SFB 823

Locally stationary multiplicative volatility modelling

Christopher Walsh, Michael Vogt

Discussion Paper

Nr. 11/2021

SFB 823

Locally Stationary Multiplicative Volatility Modelling[1]

Christopher Walsh[2] Michael Vogt[3]

TU Dortmund University of Ulm

Abstract

In this paper, we study a semiparametric multiplicative volatility model, which splits up into a nonparametric part and a parametric GARCH component. The nonparametric part is modelled as a product of a deterministic time trend component and of further components that depend on stochastic regressors. We propose a two-step procedure to estimate the model. To estimate the nonparametric components, we transform the model in order to apply the backfitting procedure used in Vogt and Walsh (2019). The GARCH parameters are estimated in a second step via quasi maximum likelihood. We show consistency and asymptotic normality of our estimators. Our results are obtained using mixing properties and local stationarity. We illustrate our method using financial data. Finally, a small simulation study illustrates a substantial bias in the GARCH parameter estimates when omitting the stochastic regressors.

Key words: Smooth Backfitting, Semiparametric, Local Stationarity, Multiplicative Volatility, GARCH.

JEL codes: C14, C22, C58

---

[1]We would like to thank Enno Mammen, Oliver Linton and Kyusang Yu for numerous helpful discussions and comments.

[2]Corresponding author. Address: TU Dortmund University, Department of Statistics, 44221 Dortmund, Germany. Email: walsh@statistik.tu-dortmund.de. Support by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG) is gratefully acknowledged.

[3]Address: Ulm University, Institute of Statistics, Helmholtzstrasse 20, 89081 Ulm, Germany. Email: m.vogt@uni-ulm.de.

1 Introduction

Given the ever-changing economic and financial environment, it is quite plausible that many financial time series behave in a nonstationary way. Especially over longer horizons, structural changes may occur. Thus, the technical assumption of stationarity is likely to be violated in many cases. This issue has been pointed out by numerous authors in recent years. In particular, it has been claimed that many interesting stylized facts of financial return and volatility series can be neatly explained by employing nonstationary models (see e.g. Mikosch and Stărică (2000, 2003, 2004)).

One way to deal with nonstationarities in financial time series is the theory on locally stationary processes. The latter has been introduced in a series of papers by Dahlhaus (1996a,b, 1997). Intuitively speaking, a process is locally stationary if over short periods of time (i.e. locally in time) it behaves approximately stationary, even though it is globally nonstationary. In recent years, many locally stationary models have been proposed in the financial time series context. Usually, these models are extensions of parametric time series models allowing for the parameters to change smoothly over time. An example is the class of ARCH processes with time-varying parameters introduced by Dahlhaus and Subba Rao (2006) and further investigated by Fryzlewicz et al. (2008) and Truquet (2017) among others.

A simple locally stationary volatility model which has been explored in a number of studies is given by the equation

$$Y_{t,T} = \tau\Big(\frac{t}{T}\Big)\varepsilon_t \quad \text{for } t = 1, \dots, T, \tag{1}$$

where $Y_{t,T}$ are log-returns, τ is a smooth deterministic function of time and $\{\varepsilon_t\}$ is a standard stationary GARCH process with $\mathbb{E}[\varepsilon_t^2] = 1$. As usual in the literature on locally stationary models, the time-varying parameter τ does not depend on real time t, but on rescaled time $\frac{t}{T}$. We comment on this feature in more detail in Section 2. Model (1) has been considered for example in Feng (2004), where the τ-function is estimated nonparametrically. Engle and Rangel (2008) work with a closely related model, where the τ-component is modelled parametrically as a flexible exponential spline function. A multivariate generalization of model (1) is studied in Hafner and Linton (2010).

Model (1) can be considered as a GARCH process with time-varying parameters, with certain restrictions imposed on the parameter functions. In particular, the unconditional volatility level $\mathbb{E}[Y_{t,T}^2]$ is given by the time-dependent function $\tau^2(t/T)$, which is allowed to vary smoothly over time. In reality, the volatility level is unlikely to change deterministically over time. Instead it reflects and varies with changes in the economic and financial environment. Therefore, the τ-function should depend on certain economic and financial variables. In model (1), these dependencies are not modelled explicitly. Instead, rescaled time serves as a catch-all for omitted explanatory variables.

These considerations show that in a more realistic version of model (1), the τ-function should depend on economic and financial influences. However, there is clearly no way to come up with a model that incorporates all relevant variables. One way to deal with this is to use rescaled time as a proxy for the omitted variables. To formalize these ideas, we propose the model

$$Y_{t,T} = \tau\Big(\frac{t}{T}, X_t\Big)\varepsilon_t, \tag{2}$$

where $Y_{t,T}$ are log-returns, $X_t = (X_t^1, \dots, X_t^d)$ is an $\mathbb{R}^d$-valued random vector of economic or financial covariates and τ is a smooth function of time and the variables X_t. As before, $\{\varepsilon_t\}$ is a standard GARCH process. To countervail the curse of dimensionality, we split up the τ-function into multiplicative components, thus yielding the model

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big)\prod_{j=1}^{d}\tau_j(X_t^j)\varepsilon_t, \tag{3}$$

where τ_0 and τ_j for $j = 1, \dots, d$ are smooth functions of time and of the regressors

X_t^j, respectively. As will be seen in Section 2, the multiplicative specification of the τ-function in (3) not only avoids the curse of dimensionality but also allows for a direct interpretation of the various components.

In the following sections, we give an in-depth theoretical treatment of model (3). The complete formulation of the model together with its assumptions is given in Section 2. In Section 3, we propose a two-step procedure to estimate both the nonparametric and the parametric components of the model. To estimate the nonparametric functions τ_j for $j = 0, \ldots, d$, we use results from Vogt and Walsh (2019) in order to extend the smooth backfitting procedure of Mammen et al. (1999) to our locally stationary stetting. Having estimates $\tilde{\tau}_j$ of the functions τ_j, we can construct approximate expressions $\tilde{\varepsilon}_t$ of the GARCH variables ε_t. This allows us to estimate the GARCH parameters of the model via approximate quasi maximum likelihood methods in a second step. Consistency and asymptotic normality of our estimators are shown in Section 4.

The contribution in this paper is twofold. From a technical point of view, we extend the asymptotic results for model (1) to a more general framework in which the τ-function depends both on rescaled time and stationary stochastic regressors. This vastly complicates both steps of the asymptotic analysis and as a result, we cannot extend existing proving techniques as provided in Hafner and Linton (2010) in a straightforward manner. In particular, novel and intricate arguments are required to derive the asymptotic behaviour of the GARCH estimates obtained in the second estimation step. In terms of volatility modelling, we introduce a flexible framework which allows to capture both nonstationarities and influences from the economic and financial environment. As the component functions τ_j in our model are completely nonparametric, we are able to explore the form of the relationship between volatility and its potential sources. Therefore, our model allows us to extend existing parametric studies on the sources of volatility as conducted e.g. in Engle and Rangel (2008) and Engle et al. (2013).

In the literature other extensions to GARCH models have been proposed that allow the incorporation of exogenous variables. For instance, Han and Kristensen (2014) linearly include a covariate in the GARCH equation and derive the asymptotic results for a quasi-maximum likelihood estimator of the unknown parameters in the stationary case and a particular nonstationary case. In order to incorporate effects of economic variables on stock market volatility a popular model class is given by GARCH-MIDAS models used for instance in Engle et al. (2013), Conrad and Loch (2014) and Asgharian et al. (2013). Typically, these models have a decomposition into two components, similarly to the decomposition in (1). One component is modelled as a GARCH process that captures short term fluctuations of volatility around a time varying long run component. The long run component is modelled as a parametric function of a finite history of realized stock market variances or some other covariate measured on a lower frequency. The number of included covariates is limited to one or two due to issues with parameter identification and stability of the proposed estimation procedure. Although these models allow for a nice interpretation of short run and long run components of volatility, the effect of an individual covariate on stock market volatility is not as easily interpreted. Furthermore, theoretical results seem to be limited to those derived in Wang and Ghysels (2015) using realized variance as the sole covariate. Finally, economic variables have been successfully used to improve predictions of stock market volatility. This has mainly been done by augmenting autoregressive models of monthly stock market realized variance with linear functions of the covariates of interest as in Christiansen et al. (2012) and Paye (2012). Mittnik et al. (2015) allow for covariates to enter an exponential ARCH model in a nonparametric way. Although their approach allows for the effect of the covariates to be flexible and interpretable, their paper is methodological and solely focused on out of sample predictive performance. In particular, they do not have any theoretical results concerning the estimated nonparametric functions.

To illustrate the usefulness of our model and to complement the technical analysis, we present an empirical example in Section 5. There, the model is applied to S&P 500 log return data using various economic and financial explanatory variables that have been deemed significant drivers of stock market volatility in previous studies. A small simulation study designed to mimic certain aspects of the application investigates the behaviour of the proposed estimation procedure in Section 6. It will be seen there, that omitting explanatory variables can lead to substantially biased GARCH parameter estimates.

2 The Model

Suppose we observe a sample of daily log-returns $Y_{t,T}$ of a financial time series and a sequence of daily $\mathbb{R}^d$-valued random stationary covariate vectors $X_t = (X_t^1, \dots, X_t^d)$ for $t = 1, \dots, T$. We assume the log-return series follows the process

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j(X_t^j)\varepsilon_t \quad \text{for } t = 1, \dots, T \tag{4}$$

with

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= w_0 + a_0 \varepsilon_{t-1}^2 + b_0 \sigma_{t-1}^2. \end{aligned}$$

Here, τ_0 and τ_j $(j = 1, \dots, d)$ are smooth nonparametric functions of time and the stochastic regressors, respectively. Furthermore, $\{\varepsilon_t\}$ is a strictly stationary GARCH process with parameters (w_0, a_0, b_0), which is assumed to be independent of the covariate process $\{X_t\}$. The residuals of the GARCH process, $\{\eta_t\}$, are assumed to be i.i.d. with zero mean and unit variance. For simplicity, we restrict attention to the GARCH(1,1) specification.

In order to conduct meaningful asymptotics, we let the function τ_0 depend on rescaled time t/T rather than on real time t. Thus, τ_0 is defined on $(0, 1]$ rather than on

$\{1, \ldots, T\}$. In the remainder of this paper, we denote rescaled time by $x_0 \in (0, 1]$. It relates to observed time $t \in \{0, \ldots, T\}$ through the mapping $t = \lfloor x_0 T \rfloor$, where the floor function $\lfloor x \rfloor$ denotes the largest integer weakly smaller than x. If we defined the function τ_0 in terms of observed time, we would not get additional information on the structure of τ_0 around a particular time point t as the sample size T increases. Within the framework of rescaled time, in contrast, the function τ_0 is observed on a finer and finer grid on the unit interval as T grows. Thus, we obtain more and more information on the local structure of τ_0 around each point x_0 in rescaled time. This is the reason why we can make meaningful asymptotic considerations within this framework. A detailed discussion of the concept of rescaled time can be found in Dahlhaus (1996a).

For a sufficiently smooth trend function τ_0, we have

$$\big|Y_{t,T} - Y_t(x_0)\big| \leq C\Big|\frac{t}{T} - x_0\Big|U_t, \tag{5}$$

where C is a constant independent of x_0, t and T, $Y_t(x_0) = \tau_0(x_0) \prod_{j=1}^{d} \tau_j(X_t^j)\varepsilon_t$, and $U_t = \prod_{j=1}^{d} \tau_j(X_t^j)\varepsilon_t$. Note that both $\{Y_t(x_0)\}$ and $\{U_t\}$ are strictly stationary processes due to the stationarity of X_t and ε_t. As $U_t = O_p(1)$, we obtain from (5) that

$$\big|Y_{t,T} - Y_t(x_0)\big| = O_p\Big(\Big|\frac{t}{T} - x_0\Big|\Big). \tag{6}$$

Therefore, if t/T is close to x_0, then $Y_{t,T}$ is close to $Y_t(x_0)$ at least in a stochastic sense. Put differently, locally in time, the process $\{Y_{t,T}\}$ is close to the stationary process $\{Y_t(x_0)\}$. In this sense, the process $\{Y_{t,T}\}$ is locally stationary.

We close this section with a remark on the interpretation of the nonparametric components of model (4). First, note that the functions $\tau_0, \ldots, \tau_d$ and the GARCH residual ε_t are only identified up to a multiplicative constant in model (4). Thus we are free to rescale them in a suitable way. Given the independence between X_t and ε_t, normalizing the components such that $\mathbb{E}[\varepsilon_t^2] = 1$ yields

$$\mathbb{E}[Y_{t,T}^2 \mid X_t] = \tau_0^2\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j^2(X_t^j). \tag{7}$$

Thus, the product of the τ-components gives the volatility at time t conditional on the covariates X_t. If we additionally scale the model components to satisfy $\mathbb{E}[\prod_{j=1}^d \tau_j^2(X_t^j)] = 1$, we obtain that

$$\mathbb{E}[Y_{t,T}^2] = \tau_0^2\Big(\frac{t}{T}\Big), \tag{8}$$

i.e. the deterministic function of time $\tau_0^2(t/T)$ gives the time-varying unconditional volatility level. In (7), $\tau_0^2(t/T)$ thus specifies the unconditional volatility level and the product of the remaining components $\prod_{j=1}^d \tau_j^2(X_t^j)$ is the multiplicative factor by which the volatility conditional on X_t deviates from the unconditional level.

3 Estimation Procedure

Next, we provide details on the two-step estimation procedure outlined in the introduction. The first step provides estimators of the nonparametric functions $\tau_0, \dots, \tau_d$. In the second step, we use the nonparametric estimates to obtain estimators of the GARCH parameters.

3.1 Estimation of the Nonparametric Model Components

In order to estimate the nonparametric functions $\tau_0, \dots, \tau_d$, we first transform the multiplicative model (4) into an additive one. Given the resulting estimators of the additive model we retrieve the estimates of the components in the multiplicative model by applying the reverse transform. Under the assumptions in Section 4, we can square the model equation (4) and take logarithms yielding

$$Z_{t,T} = m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + u_t, \tag{9}$$

where $Z_{t,T} := \log Y_{t,T}^2$, $m_j := \log \tau_j^2$ for $j = 0, \dots, d$, and $u_t := \log \varepsilon_t^2$. The model structure in (9) corresponds to the one used in Vogt and Walsh (2019) without a periodic component. Note that the functions $m_0, \dots, m_d$ in (9) are only identified up to an additive constant. To identify them, we assume that

$$\int_0^1 m_0(x_0)dx_0 = 0 \quad \text{and} \quad \int_{\mathbb{R}} m_j(x_j)p_j(x_j)dx_j = 0 \quad \text{for } j = 1, \dots, d, \tag{10}$$

where p_j is the marginal density of X_t^j. Furthermore, we normalize the error to have zero mean, $\mathbb{E}[u_t] = 0$, which introduces a constant m_c to (9), and we are left with

$$Z_{t,T} = m_c + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + u_t. \tag{11}$$

The formulation in (11) corresponds to the model setup in Vogt and Walsh (2019), where the model constant was subsumed into the periodic component. Thus, using the degenerate periodic component estimate $\tilde{m}_c = \frac{1}{T}\sum_{t=1}^{T} Z_{t,T}$, we can apply their smooth backfitting approach to estimate the nonparametric components $m_0, \dots, m_d$. Denote the resulting estimators by $\tilde{m}_0, \dots, \tilde{m}_d$. We refer the reader to section 3.2 in Vogt and Walsh (2019) for a precise definition of the estimators. In Section 4, we will give a set of sufficient conditions ensuring that the assumptions in Vogt and Walsh (2019) are fulfilled, thus allowing us to appeal directly to the asymptotic results for $\tilde{m}_c, \tilde{m}_0, \dots, \tilde{m}_d$ derived there. Finally, to get the estimators of the multiplicative components we apply the reverse transform to get

$$\tilde{\tau}_j = \sqrt{\exp(\tilde{m}_j)} \tag{12}$$

for $j = 0, \dots, d$.

3.2 Estimation of the Parametric Model Components

In order to estimate the parametric model components, suppose initially that the nonparametric components $\tau_0^2, ..., \tau_d^2$ were known. If this were the case, the GARCH variables given by

$$\varepsilon_t^2 = \frac{Y_{t,T}^2}{\tau_0^2(\frac{t}{T}) \prod_{j=1}^d \tau_j^2(X_j^t)} \tag{13}$$

would be observable and the parameters $\phi_0 := (w_0, a_0, b_0)$ could be estimated by standard quasi maximum likelihood methods using the quasi log-likelihood

$$l_T(\phi) = -\sum_{t=1}^{T} \Big(\log v_t^2(\phi) + \frac{\varepsilon_t^2}{v_t^2(\phi)} \Big) \tag{14}$$

for the parameter vector $\phi = (w, a, b)$ with

$$v_t^2(\phi) = \begin{cases} \frac{w}{1-b} & \text{for } t = 1 \\ w + a\varepsilon_{t-1}^2 + bv_{t-1}^2(\phi) & \text{for } t = 2, \dots, T \end{cases} \tag{15}$$

denoting the conditional volatility of the GARCH process with starting value $v_1^2(\phi) = w/(1-b)$. The resulting estimator from maximizing the quasi log-likelihood over the parameter space Φ is denoted by $\hat{\phi} = \arg\max_{\phi \in \Phi} l_T(\phi)$.

As the functions $\tau_0^2, \dots, \tau_d^2$ are not observed, the estimator $\hat{\phi}$ is infeasible. However, given the estimates $\tilde{\tau}_0^2, \dots, \tilde{\tau}_d^2$ from the first estimation step, we can replace the ε_t^2 by the terms

$$\tilde{\varepsilon}_t^2 = \frac{Y_{t,T}^2}{\tilde{\tau}_0^2(\frac{t}{T}) \prod_{j=1}^d \tilde{\tau}_j^2(X_j^t)} \tag{16}$$

and use these as approximations to ε_t^2 in the quasi maximum likelihood estimation.

The quasi log-likelihood then becomes

$$\tilde{l}_T(\phi) = -\sum_{t=1}^{T} \big(\log \tilde{v}_t^2(\phi) + \frac{\tilde{\varepsilon}_t^2}{\tilde{v}_t^2(\phi)} \big), \tag{17}$$

where analogously to (15),

$$\tilde{v}_t^2(\phi) = \begin{cases} \frac{w}{1-b} & \text{for } t = 1 \\ w + a\tilde{\varepsilon}_{t-1}^2 + b\tilde{v}_{t-1}^2(\phi) & \text{for } t = 2, \ldots, T \end{cases} \tag{18}$$

is the approximate conditional volatility. Our estimator $\tilde{\phi}$ of the true parameter values ϕ_0 is now defined as

$$\tilde{\phi} = \arg\max_{\phi \in \Phi} \tilde{l}_T(\phi), \tag{19}$$

where the parameter space Φ is assumed to be compact.

4 Asymptotics

Asymptotic properties for the estimators of the nonparametric components, $\tilde{\tau}_0, \ldots, \tilde{\tau}_d$, are stated in Section 4.1. The corresponding results on the asymptotic behaviour of the estimator of the GARCH parameters, $\tilde{\phi}$, are given in Section 4.2. In order to establish the asymptotic properties of our nonparametric estimators we make the following assumptions.

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

(A2) *The functions τ_j $(j = 0, \ldots, d)$ are twice (continuously) differentiable, strictly positive, and bounded away from zero with Lipschitz continuous second derivatives.*

(A3) *The processes $\{X_t\}$ and $\{\varepsilon_t\}$ are independent and the error process is normalized* *s.t.* $\mathbb{E}[\log \varepsilon_t^2] = 0$.

(A4) *The conditional volatility* σ_t^2 *is bounded away from zero and the GARCH residuals* η_t *have a density with respect to Lebesgue measure which is bounded in a neighbourhood of zero.*

(A5) *The variables* X_t *have compact support, say* $[0,1]^d$.

(A6) *The kernel* K *is bounded, has compact support* ($[-C_1, C_1]$*, say) and is symmetric about zero. Moreover, it fulfills the Lipschitz condition that there exists a positive constant* L *such that* $|K(u) - K(v)| \leq L|u - v|$.

(A7) *The density* p *of* X_t *and the densities* $p_{(0,l)}$ *of* (X_t, X_{t+l}), $l = 1, 2, \ldots,$ *are uniformly bounded. Furthermore,* p *is bounded away from zero on* $[0,1]^d$*. The first partial derivatives of* p *exist and are continuous.*

(A8) *There exists a constant* C *such that* $\mathbb{E}[|u_t|^\theta] := \mathbb{E}[|\log(\varepsilon_t^2)|^\theta] < \infty$ *for some* $\theta > \frac{8}{3}$.

(A9) *The bandwidth* h *satisfies either of the following:*

(A9a) $T^{\frac{1}{5}} h \to c_h$ *for some constant* c_h.

(A9b) $T^{\frac{1}{4}+\delta} h \to c_h$ *for some constant* c_h *and some small* $\delta > 0$.

The assumptions (A1) to (A9) ensure that the transformation used to derive the additive model in (9) is admissible and that the components in the additive model (11) satisfy the assumptions made in Vogt and Walsh (2019). Assumption (A1) restricts the nonstationarity in the model to result from the time-varying component τ_0. Assumption (A2) ensures that the nonparametric functions in (11) satisfy the smoothness conditions in Vogt and Walsh (2019). The independence assumption and normalization of the error process in (A3) ensures that the regression error in (11), u_t, is (conditionally) mean zero. Assumption (A4) along with the boundedness assumption in (A2) allows us to use the transform leading to the additive model (9).

(A5) is only needed for the second estimation step. For the first step, we could allow the support of X_t to be unbounded and estimate the functions $\tau_0, \ldots, \tau_d$ uniformly over compact subsets of the support. However, for ease of notation, we assume (A5) throughout the paper. The remaining assumptions are restatements of the corresponding assumptions made in Vogt and Walsh (2019).

The exponentially decaying mixing rates assumed in (A1) are not necessary and could be replaced by sufficiently high polynomial rates. We nevertheless make the stronger assumption (A1) to keep the notation and structure of the proofs as clear as possible. Furthermore, at the expense of additional complications in the proofs, given some modifications to assumption (A8) the independence condition in (A3) could be weakened to the requirement that almost surely $\mathbb{E}[\varepsilon_t^2|X_t] = \mathbb{E}[\varepsilon_t^2]$ and $\mathbb{E}[\log \varepsilon_t^2|X_t] = 0$, which would be satisfied if X_t and ε_t were contemporaneously independent.

In order to prove that our GARCH parameter estimators in the second estimation step are consistent and asymptotically normal, we will require the following additional assumptions.

(A10) *The parameter space* Φ *is a compact subset of* $\{\phi \in \mathbb{R}^3 \,|\, \phi = (w, a, b)$ *with* $0 < \underline{\kappa} \leq w, a \leq \overline{\kappa} < \infty$ *and* $0 \leq b < 1\}$ *with constants* $\underline{\kappa}$ *and* $\overline{\kappa}$. *The true parameter* $\phi_0 = (w_0, a_0, b_0)$ *is an interior point of* Φ *and* $a_0 + b_0 < 1$.

(A11) $\mathbb{E}[\varepsilon_t^{8+\delta}] < \infty$, *for some* $\delta > 0$.

Assumption (A10) is standard in the estimation theory for GARCH models. Note that it also implies that σ_t^2 is bounded away from zero, which was assumed in (A4). The moment condition in (A11) is needed to show asymptotic normality of the GARCH estimates.

4.1 Asymptotics for the Nonparametric Model Components

As we are mainly interested in the squared version of the estimates $\tilde{\tau}_0, \ldots, \tilde{\tau}_d$ in our multiplicative model, we will restrict ourselves to reporting results for these.

Theorem 4.1. *Suppose that conditions (A1) – (A8) hold.*

(a) Assume that the bandwidth h satisfies either (A9a) or (A9b). Then, for $I_h = [2C_1 h, 1-2C_1 h]$ and $I_h^c = [0, 2C_1 h) \cup (1-2C_1 h, 1]$,

$$\sup_{x_j \in I_h} \left| \tilde{\tau}_j^2(x_j) - \tau_j^2(x_j) \right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) \tag{20}$$

$$\sup_{x_j \in I_h^c} \left| \tilde{\tau}_j^2(x_j) - \tau_j^2(x_j) \right| = O_p(h) \tag{21}$$

for all $j = 0, \ldots, d$.

(b) Assume that the bandwidth h satisfies (A9a). Then, for any $x = (x_0, \ldots, x_d)$ with $x_0, \ldots, x_d \in (0,1)$,

$$T^{\frac{2}{5}} \begin{bmatrix} \tilde{\tau}_0^2(x_0) - \tau_0^2(x_0) \\ \vdots \\ \tilde{\tau}_d^2(x_d) - \tau_d^2(x_d) \end{bmatrix} \xrightarrow{d} N(B_{\tau^2}(x), V_{\tau^2}(x))$$

with the bias term $B_{\tau^2}(x) = [\tau_0^2(x_0)c_h^2(\beta_0(x_0) - \gamma_0), \ldots, \tau_d^2(x_d)c_h^2(\beta_d(x_d) - \gamma_d)]'$ and the covariance matrix $V_{\tau^2}(x) = \mathrm{diag}(\tau_0^4(x_0)v_0(x_0), \ldots, \tau_d^4(x_d)v_d(x_d))$. Here, $v_0(x_0) = c_h^{-1} c_K \sum_{l=-\infty}^{\infty} \gamma_u(l)$ and $v_j(x_j) = c_h^{-1} c_K \sigma^2 / p_j(x_j)$ for $j = 1, \ldots, d$ with $c_K = \int K^2(u)du$, $\gamma_u(l) = \mathrm{Cov}(u_t, u_{t+l})$ and $\sigma^2 = \mathrm{Var}(u_t)$ for $u_t = \log \varepsilon_t^2$. Furthermore, the functions $\beta_j(x_j)$ for $j = 0, \ldots, d$ as well as the constants γ_j for $j = 0, \ldots, d$ are defined exactly as in theorem 2 of Vogt and Walsh (2019).

To derive the above results, we first obtain the asymptotic properties of the estimators $\tilde{m}_0, \ldots, \tilde{m}_d$ for the components of the additively transformed model (11) and then use the smoothness of the transform $\tilde{\tau}_j^2 = \exp(\tilde{m}_j)$ for $j = 0, \ldots, d$. Our assumptions ensure that we can appeal to theorem 2 of Vogt and Walsh (2019) to get the asymptotics of the estimators $\tilde{m}_0, \ldots, \tilde{m}_d$. The main idea of the proof there is to exploit the fact that rescaled time behaves similarly to a random variable which has a uniform distribution on $(0,1]$ and is independent of the other covariates. Some details are given in Appendix A.

The rates of convergence given in Theorem 4.1(a) differ for the interior and boundary regions of the support of the covariates. In particular, the rate near the boundary in (21) is slower than in the interior (20). However, the slow convergence at the boundary does not pose a problem for the second estimation step as the size of the boundary region shrinks sufficiently fast as $T \to \infty$.

4.2 Asymptotics for the Parametric Model Components

Given the estimators for $\tau_0^2, \ldots, \tau_d^2$ from the first step, the GARCH parameters ϕ_0 are estimated by $\tilde{\phi}$ as outlined in Section 3.2. In this subsection, we look at consistency and asymptotic normality of $\tilde{\phi}$. The following theorem establishes consistency.

Theorem 4.2. *Suppose that the bandwidth h satisfies (A9a) or (A9b). In addition, let assumptions (A1) – (A8) and (A10) be fulfilled. Then $\tilde{\phi}$ is a consistent estimator of ϕ_0, i.e. $\tilde{\phi} \xrightarrow{P} \phi_0$.*

We next give a result on the limiting distribution of the GARCH estimates which shows that these are asymptotically normal.

Theorem 4.3. *Suppose that the bandwidth h satifies (A9b) and let assumptions (A1) – (A8) together with (A10) – (A11) be fulfilled. Then it holds that $\sqrt{T}(\tilde{\phi} - \phi_0) \xrightarrow{d} N(0, \Sigma)$. Details on the covariance matrix Σ can be found in Appendix B.*

The proof of asymptotic normality is the theoretically most challenging part of the paper. The details are postponed to the appendices. For now we will be content with providing an outline. By the usual Taylor expansion argument, we arrive at

$$\sqrt{T}(\tilde{\phi} - \phi_0) = -\left[\Big(\frac{1}{T} \frac{\partial^2 \tilde{l}_T(\bar{\phi}_{i,j})}{\partial \phi_i \partial \phi_j} \Big)_{1 \leq i,j \leq 3} \right]^{-1} \frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi},$$

where the term in brackets is the matrix with (i,j)-th element as stated in parenthesis and all $\bar{\phi}_i := (\bar{\phi}_{i,1}, \ldots, \bar{\phi}_{i,3})'$ are between $\tilde{\phi}$ and ϕ_0. The term in brackets can be shown to converges in probability to a nonsingular deterministic matrix. The asymptotic distribution is thus determined by the term $\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi}$, which we rewrite as

$$\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} = \underbrace{\frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi}}_{=:A_1} + \underbrace{\Big(\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} - \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi} \Big)}_{=:A_2}.$$

We will prove that this term is asymptotically normal. Asymptotic normality of A_1 can be shown by well-known results from estimation theory for GARCH models. The main challenge is to derive a stochastic expansion of the term A_2. This requires rather involved and nonstandard arguments which are presented in detail in Appendix B. In particular, we cannot just extend the arguments presented in Hafner and Linton (2010) to fit our setting. Once we have provided the expansion of A_2, we are in a position to apply a central limit theorem to the sum $A_1 + A_2$, which completes the proof. We will see that the term A_2 is itself asymptotically normal and thus contributes to the limit distribution. As a consequence, we obtain an additional term in the asymptotic variance compared to the case where we observe the GARCH errors and would only have the term A_1, thereby reflecting the additional uncertainty that results from not knowing the functions $\tau_0, \ldots, \tau_d$.

The expression for the limiting variance in Theorem 4.3, Σ, involves functions obtained from a higher order expansion of the stochastic part of the backfitting estimates $\tilde{m}_0, \ldots, \tilde{m}_d$ (see Theorem A.1 in Appendix A.1). Not only is it very complicated to calculate the exact form of these functions, it is even more challenging to give consistent estimates for them making the construction of a consistent estimate of Σ a difficult and yet unresolved problem beyond the scope of the present manuscript.

5 Application

To illustrate our model, we apply it to a sample of daily financial data spanning the period from 2nd January 1986 until 21st March 2019. The model to be estimated is given by

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big)\prod_{j=1}^{3}\tau_j(X_{t-1}^{j})\varepsilon_t, \tag{22}$$

where $\{\varepsilon_t\}$ is a GARCH(1,1) process, $Y_{t,T}$ are S&P 500 log-returns and the covariates are three different lagged interest rate spreads constructed from data obtained from the FRED database of the Federal Reserve bank of St. Louis.[1] The data used in the application are plotted in Figure 1. The top left hand panel shows the log return series. The remaining panels depict the regressor series. The top right hand panel contains the series of differences between the yields, also called the yield or credit spread, of Moody's seasoned Baa and Aaa corporate bonds. This spread can be interpreted as the risk premium of low investment grade over high investment grade corporate debt or as a measure of the credit risk of investing in low investment grade corporate debt versus high investment grade corporate debt. The regressor in the lower left hand panel is the yield spread of Moody's seasoned Aaa corporate bonds over the interest rate of 10 year constant maturity U.S. treasuries, which can similarly be interpreted as a measure of the credit risk of high investment grade corporate debt over U.S. sovereign debt. Finally, in the lower right hand corner is a measure of the slope of the yield curve given by the difference in the interest rates of 10 year and 1 year constant maturity U.S. treasuries. Although it can be argued that some of these series may be modelled as nonstationary processes, we will consider them as samples from highly persistent yet stationary processes.

The component functions $\tau_0, \ldots, \tau_3$ as well as the GARCH parameters (ω, a, b) are estimated following the procedure given in Section 3. The bandwidths of the pro-

[1]The historical prices of the S&P 500 are from Yahoo! Finance available at finance.yahoo.com. The Federal Reserve data can be obtained from https://fred.stlouisfed.org/.

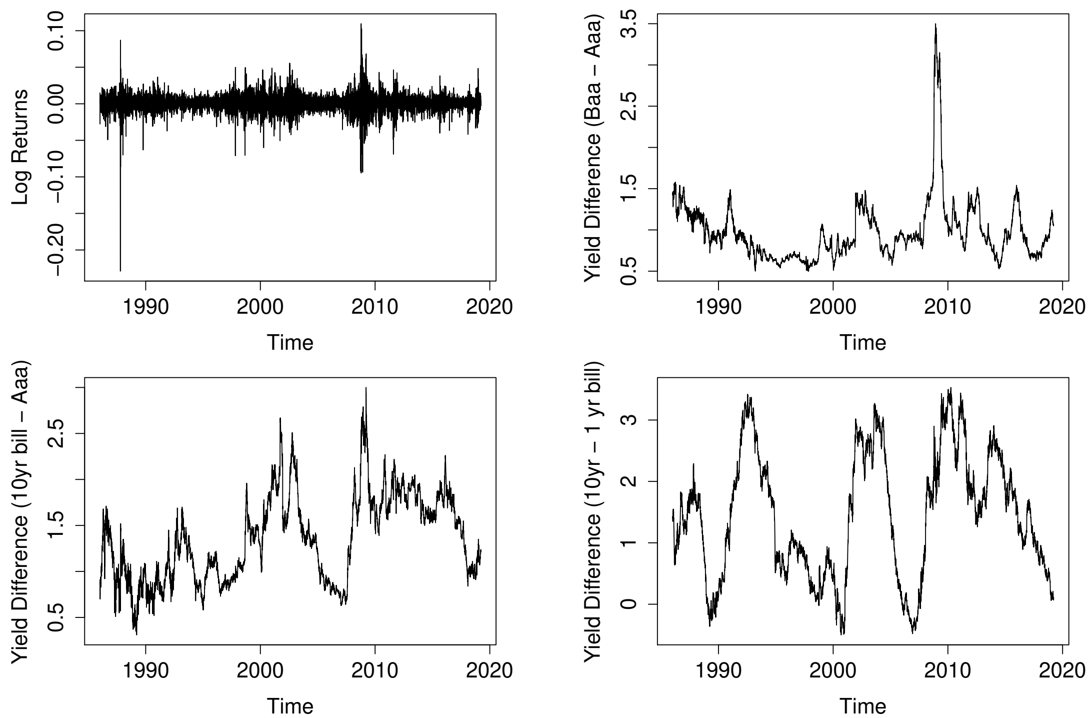

Figure 1: Data used in the application. Dependent variable: S&P 500 log returns (top left). Regressors: Yield difference between Baa and Aaa bonds (top right); Aaa bonds and 10 year Treasuries (bottom left); 10 year and 1 year Treasuries (bottom right). Sample period: 3$^{\text{rd}}$ January 1986 until 21$^{\text{st}}$ March 2019. Frequency: Daily.

cedure are selected based on iterating the plugin formula given in section of 5.2 of Vogt and Walsh (2019). In our application the iteration procedure terminated after 37 iterations with a bandwidth vector of approximately $h = (0.168, 0.192, 0.230, 0.180)'$. The estimation results for the nonparametric model components are presented in Figures 2 and 4. The solid line in Figure 2 gives (a scaled version of) the estimate $\tilde{\tau}_0^2$. The dashed lines are the pointwise 95% confidence intervals. As $\tilde{\tau}_0^2$ has been scaled in accordance with the normalization of the other component estimates discussed later on, it only estimates the time varying unconditional volatility level in (8) up to a multiplicative constant. Comparing the estimate in Figure 2 with the log return series of the S&P 500 in the top left hand panel of Figure 1 we see that the estimate captures the periods of increased log return variance around the events of Black Monday in 1987 as well as the dot-com crash in the early 2000s. Interestingly though the turbulences surrounding the recent financial crisis is not picked up by the estimate, which already suggests that the regressors have more explanatory power in the recent financial crisis. This is further exemplified in Figure 3 by comparing the estimates of time varying unconditional volatility in our model and the simpler model (1) without covariates. The solid line in Figure 3 is a rescaled version of $\tilde{\tau}_0^2$ that estimates the unconditional volatility level in our model, whereas the dashed line is the estimated unconditional volatility obtained from the simpler model (1). The main difference between the two curves in Figure 3 is that the estimated unconditional volatility level for the model without regressors does not tail off so much during the recent financial crisis. During the earlier crises, however, the difference in shape between the two curves is not so striking. Thus, indeed our regressors seem to be primarily good at explaining the recent financial crisis, which is quite plausible as our regressors mainly capture aspects of credit risk in the U.S.

Estimates of the squared components τ_j^2 for $j = 1, 2, 3$ are given as solid lines in Figure 4. The dashed lines are again the pointwise 95% confidence intervals. The estimates $\tilde{\tau}_j^2$ have been normalized such that $\tilde{\tau}_j^2(x_j^m) = 1$, where x_j^m is the median

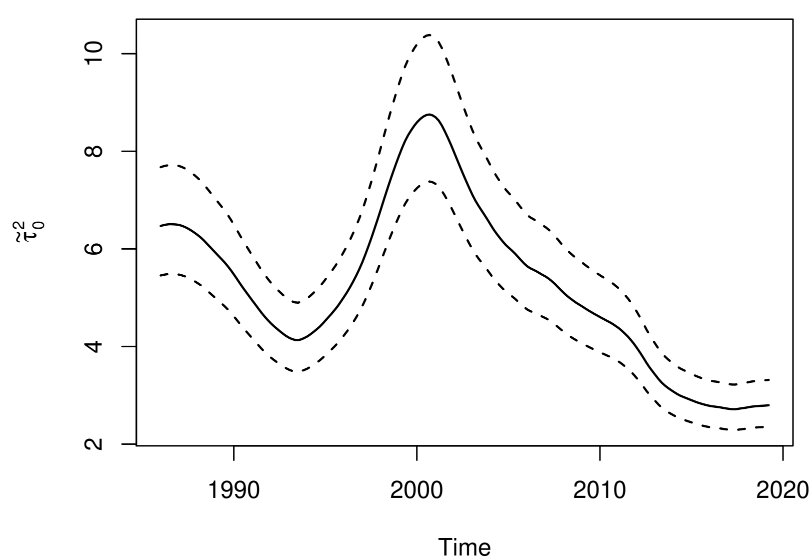

Figure 2: Estimate of squared trend component τ_0^2.

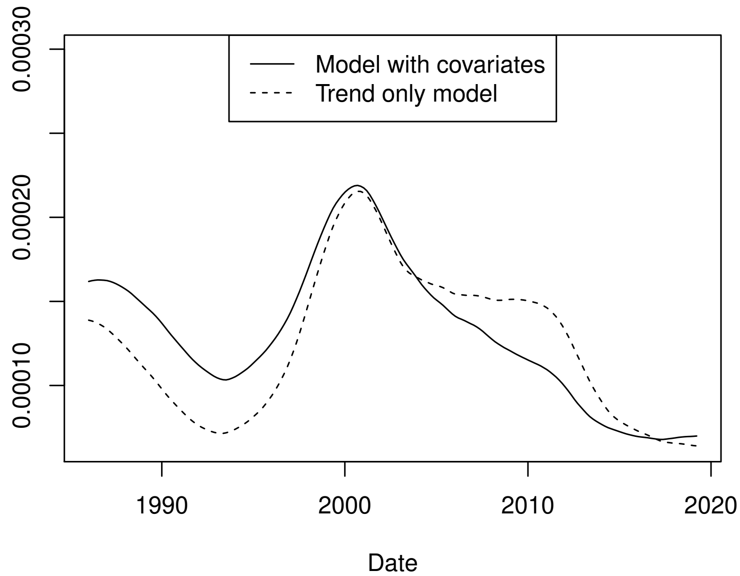

Figure 3: Time-varying unconditional volatilities for our model and the simpler model (1) without regressors.

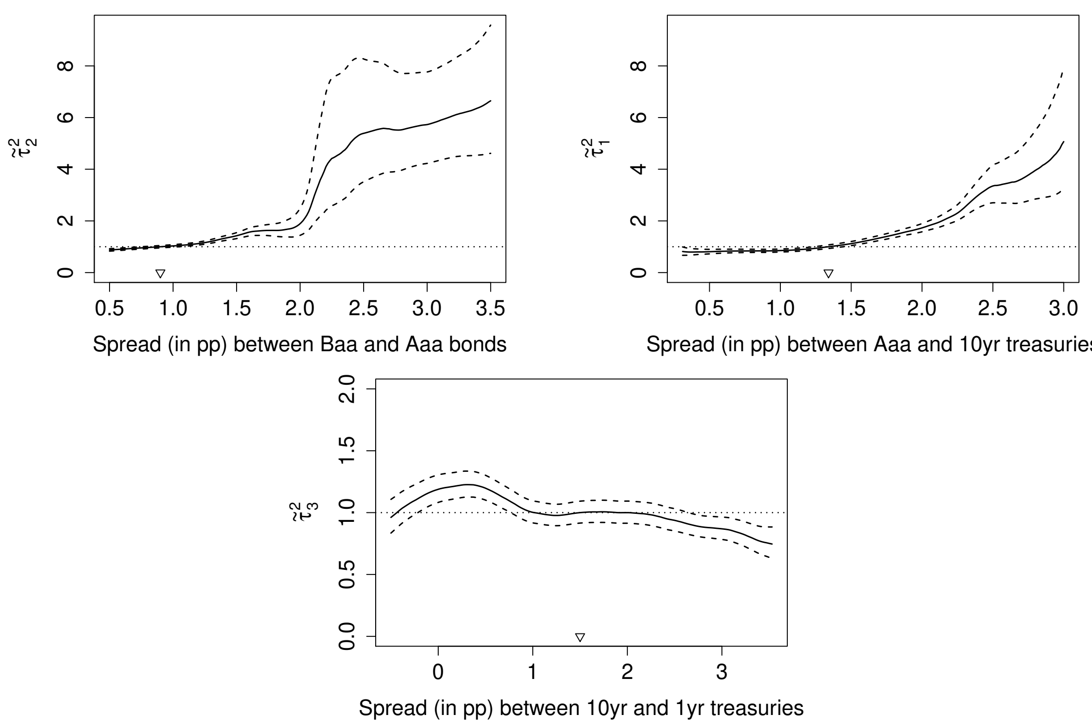

Figure 4: Estimates of τ_j^2 for $j = 1, 2, 3$ normalized to one at median value of regressors ($\bigtriangledown$). Spreads measured in percentage points.

observed realization of the j-th covariate X_t^j over the modelling period, the value of which is indicated by the triangle ($\bigtriangledown$) on the x-axis. Thus, the multiplicative effect of the j-th covariate on volatility is normalized to 1, given by the dotted line in the figure, if it takes a "normal" (i.e., its median) value. As

$$\mathbb{E}[Y_{t,T}^2 | X_t] = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{3} \tau_j(X_{t-1}^j), \tag{23}$$

the normalization allows for the estimates $\tilde{\tau}_j^2$ for $j = 1, 2, 3$ to be interpreted as the multiplicative effect of the covariate X_{t-1}^j on S&P 500 volatility. To illustrate this, let us compare volatility between two different settings: Hold all the covariates except the j-th fixed at some value x_{-j} and change the j-th regressor X_{t-1}^j from its median x_j^m to some value x_j. From (23), one can then see that the conditional volatility is changed by the factor $\tau_j^2(x_j)/\tau_j^2(x_j^m) = \tau_j^2(x_j)$ if $\tau_j^2(x_j^m)$ has been normalized to one. Consequently, the fits $\tilde{\tau}_j^2(x_j)$ estimate the factor by which the volatility level gets increased or dampened, when the j-th covariate changes from a normal value (i.e. its median) to some other more extreme value. The upper two panels in Figure 4 can be interpreted as the estimated multiplicative effect of credit risk of low over high investment grade corporate debt in the top left hand panel and of high investment grade corporate debt over U.S. sovereign debt in the top right hand panel. For ease of comparison the scale of the y-axis in both panels is the same. Both estimated effects are clearly increasing and highly nonlinear. The estimates are less precise for large regressor values as seen by the fanning out of the confidence bands. In terms of the shape of the two estimated effects, the main difference is that in the left hand panel for very large regressor values, above approximately 2 pp, the estimated effect increases quite sharply and then remains at a higher level. Although the range of the effects in both panels is quite similar it should be pointed out that the values of the credit spread between low and high investment grade corporate debt above 2 pp all occur in one burst during the recent crisis, see Figure 1. Thus, firstly, the credit risk of low over high investment grade corporate bonds seems to have been particular important in the recent crisis. Secondly, outside the recent crisis the credit risk of high investment grade corporate bonds compared to U.S. sovereign bonds has a larger effect than that of low to high investment grade corporate bonds. Finally, the lower panel in Figure 4 contains the estimate of the effect of the slope of the yield curve on volatility. The estimation precision is much more homogeneous than for the other two components. The estimated effect is also nonlinear. Somewhat surprisingly, the main deviation from linearity is due to a decrease in the estimated effect for negative regressor values, which corresponds to a so-called inverted yield curve, typically interpreted as a predictor of a recession. For upward sloping yield curves the estimates are decreasing. Thus, the steeper the upward sloping yield curve the lower the volatility. Finally, note that the scale of the y-axis in the lower panel is substantially smaller than in the upper panels. Hence the effect of the slope of the yield curve on volatility is not nearly as large as that of the measures of credit risk discussed before.

We finish our application with the estimation results for the parametric model components. In Table 1, we compare the GARCH estimates of our model with the ones obtained from the simpler model (1) and from a standard GARCH(1,1) model. The

	$\tilde{w}$	$\tilde{a}$	$\tilde{b}$	$\tilde{a}+\tilde{b}$	$\widetilde{HL}$
Standard GARCH(1,1)	0.000002	0.101	0.885	0.986	50
Model with trend	0.026	0.104	0.869	0.973	26
Model with trend and covariates	0.047	0.103	0.847	0.951	15

Table 1: GARCH parameter estimates for GARCH(1,1) and for models (1) and (22),

sum of the two estimated parameters $\tilde{a}+\tilde{b}$ reported in the penultimate column of Table 1 measures the persistence of shocks to volatility. One can see that this persistence measure decreases from 0.986 to 0.973 when accounting for time-varying unconditional volatility. This is in line with previous findings in the literature (compare e.g. Feng (2004)). Including our covariates in the model further decreases the estimated persistence to 0.951. Note that the reported decrease in persistence is quite dramatic even though it may seem rather small at first sight (compare the discussion in Lamoureux and Lastrapes (1990) and Mikosch and Stărică (2000) on this issue). To give some meaning to the numerical values of the persistence we will consider the half life of variance as in Lamoureux and Lastrapes (1990), which for a GARCH(1,1) model with parameters (ω, a, b) is defined by $HL = 1 - [\log(2)/\log(a+b)]$. The half life of volatility for the GARCH component gives the number of days it takes for a shock to the GARCH component to diminish to half its initial value. The last column of Table 1 provides the estimated half lifes for the three competing models. Allowing for time varying unconditional volatility leads to a substantial decrease of the estimated half life from 50 trading days (roughly 10 weeks) to 26 trading days. Additionally including our regressors leads to a further decrease of the estimated half life to 15 trading days, which corresponds to 3 weeks.

To sum up, our results suggest that we can explain a good deal of S&P 500 log return volatility by our model. We have also seen that the regressors we included were more important in the recent financial crisis, especially the credit spread between low and high investment grade corporate debt. The estimated effects are all highly nonlinear. Over the entire sample period the yield spread between high investment grade corporate debt and U.S. sovereign debt can be argued to have the largest effect on volatility. The effect of the slope of the yield curve shows that the volatility is lower for more upwardly sloping yield curves and sufficiently inverted yield curves. Finally, by including our regressors the persistence remaining in the GARCH component is substantially lower than in the simpler model containing only a trend component.

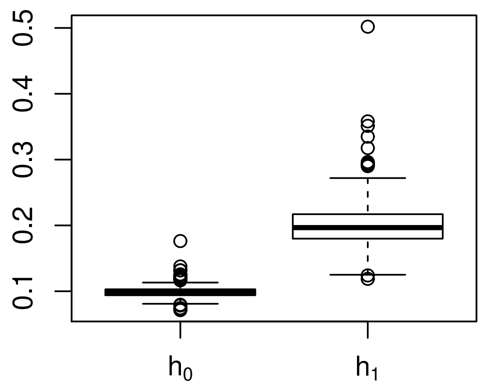

Figure 5: The distribution over the 200 simulations of the selected bandwidths. The left hand boxplot h_0 corresponds to the bandwidth associated with the estimation of the trend component. The right hand boxplot h_1 corresponds to the bandwidth for the component of the stochastic regressor.

6 Simulation

To illustrate the behaviour of our estimation method, we report the results of a small simulation study designed to mirror certain aspects of the application. The underlying data generating process we consider is given by

$$Y_{t,T} = \tau_c \tau_0 \left(\frac{t}{T}\right) \tau_1(X_t)\varepsilon_t, \tag{24}$$

where $\{\varepsilon_t\}$ is a GARCH(1,1) process with standard normal innovations and parameters $(\omega, a, b) = (0.05, 0.1, 0.85)$, thus ensuring that $\mathbb{E}[\varepsilon_t^2] = 1$. Note that the parameters are close to the estimates in the application, see Table 1. The component $\tau_1(\cdot)$ is chosen such that $\tau_{1,c}\tau_1^2(x) = 1 + 50(x - 0.5)1(x \geq 0.5)$ is piecewise linear with $\tau_{1,c} = \exp(\mathbb{E}[\log(\tau_1^2(X_t)])$. The covariate process $\{X_t\}$ is a highly persistent centred AR(1) process with standard normal innovations and AR(1) coefficient of 0.98, that is rescaled to the unit interval. The trend component $\tau_0(\cdot)$ is set equal to the estimated trend component from the application as given in (12). A (scaled) version of the trend component was displayed in Figure 2. Finally, $\tau_c = \sqrt{\tau_{0,c}\tau_{1,c}}$ with $\tau_{0,c} = \frac{1}{T}\sum_{t=1}^{T} \log(\tau_0^2(t/T))$ is a normalization constant. Note that by construction, the transformed component functions given by $m_j(\cdot) = \log(\tau_j^2(\cdot))$ for $j \in \{0,1\}$ fulfill the normalization constraints in (10).

We simulate 200 data sets of length $T = 8000$, which is close to the number of observations in the application. We run our estimation procedure on each simulated data set. In 3.5% of the cases the iterative bandwidth selection procedure had not converged within a limit of 100 iterations and the current value within the iteration was used for the estimation. Figure 5 shows that over the 200 simulations there is less variation in the chosen bandwidth for the trend function than for the other component function. Figure 6 shows the estimates of the nonparametric trend function over the

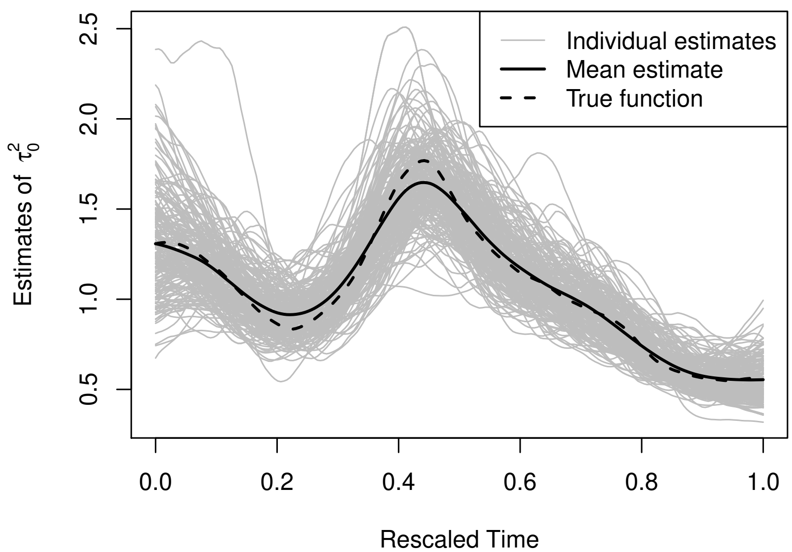

Figure 6: Each grey line is the estimate of τ_0^2 for one of the simulations. The pointwise mean of these estimates is given by the solid black line. The dashed black line is the true squared trend component.

200 simulations. We can see that the true trend function is estimated quite well as the mean over the simulations is close to the true trend. Furthermore the uncertainty of the estimate seems to be quite homogenous.

In Figure 7, we can see that the mean estimate for the stochastic regressor component is nonlinear, increasing and convex. Moreover, the shape as well as the increase in estimation uncertainty for large values of the regressor are reminiscent of the fits for

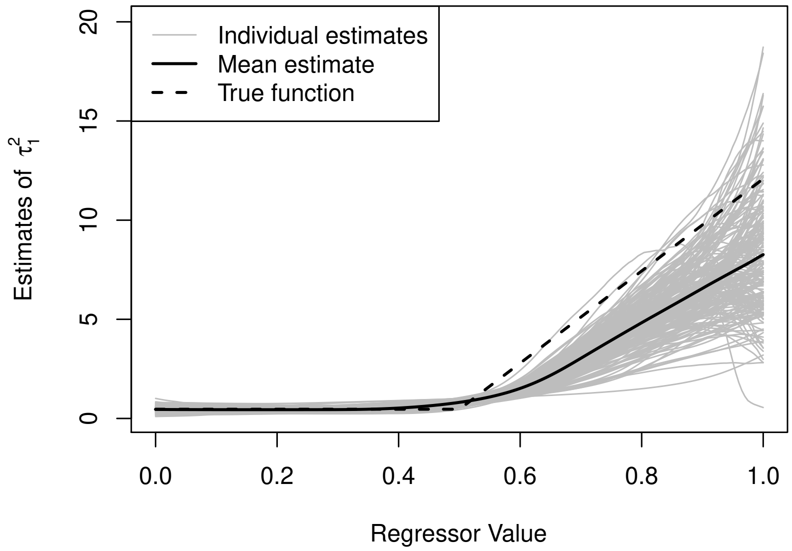

Figure 7: Each grey line is the estimate of τ_1^2 for one of the simulations. The pointwise mean of these estimates is given by the solid black line. The dashed black line is the true squared component.

the first two regressors in the application as depicted in the top panel of Figure 4. The underestimation of the increasing part of τ_1^2 can be explained by the fact that the slope of a linear regression function is underestimated by a local constant smoother when the bandwidth is increased. Support for this explanation is provided by the disappearance of the understimation, when the bandwidth for the second component is fixed at a value of 0.1.

Finally, we take a look at how well the GARCH parameters are estimated and illustrate that neglecting a nonparametric component may severely affect the parameter estimates. In comparing the estimates, we consider four settings. In the oracle model we fit a GARCH(1,1) to the actual innovation process $\{\varepsilon_t\}$, which corresponds to the case that we know the nonparametric components. In the full model case we estimate the nonparametric components τ_0 and τ_1 and fit a GARCH(1,1) process to the residuals $\tilde{\varepsilon}_t = \frac{Y_{t,T}}{\tilde{\tau}_0(t/T)\tilde{\tau}_1(X_t)}$. In the trend only case, we fit a model that erroneously omits the τ_1 component. Thus, we fit a GARCH(1,1) process to the residuals $\check{\varepsilon}_t = \frac{Y_{t,T}}{\check{\tau}_0(t/T)}$ with $\check{\tau}_0$ denoting the estimator of the nonparametric trend component τ_0. The last

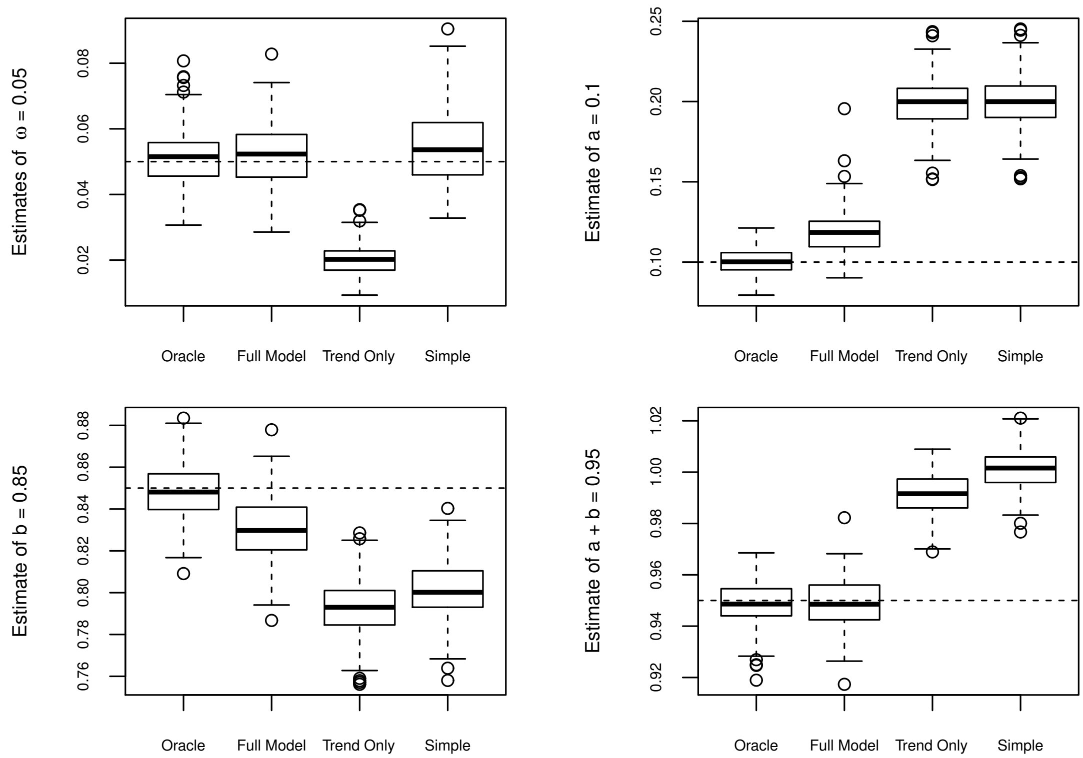

Figure 8: The distribution of the estimates for ω, a, b and the persistence $a+b$ over all 200 simulations for four different models. "Oracle" refers to the infeasible case, where the functions in the full model of (24) are known. "Full Model" is the feasible version of the model in (24). "Trend Only" refers to a model without a component function for the stochastic regressor. Lastly, "Simple" refers to the standard GARCH(1,1) without any nonparametric components.

setting is the simple model that omits both nonparametric components and fits a GARCH(1,1) process to the $Y_{t,T}$. Figure 8 provides the distribution of the estimates over the 200 simulations. We can see that in terms of estimating ω and the persistence $a + b$ the estimates from the full model are nearly as good as in the oracle case when the nonparametric components are known. Although the persistence is well estimated, the fitted GARCH model places more weight on the squared past returns and less weight on the past volatility than the true process. Lastly, in our particular setting we can see that omitting the component τ_1 leads to severely biased GARCH parameter estimates. Most notably the estimated persistence of the GARCH innovation is substantially larger, though still below that for the simple model. In fact, for some cases the bias is so severe, that the resulting estimated GARCH process is no longer covariance stationary.

7 Conclusion

We have proposed a new semiparametric volatility model, which generalizes the class of models $Y_{t,T} = \tau(\frac{t}{T})\varepsilon_t$, as for example considered in Feng (2004) and Engle and Rangel (2008). These models are able to account for nonstationarities in the volatility process. In addition, we are able to include covariates in a nonparametric way, hence allowing us to flexibly capture the effects of the financial and economic environment. We have derived the asymptotic theory both for the nonparametric and the parametric part of the model. To estimate the nonparametric model components, we have adapted the smooth backfitting approach of Mammen et al. (1999) to our nonstationary setting. Given the backfitting estimators, we were able to construct GARCH parameter estimates and to show that they are asymptotically normal. In particular, they converge at the fast parametric rate even though the nonparametric smoothers from the first step have slower nonparametric convergence rates. We concluded by illustrating the strengths of our model by applying it to financial data. In particular, our semiparametric approach allows us to estimate the form of the relationship between volatility and its potential sources. Therefore, we manage to go beyond existing parametric approaches such as in Engle and Rangel (2008) and Engle et al. (2013). Finally, we have provided simulation based evidence showing that misspecification in terms of omitting a nonparametric component can severely bias the GARCH parameter estimates.

A Appendix - results of additive model estimators

This section deals with the asymptotics for the estimators in the additive model (11). First, we will restate some results on uniform expansions for the estimators in the additive model (11) established in Vogt and Walsh (2019). These expansions will be needed to establish the asymptotic properties of the GARCH parameter estimators. Secondly, we give a brief statement on how to prove Theorem 4.1.

A.1 Stochastic expansion of estimators in the additive model

Using the modified kernel

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

where $K_h(v) = \frac{1}{h} K(\frac{v}{h})$ and the kernel function $K(\cdot)$ integrates to one, the kernel density estimators of the marginal density p_j of X_t^j and of the joint density $p_{j,k}$ of (X_t^j, X_t^k) are given by

$$\hat{p}_j(x_j) = \frac{1}{T} \sum_{t=1}^{T} K_h(x_j, X_t^j) \tag{25}$$

$$\hat{p}_{j,k}(x_j, x_k) = \frac{1}{T} \sum_{t=1}^{T} K_h(x_j, X_t^j) K_h(x_k, X_t^k). \tag{26}$$

Furthermore, the Nadaraya Watson pilot estimators for the components of the additive model (11) are given by

$$\hat{m}_j(x_j) = \frac{1}{T} \sum_{t=1}^{T} K_h(x_j, X_t^j)(Z_{t,T} - \tilde{m}_c)/\hat{p}_j(x_j), \tag{27}$$

where $\tilde{m}_c = \frac{1}{T} \sum_{t=1}^{T} Z_{t,T}$. The above are for $j = 1, \dots, d$ as well as for $j = 0$ by writing $X_t^0 = \frac{t}{T}$. The components of the smooth backfitting estimator, $\tilde{m} = \tilde{m}_0 + \cdots + \tilde{m}_d$, are characterised as the solutions to the integral equations

$$\tilde{m}_j(x_j) = \hat{m}_j(x_j) - \sum_{k \neq j} \int_0^1 \tilde{m}_k(x_k) \frac{\hat{p}_{k,j}(x_k, x_j)}{\hat{p}_j(x_j)} \, dx_k - \tilde{m}_c$$

with $\int_0^1 \tilde{m}_j(x_j)\hat{p}_j(x_j)dx_j = 0$ for $j = 0, \ldots, d$. In Vogt and Walsh (2019) it is shown that the backfitting estimators $\tilde{m}_j$ can be decomposed into a stochastic part $\tilde{m}_j^A$ and a bias part $\tilde{m}_j^B$ according to $\tilde{m}_j(x_j) = \tilde{m}_j^A(x_j) + \tilde{m}_j^B(x_j)$. The two components are defined by

$$\tilde{m}_j^S(x_j) = \hat{m}_j^S(x_j) - \sum_{k \neq j} \int_0^1 \tilde{m}_k^S(x_k) \frac{\hat{p}_{k,j}(x_k, x_j)}{\hat{p}_j(x_j)} \, dx_k - \tilde{m}_c^S \tag{28}$$

for $S = A$, B. Here, $\hat{m}_k^A$ and $\hat{m}_k^B$ denote the stochastic part and the bias part of the Nadaraya-Watson pilote estimates in (27) defined as

$$\hat{m}_j^A(x_j) = \frac{1}{T} \sum_{t=1}^T K_h(x_j, X_t^j) u_t / \hat{p}_j(x_j) \tag{29}$$

$$\hat{m}_j^B(x_j) = \frac{1}{T} \sum_{t=1}^T K_h(x_j, X_t^j) \big[(m_c - \tilde{m}_c) + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^d m_j(X_t^j) \big] / \hat{p}_j(x_j) \tag{30}$$

for $j = 0, \ldots, d$, again setting $X_t^0 = \frac{t}{T}$ to shorten the notation. Furthermore, $\tilde{m}_c^A = \frac{1}{T}\sum_{t=1}^T u_t$ and $\tilde{m}_c^B = \frac{1}{T}\sum_{t=1}^T \{m_c - \tilde{m}_c + m_0(\frac{t}{T}) + \sum_{j=1}^d m_j(X_t^j)\}$.
The first result provides a higher order expansion of the stochastic part $\tilde{m}_j^A$. The second then provides the corresponding expansion for the bias part $\tilde{m}_j^B$.

Theorem A.1. *Suppose that assumptions (A1) – (A8) apply and that the bandwidth h satisfies (A9a) or (A9b). Then uniformly for $0 \leq x_j \leq 1$,*

$$\tilde{m}_j^A(x_j) = \hat{m}_j^A(x_j) + \frac{1}{T} \sum_{t=1}^T r_{j,t}(x_j) u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big),$$

where $r_{j,t}(\cdot) := r_j(\frac{t}{T}, X_t, \cdot)$ *are absolutely uniformly bounded functions with*

$$|r_{j,t}(x_j') - r_{j,t}(x_j)| \leq C|x_j' - x_j|$$

for a constant $C > 0$.

Theorem A.2. *Suppose that (A1) – (A8) hold. If the bandwidth* h *satisfies (A9a), then*

$$\sup_{x_j \in I_h} |\tilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h^2) \tag{31}$$

$$\sup_{x_j \in I_h^c} |\tilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h) \tag{32}$$

for $j = 0, \ldots, d$. *If the bandwidth satisfies (A9b), we have*

$$\sup_{x_j \in I_h} \left| \tilde{m}_j^B(x_j) + \frac{1}{T} \sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j) \right| = O_p(h^2) \tag{33}$$

$$\sup_{x_j \in I_h^c} \left| \tilde{m}_j^B(x_j) + \frac{1}{T} \sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j) \right| = O_p(h) \tag{34}$$

for $j = 0, \ldots, d$.

For the proofs of Theorems A.1 and A.2 see Vogt and Walsh (2019).

A.2 Proof of Theorem 4.1

The results on the convergence rates and the asymptotic normality for the estimators of the additive components $\tilde{m}_0, \ldots, \tilde{m}_d$ are given in Vogt and Walsh (2019). Since $\tilde{\tau}_j^2 = \exp(\tilde{m}_j)$, Theorem 4.1(a) is an immediate consequence of these results. The joint asymptotic normality in Theorem 4.1(b) follows from the asymptotic normality of the $\tilde{m}_j$ upon applying the delta method with $g(\tilde{m}_j) = \exp(\tilde{m}_j)$ and the Cramér-Wold device.

B Appendix - Proofs of Theorems 4.2 and 4.3

This appendix contains the proofs of Theorems 4.2 and 4.3, which show consistency and asymptotic normality of our estimator for the GARCH parameters. Especially the proof of the asymptotic normality is rather involved. The major challenge is the derivation of a stochastic expansion for $\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi}$ from which we get the asymptotic normal limit. Although the general approach is as in Vogt and Walsh (2019), the arguments are substantially more difficult due to the complexity of the GARCH error in comparison to the much simpler autoregressive type error considered there. In particular, arguments from empirical process theory are needed now. The detailed arguments are collected in a series of lemmas in the supplementary material to the paper. Throughout this appendix, C denotes a finite real constant which may take a different value on each occurrence.

B.1 Auxiliary Results

To start with, we state some facts about the behaviour of the approximate GARCH variables $\tilde{\varepsilon}_t$ and of the conditional volatilities $\tilde{v}_t^2(\phi)$, which were defined in Subsection 3.2. For ease of notation, we use the shorthand $\tau(x) = \prod_{j=0}^d \tau_j(x_j)$ in what follows.

(G1) We can express $\tilde{\varepsilon}_t^2 - \varepsilon_t^2$ as

$$\tilde{\varepsilon}_t^2 - \varepsilon_t^2 = \varepsilon_t^2 \Big[\frac{\tau^2(\frac{t}{T}, X_t) - \tilde{\tau}^2(\frac{t}{T}, X_t)}{\tau^2(\frac{t}{T}, X_t)} + R_\varepsilon\Big(\frac{t}{T}, X_t\Big)\Big]$$

with $\sup_{x\in[0,1]^{d+1}} |R_\varepsilon(x)| = O_p(h^2)$.

(G2) The conditional volatility $v_t^2(\phi)$ has the expansion

$$v_t^2(\phi) = w\sum_{k=1}^{t-1} b^{k-1} + a\sum_{k=1}^{t-1} b^{k-1}\varepsilon_{t-k}^2 + b^{t-1}\frac{w}{1-b},$$

which yields that $\tilde{v}_t^2(\phi) - v_t^2(\phi) = \sum\limits_{k=1}^{t-1} ab^{k-1}(\tilde{\varepsilon}_{t-k}^2 - \varepsilon_{t-k}^2)$.

(G3) It holds that $\max_{1\le t\le T} \sup_{\phi\in\Phi} \left|\tilde{v}_t^2(\phi) - v_t^2(\phi)\right| = O_p(h)$.

(G4) It holds that $\frac{1}{\tilde{v}_t^2(\phi)} - \frac{1}{v_t^2(\phi)} = \frac{v_t^2(\phi)-\tilde{v}_t^2(\phi)}{v_t^2(\phi)v_t^2(\phi)} + R_t(\phi)$ with $\max_{1\le t\le T} \sup_{\phi\in\Phi} |R_t(\phi)| = O_p(h^2)$.

(G5) The derivatives of $v_t^2(\phi)$ with respect to the parameters w, a, and b are given by

$$\begin{aligned}
\frac{\partial v_t^2(\phi)}{\partial w} &= \sum_{k=1}^{t-1} b^{k-1} + \frac{b^{t-1}}{1-b} \\
\frac{\partial v_t^2(\phi)}{\partial a} &= \sum_{k=1}^{t-1} b^{k-1}\varepsilon_{t-k}^2 \\
\frac{\partial v_t^2(\phi)}{\partial b} &= w\Big(\sum_{k=1}^{t-1}(k-1)b^{k-2} + \frac{(t-1)b^{t-2}}{1-b} + \frac{b^{t-1}}{(1-b)^2}\Big) + a\sum_{k=1}^{t-1}(k-1)b^{k-2}\varepsilon_{t-k}^2.
\end{aligned}$$

The above facts are straightforward to verify. We thus omit the details.

B.2 Proof of Theorem 4.2

Let $l_T(\phi)$ and $\tilde{l}_T(\phi)$ be the likelihood functions introduced in (14) and (17) and define $l(\phi) = \mathbb{E}\Big[\frac{1}{T}l_T(\phi)\Big]$. By the triangle inequality,

$$\sup_{\phi\in\Phi}\Big|\frac{1}{T}\tilde{l}_T(\phi) - l(\phi)\Big| \le \sup_{\phi\in\Phi}\Big|\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T}l_T(\phi)\Big| + \sup_{\phi\in\Phi}\Big|\frac{1}{T}l_T(\phi) - l(\phi)\Big|.$$

From standard theory we know that $\sup_{\phi\in\Phi}\left|\frac{1}{T}l_T(\phi) - l(\phi)\right| = o_p(1)$ and that $l(\phi)$ is a continuous function of ϕ with a unique maximum at ϕ_0. If we can further show that

$$\sup_{\phi\in\Phi}\Big|\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T}l_T(\phi)\Big| = o_p(1), \tag{35}$$

then standard theory on M-estimation implies $\tilde{\phi} \xrightarrow{P} \phi_0$.

We will show (35) by decomposing $\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T}l_T(\phi)$ into the sum of three uniformly $o_p(1)$ terms.

$$\begin{aligned}
&\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T}l_T(\phi) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\log\tilde{v}_t^2(\phi) + \frac{\tilde{\varepsilon}_t^2}{\tilde{v}_t^2(\phi)}\Big) + \frac{1}{T}\sum_{t=1}^{T}\Big(\log v_t^2(\phi) + \frac{\varepsilon_t^2}{v_t^2(\phi)}\Big) \\
&= \frac{1}{T}\sum_{t=1}^{T}\big(\log v_t^2(\phi) - \log\tilde{v}_t^2(\phi)\big) + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\Big(\frac{\tilde{v}_t^2(\phi) - v_t^2(\phi)}{\tilde{v}_t^2(\phi)v_t^2(\phi)}\Big) + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\tilde{v}_t^2(\phi)}(\varepsilon_t^2 - \tilde{\varepsilon}_t^2) \\
&=: (A) + (B) + (C).
\end{aligned}$$

In order to prove that the three terms (A), (B), and (C) are indeed uniformly $o_p(1)$, it suffices to show that

$$\max_{1\le t\le T}\sup_{\phi\in\Phi}\big|\tilde{v}_t^2(\phi) - v_t^2(\phi)\big| = o_p(1) \tag{36}$$

$$\frac{1}{T}\sum_{t=1}^{T}\big|\tilde{\varepsilon}_t^2 - \varepsilon_t^2\big| = o_p(1) \tag{37}$$

$$v_t^2(\phi) \ge v_{\min} > 0 \quad \text{and} \quad \tilde{v}_t^2(\phi) \ge v_{\min} > 0 \quad \text{for some constant } v_{\min}. \tag{38}$$

(36) is implied by (G3). For the proof of (37), we use (G1) together with Theorem 4.1 to obtain

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\big|\tilde{\varepsilon}_t^2 - \varepsilon_t^2\big| &\le \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\Big|\frac{\tau^2(\frac{t}{T},X_t) - \tilde{\tau}^2(\frac{t}{T},X_t)}{\tau^2(\frac{t}{T},X_t)} + R_\varepsilon\Big(\frac{t}{T},X_t\Big)\Big| \\
&= O_p(h)\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 = O_p(h).
\end{aligned}$$

Finally, (38) is automatically satisfied, as by (A10)

$$v_t^2(\phi) = w\sum_{k=1}^{t-1}b^{k-1} + a\sum_{k=1}^{t-1}b^{k-1}\varepsilon_{t-k}^2 + b^{t-1}\frac{w}{1-b} \ge w \ge \underline{\kappa} > 0.$$

The same holds true for $\tilde{v}_t^2(\phi)$. $\square$

B.3 Proof of Theorem 4.3

By the usual Taylor expansion argument, we obtain

$$0 = \frac{1}{T}\frac{\partial \tilde{l}_T(\tilde{\phi})}{\partial \phi} = \frac{1}{T}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} + \Big(\frac{1}{T}\frac{\partial^2 \tilde{l}_T(\bar{\phi}_{i,j})}{\partial \phi_i \partial \phi_j}\Big)_{1\leq i,j\leq 3}(\tilde{\phi} - \phi_0),$$

where the matrix of second order partial derivatives has (i,j)-th term as stated in the parenthesis with $\bar{\phi}_i = (\bar{\phi}_{i,1},\ldots,\bar{\phi}_{i,3})'$ between ϕ_0 and $\tilde{\phi}$ for all $i = 1,\ldots,3$. Rearranging and premultiplying by $\sqrt{T}$ yields

$$\sqrt{T}(\tilde{\phi} - \phi_0) = -\left[\Big(\frac{1}{T}\frac{\partial^2 \tilde{l}_T(\bar{\phi}_{i,j})}{\partial \phi_i \partial \phi_j}\Big)_{1\leq i,j\leq 3}\right]^{-1}\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi}.$$

The proof will be completed upon showing that

$$\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} \xrightarrow{d} N(0, Q) \tag{39}$$

$$\frac{1}{T}\frac{\partial^2 \tilde{l}_T(\bar{\phi}_{i,j})}{\partial \phi_i \partial \phi_j} \xrightarrow{P} J(i,j) \quad \text{for all } 1 \leq i,j \leq 3, \tag{40}$$

where Q is some covariance matrix to be specified later on and $J(i,j)$ is the (i,j)-th element of an invertible deterministic matrix J. Thus we see that the asymptotic covariance matrix given in Theorem 4.3 is $\Sigma = J^{-1}QJ^{-1}$.

Proof of (39). Let $v_t^2 = v_t^2(\phi_0)$ and $\tilde{v}_t^2 = \tilde{v}_t^2(\phi_0)$ in order to lighten notation. Writing out the i-th element of the left hand side of (39) we get

$$\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1 - \frac{\tilde{\varepsilon}_t^2}{\tilde{v}_t^2}\Big)\frac{\partial \tilde{v}_t^2}{\partial \phi_i}\frac{1}{\tilde{v}_t^2}.$$

Thus, we obtain

$$\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} = \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial \phi_i} + \left(\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial \phi_i}\right) \tag{41}$$

with

$$\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial \phi_i} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{1}{v_t^2}\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big) \tag{A}$$

$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{\partial \tilde{v}_t^2}{\partial \phi_i}\Big(\frac{1}{v_t^2} - \frac{1}{\tilde{v}_t^2}\Big) \tag{B}$$

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{\tilde{v}_t^2}\frac{\partial \tilde{v}_t^2}{\partial \phi_i}\Big((1-\frac{\tilde{\varepsilon}_t^2}{v_t^2}) - (1-\frac{\varepsilon_t^2}{v_t^2})\Big) \tag{C}$$

$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big((1-\frac{\tilde{\varepsilon}_t^2}{v_t^2}) - (1-\frac{\tilde{\varepsilon}_t^2}{\tilde{v}_t^2})\Big)\frac{1}{\tilde{v}_t^2}\frac{\partial \tilde{v}_t^2}{\partial \phi_i}. \tag{D}$$

In what follows, we show that (A) and (B) are asymptotically negligible, whereas (C) and (D) contribute to the limiting distribution. First we establish the negligibility of the terms (A) and (B). We will only give the arguments for (A) as it is slightly more complicated. The steps to show the negligibility of (B) are analogous.

To begin, replace the truncated conditional volatilities v_t^2 by σ_t^2 to obtain

$$\begin{aligned}
(A) &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{1}{v_t^2}\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big)\\
&= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{\sigma_t^2}\Big)\frac{1}{\sigma_t^2}\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big)\\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big)\Big(\frac{1}{v_t^2} - \frac{1}{\sigma_t^2}\Big) - \varepsilon_t^2\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big)\Big(\frac{1}{(v_t^2)^2} - \frac{1}{(\sigma_t^2)^2}\Big)\Big].
\end{aligned}$$

Using (G2), we can show that $|\sigma_t^2 - v_t^2| = b^{t-1}|\sigma_1^2 - \frac{w}{1-b}|$, from which it follows that

$$(A) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underbrace{\Big(1-\frac{\varepsilon_t^2}{\sigma_t^2}\Big)}_{=(1-\eta_t^2)}\frac{1}{\sigma_t^2}\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big) + o_p(1). \tag{42}$$

Using results from empirical process theory and exploiting that $(1-\eta_t^2)$ is a martingale difference we will show in Lemma C.4 that $(A) = o_p(1)$.

Next we consider the terms (C) and (D). We restrict attention to (D), as this is the more complicated term. (C) can be treated analogously. Successively replacing the approximate expressions $\tilde{\varepsilon}_t^2$ and $\tilde{v}_t^2$ in (D) by the exact terms and using (G1) and (G3) to eliminate the resulting error yields

$$(D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t^2 \Big(\frac{v_t^2 - \tilde{v}_t^2}{v_t^2 v_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{v_t^2} + o_p(1).$$

By analogous arguments as for (A) and (B), we can further replace some of the occurrences of v_t^2 by σ_t^2 to get

$$(D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\varepsilon_t^2}{\sigma_t^2} \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + o_p(1).$$

Writing this as

$$(D) = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(1 - \frac{\varepsilon_t^2}{\sigma_t^2}\Big) \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + o_p(1),$$

one can follow analogous arguments to those used for (A) based on empirical process theory and the martingale difference structure of $(1 - \frac{\varepsilon_t^2}{\sigma_t^2}) = (1 - \eta_t^2)$ to get

$$(D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + o_p(1).$$

Defining $G_t^{[i]} := \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2}$, using (G1) – (G3) and writing $m(x) = m_c + m_0(x_0) + \ldots +$

$m_d(x_d)$ for short, we can infer that

$$
\begin{aligned}
(D) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} (\varepsilon_{t-k}^2 - \tilde{\varepsilon}_{t-k}^2) + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[\frac{\tau^2(\frac{t-k}{T}, X_{t-k}) - \tilde{\tau}^2(\frac{t-k}{T}, X_{t-k})}{\tau^2(\frac{t-k}{T}, X_{t-k})} + O_p(h^2) \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[\frac{\exp(\xi_{t-k})[m(\frac{t-k}{T}, X_{t-k}) - \tilde{m}(\frac{t-k}{T}, X_{t-k})]}{\exp(m(\frac{t-k}{T}, X_{t-k}))} \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[m\Big(\frac{t-k}{T}, X_{t-k}\Big) - \tilde{m}\Big(\frac{t-k}{T}, X_{t-k}\Big) \Big] + o_p(1),
\end{aligned}
$$

where the third equality is by a first order Taylor expansion with an intermediate point ξ_{t-k} between $m(\frac{t-k}{T}, X_{t-k})$ and $\tilde{m}(\frac{t-k}{T}, X_{t-k})$. We are now in a position to use the stochastic expansion of our estimators in the additive model, which were given in Appendix A.1. To do so, split up the difference $m(\frac{t-k}{T}, X_{t-k}) - \tilde{m}(\frac{t-k}{T}, X_{t-k})$ into its additive components and decompose the various components into their bias and stochastic parts. This yields $(D) = (D_c) - \sum_{j=0}^{d} (D_{V,j}) + \sum_{j=0}^{d} (D_{B,j}) + o_p(1)$ with

$$
\begin{aligned}
(D_c) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[(m_c - \tilde{m}_c) + \sum_{j=0}^{d} \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) \Big] \\
(D_{V,j}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \tilde{m}_j^A(X_{t-k}^j) \\
(D_{B,j}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t^{[i]} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[m_j(X_{t-k}^j) - \tilde{m}_j^B(X_{t-k}^j) - \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) \Big]
\end{aligned}
$$

for $j = 0, \ldots, d$, where for ease of notation we have used the shorthand $X_{t-k}^0 = \frac{t-k}{T}$. As in Appendix A, $\tilde{m}_j^A$ denotes the stochastic part of the backfitting estimate $\tilde{m}_j$ and $\tilde{m}_j^B$ denotes the bias part.

In Lemmas C.1 – C.3, we will show that

$$(D_c) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{c,D} u_t + o_p(1) \tag{43}$$

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{j,D}\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1) \tag{44}$$

$$(D_{B,j}) = o_p(1) \tag{45}$$

for all $j = 0, \ldots, d$ with $u_t = \log(\varepsilon_t^2)$. Here, $g_{c,D}$ is a constant which is specified in Lemma C.2 and $g_{j,D}$ for $j = 0, \ldots, d$ are functions whose exact forms are given in Lemma C.1. Using (A11), these functions are easily seen to be absolutely bounded by a constant independent of T. To summarize, we obtain that

$$(D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big[g_{c,D} + \sum_{j=0}^{d} g_{j,D}\Big(\frac{t}{T}, X_t\Big) \Big] u_t + o_p(1).$$

Repeating the arguments from above, we can derive an analogous expression for (C). We thus get that

$$(C) + (D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1)$$

with a function $g(\frac{t}{T}, X_t) = g_c + \sum_{j=0}^{d} g_j(\frac{t}{T}, X_t)$ whose additive components are absolutely bounded. Recalling that $(A) = o_p(1)$ and $(B) = o_p(1)$, we finally obtain that

$$\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} - \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1) \tag{46}$$

with an absolutely bounded function g.

We next consider the term $\frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i}$ more closely. W.l.o.g. we can take $\phi_i = a$. (The case $\phi_i = b$ runs analogously and the case $\phi_i = w$ is much easier to handle.) By similar arguments to before,

$$\frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(1 - \frac{\varepsilon_t^2}{v_t^2}\Big) \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{v_t^2} = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{1 - \eta_t^2}{\sigma_t^2}\Big) \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 + o_p(1).$$

Furthermore,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\sum_{k=1}^{t-1}b^{k-1}\varepsilon_{t-k}^2 &= \sum_{k=1}^{T-1}b^{k-1}\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\varepsilon_{t-k}^2 \\
&= \sum_{k=1}^{C_2\log T}b^{k-1}\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\varepsilon_{t-k}^2 + o_p(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\sum_{k=1}^{\min_{t,T}}b^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big) + o_p(1),
\end{aligned}$$

where $C_2 > 0$ is a sufficiently large constant and $\min_{t,T} := \min\{t-1, C_2\log T\}$. For the second equality, we have used the fact that the weights b^k converge exponentially fast to zero as $k \to \infty$. This implies that only the sums up to $C_2 \log T$ with some constant C_2 are asymptotically relevant. Summing up, we have that

$$\frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial \phi_i} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\sum_{k=1}^{\min_{t,T}}b^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big) + o_p(1). \tag{47}$$

Combining (46) and (47) yields

$$\begin{aligned}
\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} &= \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial \phi_i} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big)u_t + o_p(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big\{g\Big(\frac{t}{T}, X_t\Big)u_t - \Big(\sum_{k=1}^{\min_{t,T}}b^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\Big\} + o_p(1) \\
&=: \frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{t,T} + o_p(1),
\end{aligned}$$

i.e. the term of interest can be written as a normalized sum of random variables $U_{t,T}$ plus a term which is asymptotically negligible.

We now apply a central limit theorem for mixing arrays to the term $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{t,T}$. In particular, we employ the theorem of Francq & Zakoïan (2005), which allows the mixing coefficients of the array $\{U_{t,T}\}$ to depend on the sample size T. Verifying the conditions of this theorem, we can conclude that $\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} \to N(0, \sigma^2)$ with

$$
\begin{aligned}
\sigma^2 = \mathbb{E}\Big[\lambda_2(X_0)u_0\Big] &- 2\mathbb{E}\Big[\lambda_1(X_0)u_0\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{-k}^2\Big)\Big(\frac{1-\eta_0^2}{\sigma_0^2}\Big)\Big] \\
&+ \mathbb{E}\Big[\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{-k}^2\Big)^2\Big(\frac{1-\eta_0^2}{\sigma_0^2}\Big)^2\Big] + 2\mathbb{E}\big[\lambda_{1,1}(X_0, X_l)u_0u_l\big] \\
&- 2\mathbb{E}\Big[\lambda_1(X_0)u_0\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{l-k}^2\Big)\Big(\frac{1-\eta_l^2}{\sigma_l^2}\Big)\Big] \\
&- 2\mathbb{E}\Big[\lambda_1(X_l)u_l\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{-k}^2\Big)\Big(\frac{1-\eta_0^2}{\sigma_0^2}\Big)\Big],
\end{aligned}
$$

where we use the shorthand $\lambda_1(x) = \int_0^1 g(w,x)dw$, $\lambda_2(x) = \int_0^1 g^2(w,x)dw$, and $\lambda_{1,1}(x,x') = \int_0^1 g(w,x)g(w,x')dw$. Using the Cramer-Wold device, it is now easy to show that $\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} \to N(0, Q)$. The entries of the matrix Q can be calculated similarly to the expression σ^2. We omit the details as the formulas are rather lengthy and complicated. $\square$

Proof of (40). By straightforward but tedious calculations it can be seen that for all $i,j = 1,\ldots,3$, $\sup_{\phi\in\Phi}\left|\frac{1}{T}\frac{\partial^2 \tilde{l}_T(\phi)}{\partial\phi_i\partial\phi_j} - \frac{1}{T}\frac{\partial^2 l_T(\phi)}{\partial\phi_i\partial\phi_j}\right| = o_p(1)$. From standard estimation theory for GARCH models, we further know that for all $i,j = 1,\ldots,3$ with $\bar{\phi}_i = (\bar{\phi}_{i,1},\ldots,\bar{\phi}_{i,3})'$ between $\tilde{\phi}$ and ϕ_0, it holds that $\frac{1}{T}\frac{\partial^2 l_T(\bar{\phi}_{i,j})}{\partial\phi_j\partial\phi_j} \xrightarrow{P} J(i,j)$ with $J(i,j)$ the (i,j)-th element of some invertible deterministic matrix J. This yields (40). $\square$

C Appendix - Auxiliary results

In order to complete the proof of Theorem 4.3, we still need to show that equations (43) – (45) are fulfilled for the terms (D_c), $(D_{V,j})$ and $(D_{B,j})$ and that (A) given in (42) is asymptotically negligible. In what follows, we establish these results in a series of lemmas.

Lemma C.1. *It holds that*

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_{j,D}\Big(\frac{s}{T}, X_s\Big) u_s + o_p(1)$$

with

$$g_{j,D}\Big(\frac{s}{T}, X_s\Big) = g_{j,D}^{NW}(X_s^j) + g_{j,D}^{SBF}\Big(\frac{s}{T}, X_s\Big)$$

for $j = 0, \ldots, d$. The functions $g_{j,D}^{NW}$ and $g_{j,D}^{SBF}$ are absolutely bounded. Their exact form is given in the proof (see (52) and (55) – (57)).

Proof. We start by giving a detailed exposition of the proof for $j \neq 0$. By Theorem A.1, the stochastic part $\tilde{m}_j^A$ of the smooth backfitting estimate $\tilde{m}_j$ has the expansion

$$\tilde{m}_j^A(x_j) = \hat{m}_j^A(x_j) + \frac{1}{T} \sum_{s=1}^{T} r_{j,s}(x_j) u_s + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

uniformly in x_j, where $\hat{m}_j^A$ is the stochastic part of the Nadaraya-Watson pilot estimate and the function $r_{j,s}(\cdot) = r_j(\frac{s}{T}, X_s, \cdot)$ is Lipschitz continuous and absolutely bounded.

With this result, we can decompose $(D_{V,j})$ as follows:

$$\begin{aligned}
(D_{V,j}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \tilde{m}_j^A(X_{t-k}^j) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \hat{m}_j^A(X_{t-k}^j) \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \Big[\frac{1}{T} \sum_{s=1}^{T} r_{j,s}(X_{t-k}^j) u_s\Big] + o_p(1) \\
&=: (D_{V,j}^{NW}) + (D_{V,j}^{SBF}) + o_p(1).
\end{aligned}$$

In the following, we will give the exact arguments needed to treat $(D_{V,j}^{NW})$. The line of argument for $(D_{V,j}^{SBF})$ is essentially identical although some of the steps are easier due to the properties of the $r_{j,s}$ functions.

W.l.o.g. set $\phi_i = a$ and let $m_{i,k} = \max\{k+1, i+1\}$. Using $\partial v_t^2/\partial a = \sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2$ and $\hat{m}_j^A(x_j) = \frac{1}{T}\sum_{s=1}^T K_h(x_j, X_s^j)u_s / \frac{1}{T}\sum_{v=1}^T K_h(x_j, X_v^j)$, we get

$$(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \frac{K_h(X_{t-k}^j, X_s^j)}{\frac{1}{T}\sum_{v=1}^T K_h(X_{t-k}^j, X_v^j)} \frac{\varepsilon_{t-k}^2 \varepsilon_{t-i}^2}{\sigma_t^2 \sigma_t^2} u_s \Big]. \tag{48}$$

In a first step, we replace the sum $\frac{1}{T}\sum_{v=1}^T K_h(X_{t-k}^j, X_v^j)$ in (48) by a term which only depends on X_{t-k}^j and show that the resulting error is asymptotically negligible. Let $q_j(x_j) = \int_0^1 K_h(x_j, w)dw\, p_j(x_j)$. Furthermore define

$$\begin{aligned} B_j(x_j) &= \frac{1}{T}\sum_{v=1}^T \mathbb{E}[K_h(x_j, X_v^j)] - q_j(x_j) \\ V_j(x_j) &= \frac{1}{T}\sum_{v=1}^T \big(K_h(x_j, X_v^j) - \mathbb{E}[K_h(x_j, X_v^j)] \big). \end{aligned}$$

Notice that $\sup_{x_j\in[0,1]} |B_j(x_j)| = O_p(h)$ and $\sup_{x_j\in[0,1]} |V_j(x_j)| = O_p(\sqrt{\log T/Th})$. From the identity $\frac{1}{T}\sum_{v=1}^T K_h(x_j, X_v^j) = q_j(x_j) + B_j(x_j) + V_j(x_j)$ and a second order Taylor expansion of $(1+x)^{-1}$ we arrive at

$$\begin{aligned} \frac{1}{\frac{1}{T}\sum_{v=1}^T K_h(x_j, X_v^j)} &= \frac{1}{q_j(x_j)} \Big(1 + \frac{B_j(x_j) + V_j(x_j)}{q_j(x_j)} \Big)^{-1} \\ &= \frac{1}{q_j(x_j)} \Big(1 - \frac{B_j(x_j) + V_j(x_j)}{q_j(x_j)} + O_p(h^2) \Big) \end{aligned} \tag{49}$$

uniformly in x_j. Plugging this decomposition into (48), we obtain

$$\begin{aligned} (D_{V,j}^{NW}) = &\sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \frac{K_h(X_{t-k}^j, X_s^j)}{q_j(X_{t-k}^j)} \frac{1}{\sigma_t^2\sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 u_s \Big] \\ &- (D_{V,j}^{NW,B}) - (D_{V,j}^{NW,V}) + o_p(1) \end{aligned}$$

with

$$(D_{V,j}^{NW,B}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j) \frac{B_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 u_s \Big]$$

$$(D_{V,j}^{NW,V}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j) \frac{V_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 u_s \Big].$$

As $\sup_{x_j \in I_h} |B_j(x_j)| = O_p(h^2)$ and $\sup_{x_j \in I_h^c} |B_j(x_j)| = O_p(h)$, we can proceed similarly to the proof of Lemma C.3 later on to show that $(D_{V,j}^{NW,B}) = o_p(1)$. Next we will show that $(D_{V,j}^{NW,V}) = o_p(1)$. Let $\mathbb{E}_v[\cdot]$ denote the expectation with respect to the variables indexed by v, then

$$\begin{aligned}
\big|(D_{V,j}^{NW,V})\big| &= \Big| \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \frac{K_h(X_{t-k}^j, X_s^j)}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 \\
&\qquad\qquad \times \Big(\frac{1}{T} \sum_{v=1}^{T} (K_h(X_{t-k}^j, X_v^j) - \mathbb{E}_v[K_h(X_{t-k}^j, X_v^j)]) \Big) u_s \Big] \Big| \\
&\le \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big(\frac{1}{\sqrt{T}} \sum_{t=m_{i,k}}^{T} \Big| \frac{1}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 \Big| \\
&\qquad\qquad \times \sup_{x_j \in [0,1]} \Big| \frac{1}{T} \sum_{v=1}^{T} (K_h(x_j, X_v^j) - \mathbb{E}_v[K_h(x_j, X_v^j)]) \Big| \\
&\qquad\qquad \times \sup_{x_j \in [0,1]} \Big| \frac{1}{T} \sum_{s=1}^{T} K_h(x_j, X_s^j) u_s \Big| \Big) \\
&= O_p\Big(\frac{\log T}{Th} \Big) \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \underbrace{\Big(\frac{1}{\sqrt{T}} \sum_{t=m_{i,k}}^{T} \Big| \frac{1}{q_j^2(X_{t-k}^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 \Big| \Big)}_{= O_p(\sqrt{T}) \text{ by Markov's inequality}} \\
&= O_p\Big(\frac{\log T}{Th} \sqrt{T} \Big) = o_p(1).
\end{aligned}$$

Together with the fact that $(D_{V,j}^{NW,B}) = o_p(1)$, this yields

$$(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j) \mu_t^{i,k} u_s \Big] + o_p(1), \tag{50}$$

where we use the shorthand $\mu_t^{i,k} = (q_j(X_{t-k}^j)\sigma_t^2\sigma_t^2)^{-1}\varepsilon_{t-k}^2\varepsilon_{t-i}^2$.

In the next step, we replace the inner sum over t in (50) by a term that only depends on X_s^j and show that the resulting error can be asymptotically neglected. Define

$$\xi(X_{t-k}^j, X_s^j) := \xi_t^{i,k}(X_{t-k}^j, X_s^j) := K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k} - \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}],$$

where $\mathbb{E}_{-s}[\cdot]$ is the expectation with respect to all variables except for those depending on the index s. With the above notation at hand, we can write

$$\begin{aligned}(D_{V,j}^{NW}) = &\sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}]u_s\Big] \\ &+ (R_{V,j}^{NW}) + o_p(1),\end{aligned}$$

where

$$\begin{aligned}(R_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j)u_s\Big] \\ &= \sum_{k=1}^{C_2\log T} ab^{k-1} \sum_{i=1}^{C_2\log T} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j)u_s\Big] + o_p(1)\end{aligned} \tag{51}$$

for some sufficiently large constant $C_2 > 0$. Once we show that $(R_{V,j}^{NW}) = o_p(1)$, we are left with

$$\begin{aligned}(D_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}]u_s\Big] + o_p(1) \\ &= \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Big(\sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\frac{T-m_{i,k}}{T}\mathbb{E}_{-s}[K_h(X_{-k}^j, X_s^j)\mu_0^{i,k}]\Big)u_s + o_p(1).\end{aligned}$$

As the terms with $i, k \geq C_2 \log T$ are asymptotically negligible, we can expand the i

and k sums to infinity, which yields

$$(D_{V,j}^{NW}) = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}_{-s}[K_h(X_{-k}^j, X_s^j)\mu_0^{i,k}] \Big) u_s + o_p(1) \tag{52}$$
$$=: \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_{j,D}^{NW}(X_s^j) u_s + o_p(1)$$

with

$$\mu_0^{i,k} = \frac{1}{q_j(X_{-k}^j)} \frac{1}{\sigma_0^2 \sigma_0^2} \varepsilon_{-k}^2 \varepsilon_{-i}^2$$
$$q_j(X_{-k}^j) = \int_0^1 K_h(X_{-k}^j, w) dw \; p_j(X_{-k}^j).$$

Thus it remains to show that $(R_{V,j}^{NW}) = o_p(1)$, which requires a lot of care. We will prove that the term in square brackets in (51) is $o_p(1)$ uniformly over $i, k \leq C_2 \log T$, which yields the desired result. It is easily seen that

$$P := \mathbb{P}\Big(\max_{i,k \leq C_2 \log T} \Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big| > \delta \Big)$$
$$\leq \sum_{k=1}^{C_2 \log T} \sum_{i=1}^{C_2 \log T} \underbrace{\mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big| > \delta \Big)}_{=: P_{i,k}}$$

for a fixed $\delta > 0$. Then by Chebychev's inequality

$$P_{i,k} \leq \frac{1}{T^3 \delta^2} \sum_{s,s'=1}^{T} \sum_{t,t'=m_{i,k}}^{T} \mathbb{E}\Big[\xi(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'} \Big]$$
$$= \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \notin \Gamma_{i,k}} \mathbb{E}\Big[\xi(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'} \Big]$$
$$+ \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\Big[\xi(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'} \Big] =: P_{i,k}^1 + P_{i,k}^2,$$

where $\Gamma_{i,k}$ is the set of tuples (s, s', t, t') with $1 \leq s, s' \leq T$ and $m_{i,k} \leq t, t' \leq T$ such that one index is separated from the others. We say that an index, for instance t, is separated from the others if $\min\{|t-t'|, |t-s|, |t-s'|\} > C_3 \log T$, i.e. if it is further away from the other indices than $C_3 \log T$ for a constant C_3 to be chosen later on. We now analyse $P^1_{i,k}$ and $P^2_{i,k}$ separately.

(a) First consider $P^1_{i,k}$. If a tuple (s, s', t, t') is not an element of $\Gamma_{i,k}$, then no index can be separated from the others. Since the index t cannot be separated, there exists an index, say t', such that $|t-t'| \leq C_3 \log T$. Now take an index different from t and t', for instance s. Then by the same argument, there exists an index, say s', such that $|s-s'| \leq C_3 \log T$. As a consequence, the number of tuples $(s, s', t, t') \notin \Gamma_{i,k}$ is smaller than $CT^2(\log T)^2$ for some constant C. Using (A11), this suffices to infer that

$$\left|P^1_{i,k}\right| \leq \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\notin\Gamma_{i,k}} \frac{C}{h^2} \leq \frac{C}{\delta^2}\frac{(\log T)^2}{Th^2}.$$

Hence, $|P^1_{i,k}| \leq C\delta^{-2}(\log T)^{-3}$ uniformly in i and k.

(b) The term $P^2_{i,k}$ is more difficult to handle. We start by taking a cover $\{I_m\}_{m=1}^{M_T}$ of the compact support $[0,1]$ of X^j_{t-k}. The elements I_m are intervals of length $1/M_T$ given by $I_m = [\frac{m-1}{M_T}, \frac{m}{M_T})$ for $m = 1, \ldots, M_T - 1$ and $I_{M_T} = [1 - \frac{1}{M_T}, 1]$. The midpoint of the interval I_m is denoted by x_m. With this, we can write

$$K_h(X^j_{t-k}, X^j_s) = \sum_{m=1}^{M_T} I(X^j_{t-k} \in I_m) \times \left[K_h(x_m, X^j_s) + (K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s))\right]. \tag{53}$$

Using (53), we can further write

$$
\begin{aligned}
\xi(X_{t-k}^j, X_s^j) = \sum_{m=1}^{M_T} \Big\{ & I(X_{t-k}^j \in I_m) K_h(x_m, X_s^j) \mu_t^{i,k} \\
& - \mathbb{E}_{-s}[I(X_{t-k}^j \in I_m) K_h(x_m, X_s^j) \mu_t^{i,k}] \Big\} \\
+ \sum_{m=1}^{M_T} \Big\{ & I(X_{t-k}^j \in I_m)(K_h(X_{t-k}^j, X_s^j) - K_h(x_m, X_s^j)) \mu_t^{i,k} \\
& - \mathbb{E}_{-s}[I(X_{t-k}^j \in I_m)(K_h(X_{t-k}^j, X_s^j) - K_h(x_m, X_s^j)) \mu_t^{i,k}] \Big\} \\
=: \xi_1(X_{t-k}^j, X_s^j) + \xi_2(X_{t-k}^j, X_s^j) &
\end{aligned}
$$

and

$$
\begin{aligned}
P_{i,k}^2 = & \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\big[\xi_1(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'}\big] \\
& + \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\big[\xi_2(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'}\big] =: P_{i,k}^{2,1} + P_{i,k}^{2,2}.
\end{aligned}
$$

We first consider $P_{i,k}^{2,2}$. Set $M_T = CT(\log T)^3 h^{-3}$ and exploit the Lipschitz continuity of the kernel K to get that $|K_h(X_{t-k}^j, X_s^j) - K_h(x_m, X_s^j)| \le \frac{C}{h^2} |X_{t-k}^j - x_m|$. This gives us

$$
\begin{aligned}
\big|\xi_2(X_{t-k}^j, X_s^j)\big| \le \frac{C}{h^2} \sum_{m=1}^{M_T} \Big(& \underbrace{I(X_{t-k}^j \in I_m) |X_{t-k}^j - x_m|}_{\le I(X_{t-k}^j \in I_m) M_T^{-1}} \mu_t^{i,k} \\
& + \mathbb{E}\big[\underbrace{I(X_{t-k}^j \in I_m) |X_{t-k}^j - x_m|}_{\le I(X_{t-k}^j \in I_m) M_T^{-1}} \mu_t^{i,k} \big] \Big) \le \frac{C}{M_T h^2} \big(\mu_t^{i,k} + \mathbb{E}[\mu_t^{i,k}] \big).
\end{aligned}
\tag{54}
$$

Plugging (54) into the expression for $P_{i,k}^{2,2}$, we arrive at

$$\begin{aligned}
\left|P_{i,k}^{2,2}\right| &\leq \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \mathbb{E}\Big[\big|\xi_2(X_{t-k}^j, X_s^j)\big|\big|u_s\xi(X_{t'-k}^j, X_{s'}^j)u_{s'}\big|\Big] \\
&\leq \frac{1}{T^3\delta^2}\frac{C}{M_T h^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \underbrace{\mathbb{E}\big[(\mu_t^{i,k} + \mathbb{E}[\mu_t^{i,k}])\big|u_s\xi(X_{t'-k}^j, X_{s'}^j)u_{s'}\big|\big]}_{\leq Ch^{-1}} \leq \frac{C}{\delta^2}\frac{1}{(\log T)^3}.
\end{aligned}$$

We next turn to $P_{i,k}^{2,1}$. Write

$$P_{i,k}^{2,1} = \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \Big(\sum_{m=1}^{M_T} S_m\Big)$$

with

$$\begin{aligned}
S_m = \mathbb{E}\Big[&\big\{I(X_{t-k}^j \in I_m)K_h(x_m, X_s^j)\mu_t^{i,k} - \mathbb{E}_{-s}[I(X_{t-k}^j \in I_m)K_h(x_m, X_s^j)\mu_t^{i,k}]\big\} \\
&\times u_s\xi(X_{t'-k}^j, X_{s'}^j)u_{s'}\Big]
\end{aligned}$$

and assume that an index, w.l.o.g. t, can be separated from the others. Choosing $C_3 \gg C_2$, we get

$$\begin{aligned}
S_m &= \mathrm{Cov}\Big(I(X_{t-k}^j \in I_m)\mu_t^{i,k} - \mathbb{E}[I(X_{t-k}^j \in I_m)\mu_t^{i,k}], K_h(x_m, X_s^j)u_s\xi(X_{t'-k}^j, X_{s'}^j)u_{s'}\Big) \\
&\leq \frac{C}{h^2}(\alpha([C_3 - C_2]\log T))^{1-\frac{2}{p}} \leq \frac{C}{h^2}(a^{(C_3-C_2)\log T})^{1-\frac{2}{p}} \leq \frac{C}{h^2}T^{-C_4}
\end{aligned}$$

with some $C_4 > 0$ by Davydov's inequality, where p is chosen slightly larger than 2. Note that the above bound is independent of i and k and that we can make C_4 arbitrarily large by choosing C_3 large enough. This shows that $|P_{i,k}^{2,1}| \leq C\delta^{-2}(\log T)^{-3}$ uniformly in i and k with some constant C.

Combining (a) and (b) yields that $P \to 0$ for each fixed $\delta > 0$. This implies that

$$(R_{V,j}^{NW,V}) = o_p(1),$$

which completes the proof for the term $(D^{NW}_{V,j})$.

As stated at the beginning of the proof, the term $(D^{SBF}_{V,j})$ can be treated in exactly the same way. Following analogous arguments as above and writing $\zeta_t^{i,k} = (\sigma_t^2\sigma_t^2)^{-1}\varepsilon^2_{t-k}\varepsilon^2_{t-i}$, one obtains

$$
\begin{aligned}
(D^{SBF}_{V,j}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\mathbb{E}_{-s}[r_{j,s}(X^j_{t-k})\zeta_t^{i,k}]\ u_s\Big] + o_p(1) \qquad (55)\\
&= \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}_{-s}[r_{j,s}(X^j_{-k})\zeta_0^{i,k}]\Big)u_s + o_p(1)\\
&=: \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g^{SBF}_{j,D}\Big(\frac{s}{T}, X_s\Big)u_s + o_p(1).
\end{aligned}
$$

Finally, the proofs for $j = 0$ are very similar but somewhat simpler and are thus omitted here. For completeness we provide the functions $g^{NW}_{0,D}$ and $g^{SBF}_{0,D}$:

$$
g^{NW}_{0,D}\Big(\frac{s}{T}\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon^2_{-k}\varepsilon^2_{-i}\big]\Big)\int_0^1 \frac{K_h(\frac{s}{T}, v)}{\int_0^1 K_h(v,w)dw}dv \qquad (56)
$$

$$
g^{SBF}_{0,D}\Big(\frac{s}{T}, X_s\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon^2_{-k}\varepsilon^2_{-i}\big]\Big)\int_0^1 r_{0,s}(w)dw. \qquad (57)
$$

$\square$

Lemma C.2. *It holds that*

$$
(D_c) = \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g_{c,D}u_s
$$

with

$$
g_{c,D} = \sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\Big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon^2_{-i}\varepsilon^2_{-k}\Big].
$$

Proof. Using the fact that

$$
\tilde{m}_c = \frac{1}{T}\sum_{s=1}^{T} Z_{s,T} = m_c + \frac{1}{T}\sum_{s=1}^{T} m_0\Big(\frac{s}{T}\Big) + \sum_{j=1}^{d}\frac{1}{T}\sum_{s=1}^{T} m_j(X^j_s) + \frac{1}{T}\sum_{s=1}^{T} u_s,
$$

we arrive at

$$(D_c) = -\Big(\frac{1}{T}\sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T} u_s\Big)$$

with $G_t = \frac{\partial v_t^2}{\partial \phi_i}(\sigma_t^2\sigma_t^2)^{-1}$. Now let $m_{i,k} = \max\{k+1, i+1\}$ and assume w.l.o.g. that $\phi_i = a$. Then

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2 &= \frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2\Big)\frac{1}{\sigma_t^2\sigma_t^2}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2 \\
&= \sum_{k=1}^{C_2\log T} ab^{k-1}\sum_{i=1}^{C_2\log T} b^{i-1}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 + o_p(1)
\end{aligned}$$

with some sufficiently large constant C_2. Using Chebychev's inequality and exploiting the mixing properties of the variables involved, one can show that

$$\max_{i,k\le C_2\log T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\Big(\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 - \mathbb{E}\Big[\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2\Big]\Big) = o_p(1).$$

This allows us to infer that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2 &= \sum_{k=1}^{C_2\log T} ab^{k-1}\sum_{i=1}^{C_2\log T} b^{i-1}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\mathbb{E}\Big[\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2\Big] + o_p(1) \\
&= \sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\Big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon_{-i}^2\varepsilon_{-k}^2\Big] + o_p(1),
\end{aligned}$$

which completes the proof. □

Lemma C.3. *It holds that*

$$(D_{B,j}) = o_p(1)$$

for $j = 0, \ldots, d$.

Proof. We start by considering the case $j = 0$: Define

$$
\begin{aligned}
J_h &= \{t \in \{1, \dots, T\} : C_1 h \le \frac{t}{T} \le 1 - C_1 h\} \\
J_{h,c}^u &= \{t \in \{1, \dots, T\} : 1 - C_1 h < \frac{t}{T}\} \\
J_{h,c}^l &= \{t \in \{1, \dots, T\} : \frac{t}{T} < C_1 h\},
\end{aligned}
$$

where $[-C_1, C_1]$ is the support of K. Using the uniform convergence rates from Theorem A.2 and assuming w.l.o.g. that $\phi_i = a$, we get

$$
\begin{aligned}
|(D_{B,0})| = \Big| & \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial v_t^2}{\partial a} \frac{1}{\sigma_t^2 \sigma_t^2} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \Big[m_0\Big(\frac{t-k}{T}\Big) - \tilde{m}_0^B\Big(\frac{t-k}{T}\Big) - \frac{1}{T} \sum_{s=1}^{T} m_0\Big(\frac{s}{T}\Big) \Big] \Big| \\
&\le O_p(h) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(t-k \in J_{h,c}^l) \\
&\quad + O_p(h) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(t-k \in J_{h,c}^u) \\
&\quad + O_p(h^2) \frac{C}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(t-k \in J_h) \\
&=: (D_{B,0}^{J_{h,c}^l}) + (D_{B,0}^{J_{h,c}^u}) + (D_{B,0}^{J_h}).
\end{aligned}
$$

By Markov's inequality, $(D_{B,0}^{J_h}) = O_p(h^2 \sqrt{T}) = o_p(1)$. Recognizing that

(i) $I(t-k \in J_{h,c}^u) \le I(t \in J_{h,c}^u)$ for all $k \in \{0, \dots, t-1\}$

(ii) $\sum_{t=1}^{T} I(t \in J_{h,c}^u) \le C_1 T h$,

we get $(D_{B,0}^{J_{h,c}^u}) = O_p(h^2 \sqrt{T}) = o_p(1)$ by another appeal to Markov's inequality. This just leaves $(D_{B,0}^{J_{h,c}^l})$, which is a bit more tedious. By a change of variable $j = t - k$,

$$
\begin{aligned}
(D_{B,0}^{J_{h,c}^l}) &\le O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 \sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon_j^2 I(j \in J_{h,c}^l) \\
&= O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 I\Big(\Big[\frac{t}{2}\Big] \in J_{h,c}^l\Big)\sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon_j^2 I(j \in J_{h,c}^l) \\
&\quad + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 I\Big(\Big[\frac{t}{2}\Big] \notin J_{h,c}^l\Big)\sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon_j^2 I(j \in J_{h,c}^l) \\
&=: (A) + (B),
\end{aligned}
$$

where $[x]$ denotes the smallest integer larger than x. Realizing that $[t/2] \in J_{h,c}^l$ only if $t < 2C_1hT$, we get $(A) = O_p(h^2\sqrt{T}) = o_p(1)$ once again by Markov's inequality. In (B) we can truncate the summation over j at $[t/2]-1$, as $I(j \in J_{h,c}^l) = 0$ for $j \ge [t/2]$ if $[t/2] \notin J_{h,c}^l$. We thus obtain

$$
\begin{aligned}
(B) &\le O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 \sum_{j=1}^{[t/2]-1} ab^{t-j-1}\varepsilon_j^2 \\
&= O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{[t/2]}\sum_{i=1}^{t-1} b^{i-1}\sum_{j=1}^{[t/2]-1} ab^{t-j-1-[t/2]}\varepsilon_{t-i}^2\varepsilon_j^2.
\end{aligned}
$$

By a final appeal to Markov's inequality we arrive at

$$
(B) = O_p(h)O_p\Big(\frac{1}{\sqrt{T}}\Big) = o_p(1),
$$

thus completing the proof for $j = 0$.

Next consider the case $j \neq 0$. Similarly to before, we have

$$\begin{aligned}
|(D_{B,j})| &\leq O_p(h^2)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(X_{t-k}^j \in I_h) \\
&\quad + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(X_{t-k}^j \notin I_h) \\
&= O_p(h^2\sqrt{T}) + O_p\Big(\frac{h}{\sqrt{T}}\Big)\underbrace{\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(X_{t-k}^j \notin I_h)}_{=:R_T}
\end{aligned}$$

with $I_h = [2C_1 h, 1 - 2C_1 h]$ as defined in Theorem 4.1. Using (A11), it is easy to see that $R_T = O_p(h)$, which yields the result for $j \neq 0$. $\square$

Lemma C.4. *It holds that*

$$(A) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underbrace{\Big(1 - \frac{\varepsilon_t^2}{\sigma_t^2}\Big)}_{=(1-\eta_t^2)}\frac{1}{\sigma_t^2}\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big) + o_p(1) = o_p(1).$$

Proof. W.l.o.g. let $\phi_i = a$. With the help of (G1) and a simple Taylor expansion, we get that

$$
\begin{aligned}
\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i} &= \sum_{k=1}^{t-1} b^{k-1} \left(\tilde{\varepsilon}_{t-k}^2 - \varepsilon_{t-k}^2 \right) \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \left[\frac{\tau^2 \left(\frac{t-k}{T}, X_{t-k}\right) - \tilde{\tau}^2 \left(\frac{t-k}{T}, X_{t-k}\right)}{\tau^2 \left(\frac{t-k}{T}, X_{t-k}\right)} + R_\varepsilon \Big(\frac{t-k}{T}, X_{t-k} \Big) \right] \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \left[\frac{\exp(\xi_{t-k}) \left(m \left(\frac{t-k}{T}, X_{t-k}\right) - \tilde{m} \left(\frac{t-k}{T}, X_{t-k}\right) \right)}{\exp \left(m \left(\frac{t-k}{T}, X_{t-k}\right) \right)} \right] + O_p(h^2) \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \left[m \Big(\frac{t-k}{T}, X_{t-k} \Big) - \tilde{m} \Big(\frac{t-k}{T}, X_{t-k} \Big) \right] + O_p(h^2) \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \Bigg\{ (m_c - \tilde{m}_c) - \tilde{m}_0^A \Big(\frac{t-k}{T} \Big) - \ldots - \tilde{m}_d^A \left(X_{t-k}^d \right) \\
&\quad + \left(m_0 \Big(\frac{t-k}{T} \Big) - \tilde{m}_0^B \Big(\frac{t-k}{T} \Big) \right) + \ldots + \left(m_d \left(X_{t-k}^d \right) - \tilde{m}_d^B \left(X_{t-k}^d \right) \right) \Bigg\} \\
&\quad + O_p(h^2),
\end{aligned}
$$

where ξ_{t-k} is an intermediate point between $m(\frac{t-k}{T}, X_{t-k})$ and $\tilde{m}(\frac{t-k}{T}, X_{t-k})$. Using this together with arguments similar to those for Lemma C.3 yields that

$$
\begin{aligned}
(A) = - \sum_{k=1}^{T-1} b^{k-1} \Bigg(&\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \left(1 - \eta_t^2\right) \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \\
&\times \left\{ (m_c - \tilde{m}_c) - \tilde{m}_0^A \Big(\frac{t-k}{T} \Big) - \ldots - \tilde{m}_d^A \left(X_{t-k}^d \right) \right\} \Bigg) + o_p(1) \\
=: (A_c) - (A_0) - (A_1) - &\ldots - (A_d) + o_p(1).
\end{aligned}
$$

It is straightforward to see that $(A_c) = o_p(1)$. In what follows, we further prove that $(A_j) = o_p(1)$ for $j = 0, \ldots, d$ as well, which completes the proof.

Consider a fixed $j \in \{0, \ldots, d\}$ and let $\delta > 0$ be an arbitrarily small but fixed constant. Write

$$
(A_j) = \sum_{k=1}^{T-1} b^{k-1} \left(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \left(1 - \eta_t^2\right) \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \, \tilde{m}_j^A(X_{t-k}^j) \right) =: (A_j^{\leq}) + (A_j^{>}),
$$

where

$$(A_j^{\leq}) = \sum_{k=1}^{T-1} b^{k-1} \left(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{\leq} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \tilde{m}_j^A(X_{t-k}^j) \right)$$
$$(A_j^{>}) = \sum_{k=1}^{T-1} b^{k-1} \left(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{>} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \tilde{m}_j^A(X_{t-k}^j) \right)$$

with

$$W_t^{\leq} = \left(1-\eta_t^2\right) I(|\eta_t| \leq T^{1/48+\delta}) - \mathbb{E}[(1-\eta_t^2) I(|\eta_t| \leq T^{1/48+\delta})]$$
$$W_t^{>} = \left(1-\eta_t^2\right) I(|\eta_t| > T^{1/48+\delta}) - \mathbb{E}[(1-\eta_t^2) I(|\eta_t| > T^{1/48+\delta})].$$

We now consider the two terms $(A_j^{\leq})$ and $(A_j^{>})$ separately. We start with $(A_j^{>})$. Standard arguments for kernel estimators show that $\sup_{x_j \in [0,1]} \left|\hat{m}_j^A(x_j)\right| = O_p(\sqrt{\log T/Th})$. This together with Theorem A.1 implies that $\sup_{x_j \in [0,1]} \left|\tilde{m}_j^A(x_j)\right| = O_p(\sqrt{\log T/Th})$ as well. As $\sqrt{\log T/Th} \leq T^{-3/8+\delta}$, we can infer that

$$\begin{aligned} \left|(A_j^{>})\right| &\leq O_p\left(\sqrt{\frac{\log T}{Th}}\right) \cdot \sum_{k=1}^{T-1} b^{k-1} \frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} |W_t^{>}| \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \\ &\leq O_p(1) \underbrace{\sum_{k=1}^{T-1} b^{k-1} \frac{1}{T^{7/8-\delta}} \sum_{t=k+1}^{T} |W_t^{>}| \frac{\varepsilon_{t-k}^2}{\sigma_t^2}}_{:=(*)}. \end{aligned}$$

Moreover, since

$$\mathbb{E}\left[\left|1-\eta_t^2\right| I(|\eta_t| > T^{1/48+\delta})\right] \leq \mathbb{E}\left[\left|1-\eta_t^2\right| \frac{\eta_t^6}{T^{6(1/48+\delta)}} I(|\eta_t| > T^{1/48+\delta})\right] \leq \frac{C}{T^{1/8+6\delta}},$$

we get that $\mathbb{E}|W_t^{>}| \leq C/T^{1/8+6\delta}$. From this and Markov's inequality, it follows that $(*) = o_p(1)$ and thus $(A_j^{>}) = o_p(1)$.

We next turn to the term $(A_j^{\leq})$. Splitting $(A_j^{\leq})$ into two parts with the help of the indicators $I(\varepsilon_{t-k}^2 \leq T^{1/48+\delta})$ and $I(\varepsilon_{t-k}^2 > T^{1/48+\delta})$ and applying a similar truncation argument as above, we can show that

$$\big(A_j^{\leq}\big) = \sum_{k=1}^{T-1} b^{k-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} W_t^{\leq}\,\frac{\varepsilon_{t-k}^2}{\sigma_t^2}\,I(|\varepsilon_{t-k}|\leq T^{1/48+\delta})\,\tilde{m}_j^A\big(X_{t-k}^j\big)\Big) + o_p(1).$$

Since the weights b^{k-1} decay exponentially fast to zero, we further obtain that

$$\big(A_j^{\leq}\big) = \sum_{k=1}^{C_2\log T} b^{k-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} W_t^{\leq}\,\frac{\varepsilon_{t-k}^2}{\sigma_t^2}\,I(|\varepsilon_{t-k}|\leq T^{1/48+\delta})\,\tilde{m}_j^A\big(X_{t-k}^j\big)\Big) + o_p(1)$$

with some sufficiently large constant C_2. By Theorem A.1, it holds that uniformly in x_j,

$$\tilde{m}_j^A(x_j) = \frac{1}{T}\sum_{s=1}^{T}\left(\frac{K_h(x_j,X_s^j)}{\frac{1}{T}\sum_{v=1}^{T}K_h(x_j,X_v^j)} + r_{j,s}(x_j)\right)u_s + o_p\left(\frac{1}{\sqrt{T}}\right).$$

By the same arguments as used in the proof of Lemma C.1, we can replace the term $\frac{1}{T}\sum_{v=1}^{T}K_h(x_j,X_v^j)$ by $q_j(x_j)=\int_0^1 K_h(x_j,w)dw\,p_j(x_j)$, which yields that

$$\big(A_j^{\leq}\big) = \sum_{k=1}^{C_2\log T} b^{k-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} W_t^{\leq}\,\frac{\varepsilon_{t-k}^2}{\sigma_t^2}\,I(|\varepsilon_{t-k}|\leq T^{1/48+\delta})\,\check{m}_j^A\big(X_{t-k}^j\big)\Big) + o_p(1)$$

with

$$\check{m}_j^A(x_j) = \frac{1}{T}\sum_{s=1}^{T}\left(\frac{K_h\left(x_j,X_s^j\right)}{q_j(x_j)} + r_{j,s}(x_j)\right)u_s.$$

We can thus write $(A_j^{\leq}) = \sum_{k=1}^{C_2\log T} b^{k-1}\cdot(A_{j,k}^{\leq}) + o_p(1)$ with

$$(A_{j,k}^{\leq}) = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} W_t^{\leq}\,\frac{\varepsilon_{t-k}^2}{\sigma_t^2}\,I(|\varepsilon_{t-k}|\leq T^{1/48+\delta})\,\check{m}_j^A(X_{t-k}^j).$$

In what follows, we prove that for any fixed $\varepsilon > 0$,

$$\max_{1\leq k\leq C_2\log T}\mathbb{P}\left(\left|(A_{j,k}^{\leq})\right| > \varepsilon\right) \leq T^{-\kappa} \tag{58}$$

with some $\kappa > 0$. This implies that $\mathbb{P}(\max_{1\leq k\leq C_2\log T}|(A_{j,k}^{\leq})| > \varepsilon) \leq \sum_{k=1}^{C_2\log T}\mathbb{P}(|(A_{j,k}^{\leq})| >$

$\varepsilon) = o(1)$, that is, $\max_{1\le k\le C_2\log T} |(A^{\le}_{j,k})| = o_p(1)$. Since $(A^{\le}_j) = \sum_{k=1}^{C_2\log T} b^{k-1}\cdot(A^{\le}_{j,k}) + o_p(1) \le C\max_{1\le k\le C_2\log T}|(A^{\le}_{j,k})| + o_p(1)$, we can conclude that $(A^{\le}_j) = o_p(1)$.

It remains to prove (58). To do so, we embed the stochastic function $\check{m}^A_j$ into a class of Hölder functions: For any $\eta > 0$ and $x_j \neq x'_j$,

$$\begin{aligned}
&\left|\check{m}^A_j(x_j) - \check{m}^A_j(x'_j)\right| \Big/ \left|x_j - x'_j\right|^{1/2+\eta} \\
&\le \left|\frac{1}{T}\sum_{s=1}^{T}\frac{1}{q_j(x_j)}\left(K_h\left(x_j, X^j_s\right) - K_h\left(x'_j, X^j_s\right)\right)u_s\right| \Big/ \left|x_j - x'_j\right|^{1/2+\eta} \\
&\quad + \left|\frac{1}{T}\sum_{s=1}^{T} K_h\left(x'_j, X^j_s\right)\frac{q_j(x'_j) - q_j(x_j)}{q_j(x'_j)q_j(x_j)}u_s\right| \Big/ \left|x_j - x'_j\right|^{1/2+\eta} \\
&\quad + \left|\frac{1}{T}\sum_{s=1}^{T}\left(r_{j,s}(x_j) - r_{j,s}(x'_j)\right)u_s\right| \Big/ \left|x_j - x'_j\right|^{1/2+\eta} \\
&=: \beta_1(x_j, x'_j) + \beta_2(x_j, x'_j) + \beta_3(x_j, x'_j).
\end{aligned}$$

By standard arguments to derive uniform convergence rates for kernel estimators which can be found for example in Bosq (1998), Masry (1996) or Hansen (2008), we can show that

$$\mathbb{P}\left(\sup_{x_j, x'_j\in[0,1], x_j\neq x'_j}\left|\beta_k(x_j, x'_j)\right| > \frac{Ma_T}{6}\right) = O(T^{-\kappa})$$

for all $k = 1, 2, 3$ and some $\kappa > 0$, where $a_T = \sqrt{\log T / Th^{2+\varsigma}}$ for some small $\varsigma > 0$ and M is a sufficiently large constant. From this, it immediately follows that

$$\mathbb{P}\left(\sup_{x_j, x'_j\in[0,1], x_j\neq x'_j}\frac{\left|\check{m}^A_j(x_j) - \check{m}^A_j(x'_j)\right|}{\left|x_j - x'_j\right|^{1/2+\eta}} > \frac{Ma_T}{2}\right) = O(T^{-\kappa}). \tag{59}$$

Similarly, it can be verified that

$$\mathbb{P}\left(\sup_{x_j\in[0,1]}\left|\check{m}^A_j(x_j)\right| > \frac{Ma_T}{2}\right) = O(T^{-\kappa}). \tag{60}$$

From (59) and (60), we can conclude that with probability $1 - O(T^{-\kappa})$, the random function $\frac{1}{Ma_T}\check{m}_j^A$ is contained in the Hölder space $\mathcal{F} := C_1^{1/2+\eta}([0,1])$ which is defined as follows: For any $\alpha \in (0,1]$,

$$C_1^{\alpha}([0,1]) = \{f : [0,1] \to \mathbb{R} : f \text{ is continuous with } \|f\|_{\alpha} \leq 1\}$$

with

$$\|f\|_{\alpha} = \sup_{x\in(0,1)} |f(x)| + \sup_{x,y\in(0,1),x\neq y} \frac{|f(x)-f(y)|}{|x-y|^{\alpha}}.$$

Let $\mathcal{N}(\delta, C_1^{\alpha}([0,1]), \|\cdot\|_\infty)$ be the δ-covering number of $C_1^{\alpha}([0,1])$ endowed with the supremum norm $\|\cdot\|_\infty$. By Theorem 2.7.1 in van der Vaart and Wellner (1996), we have the bound

$$\log \mathcal{N}\left(\delta, C_1^{\alpha}([0,1]), \|\cdot\|_\infty\right) \leq K\delta^{-1/\alpha} \tag{61}$$

for any $\delta > 0$ with some fixed constant $K > 0$. We next define

$$Z_{T,k}(f) := \frac{Ma_T}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{\leq} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} I(|\varepsilon_{t-k}| \leq T^{1/48+\delta})\, f\left(X_{t-k}^j\right)$$

and note that $(A_{j,k}^{\leq}) = Z_{T,k}(\frac{1}{Ma_T}\check{m}_j^A)$. Since $\frac{1}{Ma_T}\check{m}_j^A$ is contained in the Hölder space $\mathcal{F} = C_1^{1/2+\eta}([0,1])$ with probability $1 - O(T^{-\kappa})$, it follows that

$$\mathbb{P}\left(\left|(A_{j,k}^{\leq})\right| > \varepsilon\right) \leq P\left(\sup_{f\in\mathcal{F}} |Z_{T,k}(f)| > \varepsilon\right) + O(T^{-\kappa})$$

and it remains to show that

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}} |Z_{T,k}(f)| > \varepsilon\right) \leq CT^{-\kappa}. \tag{62}$$

To do so, define $Z^{\gamma}_{T,k} := T^{\gamma} Z_{T,k}$ with $\gamma > 0$ small and write

$$\begin{aligned}
&\mathbb{P}\Big(\big|Z^{\gamma}_{T,k}(f) - Z^{\gamma}_{T,k}(g)\big| > \varepsilon\, ||f-g||_{\infty}\Big) \\
&= \mathbb{P}\Bigg(T^{\gamma} \Bigg| \frac{M a_T}{\sqrt{T}} \sum_{t=k+1}^{T} \underbrace{W_t^{\leq} \frac{\varepsilon^2_{t-k}}{\sigma^2_t} I(|\varepsilon_{t-k}| \leq T^{1/48+\delta}) \Big(f\big(X^j_{t-k}\big) - g\big(X^j_{t-k}\big)\Big)}_{=:\psi_{t,j,k}} \Bigg| > \varepsilon\, ||f-g||_{\infty}\Bigg).
\end{aligned}$$

Using the trivial bound $|\psi_{t,j,k}| \leq C T^{1/12+4\delta} ||f-g||_{\infty}$ and noting that $\{\psi_{t,j,k} : t \in \mathbb{Z}\}$ is a martingale difference sequence for any $k \geq 1$, we can show that the process $Z^{\gamma}_{T,k} = (Z^{\gamma}_{T,k}(f))_{f \in \mathcal{F}}$ has subgaussian increments. More specifically, we can apply an exponential inequality for martingale differences such as theorem 15.20 in Davidson (1994) to obtain that

$$\begin{aligned}
&\mathbb{P}\left(\big|Z^{\gamma}_{T,k}(f) - Z^{\gamma}_{T,k}(g)\big| > \varepsilon\, ||f-g||_{\infty}\right) \\
&\leq 2 \exp\left(- \frac{\varepsilon^2}{2 \sum_{t=k+1}^{T} \left(\frac{T^{\gamma} M a_T}{\sqrt{T}} C T^{1/12+4\delta}\right)^2 } \right) \\
&\leq 2 \exp\left(- \frac{\varepsilon^2}{2 (CM)^2 \left(T^{\gamma} a_T\right)^2 T^{1/6+8\delta}} \right) \leq 2 \exp\left(- \frac{\varepsilon^2}{2}\right)
\end{aligned}$$

for T large enough. Next, let $\|\cdot\|_{\psi_0}$ denote the Orlicz norm corresponding to $\psi_0(x) = \exp(x^2) - 1$. Applying a maximal inequality such as theorem 2.2.4 in van der Vaart and Wellner (1996) along with the metric entropy bound (61), we obtain that

$$\begin{aligned}
\Big\| \sup_{f \in \mathcal{F}} |Z^{\gamma}_{T,k}(f)| \Big\|_{\psi_0} &\leq \int_0^C \sqrt{K \varepsilon^{-\frac{1}{1/2+\eta}}}\, d\varepsilon = \sqrt{K} \int_0^C \varepsilon^{-\frac{1}{1+2\eta}}\, d\varepsilon \\
&= \sqrt{K} \frac{1}{1 - \frac{1}{1+2\eta}} \varepsilon^{1-\frac{1}{1+2\eta}} \Big|_0^C \leq r_0 < \infty
\end{aligned}$$

with some sufficiently large C. Hence, by Markov's inequality,

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}|Z_{T,k}(f)|>\varepsilon\right)=\mathbb{P}\left(T^{-\gamma}\sup_{f\in\mathcal{F}}|Z^{\gamma}_{T,k}(f)|>\varepsilon\right)$$
$$\leq\frac{\mathbb{E}\left[\psi_0\left(\sup_{f\in\mathcal{F}}|Z^{\gamma}_{T,k}(f)|/r_0\right)\right]}{\psi_0(\varepsilon T^{\gamma}/r_0)}\leq\frac{1}{\exp(\varepsilon^2T^{2\gamma}/r_0^2)-1},$$

which completes the proof of (62). □

References

Asgharian, H., Hou, A. J., and Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A garch-midas approach. *Journal of Forecasting*, 32:600–612.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes : estimation and prediction*. Number 110 in Lecture notes in statistics. Springer, New York, 2nd edition.

Christiansen, C., Schmeling, M., and Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economics variables. *Journal of Applied Econometrics*, 27:956–977.

Conrad, C. and Loch, K. (2014). Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, 30:1090–1114.

Dahlhaus, R. (1996a). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In Robinson, P. and Rosenblatt, M., editors, *Athens conference on applied probability and time series analysis*, volume 2. Springer.

Dahlhaus, R. (1996b). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, 62(1):139–168.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37.

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics*, 34(3):1075–1114.

Davidson, J. (1994). *Stochastic Limit Theory*. Advanced Texts in Econometrics. Oxford University Press, New York.

Engle, R. F., Ghysels, E., and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics*, 95(3):776 – 797.

Engle, R. F. and Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, 21(3):1187–1222.

Feng, Y. (2004). Simultaneously Modelling Conditional Heteroskedasticity and Scale Change. *Econometric Theory*, 20(3):563–596.

Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2008). Normalized least-squares estimation in time-varying arch models. *Annals of Statistics*, 36:742–786.

Hafner, C. M. and Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics*, 159(1):55–73.

Han, H. and Kristensen, D. (2014). Asymptotic theory for the qmle in garch-x models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics*, 32(3):416–429.

Hansen, B. E. (2008). Uniform Convergence Rates for Kernel Estimation with Dependent Data. *Econometric Theory*, 24(3):726–748.

Lamoureux, C. G. and Lastrapes, W. D. (1990). Persistence in variance, structural change, and the garch model. *Journal of Business & Economic Statistics*, 8(2):225–234.

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

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599.

Mikosch, T. and Stărică, C. (2000). Is it really long memory we see in financial returns? In Embrechts, P., editor, *Extremes and Integrated Risk Management*, pages 149–168. Risk Books.

Mikosch, T. and Stărică, C. (2003). Long-Range Dependence Effects and ARCH Modelling. In Doukhan, P., Oppenheim, G., and Taqqu, M. S., editors, *Theory and Applications of Long Range Dependence*, pages 439–459. Birkhäuser.

Mikosch, T. and Stărică, C. (2004). Non-stationarities in financial time series, the long-range dependence, and igarch effects. *The Review of Economics and Statistics*, 86(1):378–390.

Mittnik, S., Robinzonov, N., and Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of theiry impact. *Journal of Banking and Finance*, 58:1–14.

Paye, B. S. (2012). 'déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106:527–546.

Truquet, L. (2017). Parameter stability and semiparametric inference in time varying auto-regressive conditional heteroscedasticity models. *Journal of the Statistical Royal Society. Series B*, 79(5):1391–1414.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York.

Vogt, M. and Walsh, C. P. (2019). Estimating nonlinear additive models with nonstationarities and correlated errors. *Scandinavian Journal of Statistics*, 46(1):160–199.

Wang, F. and Ghysels, E. (2015). Econometric analysis of volatility component models. *Econometric Theory*, 15:362–393.