekly**		**Monthly**	
	Ante	Post	Ante	Post	Ante	Post
CF	**0.053**	**0.014**				
FT	**0.001**	0.122				
KM	0.231	0.509			0.553	**0.003**
XX	**0.041**	**0.047**			**0.075**	**0.004**
TY	0.535	0.619	0.683	**0.062**	**0.002**	0.103
US	0.537	0.254	**0.033**	0.122	**0.000**	**0.053**
NG			**0.017**	0.715		
EC			**0.000**	**0.010**	**0.000**	**0.007**
CN	**0.009**	0.946				
SY			**0.031**	0.507		

Table 7: Specification tests (local linear smooth backfitting) on subsamples where a structural break is detected according to Table 4. Null hypothesis is the linear specification in the respective component function. p-values are obtained from 10,000 bootstrap replications with residuals obtained from the null hypothesis. p-values are in bold when below 10%. Weighting function in the test statistic is the empirical density function; see Table 1 for the list of acronyms.

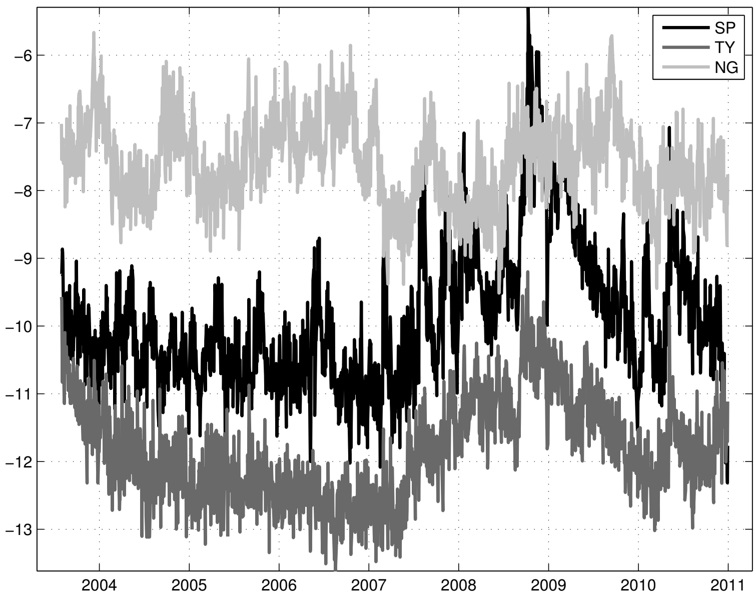

Figure 1: Daily log-RV for selected instruments: S&P500 (SP), 10yrs TNote (TY), natural gas (NG).

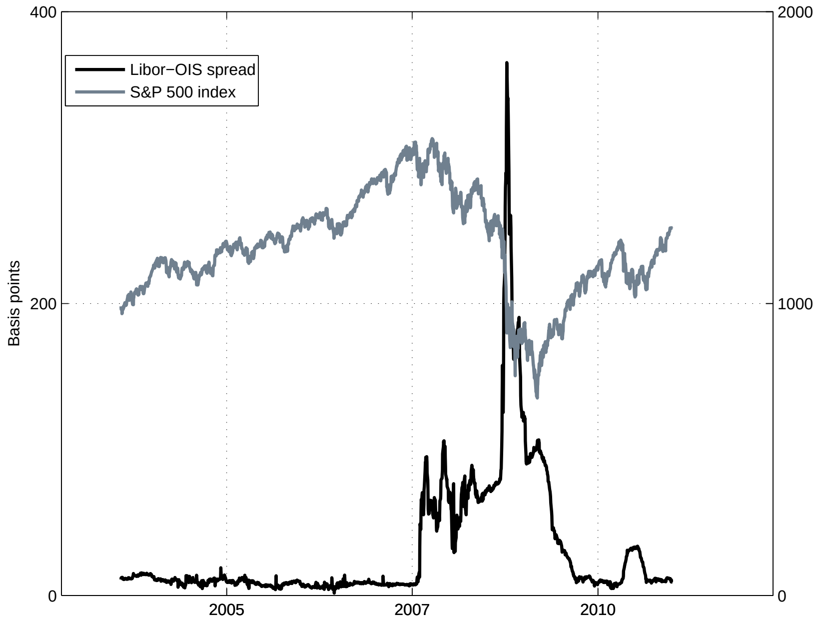

Figure 2: Difference of the 3-months USD Libor over the 3-months overnight indexed swap (left ordinate axis), S&P 500 index closing prices (right ordinate axis) from July 1, 2003, to Dec. 31, 2010. Source: Bloomberg.

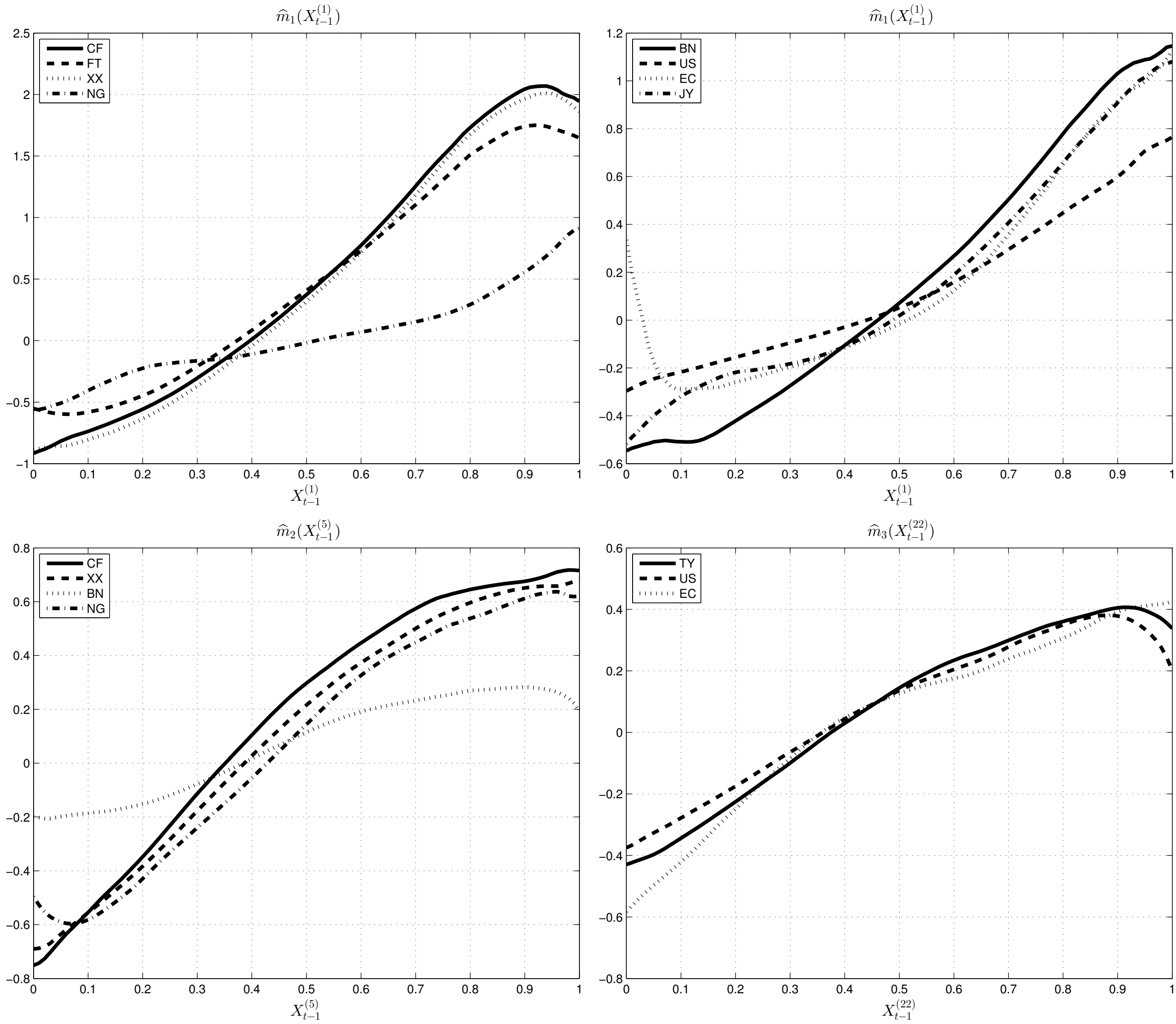

Figure 3: Selection of smooth backfitting estimates of variance component functions in the nonparametric HAR model. Daily functions (m_1) in top panel, weekly and monthly (m_2, m_3) in the lower panel; see Table 1 for the list of acronyms.

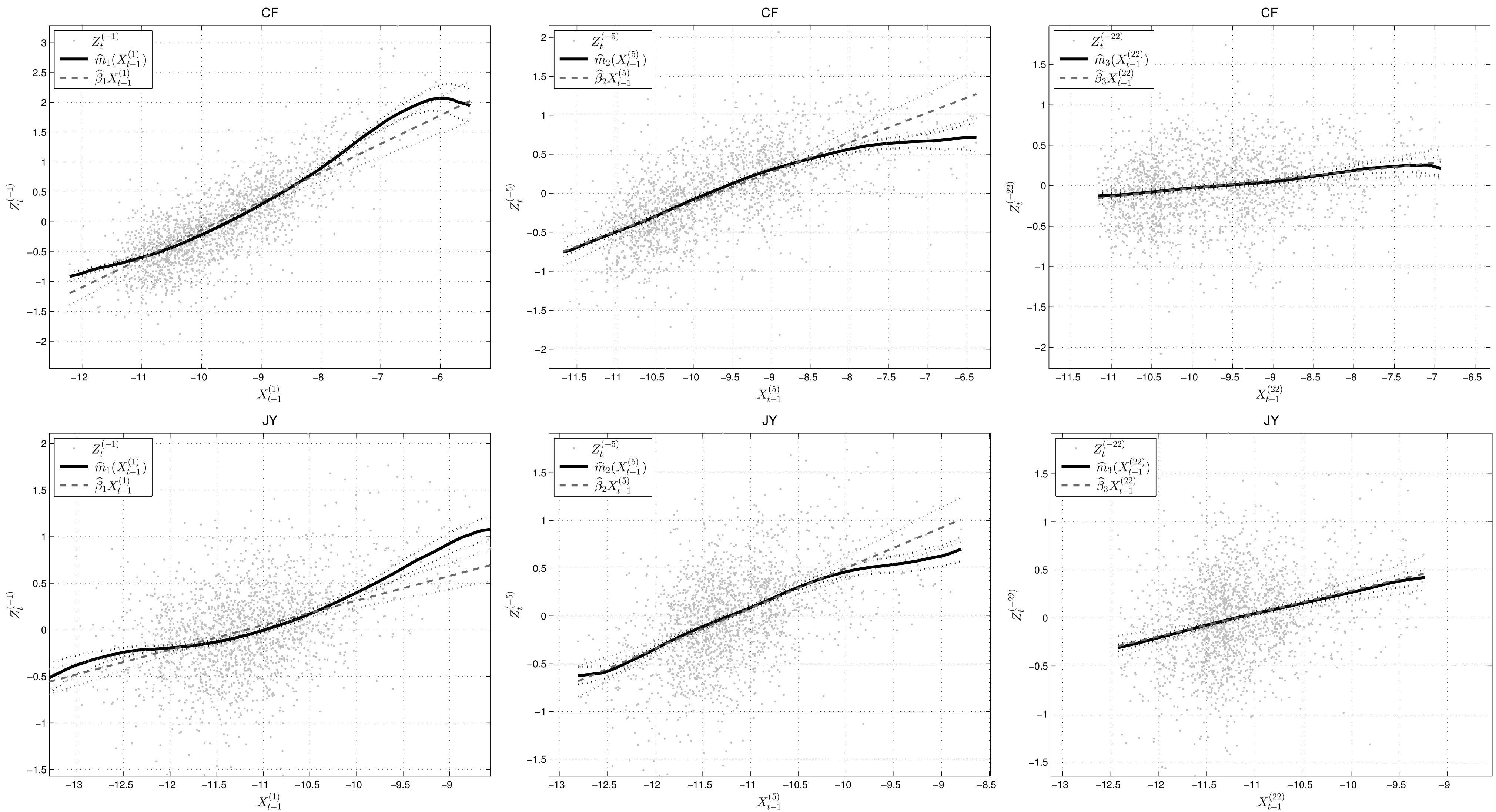

Figure 4: Fully nonparametric versus linear HAR model for CAC40 (CF, top panel) and Japenese Yen (JY, lower panel) with asymptotic confidence intervals. Like the nonparametric component functions, the parametric fits are normed such that their expectation equals zero. $Z_t^{(-\iota_j)}$ is the residual response variable after subtracting all other component functions m_i, $i \neq j$.

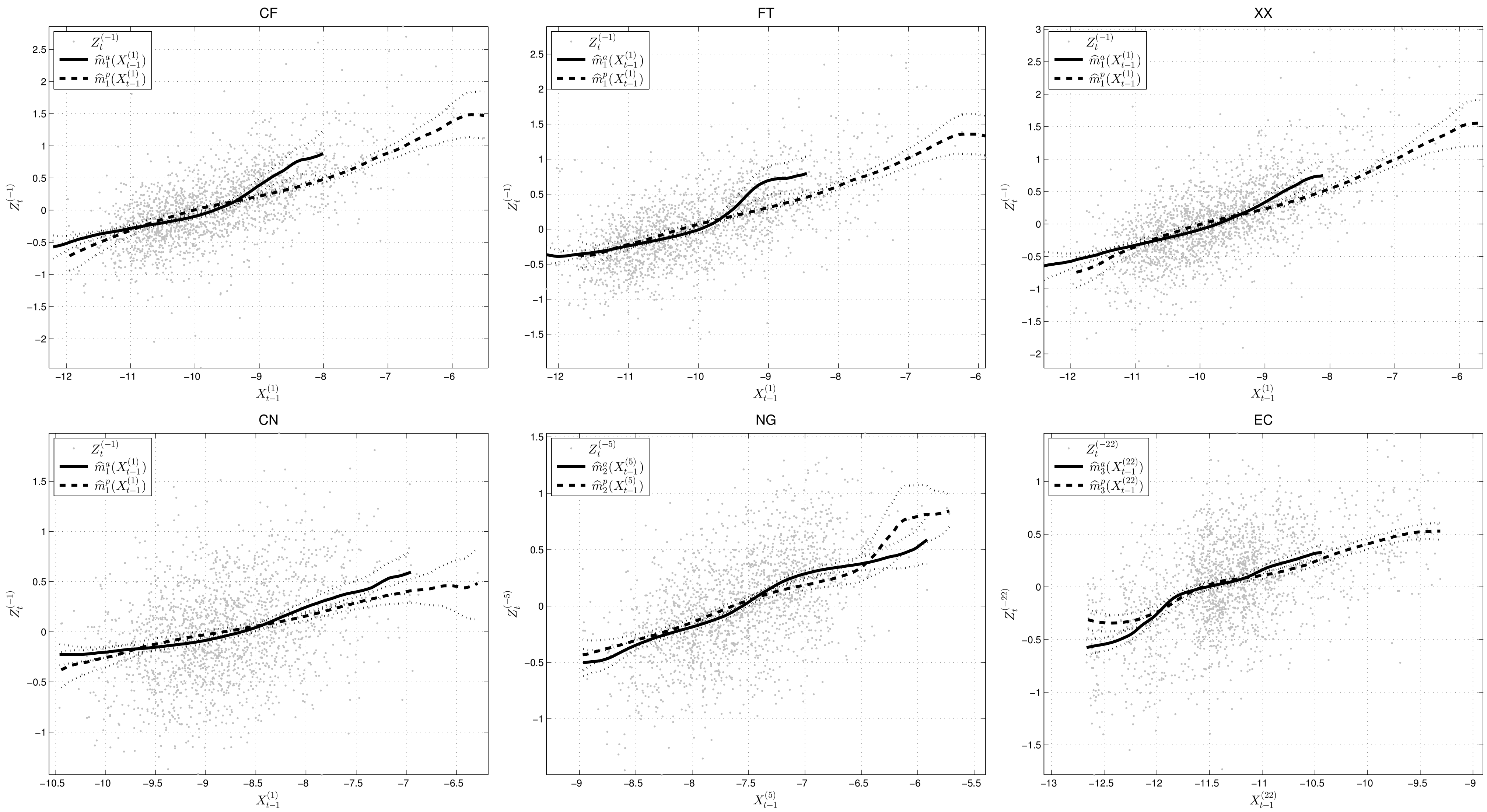

Figure 5: Selection of variance component functions estimated on the ante and post subsample. The functions are normed such that their expectation equals zero under the product density used to evaluate the structural break test. $Z_t^{(-\iota_j)}$ is the residual response variable after subtracting all other component functions m_i, $i \neq j$; see Table 1 for the list of acronyms.

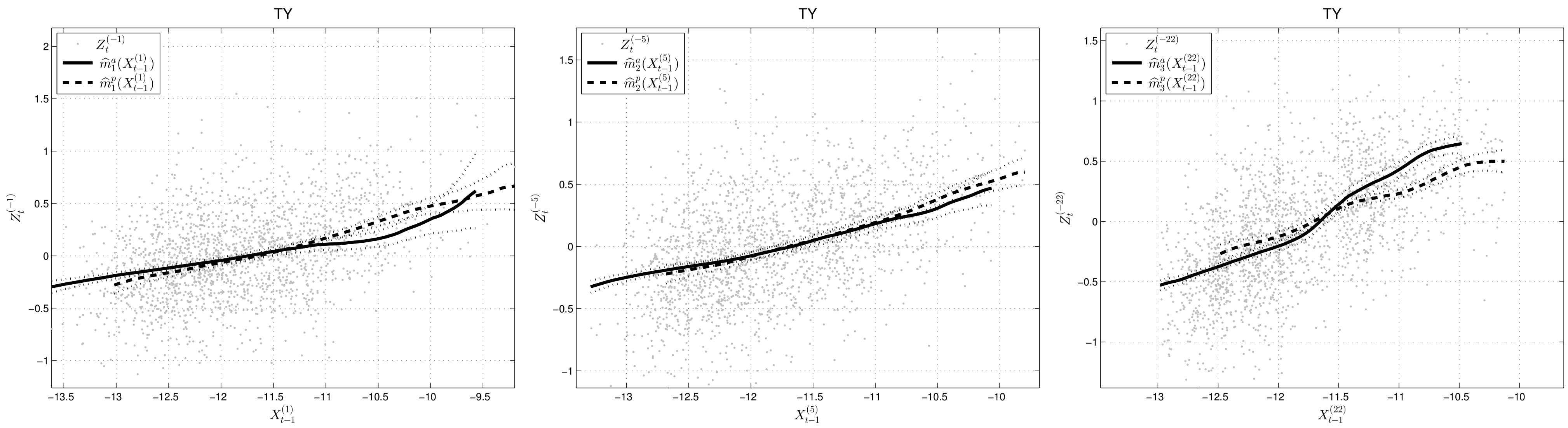

Figure 6: Fully nonparametric HAR models for 10yrs TNote (TY) for the ante and post subsample. Component functions are normed such that their expectation equals zero under the product density used to evaluate the structural break test. $Z_t^{(-\iota_j)}$ is the residual response variable after subtracting all other component functions m_i, $i \neq j$.