UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for
New Economic Thinking

Cambridge-INET Institute

Cambridge-INET Working Paper Series No: 2016/06

Cambridge Working Paper Economics: 1633

A SEMIPARAMETRIC INTRADAY GARCH MODEL

Peter Malec
(University of Cambridge)

We propose a multiplicative component model for intraday volatility. The model consists of a seasonality factor, as well as a semiparametric and parametric component. The former captures the well-documented intraday seasonality of volatility, while the latter two account for the impact of the state of the limit order book, utilizing an additive structure, and fluctuations around this state by means of a unit GARCH specification. The model is estimated by a simple and easy-to-implement approach, consisting of across-day-averaging, smooth-backfitting and QML steps. We derive the asymptotic properties of the three component estimators. Further, our empirical application based on high-frequency data for NASDAQ equities investigates non-linearities in the relationship between the limit order book and subsequent return volatility and underlines the usefulness of including order book variables for out-of-sample forecasting performance.

A Semiparametric Intraday GARCH Model*

Peter Malec†

This Version: May 2016

Abstract

We propose a multiplicative component model for intraday volatility. The model consists of a seasonality factor, as well as a semiparametric and parametric component. The former captures the well-documented intraday seasonality of volatility, while the latter two account for the impact of the state of the limit order book, utilizing an additive structure, and fluctuations around this state by means of a unit GARCH specification. The model is estimated by a simple and easy-to-implement approach, consisting of across-day-averaging, smooth-backfitting and QML steps. We derive the asymptotic properties of the three component estimators. Further, our empirical application based on high-frequency data for NASDAQ equities investigates non-linearities in the relationship between the limit order book and subsequent return volatility and underlines the usefulness of including order book variables for out-of-sample forecasting performance.

Keywords: intraday volatility, GARCH, smooth backfitting, additive models, limit order book.

JEL classification: C14, C22, C53 , C58.

1 Introduction

This paper proposes a new semiparametric model that accounts for the impact of economic covariates on the conditional volatility of intraday returns. Of particular interest are those

*For helpful comments and discussions, I thank Melanie Schienle, Oliver Linton, Ruihong Huang and the participants of the 9th International Conference on Computational and Financial Econometrics, London, December 2015, as well as the econometrics seminar at the University of Cambridge. Financial support from the Cambridge INET is gratefully acknowledged.

†Faculty of Economics, University of Cambridge. Email: pm563@cam.ac.uk. Address: Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom.

covariates that capture the state of the limit order book (LOB), which collects outstanding limit orders for a given asset with a limit order being an order to buy or sell a given number of shares at a pre-specified or better price. Several studies in empirical market microstructure suggest that especially the LOB depth, that is, the cumulative size of outstanding orders available at a given price level, plays a major role in this context. For example, Ahn et al. (2001) discover that increased depth is followed by lower return volatility. The findings of Coppejans et al. (2001) suggest that these volatility reductions might be due to a rise in the unexpected part of LOB depth. Further, Pascual and Veredas (2010) find that the effects of higher depth depend on whether the latter is available *at* or *behind* the best prices in the book, being positive in the former case and negative in the latter. More recently, Valenzuela et al. (2015) argue that it is the relative depth available at the different levels of the LOB which drives subsequent volatility.

Predictions of asset return volatility over short intraday horizons are crucial for manifold applications in financial practice. These include the development of order placement strategies, where the future volatility level is an important determinant when selecting optimal order types, the construction of intraday Value-at-Risk measures for trading desks, as well as the implementation of volatility circuit breakers by exchange operators. At the above intraday frequencies, volatility exhibits two stylised facts. First, a strong persistence, analogous to the dynamic properties of low-frequency, i.e. daily or monthly returns, typically captured by traditional GARCH models (Bollerslev, 1986; Engle, 1982). Second, a pronounced seasonality, implying that fitting the latter models to returns sampled at different intraday frequencies yields estimates that are inconsistent with the temporal aggregation results by Drost and Nijman (1993). These features gave rise to the framework proposed in Andersen and Bollerslev (1997, 1998), decomposing volatility multiplicatively into a seasonal factor and a GARCH component.

The above microstructure results motivate the incorporation of LOB covariates into the aforementioned GARCH-type models for intraday volatility. An important complication, however, is the fact that modern game-theoretical models for LOBs, such as the seminal framework by Rosu (2009), do not suggest specific functional forms for the relationship between depth and subsequent (conditional) return volatility. This problem makes a semiparametric approach necessary given that a pure nonparametric treatment would be infeasible when including several LOB characteristics in the model. A seemingly straightforward solution is to include a non- or semiparametric component accounting for the impact of the LOB state into the GARCH dynamics directly, resulting in an additive GARCH-X-type model. From a practical viewpoint, however, such a strategy is less attractive since estimation would have to proceed similarly to the approach proposed by Linton and Mammen (2005) for their semiparametric ARCH(∞) framework, which would be numerically unstable in a setting with possibly many covariates and vast high-frequency datasets. In this context, a multiplicative component structure offers a powerful, but parsimonious and computationally attractive, alternative. The idea is to augment the seasonality factor and GARCH dynamics multiplicatively by a semiparametric additive component that accounts for the impact of LOB covariates on the conditional volatility of intraday returns.

Hence, our main contributions can be summarised as follows. First, we propose a model for intraday volatility, which consists of three components with each one capturing a particular feature of the process in a parsimonious way. A seasonality factor, accounting for the intraday periodicity of volatility, a semiparametric additive component, capturing, e.g., the effect of the LOB through possibly many covariates, as well as unit-GARCH dynamics, describing fluctuations around the aforementioned previous factors and thus accounting for remaining persistence in volatility. The three components are easy to estimate by means of simple across-day averaging of squared returns, the smooth backfitting algorithm proposed by Mammen et al. (1999) and quasi-maximum likelihood, respectively. Second, we derive the asymptotic theory including rates and asymptotic normality for all component estimators. In particular, we identify the contributions of the (pre-)estimators of the seasonality and semiparametric component to the asymptotic variance of the GARCH parameter estimators on the final step. In an empirical study, we apply our methodology to high-frequency data for NASDAQ blue chips. Our findings suggest that the relationship between depth and subsequent (conditional) return volatility might be highly non-linear, underlining the importance of a semiparametric approach. Further, we conduct an out-of-sample forecasting exercise, which shows that our method significantly outperforms all benchmarks.

The proposed framework adds to the recent literature on multiplicative component GARCH models initiated by the spline-GARCH framework by Engle and Rangel (2008). The latter extends GARCH dynamics for, e.g., daily returns in a multiplicative way by a long-run trend component capturing the business cycle, resulting in a possibly non-stationary volatility model with smoothly varying unconditional variance. This approach is, e.g., refined by Hafner and Linton (2010) who estimate the long-run trend component by kernel methods instead of splines, which allows them to derive an asymptotic theory in a general setting. More recently, Han and Kristensen (2015) replace the deterministic trend by a single, possibly non-stationary covariate, yielding a semiparametric multiplicative GARCH-X. Several covariates are allowed in GARCH-MIDAS models as introduced by Engle et al. (2013) and treated theoretically in Wang and Ghysels (2015). In these models, the long-run component is driven by many lags of realised volatility estimates or various economic covariates (see, e.g., Conrad and Loch, 2015). Importantly, however, smoothness is ensured by imposing a *parametric* weighting scheme on the lag structure. Accordingly, the incorporation of a multitude of covariates into the multiplicative framework through a *semiparametric* component is a novelty in this field.

Further, our paper extends the studies on intraday GARCH models, which modify the seminal approach by Andersen and Bollerslev (1997, 1998) in several directions. Whereas in the latter framework, volatility dynamics are driven by a single intradaily *or* re-scaled daily GARCH component in addition to the seasonality factor, Engle and Sokalska (2012) allow for a daily component *and* dedicated intraday GARCH dynamics. Taylor and Xu (1997) modify the GARCH component, extending it by both lagged implied and realised volatility measures. Finally, Giot (2005) incorporates a set of LOB-related variables into the GARCH equation parametrically. However, none of the aforementioned studies captures the relationship between the LOB and conditional volatility in a flexible way without imposing (too much) functional form.

Finally, our results complement the literature on smooth backfitting of additive models. The general estimation idea put forward by Mammen et al. (1999) and its asymptotic theory have been adapted to various settings, e.g., dynamic models for realised variance estimates (Fengler et al., 2015). A framework related to ours is covered in Vogt and Walsh (2012) who study the estimation of semiparametric additive models for time series including seasonality and trend. Unlike our setting, though, the different components are combined in a purely additive fashion. The asymptotic theory presented in our study builds on the results of Mammen et al. (1999) and extends them crucially by providing rates and asymptotic normality of the smooth backfitting estimator in a setting with the semiparametric additive structure being part of a multiplicative model, subject to a pre-estimated seasonality factor and including serially dependent regression errors. In particular, the multiplicative framework does not allow for a simple application and/or modification of the results from an additive setting, but requires an independent treatment to account for novel effects.

The remainder of the paper is structured as follows. Section 2 introduces the proposed model, discusses the estimators of the various components, provides their asymptotic properties and outlines details of the implementation. In Section 3, we present an empirical application to NASDAQ data. Finally, Section 4 concludes. All proofs are deferred to Appendix A.

2 Semiparametric Estimation of Intraday Volatility

2.1 Modelling Framework

We consider D trading days and sample prices on a regular grid of N points per day. Let r_t denote the resulting i-th centred intraday log-return on day d, such that $t := (d-1)N + i$ with $i = 1, \ldots, N$ and $d = 1, \ldots, D$. We assume the multiplicative component model

$$r_t = \sqrt{s_t \, g_t \, v_t}\, \eta_t, \quad t = 1, \ldots, T, \quad T := ND. \tag{1}$$

Here, s_t captures the well-known intraday seasonality of volatility (see, e.g., Andersen and Bollerslev, 1997, 1998). Accordingly, it is periodic with period N, satisfying $s_t = s_{t+kN}$, $k \in \mathbb{N}$.

$g_t := g(x_{t-1})$ accounts for the impact of the state of the LOB observed at the beginning of the t-th return spell, which is subsumed in the vector $x_{t-1} := \left(x_{t-1}^{(1)}, \ldots, x_{t-1}^{(J)}\right)^T$. Thus, x_{t-1} can contain depths for various levels of the LOB. The fact that the latter are observed at the beginning of the given spell prevents possible endogeneity issues. For this component, we assume the semiparametric additive specification

$$g(x_{t-1}) = 1 + \sum_{j=1}^{J} g^{(j)}\left(x_{t-1}^{(j)}\right), \quad t = 1, \ldots, T, \tag{2}$$

where setting the intercept to one is necessary for the seasonality component s_t to be identified. Likewise, the unknown functions $g^{(j)}$, $j = 1, \ldots, J$, are identified only up to an additive constant. Hence, we impose the identifying restrictions

$$\int_{S_j} g^{(j)}(u)\, p^{(j)}(u)\, du = 0, \quad j = 1, \ldots J, \tag{3}$$

where $p^{(j)}$ denotes the marginal density and S_j the support of covariate j. As we discuss in more detail in Section 2.3, we require the covariates to have compact support. To keep notation straightforward, in the following we assume that $S_j = [0, 1]$, $j = 1, \ldots, J$.

v_t represents fluctuations of intraday volatility around the level set by the seasonality component and the LOB. It is specified in terms of a unit GARCH process, i.e.

$$v_t = 1 - \alpha - \beta + \alpha u_{t-1}^2 + \beta v_{t-1}, \quad u_t := r_t / \sqrt{s_t \, g_t}, \tag{4}$$

where $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. As in the case of g_t, the unit intercept ensures that the seasonality component is identified and implies that the process $\{v_t\}$ mean-reverts to an unconditional mean of one. Note that, in general, $u_t = u_t(g, s_t)$ and $v_t = v_t(\theta, g, \mathscr{S}_N)$, where $\theta := (\alpha, \beta)'$, $\theta \in \Theta$ with Θ to be specified below, and $\mathscr{S}_N := (s_1, \ldots, s_N)'$. The basic specification in (4) could be extended to allow for richer dynamics, e.g., allowing for leverage effects as in the asymmetric GARCH model by Glosten et al. (1993). However, we stick to this simple alternative to allow for a clear and accessible presentation of the resulting asymptotic theory below.

$\{\eta_t\}$ is a sequence of i.i.d. disturbances with $\mathbb{E}\big[\eta_t^2 | \mathscr{H}_{t-1}\big] = 1$, where $\mathscr{H}_t$ is the sigma algebra generated by $\{(r_t, x_t)\}$, i.e. $\mathscr{H}_t := \sigma(r_t, x_t, \ldots, r_{-\infty}, x_{-\infty})$. Summarising the above assumptions on s_t, g_t and v_t, while additionally supposing that $\mathbb{E}\Big[v_t | \mathscr{H}_{t-1}^{(x)}\Big] = 1$, with $\mathscr{H}_t^{(x)} := \sigma(x_t, \ldots, x_{-\infty})$, it immediately follows that the expected squared return conditional on past observations of the covariates satisfies $\mathbb{E}\Big[r_t^2 | \mathscr{H}_{t-1}^{(x)}\Big] = s_t$, $t = 1, \ldots, T$, implying that the general level of volatility is ultimately determined by the seasonality component.

Further and more detailed technical assumptions on the four components in (1) are discussed in Section 2.3.

2.2 Estimation

We propose to estimate the components in (1) in three steps. In an initial step, we estimate the seasonality component s_t. Subsequently, we estimate the additive functions $g^{(j)}$, $j = 1, \ldots, J$, driving g_t, and the parameter vector θ of the GARCH component v_t.

The pre-step yielding estimates of the seasonality component exploits the periodic structure of the latter. Accordingly, we compute across-day sample means of squared returns for each intraday grid point as

$$\hat{s}_t = \hat{s}_i = \frac{1}{D} \sum_{k=1}^{D} r_{(k-1)N+i}^2, \quad i = 1, \ldots, N, \tag{5}$$

where $\hat{s}_{t+kN} = \hat{s}_t$, $k \in \mathbb{N}$. This simple method can be considered as a special case with known period N of the approach proposed by Vogt and Linton (2014) and was employed in the context of the deseasonalisation of intraday returns, e.g., in Engle and Sokalska (2012).[1]

For the estimation of the remaining two components, we consider (estimated) deseasonalised returns $\hat{z}_t := r_t / \sqrt{\hat{s}_t}$. The latter are used as inputs for an estimation approach, which consists of the following two additional steps:

[1] Vogt and Linton (2014) additionally account for a smooth trend component.

1. We estimate the semiparametric additive model

$$\hat{z}_t^2 = 1 + \sum_{j=1}^{J} g^{(j)}\Big(x_{t-1}^{(j)}\Big) + \tilde{\varepsilon}_t, \quad \tilde{\varepsilon}_t := g_t\left(\frac{s_t}{\hat{s}_t}\, v_t\, \eta_t^2 - 1\right), \tag{6}$$

by smooth backfitting. Based on the estimated additive functions $\hat{g}^{(j)}$, $j = 1, \ldots, J$, we obtain $\{\hat{g}_t\}_{t=1}^{T}$ according to $\hat{g}_t := 1 + \sum_{j=1}^{J} \hat{g}^{(j)}\Big(x_{t-1}^{(j)}\Big)$. The errors $\{\tilde{\varepsilon}_t\}$ as well as their counterparts based on the true seasonality component, $\varepsilon_t := g_t\left(v_t\, \eta_t^2 - 1\right)$, $t = 1, \ldots, T$, are not serially independent. As is discussed below, this property crucially affects the choice of a suitable estimator for the additive model (6).

2. We maximize the Gaussian quasi log-likelihood function of $\{u_t\}_{t=1}^{T}$ evaluated at the estimates $\left\{\hat{u}_t^*\right\}_{t=1}^{T}$, where $\hat{u}_t^* := \hat{z}_t / \hat{g}_t = u_t(\hat{g}, \hat{s}_t)$, $t = 1, \ldots, T$, i.e.

$$\max_{\theta \in \Theta} \mathscr{L}_T\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big), \quad \mathscr{L}_T(\theta, g, \mathscr{S}_N) := \frac{1}{T} \sum_{t=1}^{T} l_t(\theta, g, \mathscr{S}_N), \tag{7}$$

$$l_t(\theta, g, \mathscr{S}_N) := -\ln v_t(\theta, g, \mathscr{S}_N) - \frac{u_t(g, s_t)^2}{v_t(\theta, g, \mathscr{S}_N)},$$

which yields the parameter estimates $\hat{\theta}$.

We refrain from extending the above estimation approach by repeating steps one and two, while starting from the "refined" series $\left\{\hat{z}_t^2 / \hat{v}_t\right\}_{t=1}^{T}$, where $\hat{v}_t := v_t\big(\hat{\theta}, \hat{g}, \hat{\mathscr{S}}_N\big)$, $t = 1, \ldots, T$. The asymptotic theory for the resulting additional estimation steps, which would be in line with the multi-step estimation approaches proposed in Hafner and Linton (2010) and Han and Kristensen (2015), could be derived analogously to the proofs of the results in Section 2.3. However, results not reported here show that, in the given setting involving pre-estimation of seasonality and smooth backfitting, the aforementioned modifications turn out to yield numerically somewhat unstable behaviour and a worse out-of-sample prediction performance in the empirical application in Section 3.[2]

We estimate $g_1, \ldots, g_J$ in (6) by smooth backfitting with the estimators minimising the smoothed least-squares criterion

$$\int_0^1 \sum_{t=1}^{T} \Big[\hat{z}_t^2 - 1 - \sum_{j=1}^{J} g^{(j)}\big(x^{(j)}\big)\Big]^2 \prod_{j=1}^{J} K_{h_j}\Big(x^{(j)}, x_{t-1}^{(j)}\Big)\, dx^{(1)} \ldots dx^{(J)}, \tag{8}$$

[2]These results are available from the author upon request.

under the constraint

$$\int_0^1 g^{(j)}(u)\,\hat{p}^{(j)}(u)\,du = 0, \quad j = 1, \ldots, J. \tag{9}$$

Here, $\hat{p}^{(j)}$ is the kernel estimator of the density of the j-th covariate, i.e.

$$\hat{p}^{(j)}\big(x^{(j)}\big) = \frac{1}{T}\sum_{t=1}^{T} K_{h_j}\Big(x^{(j)}, x_{t-1}^{(j)}\Big), \quad j = 1, \ldots, J, \tag{10}$$

while $K_{h_j}(u,v)$ denotes a modified kernel defined as

$$K_{h_j}(u,v) := \frac{K_{h_j}(u-v)}{\int_0^1 K_{h_j}(z-v)\,dz}, \quad j = 1, \ldots, J, \tag{11}$$

with $K_h(u-v) := h^{-1}K\big[h^{-1}(u-v)\big]$, where K integrates to one on its support. The modification of the kernel in (11) ensures that the estimate $\hat{p}^{(j)}\big(x^{(j)}\big)$ integrates to one over the compact support. The latter property is required for the asymptotic theory of the smooth backfitting estimator. See Section 2.3 for details. Importantly, Mammen et al. (1999) show that the minimisation of the criterion (8) is tantamount to a projection of the data, i.e. in our case $\hat{z}_t^2$, $t = 1, \ldots, T$, onto the space of additive functions $\sum_{j=1}^{J} g^{(j)}\big(x^{(j)}\big)$ with respect to a particular semi-norm. Thus, if the additive model is not correct, smooth backfitting estimates the best additive fit to the non-additive model.

Fixing the value of $x^{(j)}$ and minimising the criterion (8) with respect to the estimator $\hat{g}^{(j)}\big(x^{(j)}\big)$ after some simplifications yields the solution

$$\hat{g}^{(j)}\big(x^{(j)}\big) = \hat{g}_{NW}^{(j)}\big(x^{(j)}\big) - 1 - \sum_{k\neq j}\int_0^1 \frac{\hat{p}^{(j,k)}\big(x^{(j)}, x^{(k)}\big)}{\hat{p}^{(j)}\big(x^{(j)}\big)}\hat{g}^{(k)}\big(x^{(k)}\big)\,dx^{(k)}, \tag{12}$$

where $\hat{g}_{NW}^{(j)}\big(x^{(j)}\big)$ is the univariate Nadaraya-Watson estimator of the component function $g^{(j)}\big(x^{(j)}\big)$ and $\hat{p}^{(j,k)}\big(x^{(j)}, x^{(k)}\big)$ denotes the kernel estimator of the joint density of the j-th and k-th covariate, $p^{(j,k)}$, such that

$$\hat{g}_{NW}^{(j)}\big(x^{(j)}\big) = \left[\sum_{t=1}^{T} K_{h_j}\Big(x^{(j)}, x_{t-1}^{(j)}\Big)\right]^{-1}\sum_{t=1}^{T} K_{h_j}\Big(x^{(j)}, x_{t-1}^{(j)}\Big)\hat{z}_t^2, \tag{13}$$

$$\hat{p}^{(j,k)}\big(x^{(j)}, x^{(k)}\big) = \frac{1}{T}\sum_{t=1}^{T} K_{h_j}\Big(x^{(j)}, x_{t-1}^{(j)}\Big) K_{h_k}\Big(x^{(k)}, x_{t-1}^{(k)}\Big). \tag{14}$$

Condition (12) gives rise to an iterative algorithm, allowing for a straightforward implementation of the smooth backfitting estimator. We refer to Mammen et al. (1999) for details.

Possible alternatives to the above smooth backfitting approach would be either "traditional" backfitting (see Buja et al., 1989; Hastie and Tibshirani, 1990) or marginal integration (see Linton and Nielsen, 1995; Newey, 1994; Tjostheim and Auestad, 1994). The major shortcoming of the backfitting estimator in the given context is that its asymptotic analysis requires the regression errors of the underlying semiparametric additive model to be independent (see Opsomer, 2000; Opsomer and Ruppert, 1997). As was discussed above, this assumption is not satisfied in framework (6). Moreover, stronger assumptions on the joint density of the covariates are needed than for the study of the smooth backfitting estimator. Finally, Nielsen and Sperlich (2005) illustrate by simulations that the latter, in comparison to the classical backfitting algorithm, is more robust to degenerated designs and a large number of additive components. This finding can be explained by the fact that the iteration equation resulting from (12) is a smoothed, and hence stabilised, version of its "traditional" backfitting counterpart. Marginal integration, on the other hand, suffers from the fact that it does not generally achieve optimal rates. To improve efficiency, two-stage estimation methods as proposed in Linton (1997) and Fan et al. (1998) are required.

2.3 Asymptotic Theory

We begin by summarising the technical assumptions for the components in (1). Subsequently, we present asymptotic results for the estimators of the seasonality component, the functions inside the semiparametric component and the parameters of the GARCH component.

We impose the following conditions on the seasonality component, component functions of the additive component, as well as the kernels employed to estimate the latter:

Assumption 1

(a) The kernel K is bounded with compact support (e.g. $[-C,C]$) and satisfies $|K(u)-K(w)| \leq C_1\,|u-v|$ for some constant $C_1 < \infty$ and all $x, y \in [-C,C]$.

(b) X_t has compact support (e.g. $[0,1]^J$). 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. p is bounded away from zero on $[0,1]^J$. The first partial derivatives of the density p exist and are continuous.

(c) The second partial derivatives of the functions $g^{(j)}$, $j = 1,\ldots,J$, exist and are Lipschitz continuous.

(d) For some $\gamma > 16/3$, $C_1 < \infty$ and $\bar{l} \in \mathbb{N}$, $\mathbb{E}[|\eta_t|^\gamma | X_{t-1}] < C_1$ and $\mathbb{E}[|\eta_t \eta_{t+l}| | X_{t-1}, X_{t-1+l}] < C_1 \; \forall \, l > \bar{l}$.

(e) For some $\rho > 2$, $\mathbb{E}\big[X_t^{2\rho}\big] < \infty$.

(f) The conditional densities $p_{X_t|Z_t}$ and $p_{X_t,X_{t+l}|Z_t,Z_{t+l}}$, $l = 1, 2, \ldots$, exist and are bounded from above.

(g) $\{(\eta_t, X_t)\}$ is strictly stationary and strongly mixing with mixing coefficients $\alpha(i) \leq C_1 m^i$ for some $m < 1$.

(h) s_t is bounded away from zero for $t = 1, \ldots, N$.

(i) $g(x)$ is bounded away from zero uniformly in $x \in [0,1]^J$.

Assumptions 1 are standard and widely in line with Mammen et al. (1999). The compact support in part (b) is necessary to obtain uniform expansions of the so-called "stochastic" and "bias" part of the smooth backfitting estimators, respectively. If the covariates are strictly stationary as in part (g), this type of assumption is not restrictive. In empirical applications, the lower and upper bound of the support can be set equal, e.g., to the sample minimum and maximum of the given covariate. The exponential mixing rate in part (g) is not necessary and mainly chosen for notational convenience. Somewhat slower and more general rates are possible at the cost of more tedious notation (see, e.g., Mammen et al., 1999) as well as more restrictive moment assumptions for the GARCH component below.

The next set of assumptions ensures consistency and asymptotic normality of the GARCH parameter estimators $\hat{\theta}$, where we denote the true parameter vector by $\theta_0 := (\alpha_0, \beta_0)'$:

Assumption 2

(a) $\theta_0 \in \Theta$ with Θ being compact.

(b) $\alpha + \beta < 1$ for each $\theta \in \Theta$.

(c) η_t^2 has a non-degenerate distribution with $\mathbb{E}\big[\eta_t^2 | \mathscr{H}_{t-1}\big] = 1$ and $\mathbb{E}[\eta_t | \mathscr{H}_{t-1}] = 0$.

(d) If $\beta_0 > 0$, $\alpha_0 > 0$.

(e) $\mathbb{E}\Big[\big(\beta_0 + \alpha_0 \eta_t^2\big)^5\Big] < 1$ and $\mathbb{E}\Big[\big(\beta_0 + \alpha_0 \eta_t^2\big)^2 | \mathscr{H}^{(x,j)}\Big] < 1$, where $\mathscr{H}^{(x,j)}$ is the sigma algebra generated by $\big\{x_{\infty}^{(j)}, \ldots, x_{-\infty}^{(j)}\big\}$, $j = 1, \ldots, J$.

(f) θ_0 is in the interior of Θ.

Parts (a)-(d) of Assumption 2 ensure consistency, while parts (e) and (f) are additionally required for asymptotic normality. These assumptions roughly follow those made in Francq and Zakoian (2004). Assumption 2(b) guarantees strict stationarity and strong mixing with exponential rate of v_t. Our formulation is somewhat stronger than the corresponding Assumption (A2) in Francq and Zakoian (2004), but ensures that the true intercept is strictly positive and that v_t is bounded away from zero uniformly in $\theta \in \Theta$. Assumption 2(e) is additionally introduced and, by an adaptation of Corollary 6 from Carrasco and Chen (2002), guarantees that $\mathbb{E}\big[u_t^{10}\big] < \infty$ and $\mathbb{E}\big[u_t^4 | \mathscr{H}^{(x,j)}\big] < \infty$, $j = 1, \dots, J$, respectively. Hence, Assumption 2(e) is clearly stronger than its counterpart in the parametric GARCH case, requiring only the existence of fourth unconditional moments of the GARCH disturbances (Assumption (A6) in Francq and Zakoian (2004)). In the given semiparametric *and* multiplicative framework, however, this strengthening of the assumptions is necessary to bound certain terms in the contribution of the smooth backfitting estimators in (6) to the asymptotic variance of the GARCH parameter estimator. More precisely, to account for the variance part of the smooth backfitting estimator, we only need that $\mathbb{E}\big[u_t^{8+\delta}\big] < \infty$, $\delta > 0$, but as is well-known, the computation of odd-order unconditional moments implied by a GARCH process is extremely difficult (see, e.g., Francq and Zakoian, 2010, Section 2.4).[3] Finally, note that we cannot loosen Assumption 2(e) any further as this would have to be compensated by an even faster decay rate of the mixing coefficients, which is already taken to be exponential according to Assumption 1(g).[4]

Our first basic result provides asymptotic normality for the estimator of the seasonality component s_t in (5).

Theorem 1. *Let Assumptions 1(d)-(h) and 2(b) be fulfilled. Then*

$$\sqrt{T}\,(\hat{s}_t - s_t) \xrightarrow{d} \mathcal{N}\left(0, N s_t^2 \sum_{j=-\infty}^{\infty} \mathbb{C}\mathrm{ov}\big[q_t, q_{t-j}\big]\right), \quad t = 1, \dots, N, \tag{15}$$

where $q_t := z_t^2 - 1$.

The straightforward proof of Theorem 1 proceeds along standard lines. See Appendix A for details. Note that due the finite number of indices across which s_t may differ, i.e. $t = 1, \dots, N$, Theorem 1 directly implies that $\sup_{t \in \{1,\dots,T\}} |\hat{s}_t - s_t| = O_p\big(T^{-1/2}\big)$. Further, for the purpose of inference in empirical applications, the asymptotic variance in (15) could be estimated by applying a standard long-run variance estimator to $\{\hat{q}_t\}$.

[3]Trivially, odd-order moments will be zero when they exist if we assume that the distribution of η_t is symmetric.

[4]More precisely, Assumption 2(e) follows from an application of Davydov's lemma in order to bound the absolute (auto-)covariances of a strong mixing process (see, e.g., Hall and Heyde, 1980, Corollary A2). See Appendix A for details.

The next theorem provides uniform consistency and asymptotic normality of the smooth backfitting estimators $\hat{g}^{(j)}$, $j = 1, \ldots, J$. For ease of notation, we set $h = h_1 = \ldots = h_J$.

Theorem 2. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that $h \asymp T^{-(1/4+\epsilon)}$ for some $\epsilon > 0$. Then*

$$\sup_{x^{(j)} \in [0,1]} \left| \hat{g}^{(j)}\big(x^{(j)}\big) - g^{(j)}\big(x^{(j)}\big) \right| = O_p(h), \quad j = 1, \ldots, J. \tag{16}$$

In addition, for $x^{(j)} \in (0,1)$,

$$\sqrt{Th}\left(\hat{g}^{(j)}\big(x^{(j)}\big) - g^{(j)}\big(x^{(j)}\big)\right) \xrightarrow{d} \mathcal{N}\left(0, \kappa_2 \frac{\sigma_j^2\big(x^{(j)}\big)}{p^{(j)}\big(x^{(j)}\big)}\right), \quad j = 1, \ldots, J, \tag{17}$$

where $\kappa_2 := \int K(u)^2 du$ and $\sigma_{z,j}^2\big(x^{(j)}\big) := \mathbb{V}\big[z_t^2 - g(X)\,|X^{(j)} = x^{(j)}\big]$.

The proof of Theorem 2 consists of two main elements. First, the pre-estimation of the seasonality component needs to be accounted for. The corresponding estimation error is included in the bias part of the smooth backfitting estimator. Then, a uniform expansion for the latter needs to be formulated, which crucially exploits the periodic nature of the seasonality component and its estimator. Second, a uniform expansion for the stochastic part of the smooth backfitting estimator originally proposed in Mammen and Park (2005) has to be extended to allow for strongly mixing data. Details are provided in Appendix A. Importantly, the bandwidth of order $T^{-(1/4+\epsilon)}$ implies that we undersmooth compared to the MSE-optimal choice for the univariate estimators $\hat{g}_{NW}^{(j)}$, i.e. $h = O\big(T^{-1/5}\big)$ used in Mammen et al. (1999). As we discuss below, this modification is required for the asymptotic normality of the GARCH parameter estimator $\hat{\theta}$ in the following step. Obviously, if we are merely interested in the component functions $g^{(j)}$, $j = 1, \ldots, J$, and point estimates of θ, MSE-optimal smoothing can be retained. The latter additionally introduces an asymptotic bias term of order h^2, representing the bias leading term of the J-dimensional Nadaraya-Watson estimator projected on the space of additive functions. See Mammen et al. (1999) for details.

The final theorems in this section grant consistency and asymptotic normality to the GARCH parameter estimator $\hat{\theta}$.

Theorem 3. *Let Assumption 1 and Assumption 2(a)-(d) be fulfilled. Further, assume that $h \asymp T^{-(1/4+\epsilon)}$ for some $\epsilon > 0$. Then $\hat{\theta} \to \theta_0$ almost surely.*

Theorem 4. *Let Assumption 1 and Assumption 2 be fulfilled. Further, assume that $h \asymp T^{-(1/4+\epsilon)}$ for some $\epsilon > 0$. Then*

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\Big(0, \mathcal{H}(\theta_0, g, \mathcal{S}_N)^{-1}\, \Sigma(\theta_0, g, \mathcal{S}_N)\, \mathcal{H}(\theta_0, g, \mathcal{S}_N)^{-1}\Big),$$

where $\mathcal{H}(\theta_0, g, \mathcal{S}_N) := \mathbb{E}\Big[\frac{\partial^2 l_t(\theta, g, \mathcal{S}_N)}{\partial\theta\partial\theta^T}|_{\theta=\theta_0}\Big]$, $\Sigma(\theta_0, g, \mathcal{S}_N) := \sum_{j=-\infty}^{\infty} \mathbb{C}\mathrm{ov}\big[y_{t,T}, y_{t-j,T}\big]$ *and* $y_{t,T} = y_{t,T}(\theta_0, g, \mathcal{S}_N)$ *is defined in (54) in Appendix A.*

The proofs of both theorems can be found in Appendix A. For Theorem 3, the approximation error in the likelihood is decomposed into the errors due to the pre-estimation of the seasonality and semiparametric component, respectively, while using the consistency of the two latter component estimators, as well as standard results from Francq and Zakoian (2004). As in Theorem 2, the bandwidth could also be chosen in the (univariate) MSE-optimal way, which is not the case for Theorem 4. The reason is that the proof of the latter is based on considering the score evaluated at the true parameter vector θ_0 as a functional in the semiparametric component g. For this functional evaluated at the pre-estimate $\hat{g}$, we conduct a first-order expansion around the true semiparametric component g. The undersmoothing condition is then necessary to ensure that the corresponding remainder term, which is of order $O_p\big(h^2\big)$ becomes $o_p\big(T^{-1/2}\big)$. Importantly, the definition of $y_{t,T}$, $t = 1, \ldots, T$, in (54) in Appendix A demonstrates that the asymptotic variance of the GARCH parameter estimator is affected by the pilot estimates of the seasonality and semiparametric component, implying that the estimator does not possess an oracle property in the general framework considered. That is, in addition to the usual GARCH(1, 1) score in the Gaussian QML setting, which would drive the asymptotic variance given an oracle result, we can identify the contributions of the stochastic and bias part of the smooth backfitting estimators in (6) as well as the asymptotic variance of the seasonality component estimator (5). Accordingly, relying on standard asymptotic results for the QML estimator of GARCH(1, 1) models, e.g., by employing Bollerslev and Wooldridge (1992) standard errors, will generally lead to wrong inference in practice.

The rather complicated structure of the asymptotic variance in Theorem 4 has further implications for the ability to conduct inference in empirical applications. Here, in particular, the presence of weighting functions following from a uniform expansion of the stochastic part of the smooth backfitting estimators makes the derivation of a consistent estimator of $\Sigma(\theta_0, g, \mathcal{S}_N)$, i.e. the long-run variance of $\{y_{t,T}\}$, challenging in general. Thus, a more viable strategy to obtain confidence intervals in practice is to implement a valid re-sampling scheme. A simple algorithm is proposed in Section 2.4 below.

2.4 Feasible Implementation

To implement the component estimators introduced in Section 2.2 and conduct valid inference in practice, two remaining open questions need to be addressed. First, the choice of the bandwidths h_j for the smooth backfitting estimators $\hat{g}^{(j)}$, $j = 1, \ldots, J$, in (6). Second, a suitable bootstrap scheme for the GARCH parameters θ estimated in (7).

To determine the bandwiths h_j, $j = 1, \ldots, J$, in a data-driven way, we follow an approach proposed by Mammen and Park (2005). The later relies on the minimisation of a penalised sum of squares (PLS) criterion based on the smooth backfitting residuals with the penalty preventing in-sample over-fitting caused by "too small" bandwidths. Mammen and Park (2005) show that the resulting bandwidths are asymptotically equivalent to those minimising the average weighted squared error (up to an additive constant).[5]

In our framework, we need to modify the above method in two ways. First, we have to ensure that the estimates of the semiparametric component, $\hat{g}_t$, $t = 1, \ldots, T$, in (6) are strictly positive. Second, we need to satisfy the undersmoothing condition $h = O\left(T^{-(1/4+\epsilon)}\right)$, $\epsilon > 0$, in settings where we are interested in the asymptotic distribution of the GARCH parameter estimator $\hat{\theta}$ according to Theorem 4. This is achieved by introducing an additional penalty for negative estimates and employing undersmoothing bandwidths when checking for the latter. Thus modified, the bandwidth selection problem becomes

$$\hat{h}_1, \ldots, \hat{h}_J = \underset{h_1, \ldots, h_J}{\arg\min} \, \mathrm{RSS}(h_1, \ldots, h_J) \left\{ 1 + \frac{2K(0)}{T} \left[\sum_{j=1}^{J} \frac{1}{h_j} \right] + \sum_{t=1}^{T} \mathbb{1}_{\{\hat{g}_t^* \leq 0\}} \Delta \right\}, \tag{18}$$

where $\mathrm{RSS}(h_1, \ldots, h_J) := \sum_{t=1}^{T} \left[\hat{z}_t^2 - 1 - \sum_{j=1}^{J} \hat{g}^{(j)}\left(x_{t-1}^{(j)}\right) \right]^2 / T$, the estimates $\hat{g}_t^*$, $t = 1, \ldots, T$, are based on the undersmoothing bandwidths $h_j^* = h_j T^{-(1/20+\epsilon)}$ with ϵ "small", $j = 1, \ldots, J$, while Δ denotes a "large" positive constant. Finally, the bandwidths resulting from (18) are deflated according to $\hat{h}_j^* = \hat{h}_j T^{-(1/20+\epsilon)}$, $j = 1, \ldots, J$.

To reduce the computational burden and, in particular, allow for the use in a rolling-window forecasting application (see below), we implement the above procedure using a slightly modified version of the algorithm proposed in Mammen and Park (2005). The steps are as follows:

Step 0 Initialise $h_1, \ldots, h_J$ to the least-squares cross validation bandwidths for the univariate Nadaraya-Watson estimators $\hat{g}_{NW}^{(j)}$, $j = 1, \ldots, J$, in (13).

[5] Mammen and Park (2005) also propose plug-in procedures for bandwidth choice. They argue, however, that these are sub-optimal whenever one is interested in maximising the overall fit of the additive model instead of specific component functions.

*S*tep 1 Compute the above modified PLS criterion over a grid for the bandwidth h_j, $j = 1,\dots,J$, while holding the other bandwidths $h_1,\dots h_{j-1},h_{j+1},\dots,h_J$ constant.

*S*tep 2 Replace the starting value for h_j by the corresponding minimising grid point $\tilde{h}_j$, $j = 1,\dots,J$.

*S*tep 3 Repeat steps 1 and 2 starting from the bandwidths $\tilde{h}_1,\dots,\tilde{h}_J$. Iterate in that way until a pre-specified convergence criterion is met.

In Step 1, we consider an equidistant grid of 50 points around the initial cross-validation bandwidths.

As was pointed out in Section 2.3, inference associated with the GARCH parameter estimator $\hat{\theta}$ from (7) requires an appropriate bootstrap approach. The predominant method in a GARCH framework is the block bootstrap for the GARCH (quasi-) likelihood proposed by Goncalves and White (2004) and further refined by Corradi and Iglesias (2008). However, the latter method is based on re-sampling likelihood contributions and thus appears less suitable for our setting, in which the bootstrap is supposed to account for the impact of the pre-estimation of the seasonality and semiparametric component on the variability of the GARCH parameter estimator. Hence, we instead take a practical stance and implement a residual bootstrap in the spirit of Pascual et al. (2006) to obtain percentile confidence intervals for θ.[6] For that purpose, we *condition* on the covariate sample $\{x_t\}_{t=1}^T$. A similar type of residual bootstrap conditional on the covariates has been applied in the context of semiparametric additive models, e.g., by Fan and Jiang (2005). To summarise, our re-sampling algorithm consists of the following steps:

*S*tep 0 Fix the bandwidths at their values determined from the data according to (18), i.e. $\hat{h}_j^*$, $j = 1,\dots,J$.

*S*tep 1 Obtain a sample $\left\{\eta_t^*\right\}_{t=1}^T$ by re-sampling with replacement from the empirical distribution of the centred residuals $\hat{\eta}_t - T^{-1}\sum_{t=1}^T \hat{\eta}_t$, where $\hat{\eta}_t = r_t/\sqrt{\hat{s}_t\hat{g}_t\hat{v}_t}$ and with $\hat{s}_t$, $\hat{g}_t$ as well as $\hat{v}_t$, $t = 1,\dots,T$, being the estimates of the three model components based on the original data.

*S*tep 2 Obtain the sample $\left\{u_t^*\right\}_{t=1}^T$ from the recursion

$$v_t^* = 1 - \hat{\alpha} - \hat{\beta} + \hat{\alpha}\, u_{t-1}^{*2} + \hat{\beta}\, v_{t-1}^*, \quad u_t^* := \sqrt{v_t^*}\,\eta_t^*, \tag{19}$$

[6] A residual bootstrap has also been considered in the context of GARCH specification testing by Hidalgo and Zaffaroni (2007).

where $\hat{\alpha}$ and $\hat{\beta}$ are parameter estimates based on the original data. From this, generate the (conditional) bootstrap sample $\left\{\left(r_t^*, x_t\right)\right\}_{t=1}^T$, where $r_t^* = u_t^* \sqrt{\hat{s}_t \hat{g}_t}$, $t = 1, \ldots, T$.

*S*tep 3 Use $\left\{\left(r_t^*, x_t\right)\right\}_{t=1}^T$ to compute the GARCH parameter estimates $\hat{\theta}^*$ following the estimation steps (5), (6) and (7).

*S*tep 4 Repeat steps 1–3 B times and obtain the limits of the desired percentile intervals for α and β as the corresponding percentiles of $\left\{\hat{\alpha}_b^*\right\}_{b=1}^B$ and $\left\{\hat{\beta}_b^*\right\}_{b=1}^B$, respectively.

It is well-known that percentile intervals may exhibit a worse coverage accuracy compared to percentile t-intervals for small sample sizes (see, e.g., Inoue and Kilian, 2002). However, the latter issue should not be relevant in the given high-frequency setting with about 8000 observations. Finally, note that establishing the asymptotic validity of the above bootstrap approach in the given framework is clearly beyond the scope of this paper and deserves a study on its own. Hence, we leave it for future research.

3 Empirical Application

3.1 Data

We consider mid-prices and further LOB information for four of the most liquid stocks traded on NASDAQ: Apple (AAPL), Google (GOOG), Facebook (FB) and Microsoft (MSFT). The sample period is from January to May 2015. The LOB data is originally retrieved up to level 50 with the source being the LOBSTER database.[7] The latter reconstructs the LOB from a message stream, which is part of NASDAQ's historical TotalView-ITCH data and contains all limit order submissions, cancellations and executions on each trading day (see Huang and Polak, 2011). A crucial advantage of the resulting datasets, e.g., compared to similar ones sampled from the Trade and Quote (TAQ) database, is the fact that recording errors are virtually non-existent, such that cleaning procedures, as e.g., proposed in Brownlees and Gallo (2006) are not necessary.

We aggregate the original data, which is sampled in event time, i.e. whenever an order event changes the book on the first 50 levels, to a regular frequency of five minutes using previous-tick interpolation (e.g. Dacorogna et al., 2001). This choice of frequency is consistent, e.g., with Andersen and Bollerslev (1997) and Andersen and Bollerslev (1998). We compute mid-prices as $P_t := (A_t + B_t)/2$, $t = 1, \ldots, T$, where A_t and B_t denote the best ask and bid prices at the end of the t-th five-minute interval, respectively, and obtain the corresponding log-returns. In order to prevent having to account for overnight effects,

[7] See `https://lobsterdata.com/`.

we exclude overnight returns, which is common in the literature (see, e.g., Engle and Sokalska, 2012). The resulting daily number of observations is $N = 78$. Finally, we centre the log-returns by subtracting their sample means.

We measure the state of the LOB at the beginning of each return spell by the sum of bid and ask depth at successive LOB levels relative to the total depth in the book, i.e.,

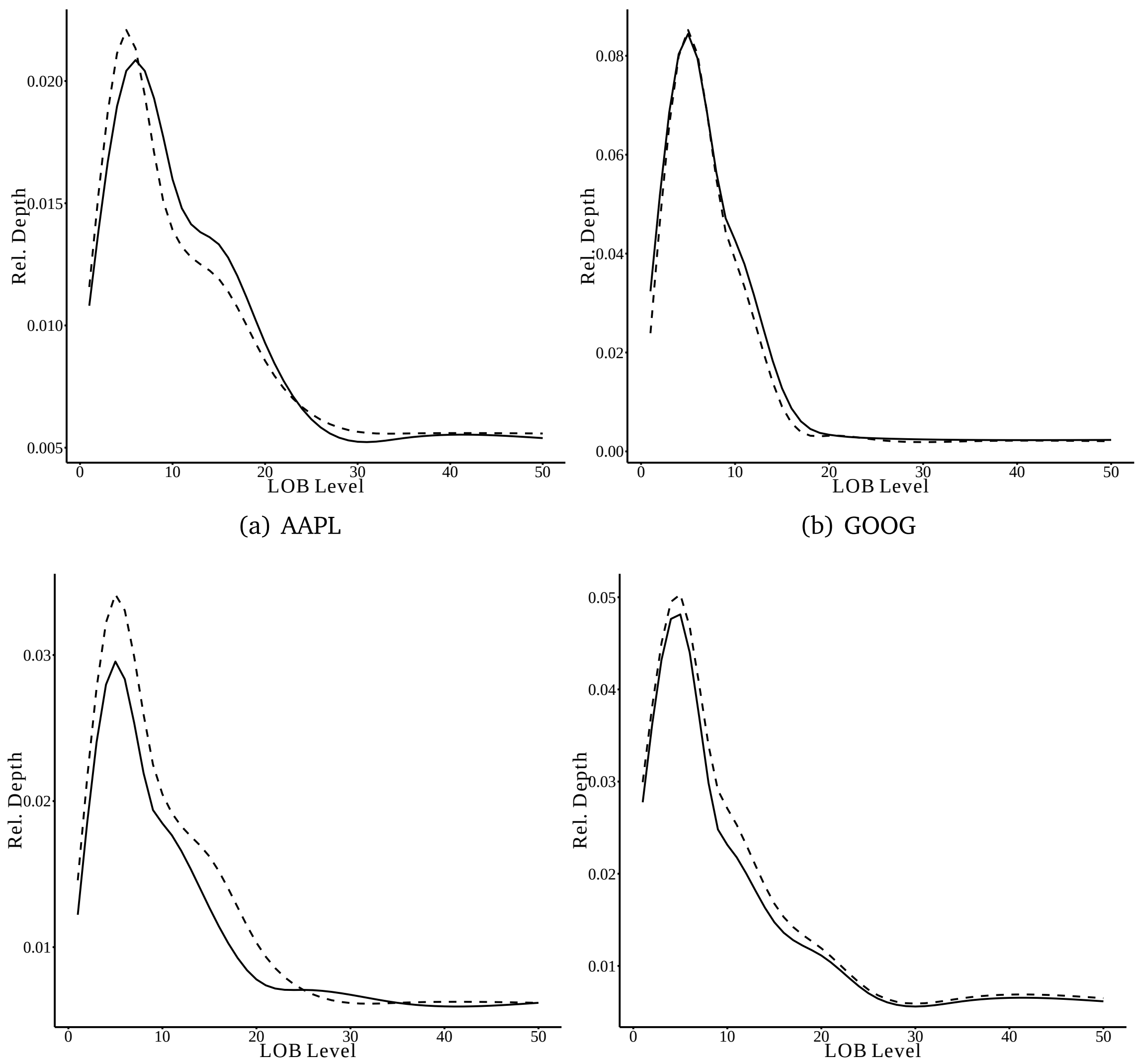
$$\text{RDEPTH}_{t-1}^{(l)} := \frac{\text{ADEPTH}_{t-1}^{(l)} + \text{BDEPTH}_{t-1}^{(l)}}{\sum_{k=1}^{L_{\max}} \left(\text{ADEPTH}_{t-1}^{(k)} + \text{BDEPTH}_{t-1}^{(k)}\right)}, \quad l = 1, \ldots, L_{\max},\ t = 1, \ldots, T, \quad (20)$$

where $\text{ADEPTH}_{t-1}^{(l)}$ and $\text{BDEPTH}_{t-1}^{(l)}$ denote the ask and bid depth on LOB level l at the beginning of the t-th spell, respectively, while $L_{\max}$ is the maximum LOB level considered. This strategy is motivated by the results in Valenzuela et al. (2015) who show that the resulting *relative* depths have a considerable predictive power regarding subsequent intraday volatility and, in particular, outperform methods based on plain depths. A further advantage of this approach is that it automatically normalises depths to the unit interval and accounts for their seasonal pattern as documented, e.g., in Ahn et al. (2001). Exemplarily, Figure 1 depicts medians of relative bid and ask depths up to level 50. In all cases, the bid and ask depth appears to be concentrated roughly on the first 25 levels. Hence, we set $L_{\max} = 25$ in order to keep the analysis parsimonious. Finally, we account for the fact that, in particular on the first levels, depth data can be rather noisy, e.g., due to "pinging" strategies for the detection of hidden depth (see, e.g., Hautsch and Huang, 2012). Hence, we average relative depth over five levels, such that $\text{RDEPTH}_{t-1}^{(1-5)} := \frac{1}{5}\sum_{l=1}^{5} \text{RDEPTH}_{t-1}^{(l)}$ with $\text{RDEPTH}_{t-1}^{(6-10)}$, $\text{RDEPTH}_{t-1}^{(11-15)}$ and $\text{RDEPTH}_{t-1}^{(16-20)}$ defined analogously for $t = 1, \ldots, T$. Note that we skip the average relative depth for levels 21 to 25 in order to mitigate possible multicollinearity issues. Hence, we consider the covariate vector $x_{t-1} = \left(\text{RDEPTH}_{t-1}^{(1-5)}, \text{RDEPTH}_{t-1}^{(6-10)}, \text{RDEPTH}_{t-1}^{(11-15)}, \text{RDEPTH}_{t-1}^{(16-20)}\right)^T$, $t = 1, \ldots, T$.

Table 1 reports some descriptive statistics for log returns as well as relative depths on levels $1-5$ and $16-20$.[8] The return distribution exhibits the well-known excess kurtosis in all four cases, while it is left-skewed for AAPL and FB and slightly right-skewed for GOOG and MSFT. Relative depth is concentrated on the first five levels as opposed to deeper in the book. At the same time, however, depth *variability* does not overly depend on the LOB level.

Ljung-Box statistics for squared log returns and relative depths on levels $1-5$ and $16-20$ are shown in Table 6 in Appendix B. In line with literature, we observe the typical volatility clustering behaviour with the Ljung-Box statistics being highly significant across

[8] Descriptive statistics for relative depth on levels $6-10$ and $11-15$ have been omitted to conserve space, but are available on request. The same holds for the Ljung-Box statistics in Table 6 in Appendix B discussed below.



(a) AAPL

(b) GOOG

(c) FB

(d) MSFT

Figure 1: Medians of Relative Bid and Ask Depths. Solid (dashed) line: median relative ask (bid) depth. Relative ask depth for level l: $\text{RADEPTH}_t^{(l)} := \frac{\text{ADEPTH}_t^{(l)}}{\sum_{k=1}^{50} \text{ADEPTH}_t^{(k)}}$, $l = 1, \ldots, 50$, $t = 1, \ldots, T$. Relative bid depth defined analogously. Figure depicts medians computed over $t = 1, \ldots, T$.

Table 1: Descriptive Statistics. Log returns (LRET) reported in percentage points. DEP_{1-5} and DEP_{16-20}: average relative depth on levels 1 to 5 and 16 to 20, respectively. AVG: sample mean. SD: standard deviation. $P_{0.05}$: 0.05-percentile. Results for (average) relative depth on levels 6 to 10 and 11 to 15 omitted due to space constraints, but available upon request.

	AAPL				GOOG			
	LRET	LRET^2	DEP_{1-5}	DEP_{16-20}	LRET	LRET^2	DEP_{1-5}	DEP_{16-20}
AVG	0.000	0.017	0.044	0.023	0.000	0.013	0.042	0.026
SD	0.131	0.047	0.017	0.013	0.116	0.034	0.013	0.009
$P_{0.05}$	-0.207	0.000	0.020	0.009	-0.183	0.000	0.024	0.014
$P_{0.95}$	0.197	0.072	0.070	0.050	0.183	0.054	0.065	0.045
SKEW	-0.441				0.016			
KURT	8.431				7.364			
#OBS	7854	7854	7854	7854	7854	7854	7854	7854

	FB				MSFT			
	LRET	LRET^2	DEP_{1-5}	DEP_{16-20}	LRET	LRET^2	DEP_{1-5}	DEP_{16-20}
AVG	0.000	0.019	0.049	0.032	0.000	0.013	0.069	0.027
SD	0.139	0.056	0.018	0.017	0.116	0.036	0.016	0.012
$P_{0.05}$	-0.214	0.000	0.020	0.012	-0.189	0.000	0.042	0.014
$P_{0.95}$	0.216	0.078	0.075	0.071	0.187	0.056	0.093	0.055
SKEW	-0.235				0.065			
KURT	9.536				8.112			
#OBS	7854	7854	7854	7854	7854	7854	7854	7854

all lags and stocks. The same is true for relative depths, while the relative persistence of depth on levels $1-5$ vs. $16-20$ depends on the given asset. Relative depth is considerably more persistent near the top of the book compared to the deeper levels for AAPL and MSFT, while the opposite holds in the case of GOOG and FB. The sample autocorrelation functions displayed in Figure 2 confirm the persistence of squared returns. Moreover, they indicate the well-known seasonality patterns in volatility.

Finally, Table 7 in Appendix B reports the results of ADF tests for unit-roots in the relative depth series. For all stocks and depth series, the null of a unit-root can be easily rejected at all conventional levels. Accordingly, the assumption of a stationary covariate process made in Section 2.3 appears to be reasonable in the given application scenario.

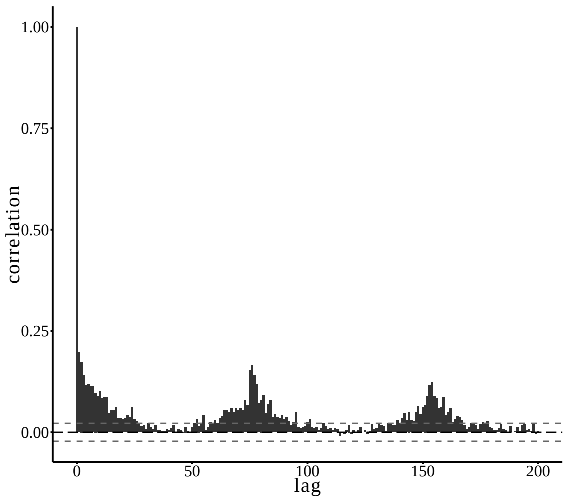
3.2 Intraday Volatility Dynamics and Order Book Depth

As the first part of the empirical study, we consider the entire sample. For that purpose, we estimate our component model following the estimation steps (5), (6) and (7) in Section 2.2. For the smooth backfitting estimators $\hat{g}^{(j)}$, $j=1,\ldots,J$, in (6), we set the "base kernel" K to be the Epanechnikov kernel. The compact support of the latter is in line with our Assumption 1(a).

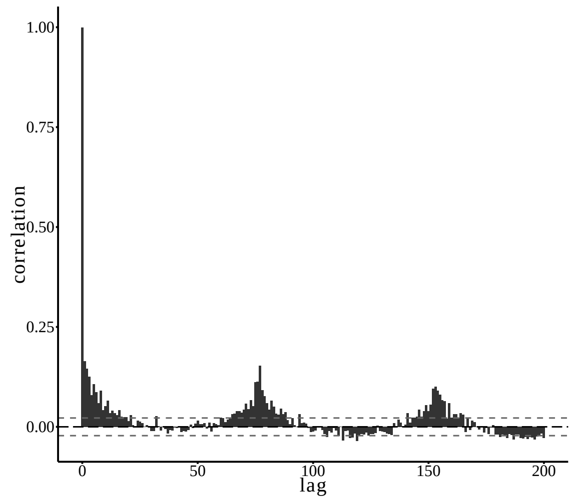
We start the discussion of the empirical results with the estimates of the seasonality component. The latter are depicted in Figure 3, indicating that the seasonality component exhibits the L-shape that is typical for equities, thus confirming the results, e.g., in Engle and Sokalska (2012). Accordingly, volatility is higher at the opening, lower around midday and then increases somewhat before the close of trading.

We can proceed with the semiparametric component. Table 2 reports the bandwidths for the smooth backfitting estimators that were obtained by the modified PLS approach discussed in Section 2.4. Most importantly, all bandwidths have a comparable level of magnitude, implying that none of the covariates is being "smoothed out". Figures 4 and 5 display the resulting estimates of the component functions together with approximative 95% confidence intervals constructed following Theorem 2. The latter require estimates of the variances $\sigma^2_{z,j}\big(x^{(j)}\big)$, $j=1\ldots,J$, which are obtained by a Nadaraya-Watson regression of the squared backfitting residuals from (6) on $x^{(j)}$ using least-squares cross-validation bandwidths. To focus on the difference in terms of the impact on subsequent volatility between depth near the top and deeper in the book, only the estimates corresponding to relative depths on LOB levels $1-5$, $\hat{g}_1$, and $16-20$, $\hat{g}_4$, are reported.[9]

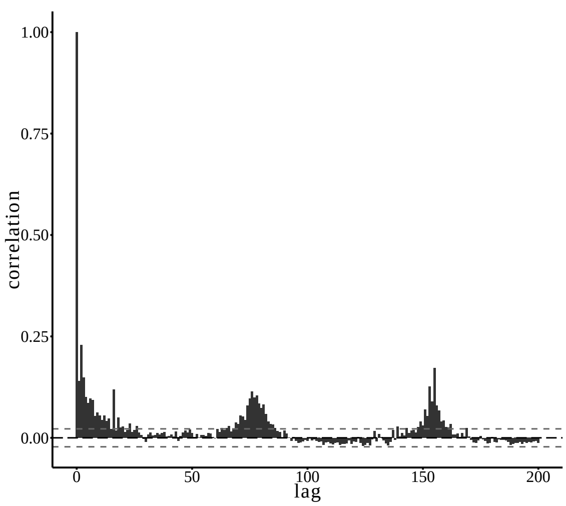
[9]The estimates corresponding to relative depths on levels $6-10$, $\hat{g}_2$, and $11-15$, $\hat{g}_3$, are available upon request.



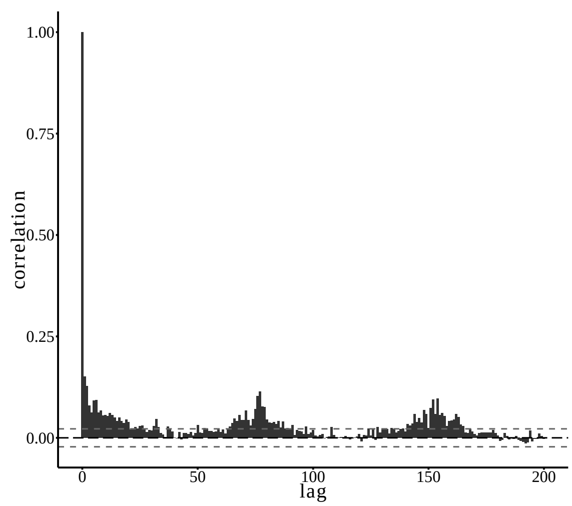
(a) AAPL



(b) GOOG

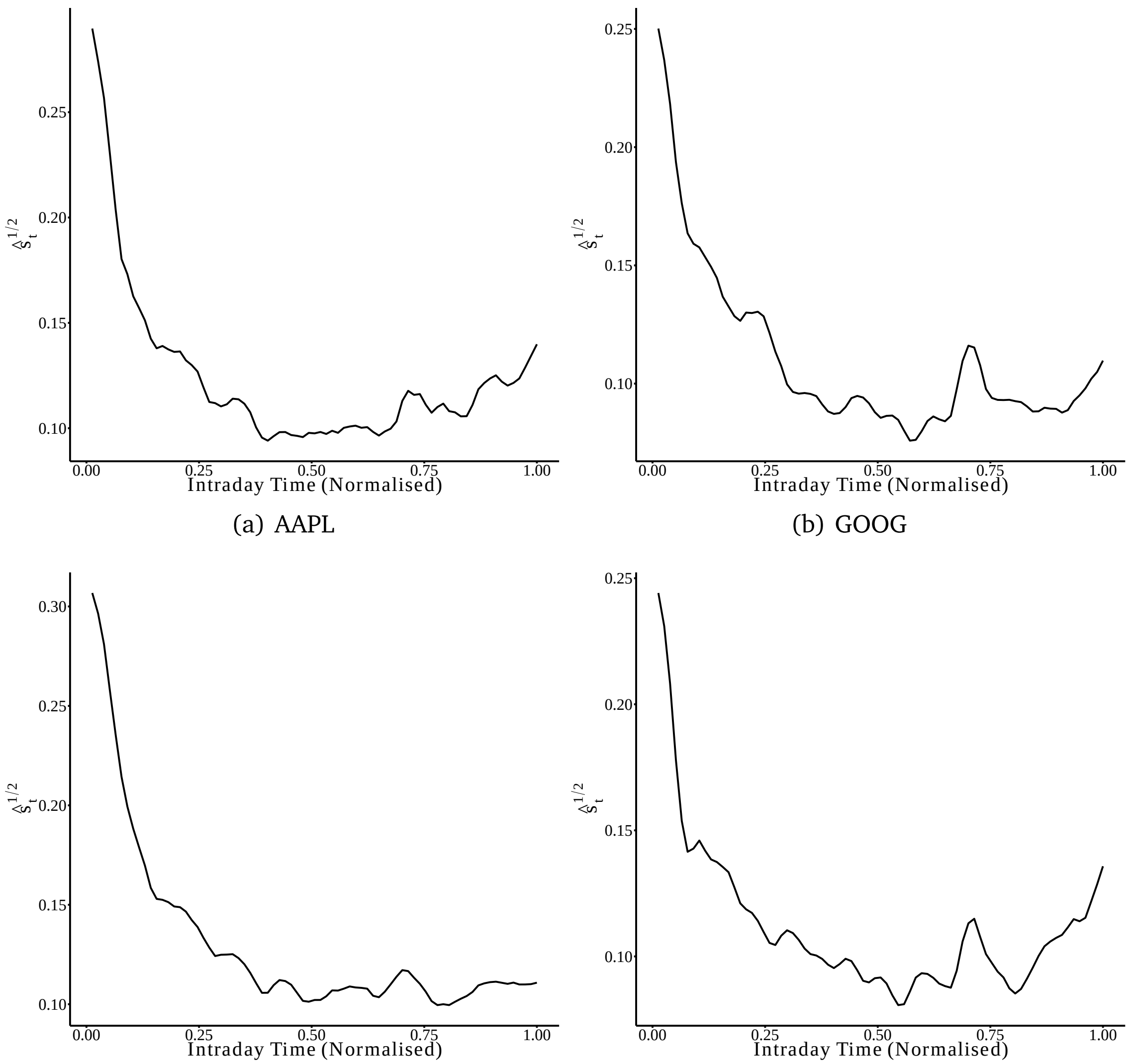


(c) FB



(d) MSFT

Figure 2: Sample Autocorrelation Functions of Squared Log Returns. Horizontal lines: limits of 95% confidence intervals ($\mp 1.96/\sqrt{T}$).



(a) AAPL

(b) GOOG

(c) FB

(d) MSFT

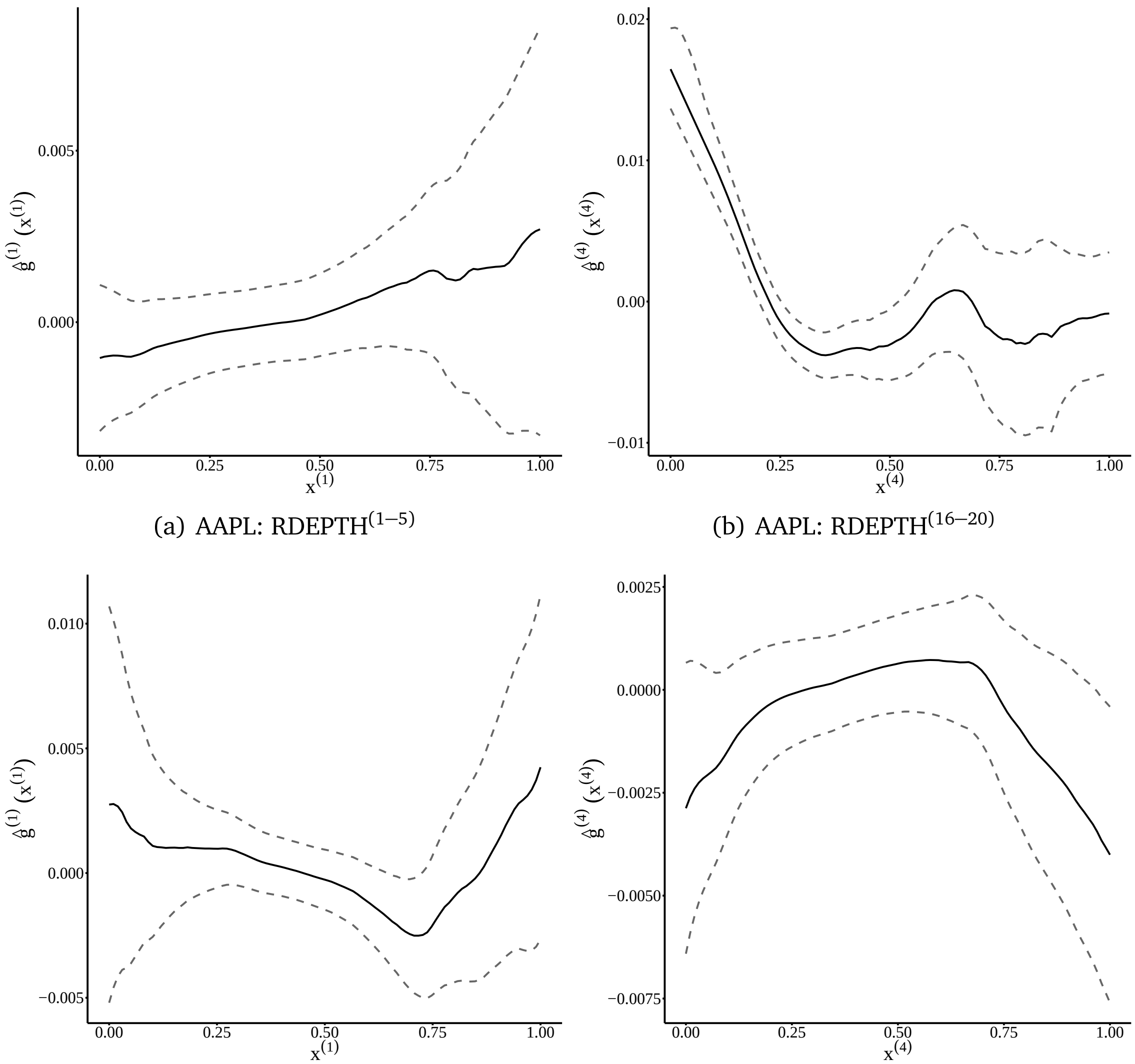
Figure 3: Estimates of Seasonality Component s_t

Table 2: Bandwidths for SBF Estimator of Additive Component. Bandwidths determined by minimising modified PLS criterion (18).

	AAPL	GOOG	FB	MSFT
RDEPTH$^{(1-5)}$	0.354	0.586	0.692	0.193
RDEPTH$^{(6-10)}$	0.308	0.399	0.162	0.179
RDEPTH$^{(11-15)}$	0.403	0.284	0.299	0.465
RDEPTH$^{(16-20)}$	0.241	0.588	0.695	0.242

The main findings are as follows. First, the functional form is highly non-linear in all cases but one. That is, only in the case of $\hat{g}_1$ for AAPL, the confidence region associated with the fit would allow for a linear relationship, underlining the importance of a flexible semiparametric approach for uncovering the impact of relative order book depths on intraday volatility. Second, for all four stocks, there is a pronounced difference between the shape of $\hat{g}_1$ and $\hat{g}_4$, suggesting that the depth-volatility relationship crucially depends on the given LOB level. E.g. for AAPL, increasing relative depth on levels $1-5$ has a positive effect on subsequent volatility. However, the impact is significantly negative when analysing levels $16-20$ as $\hat{g}_4$ first decreases sharply and flattens out eventually. This result is in line with, e.g., Pascual and Veredas (2010) and might be explained by a reduction of the sizes of less aggressive limit orders when an increase in volatility is expected. In the case of the remaining three stocks, the relationships are more complex. $\hat{g}_1$ is mostly convex, being predominantly decreasing (increasing) for lower (higher) relative depths. Hence, the results of Ahn et al. (2001) for depth at the top of the book can be confirmed only for smaller values. Interestingly, in the case of GOOG and MSFT, $\hat{g}_4$ is roughly inverted and thus concave. Up to intermediate levels of relative depth, this finding is in line with Valenzuela et al. (2015) and could be caused by market participants interpreting a concentration of depth deeper in the book as an indication of mispricing, which in turn leads to price adjustments resulting in a higher volatility.

Estimates of the parameters of the GARCH component are reported in Table 3. Unsurprisingly, the parameter estimates are well in line with literature (see, e.g., Andersen and Bollerslev, 1997). They translate into half-lives for a volatility shock from around 36 five-minute intervals, i.e. three hours, for GOOG to 73 intervals or roughly six hours for MSFT. More interestingly, Table 3 also reports percentile confidence intervals computed using the bootstrap algorithm introduced in Section 2.4 with $B = 5000$ as well as corresponding "naive" asymptotic confidence intervals based on Bollerslev and Wooldridge (1992) standard errors. For all four stocks and both parameters, the percentile intervals are markedly wider than their naive counterparts. This result underlines the importance of a



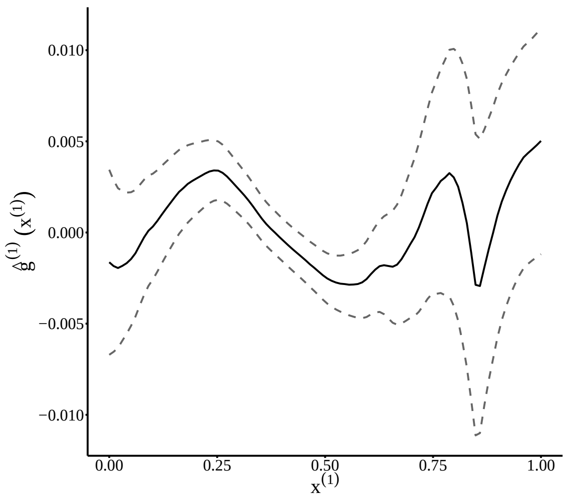
(a) AAPL: RDEPTH$^{(1-5)}$

(b) AAPL: RDEPTH$^{(16-20)}$

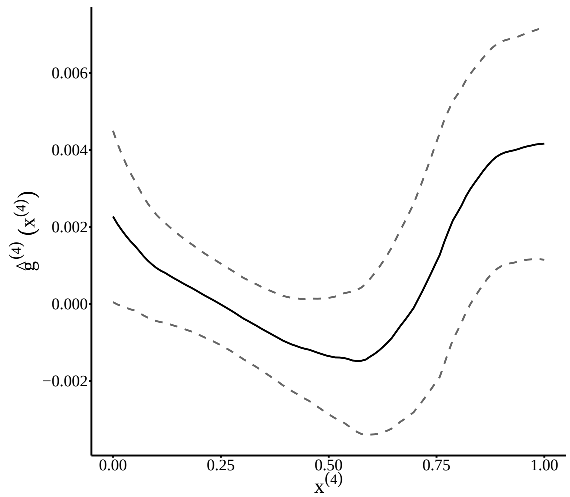
(c) GOOG: RDEPTH$^{(1-5)}$

(d) GOOG: RDEPTH$^{(16-20)}$

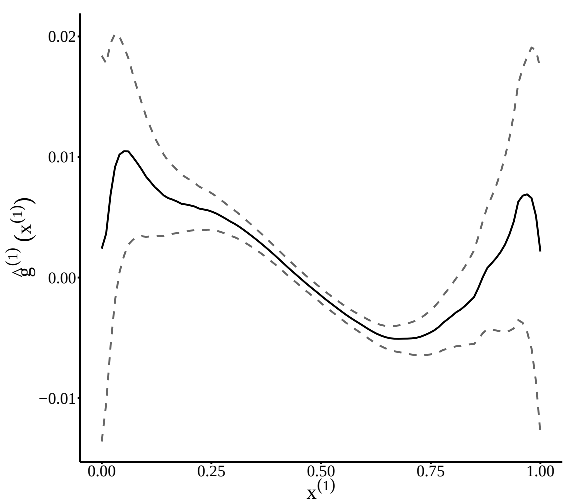
Figure 4: Estimates of Additive Component Functions $g^{(1)}$ and $g^{(4)}$ (AAPL & GOOG). Dashed lines: approximative point-wise 95% confidence intervals according to Theorem 2. Estimates of variances $\sigma^2_{z,j}\left(x^{(j)}\right)$, $j = 1 \ldots, J$, obtained by Nadaraya-Watson regression of squared backfitting residuals from (6) on $x^{(j)}$ using least-squares cross-validation bandwidths. Estimates of $g^{(2)}$ and $g^{(3)}$ available upon request.



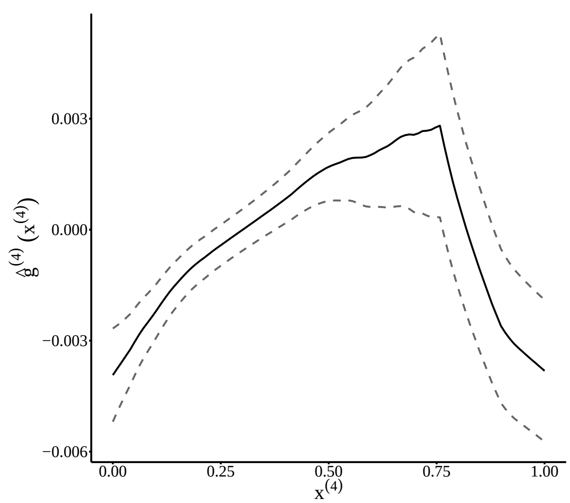
(a) FB: $\text{RDEPTH}^{(1-5)}$



(b) FB: $\text{RDEPTH}^{(16-20)}$



(c) MSFT: $\text{RDEPTH}^{(1-5)}$



(d) MSFT: $\text{RDEPTH}^{(16-20)}$

Figure 5: Estimates of Additive Component Functions $g^{(1)}$ and $g^{(4)}$ (FB & MSFT). Dashed lines: approximative point-wise 95% confidence intervals according to Theorem 2. Estimates of variances $\sigma^2_{z,j}\big(x^{(j)}\big)$, $j = 1 \ldots, J$, obtained by Nadaraya-Watson regression of squared backfitting residuals from (6) on $x^{(j)}$ using least-squares cross-validation bandwidths. Estimates of $g^{(2)}$ and $g^{(3)}$ available upon request.

Table 3: Estimation Results for GARCH Component. CI_L^b and CI_U^b: lower and upper bound of bootstrapped confidence interval with $B = 5000$. CI_L^a and CI_U^a: lower and upper bound of "naive" asymptotic confidence interval based on Bollerslev and Wooldridge (1992) standard errors. $\widehat{\mathrm{HL}}$: half-life in five-minute intervals.

	AAPL		GOOG		FB		MSFT	
	α	β	α	β	α	β	α	β
Est.	0.067	0.919	0.044	0.937	0.049	0.935	0.042	0.948
CI_L^b	0.054	0.902	0.033	0.916	0.037	0.920	0.033	0.936
CI_U^b	0.082	0.935	0.056	0.953	0.058	0.949	0.051	0.959
CI_L^a	0.066	0.917	0.042	0.932	0.046	0.931	0.041	0.946
CI_U^a	0.069	0.922	0.047	0.941	0.051	0.938	0.044	0.951
$\widehat{\mathrm{HL}}$	51.677		35.993		41.433		72.672	

Table 4: Ljung-Box Statistics of Residuals. $\hat{z}_t^2$, $\hat{u}_t^2$ and $\hat{\eta}_t^2$: squared returns successively standardised by estimated seasonality ($\hat{s}_t$), additive ($\hat{g}_t$) and GARCH component ($\hat{v}_t$). LB_l: Ljung-Box statistic associated with l lags. 5% (1%) critical values associated with lag lengths 30, 100 and 200: 43.78 (50.89), 124.34 (135.81) and 233.99 (249.45), respectively.

	AAPL			GOOG		
	$\hat{z}_t^2$	$\hat{u}_t^2$	$\hat{\eta}_t^2$	$\hat{z}_t^2$	$\hat{u}_t^2$	$\hat{\eta}_t^2$
LB_{30}	3869.650	2519.534	42.454	988.371	754.868	35.475
LB_{100}	5983.760	3826.552	96.437	1750.869	1344.412	91.134
LB_{200}	7748.654	4914.765	198.721	1940.818	1492.775	157.896
	FB			MSFT		
	$\hat{z}_t^2$	$\hat{u}_t^2$	$\hat{\eta}_t^2$	$\hat{z}_t^2$	$\hat{u}_t^2$	$\hat{\eta}_t^2$
LB_{30}	2701.177	1959.236	28.819	1464.540	1096.649	26.279
LB_{100}	4241.013	3085.410	107.322	2459.489	1929.376	108.898
LB_{200}	4486.168	3255.691	184.377	3308.115	2643.356	190.624

re-sampling approach and shows that relying on standard asymptotic results while ignoring the implications of Theorem 4 could lead to incorrect inference.

Table 4 shows how much of the variability in squared returns the different components can explain. For that purpose, Ljung-Box statistics for squared returns standardised by the estimated seasonality component, additionally the semiparametric component, and, finally, also the GARCH component are reported. For all stocks except MSFT, deseasonalised squared returns are more persistent than their raw counterparts, which is a common effect in equity data (see, e.g., Andersen and Bollerslev, 1997). Standardisation by the semiparametric component implies a noticeable reduction in the magnitude of the Ljung-Box statistics. Finally, removing the GARCH component eliminates the persistence with all statistics being insignificant at the 5% level.

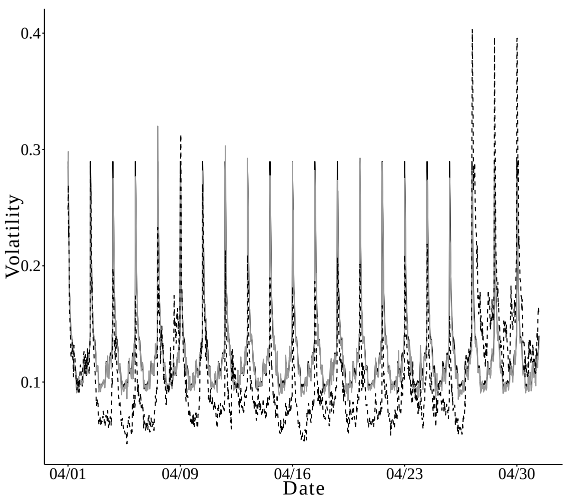
Finally, we can briefly examine the evolution of the three volatility components. Figure 6 depicts the time series of the estimates of the seasonality component, $\hat{s}_t$, seasonality including the semiparametric component, $\hat{s}_t\,\hat{g}_t$, as well as the product of all three components, $\hat{s}_t\,\hat{g}_t\,\hat{v}_t$, for April. Evidently, the dominating feature is the intraday pattern captured by $\hat{s}_t$. At the same time, the product $\hat{s}_t\,\hat{g}_t$ tracks the latter quite closely. However, the component $\hat{v}_t$ introduces noticeable movements around the volatility level set by the previous two components. To demonstrate the behaviour on particular days by means of an example, Figure 7 additionally displays the time series of $\hat{g}_t$ as well as the relative depth on levels $1-5$ and $16-20$ for AAPL in the last trading week of April. On April 27, Apple Inc. made a scheduled earnings announcement at 5 pm EST, i.e. after the close of trading at NASDAQ. Figure 7 shows that, on this day, relative depth was concentrated deeper in the book as opposed to the first five levels, suggesting that rather passive orders were being submitted prior to the announcement. Due to the estimates of the component functions g_1 and g_4 being predominantly increasing and decreasing, respectively, as reported in Figure 4, the aforementioned effect translates into a generally lower level of the component estimate $\hat{g}_t$ on April 27.

3.3 Out-of-Sample Forecasting Study

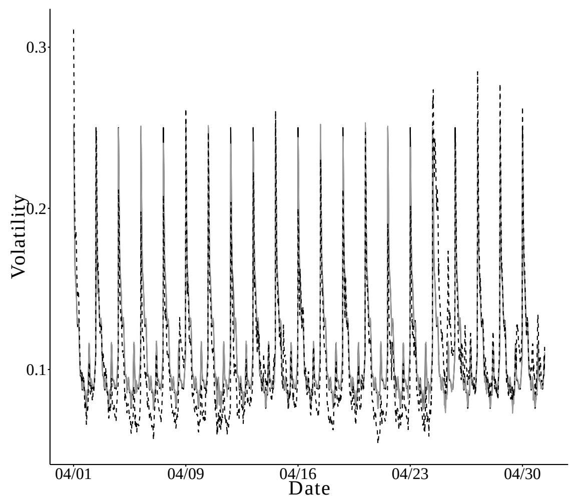
To assess the usefulness of our model for the prediction of intraday volatility, we conduct a simple out-of-sample forecasting study. For that purpose, we consider rolling estimation windows of one month, i.e. 1560 five-minute intervals, and compute one-step ahead volatility forecasts. A similar study based on rolling windows of two months yields qualitatively similar results. The latter are available upon request.

The estimation of the model components in the rolling window setting deviates in two ways from the approach outlined in Section 2.4. That is, when selecting the bandwidths for the smooth backfitting estimators of the component functions $g^{(j)}$, $j = 1, \ldots, J$, by minimising the criterion in (18), we use the estimated bandwidths from the previous window as starting values, which reduces computation time considerably. Second, since we only need point estimates of the GARCH parameters θ in a forecasting application, we omit the undersmoothing steps in the above bandwidth selection algorithm.

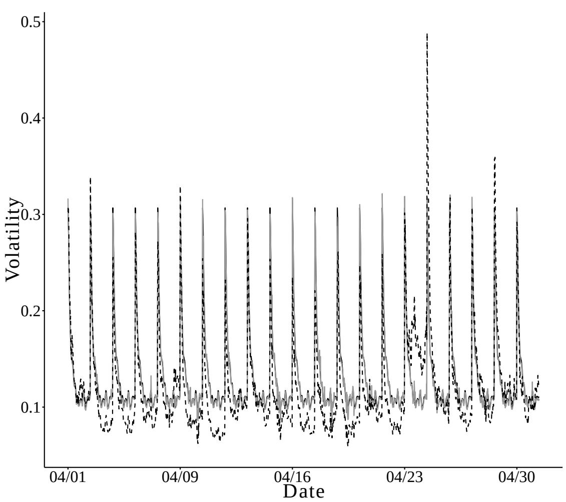
In our framework, the construction of one-step ahead forecasts of conditional variances $\varsigma^2_{t+1} := \mathbb{E}\big[r^2_{t+1}|\mathscr{H}_t\big]$, with the filtration $\mathscr{H}_t$ defined as in Section 2.1, $t = 1, \ldots, T-1$, consists of the following steps. First, we predict the seasonality component by selecting the estimate $\hat{s}_i$, $i = 1, \ldots, 78$, corresponding to the intraday return spell $t+1$, $\hat{s}_{t+1|t}$, $t = 1, \ldots, T-1$. Second, we employ the covariate vector at the beginning of the spell $t+1$, x_t, to compute the forecast of the semiparametric component according to (6), $\hat{g}_{t+1|t}$, $t = 1, \ldots, T-1$,



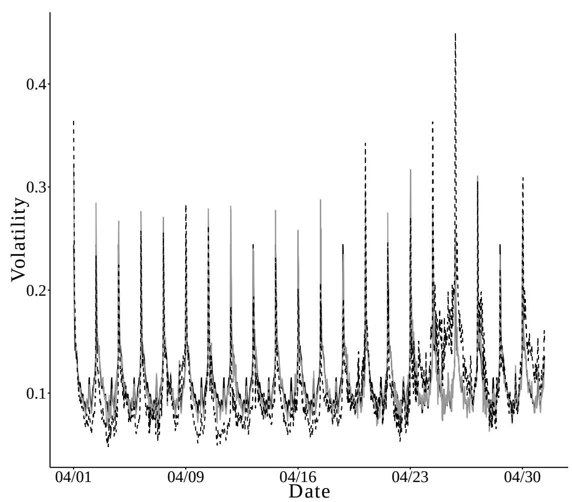
(a) AAPL



(b) GOOG

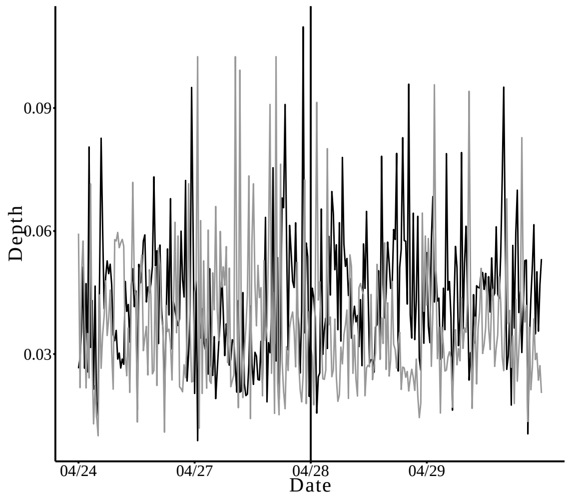


(c) FB

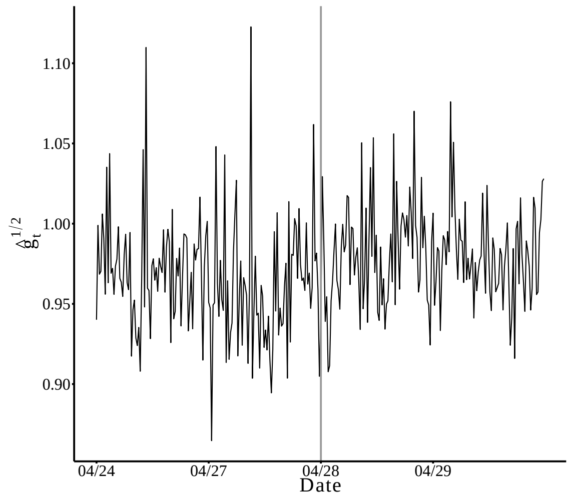


(d) MSFT

Figure 6: Time Series of Volatility Component Estimates (April). Solid black line: $\sqrt{\hat{s}_t}$. Solid grey line: $\sqrt{\hat{s}_t\ \hat{g}_t}$. Dashed black line: $\sqrt{\hat{s}_t\ \hat{g}_t\ \hat{v}_t}$. Volatility reported in percentage points.



(a) RDEPTH$^{(1-5)}$ and RDEPTH$^{(16-20)}$



(b) $\sqrt{\hat{g}_t}$

Figure 7: Relative Depths and Semiparametric Component $\hat{g}_t$ for AAPL in last trading week of April. Black (grey) line in left plot: RDEPTH$^{(1-5)}$ (RDEPTH$^{(16-20)}$). $\sqrt{g_t}$ in right plot reported in percentage points. Vertical lines indicate end of trading on April 27, 2015.

using the fitted functions $\hat{g}^{(j)}$, $j = 1, \ldots, J$. Finally, we compute the forecast of the GARCH component following the dynamics (4) using the parameter estimates $\hat{\theta}$ and the series of estimated standardised returns $\left\{\hat{u}_t^*\right\}$, yielding $\hat{v}_{t+1|t}$, $t = 1, \ldots, T-1$. The final forecast of ς_{t+1}^2 is thus obtained according to $\hat{\varsigma}_{t+1|t}^2 = \hat{s}_{t+1|t}\, \hat{g}_{t+1|t}\, \hat{v}_{t+1|t}$, $t = 1, \ldots, T-1$.

We compare the performance of our model to the following benchmarks. First, we consider a variant of the intraday component GARCH by Engle and Sokalska (2012), which is equivalent to our framework, but neglects the semiparametric component, i.e.

$$r_t = \sqrt{s_t\, v_t^{(1)}}\, \eta_t^{(1)}, \quad v_t^{(1)} = 1 - \alpha - \beta + \alpha \frac{r_{t-1}^2}{s_{t-1}} + \beta v_{t-1}^{(1)}, \quad t = 1, \ldots, T, \tag{21}$$

where the disturbances $\eta_t^{(1)}$ are defined analogously to η_t, $t = 1, \ldots, T$, in Section 2.1. The seasonality component s_t and parameters of the GARCH component $v_t^{(1)}$ are estimated as in (5) and (7), respectively, replacing u_t by z_t, $t = 1, \ldots, T$, in the latter case. One-step ahead forecasts of these two components are computed analogously to above, implying the combined forecast $\hat{\varsigma}_{t+1|t}^{(1),2} = \hat{s}_{t+1|t}\, \hat{v}_{t+1|t}^{(1)}$, $t = 1, \ldots, T-1$. Note that, unlike Engle and Sokalska (2012), we do not additionally include a separate *daily* volatility component (e.g. a daily GARCH) as the latter deteriorates the forecasting performance in our application.

As a second benchmark, we implement an intraday version of the trend GARCH model proposed by Hafner and Linton (2010), which compared to the previous approach, introduces an additional long-term trend component, i.e.,

$$r_t = \sqrt{s_t \, \tau_t \, v_t^{(2)}} \, \eta_t^{(2)}, \qquad \tau_t = \tau(t/T), \tag{22}$$
$$v_t^{(2)} = 1 - \alpha - \beta + \alpha \frac{r_{t-1}^2}{s_{t-1}\tau_{t-1}} + \beta v_{t-1}^{(2)}, \qquad t = 1, \ldots, T.$$

To estimate the trend component τ_t and the GARCH parameters in $v_t^{(2)}$, we consider the first estimation step of the algorithm proposed in Hafner and Linton (2010) as a direct implementation of their additional efficient step in the given intraday setting turns out to be harmful in terms of forecasting precision. Accordingly, τ_t is estimated by the Nadaraya-Watson regression

$$\hat{\tau}_{NW}(w) = \left\{ \sum_{t=1}^{T} K\left[h^{-1}(w - t/T)\right] \right\}^{-1} \sum_{t=1}^{T} K\left[h^{-1}(w - t/T)\right] \frac{r_t^2}{\hat{s}_t}, \tag{23}$$

where, following Hafner and Linton (2010), we choose the quartic kernel for K and set the bandwidth as $h = 0.05$. The GARCH parameters are then estimated analogously to (7), but with the estimate of the semiparametric component, $\hat{g}_t$, replaced by $\hat{\tau}_t := \hat{\tau}_{NW}(t/T)$. Forecasts of the seasonality and GARCH component are computed as before, while for the trend component, we proceed as in Section 7.1 of Hafner and Linton (2010), setting $\hat{\tau}_{t+1|t} = \hat{\tau}_{NW}(1)$ to obtain $\hat{\varsigma}_{t+1|t}^{(2),2} = \hat{s}_{t+1|t} \, \hat{\tau}_{t+1|t} \, \hat{v}_{t+1|t}^{(2)}$, $t = 1, \ldots, T-1$.

Our final benchmark includes the covariates x_{t-1}, $t = 1, \ldots, T$, into the GARCH dynamics on an additive basis. Hence, the resulting intraday GARCH-X model can be obtained as an extension of the first benchmark above to

$$r_t = \sqrt{s_t \, v_t^{(3)}} \, \eta_t^{(3)}, \quad v_t^{(3)} = \omega + \alpha \frac{r_{t-1}^2}{s_{t-1}} + \beta v_{t-1}^{(3)} + \delta^T x_{t-1}, \quad t = 1, \ldots, T, \tag{24}$$

where, due to the presence of the covariates, we refrain from imposing the unit-mean constraint on the GARCH component $v_t^{(3)}$ directly. The extended parameter vector of the latter, $\theta_x := (\omega, \alpha, \beta, \delta)^T$, is estimated as before, i.e. by adapting estimation step (7). The computation of the forecast $\hat{v}_{t+1|t}^{(3)}$ now additionally requires the covariates at the beginning of the $(t+1)$-th spell, x_t, and yields $\hat{\varsigma}_{t+1|t}^{(3),2} = \hat{s}_{t+1|t} \, \hat{v}_{t+1|t}^{(3)}$, $t = 1, \ldots, T-1$.

To evaluate the forecasting performance of our model against the benchmarks, we require a proxy for the conditional variances ς_t^2, $t = 1, \ldots, T$. In order to maximise the efficiency of the latter, we consider quote data corresponding to our LOB sample, which is updated whenever the first level is affected by an order event, and obtain the resulting mid-prices. For each return spell, we then compute a realised variance measure that is robust to the market microstructure noise which dilutes the underlying efficient price process and makes plain realised variance estimates generally biased. On overview of available methods is provided in Andersen et al. (2008). In our *ultra-high frequency* setting, we additionally need to account for the fact that the noise process my exhibit serial dependence (see, e.g., Hansen and Lunde, 2006). Accordingly, we employ an extension of the popular pre-averaging method by Jacod et al. (2009) that was proposed by Hautsch and Podolskij (2013) and is robust to the aforementioned effect, while achieving rate optimality.[10] Following Hautsch and Podolskij (2013), we set the proportionality parameter which determines the length of the required moving windows to 0.4 and estimate the long-run variance of the noise process according to the approach they suggest in Section 3 of their paper.[11] Finally, based on the resulting pre-averaging estimates for each return spell, $\tilde{\varsigma}_t^2$, $t = 1, \ldots, T$, we compute losses according to the QLIKE loss function proposed by Patton (2011) as this loss function allows for a consistent ranking of conditional variance forecasts when an imperfect proxy for the unobserved conditional variance is used for evaluation purposes.

Table 5 reports the average losses for all four competing models. In addition, results of two forecast evaluation tests are provided. First, consistent p-values of the test for superior predictive ability (SPA) proposed by Hansen (2005), which are based on the stationary bootstrap by Politis and Romano (1994) with 5000 replications. The null hypothesis is that none of the competitors yields a lower expected loss than the proposed model. Second, p-values for the proposed model implied by the model confidence set (MCS) approach proposed by Hansen et al. (2011) employing the same bootstrap setting. A $(1-\alpha)$-MCS contains the set of models, which are "best" according to a given criterion, with probability no less than $1-\alpha$. If a model has MCS p-value $\hat{p}$, it is contained in all MCS for which $\alpha \leq \hat{p}$. For completeness, Table 5 also reports p-values of the *pair-wise* Diebold and Mariano (1995) and West (1996) test for equal expected loss of the forecasts based on the proposed model vs. each competitor. The major finding is that the proposed model is not significantly outperformed according to the SPA test in all cases but one. In these cases, it yields the lowest average loss, while being contained even in the finest MCS. Only for MSFT, the null hypothesis that all competitors imply an equal or higher expected loss can be rejected at conventional levels, while the model is contained in MCS with confidence level 0.985 and higher only. However, the proposed model still yields a considerably lower average loss than the (additive) intraday GARCH-X model. As the latter benchmark is associated with

[10] The main idea behind the estimator is to obtain weighted moving averages of the returns and, subsequently, compute a suitably bias-corrected and re-scaled sum of squares of the latter. See Jacod et al. (2009) and Hautsch and Podolskij (2013) for details.

[11] Somewhat smaller or larger values of the proportionality parameter yield qualitatively similar forecast evaluation results.

Table 5: Out-of-Sample Forecasting Results. Based on rolling estimation windows of one month (1560 five-minute intervals). Intra-C-GARCH: intraday component GARCH from (21). Intra-T-GARCH: intraday trend GARCH from (22). Intra-GARCH-X: intraday additive GARCH-X from (24). $\bar{L}$: average loss based on QLIKE loss function by Patton (2011) with conditional variances proxied by pre-averaging estimates following Hautsch and Podolskij (2013). DMW: p-value of two-sided Diebold and Mariano (1995) and West (1996) test vs. proposed model. SPA_c: consistent p-values of test for superior predictive ability by Hansen (2005) based on stationary bootstrap with 5000 replications. H_0: proposed model is not outperformed by any competitor. MCS_{Max}: model confidence set p-value of proposed model following Hansen et al. (2011) based on T_{Max} statistic and stationary bootstrap with 5000 replications.

	AAPL		GOOG		FB		MSFT	
	$\bar{L}$	DMW	$\bar{L}$	DMW	$\bar{L}$	DMW	$\bar{L}$	DMW
Proposed Model	0.184		0.312		0.184		0.417	
Intra-C-GARCH	0.234	0.000	0.357	0.000	0.202	0.024	0.404	0.392
Intra-T-GARCH	0.229	0.000	0.390	0.000	0.225	0.000	0.384	0.002
Intra-GARCH-X	0.247	0.000	0.427	0.000	0.261	0.000	0.541	0.000
SPA_c	1.000		1.000		1.000		0.004	
MCS_{Max}	1.000		1.000		1.000		0.015	

the highest average loss for all stocks, including relative depth covariates into the GARCH equation in an additive manner appears to be rather disadvantageous in terms of forecasting performance. In contrast, the above results suggest that augmenting an intraday GARCH framework multiplicatively by means of a semiparametric component to capture LOB depth information may improve out-of-sample forecasting precision.

4 Conclusion

In this paper, we propose a semiparametric component model for the conditional volatility of intraday returns. The model consists of a diurnal factor, accounting for the intraday seasonality of volatility, a semiparametric additive component, allowing for the impact of many covariates, e.g., originating from the LOB, as well as unit-GARCH dynamics, which represent fluctuations of volatility around the level set by the seasonality and semiparametric component. The estimators of all three components are easy-to-implement in practice. We provide a comprehensive asymptotic theory for these estimators, including uniform convergence rates and asymptotic distributions.

In an empirical study, we apply the above methods to blue chips traded on NASDAQ. Our results underline the practical relevance of a semiparametric model for intraday volatility. First, we find that the latter may depend on LOB depth in a highly non-linear fashion.

Second, our model significantly outperforms all relevant benchmarks in an out-of-sample forecasting exercise.

Possible extensions and applications of the framework we introduced are manifold. For example, one could study the case of non-stationary covariates entering the model trough the semiparametric component. Moreover, a test for the functional form of the impact of particular covariates on subsequent conditional return volatility might be of interest.

References

AHN, H.-J., K.-H. BAE, AND K. CHAN (2001): "Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong," *The Journal of Finance*, 56, 767–788.

ANDERSEN, T. G. AND T. BOLLERSLEV (1997): "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance*, 4, 115–158.

——— (1998): "Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies," *Journal of Finance*, 53, 219–265.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. DIEBOLD (2008): "Parametric and Nonparametric Measurement of Volatility," in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia and L. P. Hansen, Amsterdam: North Holland.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307 – 327.

BOLLERSLEV, T. AND J. M. WOOLDRIDGE (1992): "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric Reviews*, 11, 143–172.

BRADLEY, R. C. (1983): "Approximation theorems for strongly mixing random variables." *Michigan Math. J.*, 30, 69–81.

BROWNLEES, C. AND G. GALLO (2006): "Financial econometric analysis at ultra-high frequency: Data handling concerns," *Computational Statistics & Data Analysis*, 51, 2232 – 2245.

BUJA, A., T. HASTIE, AND R. TIBSHIRANI (1989): "Linear smoothers and additive models (with discussion)," *Ann. of Statist.*, 17, 453–510.

CARRASCO, M. AND X. CHEN (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 17–39.

CONRAD, C. AND K. LOCH (2015): "Anticipating Long-Term Stock Market Volatility," *Journal of Applied Econometrics*, 30, 1090–1114.

COPPEJANS, M., I. DOMOWITZ, AND A. MADHAVAN (2001): "Liquidity in an Automated Auction," .

CORRADI, V. AND E. M. IGLESIAS (2008): "Bootstrap refinements for {QML} estimators of the GARCH(1,1) parameters," *Journal of Econometrics*, 144, 500 – 510.

DACOROGNA, M., R. GENCAY, U. MULLER, R. OLSEN, AND O. PICTET (2001): *An Intoduction to High-Frequency Finance*, San Diego, CA: Academic Press.

DIEBOLD, F. X. AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

DROST, F. C. AND T. E. NIJMAN (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.

ENGLE, R., E. GHYSELS, AND B. SOHN (2013): "Stock market volatility and macroeconomic fundamentals," *Review of Economics and Statistics*, 95, 776–797.

ENGLE, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

ENGLE, R. F. AND J. G. RANGEL (2008): "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187–1222.

ENGLE, R. F. AND M. E. SOKALSKA (2012): "Forecasting intraday volatility in the US equity market. Multiplicative component GARCH," *Journal of Financial Econometrics*, 10, 54–83.

FAN, J., W. HARDLE, AND E. MAMMEN (1998): "Direct Estimation of Low-Dimensional Components in Additive Models," *The Annals of Statistics*, 26, 943–971.

FAN, J. AND J. JIANG (2005): "Nonparametric Inferences for Additive Models," *Journal of the American Statistical Association*, 100, 890–907.

FENGLER, M., E. MAMMEN, AND M. VOGT (2015): "Specification and structural break tests for additive models with applications to realized variance data," *Journal of Econometrics*, 188, 196–218.

FRANCQ, C. AND J.-M. ZAKOIAN (2004): "Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes," *Bernoulli*, 10, 605–637.

——— (2010): *GARCH Models*, Chichester: John Wiley & Sons Ltd.

GIOT, P. (2005): "Market risk models for intraday data," *The European Journal of Finance*, 11, 309–324.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): " On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

GONCALVES, S. AND H. WHITE (2004): "Maximum likelihood and the bootstrap for nonlinear dynamic models," *Journal of Econometrics*, 119, 199 – 219.

HAFNER, C. M. AND O. LINTON (2010): "Efficient estimation of a multivariate multiplicative volatility model," *Journal of Econometrics*, 159, 55 – 73.

HALL, P. AND C. HEYDE (1980): *Martingale Limit Theory and its Application*, Probability and Mathematical Statistics, New York: Academic Press.

HAN, H. AND D. KRISTENSEN (2015): "Semiparametric Multiplicative GARCH-X Model: Adopting Economics Variables to Explain Volatility," Tech. rep., Sungkyunkwan University.

HANSEN, P. R. (2005): "A Test for Superior Predictive Ability," *Journal of Business & Economic Statistics*, 23, 365–380.

HANSEN, P. R. AND A. LUNDE (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business & Economic Statistics*, 24, 127–161.

HANSEN, P. R., A. LUNDE, AND J. M. NASON (2011): "The Model Confidence Set," *Econometrica*, 79, 453–497.

HASTIE, T. J. AND R. J. TIBSHIRANI (1990): *Generalized Additive Models*, Chapman and Hall, London.

HAUTSCH, N. AND R. HUANG (2012): "On the Dark Side of the Market: Identifying and Analyzing Hidden Order Placements," Tech. rep., SFB 649 discussion paper.

HAUTSCH, N. AND M. PODOLSKIJ (2013): "Preaveraging-Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence," *Journal of Business & Economic Statistics*, 31, 165–183.

HIDALGO, J. AND P. ZAFFARONI (2007): "A goodness-of-fit test for models," *Journal of Econometrics*, 141, 973 – 1013.

HUANG, R. AND T. POLAK (2011): "LOBSTER: Limit Order Book Reconstruction System," Technical report, Humboldt-Universität zu Berlin.

IBRAGIMOV, I. A. AND Y. V. LINNIK (1971): *Independent and Stationary Sequences of Random Variables*, Groningen: Wolters-Noordhoff.

INOUE, A. AND L. KILIAN (2002): "Bootstrapping Autoregressive Processes with Possible Unit Roots," *Econometrica*, 70, 377–391.

JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Microstructure Noise in the Continous Case: the Pre-Averaging Approach," *Stochastic Processes and their Applications*, 119, 2803–2831.

LINTON, O. AND E. MAMMEN (2005): "Estimating Semiparametric ARCH(∞) Models by Kernel Smoothing Methods," *Econometrica*, 73.

LINTON, O. AND J. P. NIELSEN (1995): "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, 82, 93–100.

LINTON, O. B. (1997): "Efficient Estimation of Additive Nonparametric Regression Models," *Biometrika*, 84, 469–473.

MAMMEN, E., O. LINTON, AND J. NIELSEN (1999): "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions," *Ann. Statist.*, 27, 1443–1490.

MAMMEN, E. AND B. U. PARK (2005): "Bandwidth selection for smooth backfitting in additive models," *Ann. Statist.*, 33, 1260–1294.

MASRY, E. (1996): "Multivariate local polynomial regression for time series: uniform strong consistency and rates," *Journal of Time Series Analysis*, 17, 571–600.

NEWEY, W. K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.

NIELSEN, J. AND S. SPERLICH (2005): "Smooth backfitting in practice," *J. Roy. Statist. Soc.B*, 67, 43–61.

OPSOMER, J. D. (2000): "Asymptotic properties of backfitting estimators," *J. Mult. Anal.*, 73, 166–179.

OPSOMER, J. D. AND D. RUPPERT (1997): "Fitting a bivariate additive model by local polynomial regression," *Ann. Statist.*, 25, 186 – 211.

PASCUAL, L., J. ROMO, AND E. RUIZ (2006): "Bootstrap prediction for returns and volatilities in GARCH models," *Computational Statistics & Data Analysis*, 50, 2293 – 2312.

PASCUAL, R. AND D. VEREDAS (2010): "Does the Open Limit Order Book Matter in Explaining Informational Volatility?" *Journal of Financial Econometrics*, 8, 57–87.

PATTON, A. J. (2011): "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, 160, 246 – 256, realized Volatility.

PHILIPP, W. (1969): "The central limit problem for mixing sequences of random variables," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 12, 155–171.

POLITIS, D. N. AND J. P. ROMANO (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

ROSU, I. (2009): "A Dynamic Model of the Limit Order Book," *Review of Financial Studies*, 22, 4601–4641.

TAYLOR, S. J. AND X. XU (1997): "The incremental volatility information in one million foreign exchange quotations," *Journal of Empirical Finance*, 4, 317–340.

TJOSTHEIM, D. AND B. H. AUESTAD (1994): "Nonparametric Identification of Nonlinear Time Series: Projections," *Journal of the American Statistical Association*, 89, 1398–1409.

VALENZUELA, M., I. ZER, P. FRYZLEWICZ, AND T. RHEINLÄNDER (2015): "Relative Liquidity and Future Volatility," *Journal of Financial Markets*, 24, 25–48.

VOGT, M. AND O. LINTON (2014): "Nonparametric estimation of a periodic sequence in the presence of a smooth trend," *Biometrika*, 101, 121–140.

VOGT, M. AND C. WALSH (2012): "Estimating Nonlinear Additive Models with Nonstationarities and Correlated Errors," Tech. rep., University of Cambridge.

WANG, F. AND E. GHYSELS (2015): "Econometric Analysis of Volatility Component Models," *Econometric Theory*, 31, 362–393.

WEST, K. (1996): "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067–84.

A Proofs

Proof of Theorem 1

The proof is a simplified version of the proof of Theorem 2 in Vogt and Linton (2014) as the period N is known. Due to $s_t = s_{(d-1)N+i}$ and $s_t = s_{t+kN}$, $k \in \mathbb{N}$, we can write

$$\hat{s}_t - s_t = s_t \frac{1}{D} \sum_{k=1}^{D} \left[z_{(k-1)N+i}^2 - 1 \right] = s_t \frac{1}{D} \sum_{k=1}^{D} q_{(k-1)N+i}. \tag{25}$$

Since $\{q_t\}$ is strictly stationary and strongly mixing by Assumption 1(g) with the mixing coefficients decaying quickly enough, we can apply, e.g., the CLT from Theorem 18.5.3 in Ibragimov and Linnik (1971) to $\frac{\sqrt{T}}{D} \sum_{k=1}^{D} q_{(k-1)N+i}$, recalling that $T := ND$.

□

Proof of Theorem 2

We follow Mammen et al. (1999) in decomposing the estimators of the additive functions, $\hat{g}^{(j)}$, into a "stochastic" and "bias" part, i.e. $\hat{g}^{(j)} = \hat{g}^{(j),A} + \hat{g}^{(j),B}$. Analogous to (12), $\hat{g}^{(j),s}$, $s = A, B$, is defined as the solution to

$$\hat{g}^{(j),s}\left(x^{(j)}\right) = \hat{g}_{NW}^{(j),s}\left(x^{(j)}\right) - 1 - \sum_{k \neq j} \int_0^1 \frac{\hat{p}^{(j,k)}\left(x^{(j)}, x^{(k)}\right)}{\hat{p}^{(j)}\left(x^{(j)}\right)} \hat{g}^{(k),s}\left(x^{(k)}\right) dx^{(k)}, \tag{26}$$

where, similarly to above, $\hat{g}_{NW}^{(j),A}$ and $\hat{g}_{NW}^{(j),B}$ are the stochastic and bias part of the univariate Nadaraya-Watson estimator, respectively, with

$$\hat{g}_{NW}^{(j),A}\left(x^{(j)}\right) = \left[\sum_{t=1}^{T} K_h\left(x^{(j)}, x_{t-1}^{(j)}\right) \right]^{-1} \sum_{t=1}^{T} K_h\left(x^{(j)}, x_{t-1}^{(j)}\right) \varepsilon_t, \tag{27}$$

$$\begin{aligned} \hat{g}_{NW}^{(j),B}\left(x^{(j)}\right) = & \left[\sum_{t=1}^{T} K_h\left(x^{(j)}, x_{t-1}^{(j)}\right) \right]^{-1} \sum_{t=1}^{T} K_h\left(x^{(j)}, x_{t-1}^{(j)}\right) [g(x_{t-1}) \\ & + \Delta_{t,T} + o_p\left(T^{-1}\right)], \end{aligned} \tag{28}$$

where $\varepsilon_t := g_t\left(v_t \eta_t^2 - 1\right)$ and $\Delta_{t,T} := \frac{s_t - \hat{s}_t}{s_t} z_t \left[1 + \frac{1}{2} \frac{s_t - \hat{s}_t}{s_t}\right]$. $\Delta_{t,T} + o_p\left(T^{-1}\right)$ is the error $\tilde{\varepsilon}_t - \varepsilon_t$, caused by the pre-estimation of the seasonality component s_t, and following from a simple second order expansion of $\hat{s}_t^{-1}$ around s_t.

For the stochastic part $\hat{g}^{(j),A}$, we obtain a higher order uniform expansion by applying Lemma A.1, which is a simple generalisation of Theorem 6.1 from Mammen and Park (2005). Similarly, a uniform expansion for the bias part $\hat{g}^{(j),B}$ is given in Lemma A.2 below.

Combining the latter with (29) completes the proof after some straightforward calculations. $\square$

The first two lemmas grant uniform expansions for the stochastic and bias part of the smooth backfitting estimators, respectively.

Lemma A.1. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that $h = O\big(T^{-(1/4+\epsilon)}\big)$ for some $\epsilon > 0$. Then, for $j = 1, \dots, J$,*

$$\sup_{x^{(j)} \in [0,1]} \left| \hat{g}^{(j),A}\big(x^{(j)}\big) - \hat{g}^{(j),A}_{NW}\big(x^{(j)}\big) - \overline{f^{(j)}_T}\big(x^{(j)}\big) \right| = o_p\big(T^{-1/2}\big),\ j = 1, \dots, J, \tag{29}$$

where $\overline{f^{(j)}_T}\big(x^{(j)}\big) := \frac{1}{T}\sum_{t=1}^{T} f^{(j)}_t\big(x^{(j)}\big)\,\varepsilon_t$ with $f^{(j)}_t$, $t = 1, \dots, T$, being uniformly absolutely bounded functions satisfying[12]

$$\left| f^{(j)}_t(x^*) - f^{(j)}_t(x) \right| \leq C\,|x^* - x|\,, \quad C > 0. \tag{30}$$

Proof. Lemma A.1 generalises Theorem 6.1 from Mammen and Park (2005), which is restricted to the i.i.d. case, to our more general setting. We can directly follow their proof with slight adjustments being necessary to show their claim (6.16). For the latter, we need to use the uniform convergence results for the uni- and bivariate kernel density estimators from Lemma A.8. Further, for the proof of their claim (6.22), we can follow the arguments in the proof of Theorem 2 in Masry (1996) based on using the strong approximation theorem by Bradley (1983) to approximate a sequence of random variables by independent random variables. $\square$

Lemma A.2. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that $h = O\big(T^{-(1/4+\epsilon)}\big)$ for some $\epsilon > 0$. Then, for $j = 1, \dots, J$,*

$$\sup_{x^{(j)} \in I_h} \left| \hat{g}^{(j),B}\big(x^{(j)}\big) - g^{(j)}\big(x^{(j)}\big) + \frac{1}{T}\sum_{t=1}^{T} g^{(j)}\Big(x^{(j)}_{t-1}\Big) \right| = O_p\big(h^2\big), \tag{31}$$

$$\sup_{x^{(j)} \in [0,1]} \left| \hat{g}^{(j),B}\big(x^{(j)}\big) - g^{(j)}\big(x^{(j)}\big) + \frac{1}{T}\sum_{t=1}^{T} g^{(j)}\Big(x^{(j)}_{t-1}\Big) \right| = O_p(h), \tag{32}$$

where $I_h := [2C_1 h, 1 - 2C_1 h]$, $C_1 > 0$.

Proof. Since Lemma A.2 directly follows from Theorem 3 in Mammen et al. (1999), we need to check their conditions (A1)-(A6), (A8) and (A9). The latter can be done analogously to the proof of Theorem 4 in the above paper, using the uniform convergence results collected

[12]For the remainder of this appendix, expressions such as C, C_1 etc. denote non-stochastic positive constants.

in Lemma A.8. To verify (A9) in Mammen et al. (1999), their claims (112) and (113) need to be replaced with Lemma A.3 below, which provides uniform expansions for the bias part of the univariate Nadaraya-Watson estimator in our setting. The validity of (A9) given an undersmoothing bandwidth $h << T^{-1/5}$ further requires the presence of the $O_p\big(T^{-1/2}\big)$ term $\frac{1}{T}\sum_{t=1}^{T} g^{(j)}\big(x_{t-1}^{(j)}\big)$ in (31). The latter augments the additive constant $\gamma_{T,j}$ in Theorem 3 in Mammen et al. (1999), which is of order $O_p\big(h^2\big)$ for $x^{(j)} \in I_h$ and $O_p(h)$ else, respectively. This modification ensures that the accordingly adjusted version of their claim (114) holds. □

The lemma below provides uniform expansions for the bias part of the univariate Nadaraya-Watson estimator.

Lemma A.3. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that* $h = O\big(T^{-(1/4+\epsilon)}\big)$ *for some* $\epsilon > 0$. *Then,*

$$\sup_{x^{(j)} \in I_h} \left| \hat{g}_{NW}^{(j),B}\big(x^{(j)}\big) - \hat{\mu}_T^{(j)}\big(x^{(j)}\big) + \overline{g_{N,T}^{(j),*}}\big(x^{(j)}\big) \right| = o_p\big(h^2\big), \tag{33}$$

$$\sup_{x^{(j)} \in [0,1]} \left| \hat{g}_{NW}^{(j),B}\big(x^{(j)}\big) - \hat{\mu}_T^{(j)}\big(x^{(j)}\big) + \overline{g_{N,T}^{(j),*}}\big(x^{(j)}\big) \right| = o_p(h), \tag{34}$$

where

$$\begin{aligned} \overline{g_{N,T}^{(j),*}}\big(x^{(j)}\big) := \frac{1}{N} \sum_{i=1}^{N} & \left(1 - \frac{N}{T} \sum_{d=1}^{T/N} z_{(d-1)N+i}\right) \\ & \left(2 - \frac{N}{T} \sum_{d=1}^{T/N} z_{(d-1)N+i}\right) g^{(j),\mathbb{E}}\big(x^{(j)}\big), \end{aligned} \tag{35}$$

with $g^{(j),\mathbb{E}}\big(x^{(j)}\big) := \mathbb{E}\left[g(X_{t-1}) \,\middle|\, X_{t-1}^{(j)} = x^{(j)} \right]$ *and*

$$\begin{aligned} \mu_T^{(j)}\big(x^{(j)}\big) := \alpha_T^{(0)} + \alpha_T^{(j)}\big(x^{(j)}\big) + \sum_{k \neq j} \int_0^1 \alpha_T^{(k)}\big(x^{(k)}\big) \frac{\hat{p}^{(j,k)}\big(x^{(j)}, x^{(k)}\big)}{\hat{p}^{(j)}\big(x^{(j)}\big)} dx^{(k)} \\ + \Delta_T \int_0^1 \beta(x) \frac{p(x)}{p^{(j)}\big(x^{(j)}\big)} dx^{(-j)}, \end{aligned} \tag{36}$$

where $x^{(-j)}$ denotes the vector x without element $x^{(j)}$, $\Delta_T = h^2$ for $x^{(j)} \in I_h$ and $\Delta_T = h$ else, $\alpha_T^{(0)} = 0$, while

$$\alpha_T^{(j)}\big(x^{(j)}\big) := g^{(j)}\big(x^{(j)}\big) + \frac{g^{(j)\prime}\big(x^{(j)}\big) \int_0^1 K_h\big(x^{(j)}, u\big)\big(u - x^{(j)}\big)\, du}{\int_0^1 K_h\big(x^{(j)}, v\big)\, dv}, \tag{37}$$

$$\beta(x) := \sum_{j=1}^{J} \left[g^{(j)\prime}\big(x^{(j)}\big) \frac{\partial}{\partial x^{(j)}} \ln p(x) + \frac{1}{2} g^{(j)\prime\prime}\big(x^{(j)}\big) \right] \int_0^1 u^2 K(u)\, du. \tag{38}$$

Proof. We focus on the case $x^{(j)} \in I_h$. From (28), we obtain

$$\begin{aligned}
\hat{g}_{NW}^{(j),B}\big(x^{(j)}\big) &= \hat{p}^{(j)}\big(x^{(j)}\big)^{-1} \frac{1}{T} \sum_{t=1}^{T} K_h\big(x^{(j)}, x_{t-1}^{(j)}\big)\, g(x_{t-1}) \\
&\quad + \hat{p}^{(j)}\big(x^{(j)}\big)^{-1} \frac{1}{T} \sum_{t=1}^{T} K_h\big(x^{(j)}, x_{t-1}^{(j)}\big) \frac{s_t - \hat{s}_t}{s_t} z_t \\
&\quad + \hat{p}^{(j)}\big(x^{(j)}\big)^{-1} \frac{1}{T} \sum_{t=1}^{T} K_h\big(x^{(j)}, x_{t-1}^{(j)}\big) \frac{1}{2} \frac{(s_t - \hat{s}_t)^2}{s_t^2} z_t + o_p\big(h^2\big), \\
&=: \hat{g}_{NW}^{(j),B,0}\big(x^{(j)}\big) + \hat{g}_{NW}^{(j),B,1}\big(x^{(j)}\big) + \hat{g}_{NW}^{(j),B,2}\big(x^{(j)}\big) + o_p\big(h^2\big),
\end{aligned} \tag{39}$$

uniformly in $x^{(j)}$, where we used that $o_p\big(T^{-1}\big) = o_p\big(h^2\big)$ for $h = O\big(T^{-(1/4+\epsilon)}\big)$. For the second term in (39), we have

$$\begin{aligned}
\hat{g}_{NW}^{(j),B,1}\big(x^{(j)}\big) &= \frac{1}{N} \sum_{i=1}^{N} \frac{s_i - \hat{s}_i}{s_i} \frac{\frac{1}{D} \sum_{d=1}^{D} K_h\Big(x^{(j)}, x_{(d-1)N+i-1}^{(j)}\Big) z_{(d-1)N+i}}{\hat{p}^{(j)}\big(x^{(j)}\big)}, \\
&= \frac{1}{N} \sum_{i=1}^{N} \frac{s_i - \hat{s}_i}{s_i} \left[g^{(j),\mathbb{E}}\big(x^{(j)}\big) + O_p\left(\sqrt{\frac{\ln T}{Th}}\right) + O_p\big(h^2\big) \right], \\
&= \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{1}{D} \sum_{d=1}^{D} \underbrace{g_{(d-1)N+i}\, \nu_{(d-1)N+i}\, \eta_{(d-1)N+i}^2}_{z_{(d-1)N+i}} \right) g^{(j),\mathbb{E}}\big(x^{(j)}\big) + o_p\big(h^2\big),
\end{aligned} \tag{40}$$

uniformly in $x^{(j)}$, where for the second equality we have used that $\mathbb{E}\Big[Z_t \Big| X_{t-1}^{(j)} = x^{(j)} \Big] = \mathbb{E}\Big[\mathbb{E}[Z_t | X_{t-1}] \Big| X_{t-1}^{(j)} = x^{(j)} \Big] = \mathbb{E}\Big[g(X_{t-1}) \Big| X_{t-1}^{(j)} = x^{(j)} \Big] = g^{(j),\mathbb{E}}\big(x^{(j)}\big)$. We can analogously show that the third term in (39) satisfies

$$\hat{g}_{NW}^{(j),B,2}\big(x^{(j)}\big) = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{1}{D} \sum_{d=1}^{D} z_{(d-1)N+i} \right)^2 g^{(j),\mathbb{E}}\big(x^{(j)}\big) + o_p\big(h^2\big), \tag{41}$$

uniformly in $x^{(j)}$. Finally, directly following the proof of claims (112) and (113) in Mammen et al. (1999) we can show for the first term in (39) that

$$\sup_{x^{(j)}\in I_h}\left|\hat{g}_{NW}^{(j),B,0}\left(x^{(j)}\right)-\hat{\mu}_T^{(j)}\left(x^{(j)}\right)\right|=o_p\left(h^2\right). \tag{42}$$

Combining (42), (40) and (41) completes the proof. The proof for the case $x^{(j)}\notin I_h$ proceeds analogously. □

Proof of Theorem 3

We decompose the deviation of the quasi-likelihood evaluated at the estimated standardised returns, $\mathscr{L}_T\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)$, from its unobservable true counterpart $\mathscr{L}_T(\theta,g,\mathscr{S}_N)$ as

$$\begin{aligned}\sup_{\theta\in\Theta}\left|\mathscr{L}_T\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)-\mathscr{L}_T(\theta,g,\mathscr{S}_N)\right| &\leq \sup_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\left|\ln v_t\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)-\ln v_t(\theta,g,\mathscr{S}_N)\right| \\ &\quad+\sup_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\left|\frac{u_t^2(\hat{g},\hat{s}_t)}{v_t\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)}-\frac{u_t^2(g,s_t)}{v_t(\theta,g,\mathscr{S}_N)}\right|, \\ &=: \sup_{\theta\in\Theta}|L_1(\theta)|+\sup_{\theta\in\Theta}|L_2(\theta)|.\end{aligned} \tag{43}$$

For the first term in (43), we have

$$\begin{aligned}\sup_{\theta\in\Theta}|L_1(\theta)| &\leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left|\ln v_t\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)-\ln v_t(\theta,g,\mathscr{S}_N)\right|, \\ &\leq \frac{C}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left|v_t\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)-v_t(\theta,g,\mathscr{S}_N)\right|, \\ &\leq O_p(h)=o_p(1),\end{aligned} \tag{44}$$

with $0<C<\infty$ a real constant and where, for the second inequality, we have used the mean-value theorem and Assumption 2(b), while the third inequality follows from Lemma A.12. Further, for the second term in (43), we obtain

$$\begin{aligned}\sup_{\theta\in\Theta}|L_2(\theta)| &\leq \frac{1}{T}\sum_{t=1}^{T}u_t^2\sup_{\theta\in\Theta}\left|\frac{s_t g_t}{\hat{s}_t\hat{g}_t\hat{v}_t\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)}-\frac{1}{v_t(\theta,g,\mathscr{S}_N)}\right|, \\ &\leq \frac{1}{T}\sum_{t=1}^{T}u_t^2\left(\sup_{\theta\in\Theta}\left|\frac{1}{v_t(\theta,g,\mathscr{S}_N)}-\frac{1}{v_t\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)}\right|\right. \\ &\quad+\sup_{\theta\in\Theta}\left|\frac{1}{v_t(\theta,g,\mathscr{S}_N)}-\frac{g_t}{\hat{g}_t v_t\left(\theta,\hat{g},\hat{\mathscr{S}}_N\right)}\right|\end{aligned} \tag{45}$$

$$
+ \sup_{\theta \in \Theta} \left| \frac{g_t}{\hat{g}_t v_t(\theta, g, \mathscr{S}_N)} - \frac{s_t g_t}{\hat{s}_t \hat{g}_t v_t\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big)} \right| \Bigg),
$$

$$
= O_p(h) + O_p(h) + O_p\big(T^{-1/2}\big) = O_p(h) = o_p(1),
$$

where the third inequality holds due to Lemma A.11 and Assumptions 1(h), 1(i) and 2(b). We have thus shown that

$$
\sup_{\theta \in \Theta} \left| \mathscr{L}_T\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big) - \mathscr{L}_T(\theta, g, \mathscr{S}_N) \right| \leq o_p(1). \tag{46}
$$

We can now complete the proof using standard results on the strong consistency of the Gaussian quasi-maximum likelihood estimator of GARCH processes, e.g., Theorem 2.1 in Francq and Zakoian (2004).

□

Proof of Theorem 4

We begin with the usual Taylor expansion of the first-order condition for a local maximum of the quasi-likelihood function $\mathscr{L}_T\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big)$ around the true parameter vector θ_0, yielding

$$
T^{1/2}\big(\hat{\theta} - \theta_0\big) = -\big[\mathscr{H}\big(\bar{\theta}, \hat{g}, \hat{\mathscr{S}}_N\big)\big]^{-1} T^{1/2} \mathscr{G}_T\big(\theta_0, \hat{g}, \hat{\mathscr{S}}_N\big), \tag{47}
$$

where $\mathscr{G}_T\big(\theta_0, \hat{g}, \hat{\mathscr{S}}_N\big) := T^{-1} \sum_{t=1}^T \partial/(\partial\theta)\, l_t\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big)|_{\theta=\theta_0}$ is the score, $\mathscr{H}_T\big(\bar{\theta}, \hat{g}, \hat{\mathscr{S}}_N\big) := T^{-1} \sum_{t=1}^T \partial^2/\big(\partial\theta\partial\theta^T\big)\, l_t\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big)|_{\theta=\bar{\theta}}$ denotes the Hessian, while $\bar{\theta} := \lambda\,\hat{\theta} + (1-\lambda)\,\theta_0$ for $\lambda \in [0,1]$. Since the score in (47) is based on both the pre-estimated seasonality and semiparametric component, it can be decomposed as

$$
\begin{aligned}
\mathscr{G}_T\big(\theta_0, \hat{g}, \hat{\mathscr{S}}_N\big) &= \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) + \mathscr{G}_T(\theta_0, \hat{g}, \mathscr{S}_N) - \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) \\
&\quad + \mathscr{G}_T\big(\theta_0, \hat{g}, \hat{\mathscr{S}}_N\big) - \mathscr{G}_T(\theta_0, \hat{g}, \mathscr{S}_N), \\
&=: \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) + \Delta_T^{(\mathscr{G},g)} + \Delta_T^{(\mathscr{G},s)},
\end{aligned} \tag{48}
$$

where $\mathscr{G}_T(\theta_0, g, \mathscr{S}_N) = -T^{-1} \sum_{t=1}^T v_t'/v_t\,\big(1-\eta^2\big) =: T^{-1} \sum_{t=1}^T y_t^{(0)}$ denotes the "true" score with $v_t' = v_t'(\theta_0, g, \mathscr{S}_N) := \partial/(\partial\theta)\, v_t(\theta_0, g, \mathscr{S}_N)$ given in Lemma A.9. For the third term in (48), we obtain by Lemma A.4 below that

$$
\Delta_T^{(\mathscr{G},S)} = \frac{1}{T} \sum_{t=1}^T y_t^{(S)} + o_p\big(T^{-1/2}\big), \tag{49}
$$

where $y_t^{(S)}$, $t = 1,\ldots,T$, is defined in Lemma A.4. For the second term in (48), we consider $\mathscr{G}_T(\theta_0, \hat{g}, \mathscr{S}_N)$ as a functional in g evaluated at $\hat{g}$ and expand it around the true semiparametric component g, such that

$$
\begin{aligned}
\Delta_T^{(\mathscr{G},g)} &= \nabla_{\hat{g}-g}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) + \mathscr{R}_T, \\
&= \nabla_{\hat{g}^A}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) + \nabla_{\hat{g}^B-g}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) + \mathscr{R}_T,
\end{aligned}
\tag{50}
$$

where $\nabla_{\hat{g}-g}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) := \partial/(\partial\, \nu)\, \mathscr{G}_T(\theta_0, g + \nu\,(\hat{g} - g)\,, \mathscr{S}_N)|_{\nu=0}$, while for the second equality we have used the decomposition (26) with $\hat{g}^s := \sum_{j=1}^{J} \hat{g}^{(j),s}$, $s = A, B$. Further, the remainder satisfies $\mathscr{R}_T = O_p\big(\|\hat{g} \; - g\|^2_{\infty,[0,1]}\big) = O_p\big(h^2\big) = o_p\big(T^{-1/2}\big)$ with the last equality following from the undersmoothing bandwidth $h = O\big(T^{-(1/4+\epsilon)}\big)$.[13] For the first two terms in (50), Lemmas A.5 and A.6 imply

$$
\nabla_{\hat{g}^A}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) = \frac{1}{T}\sum_{t=1}^{T} y_{t,T}^{(g,A)} + o_p\big(T^{-1/2}\big), \tag{51}
$$

$$
\nabla_{\hat{g}^B-g}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) = \frac{1}{T}\sum_{t=1}^{T} y_{t}^{(g,B)} + o_p\big(T^{-1/2}\big), \tag{52}
$$

with $y_{t,T}^{(g,A)}$ and $y_t^{(g,B)}$, $t = 1,\ldots,T$, defined in Lemmas A.5 and A.6, respectively. Finally, by combining results (48)-(52), we can re-write expansion (47) as

$$
\begin{aligned}
T^{1/2}\big(\hat{\theta} - \theta_0\big) = &-\big[\mathscr{H}_T\big(\bar{\theta}, \hat{g}, \hat{\mathscr{S}}_N\big)\big]^{-1}\, T^{-1/2} \sum_{t=1}^{T} \underbrace{\Big(y_t^{(0)} + y_{t,T}^{(g,A)} + y_t^{(g,B)} + y_t^{(S)}\Big)}_{:=y_{t,T}} \\
&+ \big[\mathscr{H}_T\big(\bar{\theta}, \hat{g}, \hat{\mathscr{S}}_N\big)\big]^{-1}\, T^{1/2}\, \mathscr{R}_T + o_p(1),
\end{aligned}
\tag{53}
$$

where

$$
\begin{aligned}
y_{t,T} := &-\frac{v_t'}{v_t}\big(1 - \eta_t^2\big) - \sum_{j=1}^{J}\left[\alpha_0 \sum_{l=0}^{\infty} w_{2,j,l}\Big(x_{t-1}^{(j)}\Big) + w_{3,j}\Big(x_{t-1}^{(j)}\Big)\right]\varepsilon_t \\
&- \left\{\alpha_0 \sum_{l=0}^{\infty} \beta_0^l\, \mathbb{E}\left[\frac{u_{t-(l+1)}^2}{g_{t-(l+1)}}\frac{v_t'}{v_t^2}\right] - \mathbb{E}\left[\frac{v_t'}{v_t\, g_t}\right]\right\} \sum_{j=1}^{J} g^{(j)}\Big(x_{t-1}^{(j)}\Big) \\
&- \left\{\mathbb{E}\left[\frac{v_t'}{v_t}\right] - \alpha_0 \sum_{l=0}^{\infty} \beta_0^l\, \mathbb{E}\left[u_{t-(l+1)}^2 \frac{v_t'}{v_t}\right]\right\}\big(z_t^2 - 1\big),
\end{aligned}
\tag{54}
$$

[13] $\|f\|_{\infty,[0,1]}$ denotes the sup norm of a function f over the unit interval, i.e. $\|f\|_{\infty,[0,1]} := \sup_{x\in[0,1]} |f(x)|$.

with $w_{2,j,l}\left(x_{t-1}^{(j)}\right)$ and $w_{3,j}\left(x_{t-1}^{(j)}\right)$ defined in Lemma A.5. We note that, first, $\mathscr{H}_T\left(\bar{\theta},\hat{g},\hat{\mathscr{S}}_N\right)=\mathscr{H}(\theta_0,g,\mathscr{S}_N)+o_p(1)$ according to Lemma A.7. Second, $\left\{T^{-1/2}y_{t,T}\right\}$ constitutes a triangular array which is strongly mixing with exponential rate and with centred elements. Hence, we can apply, e.g., the CLT from Theorem 6 in Philipp (1969) to $\sum_{t=1}^T T^{-1/2}y_{t,T}$. The latter, together with the Slutzky theorem completes the proof.

□

The three Lemmas below provide expansions for the errors in the score due to the pre-estimation of the seasonality and semiparametric component. In the latter case, we analyse the contribution of the stochastic and bias part of the estimator separately.

Lemma A.4. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that $h=O\left(T^{-(1/4+\epsilon)}\right)$ for some $\epsilon>0$. Then,*

$$\begin{aligned}\Delta_T^{(\mathscr{G},S)}&=\frac{1}{T}\sum_{t=1}^T\left\{\alpha_0\sum_{l=0}^{\infty}\beta_0^l\,\mathbb{E}\left[u_{t-(l+1)}^2\frac{v_t'}{v_t}\right]-\mathbb{E}\left[\frac{v_t'}{v_t}\right]\right\}\left(z_t^2-1\right)+o_p\left(T^{-1/2}\right),\\&=:\frac{1}{T}\sum_{t=1}^T y_t^{(S)}+o_p\left(T^{-1/2}\right).\end{aligned}\tag{55}$$

Proof. We decompose $\Delta_T^{(\mathscr{G},S)}$ as

$$\begin{aligned}\Delta_T^{(\mathscr{G},S)}=&-\frac{1}{T}\sum_{t=1}^T\frac{v_t'\left(\theta_0,\hat{g},\hat{\mathscr{S}}\right)-v_t'(\theta_0,\hat{g},\mathscr{S})}{v_t\left(\theta_0,\hat{g},\hat{\mathscr{S}}\right)}\left(1-\frac{r_t^2}{\hat{s}_t\hat{g}_t v_t\left(\theta_0,\hat{g},\hat{\mathscr{S}}\right)}\right)\\&-\frac{1}{T}\sum_{t=1}^T v_t'(\theta_0,\hat{g},\mathscr{S})\left(\frac{1}{v_t\left(\theta_0,\hat{g},\hat{\mathscr{S}}\right)}-\frac{1}{v_t(\theta_0,\hat{g},\mathscr{S})}\right)\left(1-\frac{r_t^2}{\hat{s}_t\hat{g}_t v_t\left(\theta_0,\hat{g},\hat{\mathscr{S}}\right)}\right)\\&-\frac{1}{T}\sum_{t=1}^T\frac{v_t'(\theta_0,\hat{g},\mathscr{S})}{v_t(\theta_0,\hat{g},\mathscr{S})}\left(\frac{1}{s_t}-\frac{1}{\hat{s}_t}\right)\frac{r_t^2}{\hat{g}_t v_t\left(\theta_0,\hat{g},\hat{\mathscr{S}}\right)}\\&-\frac{1}{T}\sum_{t=1}^T\frac{v_t'(\theta_0,\hat{g},\mathscr{S})}{v_t(\theta_0,\hat{g},\mathscr{S})}\left(\frac{1}{v_t(\theta_0,\hat{g},\mathscr{S})}-\frac{1}{v_t\left(\theta_0,\hat{g},\hat{\mathscr{S}}\right)}\right)\frac{u_t^2}{\hat{g}_t},\\&=:\Delta_T^{(\mathscr{G},S,1)}+\Delta_T^{(\mathscr{G},S,2)}+\Delta_T^{(\mathscr{G},S,3)}+\Delta_T^{(\mathscr{G},S,4)}.\end{aligned}\tag{56}$$

We begin with the third term, for which we can exploit Lemma A.12, such that

$$\begin{aligned}\Delta_T^{(\mathscr{G},S,3)}&=-\frac{1}{T}\sum_{t=1}^T\frac{v_t'+O_p(h)}{v_t+O_p(h)}\frac{\hat{s}_t-s_t}{\left(s_t+O_p\left(T^{-1/2}\right)\right)}\frac{u_t^2}{\left(g_t+O_p(h)\right)\left(v_t+O_p(h)\right)},\\&=-\frac{1}{T}\sum_{t=1}^T\frac{v_t'}{v_t}\eta_t^2\frac{\hat{s}_t-s_t}{s_t}+o_p\left(T^{-1/2}\right),\end{aligned}\tag{57}$$

$$
\begin{aligned}
&= -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{s}_i - s_i}{s_i}\,\frac{1}{D}\sum_{d=1}^{D}\frac{v'_{(d-1)N+i}}{v_{(d-1)N+i}}\,\eta^2_{(d-1)N+i} + o_p\big(T^{-1/2}\big),\\
&= -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{s}_i - s_i}{s_i}\,\mathbb{E}\bigg[\frac{v'_t}{v_t}\bigg] + o_p\big(T^{-1/2}\big),\\
&= -\frac{1}{N}\sum_{i=1}^{N}\bigg(\frac{1}{D}\sum_{d=1}^{D} z^2_{(d-1)N+i} - 1\bigg)\,\mathbb{E}\bigg[\frac{v'_t}{v_t}\bigg] + o_p\big(T^{-1/2}\big),\\
&= -\bigg(\frac{1}{T}\sum_{t=1}^{T} z_t^2 - 1\bigg)\,\mathbb{E}\bigg[\frac{v'_t}{v_t}\bigg] + o_p\big(T^{-1/2}\big).
\end{aligned}
$$

For the second equality, we have repeatedly used that $\big(\zeta + O_p(a_T)\big)^{-1} = \zeta^{-1} + O_p(a_T)$ with $\zeta = O_p(1)$ and $a_T = o(1)$, Assumptions 1(h), 1(i) and 2(b), as well as the fact that $\sup_{t=1,\ldots,N}|\hat{s}_t - s_t| = O_p\big(T^{-1/2}\big)$ by Theorem 1. To obtain the fourth equality, we again use the latter result and further apply the CLT from Theorem 18.5.3 in Ibragimov and Linnik (1971) to $\frac{1}{D}\sum_{d=1}^{D}\frac{v'_{(d-1)N+i}}{v_{(d-1)N+i}}\,\eta^2_{(d-1)N+i}$, while exploiting that $\{\eta_t\}$ is i.i.d. with unit mean. The remaining three terms in (56) can be handled analogously using the results from Lemma A.12 and the parametric uniform rate of $\hat{s}_t$, $t = 1,\ldots,N$. Accordingly, we can show that

$$
\Delta_T^{(\mathscr{G},S,1)} = o_p\big(T^{-1/2}\big), \tag{58}
$$

$$
\Delta_T^{(\mathscr{G},S,2)} = o_p\big(T^{-1/2}\big), \tag{59}
$$

$$
\Delta_T^{(\mathscr{G},S,4)} = \frac{1}{T}\sum_{t=1}^{T}\alpha_0\sum_{l=0}^{\infty}\beta_0^l\,\mathbb{E}\bigg[u^2_{t-(l+1)}\frac{v'_t}{v_t}\bigg]\big(z_t^2 - 1\big) + o_p\big(T^{-1/2}\big), \tag{60}
$$

which completes the proof. □

Lemma A.5. *Let Assumptions 1 as well as 2(b) and 2(e) be fulfilled. Further, assume that $h = O\big(T^{-(1/4+\epsilon)}\big)$ for some $\epsilon > 0$. Then,*

$$
\begin{aligned}
\nabla_{\hat{g}^A}\mathscr{G}_T(\theta_0, g, \mathscr{S}_N) = -\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\bigg[&\alpha_0\sum_{l=0}^{\infty} w_{2,j,l}\big(x_{t-1}^{(j)}\big) \\
&+ w_{3,j}\big(x_{t-1}^{(j)}\big)\bigg]\varepsilon_t + o_p\big(T^{-1/2}\big),
\end{aligned} \tag{61}
$$

$$
=: \frac{1}{T}\sum_{t=1}^{T} y_{t,T}^{(g,A)} + o_p\big(T^{-1/2}\big),
$$

where, for $j = 1,\ldots,J$ and $l = 0,1,\ldots,$

$$
w_{2,j,l}\big(x_{t-1}^{(j)}\big) := w_{2,j,l}^{(1)}\big(x_{t-1}^{(j)}\big) + w_{2,j,l}^{(2)}\big(x_{t-1}^{(j)}\big), \tag{62}
$$

$$w_{2,j,l}^{(1)}\left(x_{t-1}^{(j)}\right) := \mathbb{E}_{(t)}\left[\frac{K_h\left(x_{s-(l+1)}^{(j)}, x_{t-1}^{(j)}\right)\frac{v_s'}{v_s^2}\frac{u_{s-(l+1)}^2}{g_{s-(l+1)}}}{\int_0^1 K_h\left(x_{s-(l+1)}^{(j)}, z\right)dz\, p^{(j)}\left(x_{s-(l+1)}^{(j)}\right)}\right], \tag{63}$$

$$w_{2,j,l}^{(2)}\left(x_{t-1}^{(j)}\right) := \mathbb{E}_{(t)}\left[f_t^{(j)}\left(x_{s-(l+1)}^{(j)}\right)\frac{v_s'}{v_s^2}\frac{u_{s-(l+1)}^2}{g_{s-(l+1)}}\right], \tag{64}$$

and

$$w_{3,j}\left(x_{t-1}^{(j)}\right) := w_{3,j}^{(1)}\left(x_{t-1}^{(j)}\right) + w_{3,j}^{(2)}\left(x_{t-1}^{(j)}\right), \tag{65}$$

$$w_{3,j}^{(1)}\left(x_{t-1}^{(j)}\right) := \mathbb{E}_{(t)}\left[\frac{K_h\left(x_{s-1}^{(j)}, x_{t-1}^{(j)}\right)\frac{v_s'}{v_s g_s}}{\int_0^1 K_h\left(x_{s-1}^{(j)}, z\right)dz\, p^{(j)}\left(x_{s-1}^{(j)}\right)}\right], \tag{66}$$

$$w_{3,j}^{(2)}\left(x_{t-1}^{(j)}\right) := \mathbb{E}_{(t)}\left[f_t^{(j)}\left(x_{s-1}^{(j)}\right)\frac{v_s'}{v_s g_s}\right], \tag{67}$$

with $\mathbb{E}_{(t)}$ denoting the expectation with respect to all random variables except those indexed by t and the functions $f_t^{(j)}$ defined in Lemma A.1.

Proof. $\nabla_{\hat{g}^A}\mathscr{G}_T(\theta_0, g, \mathscr{S}_N) := \frac{\partial}{\partial \nu}\mathscr{G}_T\left(\theta_0, g + \nu\,\hat{g}^A, \mathscr{S}_N\right)|_{\nu=0}$ satisfies

$$\begin{aligned}
\nabla_{\hat{g}^A}\mathscr{G}_T(\theta_0, g, \mathscr{S}_N) &= \frac{1}{T} - \sum_{j=1}^{J}\sum_{t=1}^{T}\frac{\nabla_{\hat{g}^{(j),A}} v_t'(\theta_0, g, \mathscr{S}_N)}{v_t(\theta_0, g, \mathscr{S}_N)}\left(1-\eta_t^2\right) \\
&\quad - \frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\nabla_{\hat{g}^{(j),A}} v_t(\theta_0, g, \mathscr{S}_N)\frac{v_t'(\theta_0, g, \mathscr{S}_N)}{v_t^2(\theta_0, g, \mathscr{S}_N)}\left(2\eta_t^2 - 1\right) \\
&\quad - \frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{v_t'(\theta_0, g, \mathscr{S}_N)}{v_t(\theta_0, g, \mathscr{S}_N)}\frac{\eta_t^2}{g_t}\,\hat{g}^{(j),A}\left(x_{t-1}^{(j)}\right), \\
&:= \nabla_{\hat{g}^A}\mathscr{G}_T^{(1)} + \nabla_{\hat{g}^A}\mathscr{G}_T^{(2)} + \nabla_{\hat{g}^A}\mathscr{G}_T^{(3)}.
\end{aligned} \tag{68}$$

We begin with the third term. By Lemma A.1, the latter can be decomposed as

$$\begin{aligned}
\mathscr{G}_T^{(3)} &= -\frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{v_t'\eta_t^2}{v_t g_t}\,\hat{g}_{NW}^{(j),A}\left(x_{t-1}^{(j)}\right) - \frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{v_t'\eta_t^2}{v_t g_t}\,\overline{f_T^{(j)}}\left(x_{t-1}^{(j)}\right) + o_p\left(T^{-1/2}\right), \\
&=: \mathscr{G}_T^{(3,1)} + \mathscr{G}_T^{(3,2)} + o_p\left(T^{-1/2}\right).
\end{aligned} \tag{69}$$

We initially focus on the first term, which using the definition (27), satisfies

$$\mathscr{G}_T^{(3,1)} = -\sum_{j=1}^{J}\frac{1}{T}\sum_{s=1}^{T}\varepsilon_s\frac{1}{T}\sum_{t=1}^{T}K_h\left(x_{t-1}^{(j)}, x_{s-1}^{(j)}\right)q_t\left[\frac{1}{T}\sum_{s=1}^{T}K_h\left(x_{t-1}^{(j)}, x_{s-1}^{(j)}\right)\right]^{-1}, \tag{70}$$

$$= -\sum_{j=1}^{J} \frac{1}{T} \sum_{s=1}^{T} \varepsilon_s \frac{1}{T} \sum_{t=1}^{T} K_h\Big(x_{t-1}^{(j)}, x_{s-1}^{(j)}\Big) \frac{q_t}{\tilde{p}^{(j)}\Big(x_{t-1}^{(j)}\Big)} + o_p\Big(T^{-1/2}\Big),$$

where $q_t := v_t' \eta_t^2 / (v_t g_t)$, while $\tilde{p}^{(j)}\big(x^{(j)}\big)$ is defined in Lemma A.8. For the second equality, we have exploited that $T^{-1}\sum_{s=1}^{T} K_h\Big(x_{t-1}^{(j)}, x_{s-1}^{(j)}\Big) = \tilde{p}^{(j)}\Big(x_{t-1}^{(j)}\Big) + O_p(h)$ by Lemma A.8, conducted an expansion of the inverse of the latter term as in the proof of Lemma A.4 and applied standard results on uniform convergence rates from Masry (1996) to bound the resulting remainder term. As the next step, we show that

$$\mathscr{G}_T^{(3,1)} = -\sum_{j=1}^{J} \frac{1}{T} \sum_{s=1}^{T} \varepsilon_s \, \mathbb{E}_{(s)} \left[K_h\Big(x_{t-1}^{(j)}, x_{s-1}^{(j)}\Big) \frac{q_t}{\tilde{p}^{(j)}\Big(x_{t-1}^{(j)}\Big)} \right] + o_p\Big(T^{-1/2}\Big), \tag{71}$$

which requires $\Delta\mathscr{G}_T^{(3,1)} := \sum_{j=1}^{J} \Delta\mathscr{G}_T^{(3,1,j)} = o_p\Big(T^{-1/2}\Big)$, $\Delta\mathscr{G}_T^{(3,1,j)} := \Big(\Delta\mathscr{G}_{T,\alpha}^{(3,1,j)}, \Delta\mathscr{G}_{T,\beta}^{(3,1,j)}\Big)'$, where

$$\begin{aligned} \Delta\mathscr{G}_{T,k}^{(3,1,j)} := & -\frac{1}{T} \sum_{s=1}^{T} \varepsilon_s \frac{1}{T} \sum_{t=1}^{T} \left\{ K_h\Big(x_{t-1}^{(j)}, x_{s-1}^{(j)}\Big) \frac{q_t^{(k)}}{\tilde{p}^{(j)}\Big(x_{t-1}^{(j)}\Big)} \right. \\ & \left. -\mathbb{E}_{(s)} \left[K_h\Big(x_{t-1}^{(j)}, x_{s-1}^{(j)}\Big) \frac{q_t^{(k)}}{\tilde{p}^{(j)}\Big(x_{t-1}^{(j)}\Big)} \right] \right\}, \\ =: & -\frac{1}{T} \sum_{s=1}^{T} \varepsilon_s \frac{1}{T} \sum_{t=1}^{T} \rho_{j,t,s}^{(k)}, \end{aligned} \tag{72}$$

with $q_t^{(k)} := v_{k,t}' \eta_t^2 / (v_t g_t)$, $k = \alpha, \beta$. Thus, we need to prove that $\mathbb{E}\Big[T\Big(\Delta\mathscr{G}_{T,k}^{(3,1,j)}\Big)^2\Big] = o(1)$, $k = \alpha, \beta$.

We begin with $\Delta\mathscr{G}_{T,\alpha}^{(3,1,j)}$. $\Delta\mathscr{G}_{T,\beta}^{(3,1,j)}$ can be treated analogously. For this purpose, note that we can decompose $\mathbb{E}\Big[T\Big(\Delta\mathscr{G}_{T,\alpha}^{(3,1,j)}\Big)^2\Big]$ as

$$\begin{aligned} \mathbb{E}\Big[T\Big(\Delta\mathscr{G}_{T,\alpha}^{(3,1,j)}\Big)^2\Big] = & \frac{1}{T^3} \sum_{(t,s,l,m)\in Z} \mathbb{E}\Big[\varepsilon_s \varepsilon_l \rho_{j,t,s}^{(\alpha)} \rho_{j,l,m}^{(\alpha)}\Big] \\ & + \frac{1}{T^3} \sum_{(t,s,l,m)\notin Z} \mathbb{E}\Big[\varepsilon_s \varepsilon_l \rho_{j,t,s}^{(\alpha)} \rho_{j,l,m}^{(\alpha)}\Big], \\ =: & E_1^{(\alpha,j)} + E_2^{(\alpha,j)}, \end{aligned} \tag{73}$$

where $Z := \{(t,s,l,m) | \delta(\{t,s,l,m\}) \geq \delta_0\}$, with $\delta(M) := \inf_{x,y\in M, x\neq y} |x-y|$, while $\delta_0 = \delta_0(T) > 0$. Bounding the second term in (73) is straightforward by applying the Cauchy-

Schwarz inequality, Lemma A.9, Assumptions 1(b), 1(i) and 2(b), as well as noting that $\sup_{u,v\in[0,1]} K_h(u,v) = O\left(h^{-1}\right)$ and $\mathbb{E}\left[\varepsilon_t^4\right] < \infty$. Thus,

$$E_2^{(\alpha,j)} \leq C\frac{T^2\delta_0^2}{T^3h^2} = o(1), \tag{74}$$

provided that $\delta_0 = o\left(T^{1/2}h\right)$.

The first term in (73) is somewhat more difficult to handle. We decompose it as

$$\begin{aligned} E_1^{(\alpha,j)} &= \sum_{n\in\{t,s,l,m\}} \frac{1}{T^3} \sum_{(t,s,l,m)\in Z_n} \mathbb{E}\left[\varepsilon_s\varepsilon_l\rho_{j,t,s}^{(\alpha)}\rho_{j,l,m}^{(\alpha)}\right] \\ &=: \sum_{n\in\{t,s,l,m\}} E_{1,n}^{(\alpha,j)}, \end{aligned} \tag{75}$$

where $Z_n := \{(t,s,l,m)|\delta_n(\{t,s,l,m\}) \geq \delta_0,\, Z_n\cap Z_{\tilde{n}} = \emptyset\, \forall\, \tilde{n}\neq n\}$, $n \in \{t,s,l,m\}$, with $\delta_x(M) := \inf_{y\in M, y\neq x}|x-y|$. We focus on $E_{1,t}^{(\alpha,j)}$ with the treatment of the remaining terms in (75) being completely analogous. We can now apply the standard covering argument from the proof of Theorem 2 in Masry (1996) to the compact support of $x_t^{(j)}$, $[0,1]$. Thus, we cover the latter by $L(T)$ equidistant cubes I_k with centres x_k and length $L(T)^{-1}$, $k = 1,\ldots,L(T)$. This approach allows us to decompose $E_{1,t}^{(\alpha,j)}$ according to

$$\begin{aligned} E_{1,t}^{(\alpha,j)} &= \frac{1}{T^3}\sum_{(t,s,l,m)\in Z_n}\sum_{k=1}^{L(T)}\mathbb{E}\left[\varepsilon_s\varepsilon_l\rho_{j,l,m}^{(\alpha)}\left\{\mathbb{I}_{\left\{x_{t-1}^{(j)}\in I_k\right\}}K_h\left(x_k,x_{s-1}^{(j)}\right)\frac{q_t^{(\alpha)}}{\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)}\right.\right. \\ &\quad \left.\left.-\mathbb{E}_{(s)}\left[\mathbb{I}_{\left\{x_{t-1}^{(j)}\in I_k\right\}}K_h\left(x_k,x_{s-1}^{(j)}\right)\frac{q_t^{(\alpha)}}{\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)}\right]\right\}\right] \\ &\quad + \frac{1}{T^3}\sum_{(t,s,l,m)\in Z_n}\sum_{k=1}^{L(T)}\mathbb{E}\left[\varepsilon_s\varepsilon_l\rho_{j,l,m}^{(\alpha)}\left\{\mathbb{I}_{\left\{x_{t-1}^{(j)}\in I_k\right\}}\left[K_h\left(x_{t-1}^{(j)},x_{s-1}^{(j)}\right)-K_h\left(x_k,x_{s-1}^{(j)}\right)\right]\frac{q_t^{(\alpha)}}{\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)}\right.\right. \\ &\quad \left.\left.-\mathbb{E}_{(s)}\left[\mathbb{I}_{\left\{x_{t-1}^{(j)}\in I_k\right\}}\left[K_h\left(x_{t-1}^{(j)},x_{s-1}^{(j)}\right)-K_h\left(x_k,x_{s-1}^{(j)}\right)\right]\frac{q_t^{(\alpha)}}{\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)}\right]\right\}\right], \\ &=: E_{1,t,1}^{(\alpha,j)} + E_{1,t,2}^{(\alpha,j)}. \end{aligned} \tag{76}$$

For the second term, we note that, due to the Lipschitz continuity of $K(u)$ granted by Assumption 1(a), we have $\left|K_h\left(x_{t-1}^{(j)},x_{s-1}^{(j)}\right)-K_h\left(x_k,x_{s-1}^{(j)}\right)\right| \leq C\left|x_{t-1}^{(j)}-x_k\right|h^{-2} \leq CL(T)^{-1}h^{-2}$

if $x_{t-1}^{(j)} \in I_k$, $t,s = 1,\ldots,T$, $k = 1,\ldots,L(T)$. This yields

$$\left|E_{1,t,2}^{(\alpha,j)}\right| \leq \frac{C}{T^3 h^2 L(T)} \sum_{(t,s,l,m)\in Z_n} \mathbb{E}\left[|\varepsilon_s||\varepsilon_l|\left|\rho_{j,l,m}^{(\alpha)}\right| \left\{\left|\frac{q_t^{(\alpha)}}{\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)}\right| \right.\right. \tag{77}$$
$$\left.\left. + \ \mathbb{E}\left[\left|\frac{q_t^{(\alpha)}}{\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)}\right|\right]\right\}\right],$$
$$\leq C\frac{T}{h^3 L(T)},$$

where, for the second inequality, we have applied Assumptions 1(b), 1(i) and 2(b), as well as Lemma A.9 and the fact that $\sup_{u,v\in[0,1]} K_h(u,v) = h^{-1}$. Thus, $E_{1,t,2}^{(\alpha,j)} = o(1)$ if $L(T)^{-1} = o\left(T^{-1}h^3\right)$, e.g., $L(T) = O\left(Th^{-3}\ln T\right)$. Likewise, the first term in (76) satisfies

$$E_{1,t,1}^{(\alpha,j)} = TL(T)\mathbb{C}\mathrm{ov}\left[K_h\left(x_k, x_{s-1}^{(j)}\right)\varepsilon_s\varepsilon_l\rho_{j,l,m}^{(\alpha)}, \mathbb{I}_{\left\{x_{t-1}^{(j)}\in I_k\right\}}\frac{q_t^{(\alpha)}}{\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)}\right], \tag{78}$$
$$\leq CTL(T)\,d^{\delta_0\left(1-\kappa^{-1}-\gamma^{-1}\right)}\,\mathbb{E}\left[K_h\left(x_k, x_{s-1}^{(j)}\right)^\kappa |\varepsilon_s|^\kappa |\varepsilon_l|^\kappa \left|\rho_{j,l,m}^{(\alpha)}\right|^\kappa\right]^{1/\kappa}$$
$$\times\,\mathbb{E}\left[\mathbb{I}_{\left\{x_{t-1}^{(j)}\in I_k\right\}}\frac{\left|q_t^{(\alpha)}\right|^\gamma}{\left(\tilde{p}^{(j)}\left(x_{t-1}^{(j)}\right)\right)^\gamma}\right]^{1/\gamma},$$
$$\leq C\frac{TL(T)}{h^2}\,d^{\delta_0\left(1-\kappa^{-1}-\gamma^{-1}\right)}\,\mathbb{E}[|\varepsilon_s|^\kappa|\varepsilon_l|^\kappa]^{1/\kappa}\,\mathbb{E}\left[\left|v'_{\alpha,t}\right|^\gamma\right]^{1/\gamma},$$

with $0 < d < 1$ and $0 < \kappa^{-1} + \gamma^{-1} < 1$. For the first inequality, we have applied Davydov's Lemma (see, e.g., Hall and Heyde, 1980, Corollary A2), while the second one follows from arguments analogous to those used in (77).

In order to bound the two expectations in (78) , we note that, by Lemma A.9 and the Minkowski inequality,

$$\mathbb{E}\left[\left|v'_{\alpha,t}\right|^\gamma\right]^{1/\gamma} \leq C_1 + C_2\,\mathbb{E}\left[|u_t|^{2\gamma}\right]^{1/\gamma}. \tag{79}$$

Likewise, the Cauchy-Schwarz inequality and the stationarity of $\{\varepsilon_t\}$ imply $\mathbb{E}[|\varepsilon_s|^\kappa|\varepsilon_l|^\kappa]^{1/\kappa} \leq \mathbb{E}\left[|\varepsilon_t|^{2\kappa}\right]^{1/\kappa}$, while by the Hölder inequality,

$$\mathbb{E}\left[|\varepsilon_t|^{2\kappa}\right]^{1/\kappa} \leq \mathbb{E}\left[\left|u_t^2 - 1\right|^{2\kappa\zeta}\right]^{1/(\kappa\zeta)}\mathbb{E}\left[g_t^{2\kappa\tau}\right]^{1/(\kappa\tau)}, \tag{80}$$
$$\leq \left(1 + \mathbb{E}\left[|u_t|^{4\kappa\zeta}\right]^{1/(\kappa\zeta)}\right)\mathbb{E}\left[g_t^{2\kappa\tau}\right]^{1/(\kappa\tau)},$$

with $\zeta^{-1} + \tau^{-1} = 1$ and where the second inequality follows from another application of the Minkowski inequality. For the two above moments to exist and be finite, we thus need that $\mathbb{E}[|u_t|^{\varphi}] < \infty$, where $\varphi := \max(2\gamma, 4\kappa\zeta)$. Then, (80) implies $\mathbb{E}\big[|\varepsilon_t|^{2\kappa}\big] < \infty$ as $\mathbb{E}\big[g_t^{2\kappa\tau}\big] < \infty$ for finite κ and τ by Assumptions 1(c) and 1(b). Accordingly, we can choose ζ as a rational number arbitrarily close to (but strictly greater than) one, while $\tau = (1-\zeta^{-1})^{-1}$. Then, setting $\kappa = 3/2 + \epsilon_1$ and $\gamma = 3 + \epsilon_2$, $\epsilon_1, \epsilon_2 > 0$, the above condition requires $\mathbb{E}\big[|u_t|^{6+\epsilon}\big] < \infty$, which is granted by Assumption 2(e). Using these results in (78), as well as choosing δ_0 from (73) as $\delta_0 = O(\ln T)$, yields by simple algebra

$$
\begin{aligned}
E_{1,t,1}^{(\alpha,j)} &\leq C\frac{TL(T)}{h^2}\, d^{C' \ln T\left(1-\kappa^{-1}-\gamma^{-1}\right)}, \\
&= C\frac{TL(T)}{h^2}\, T^{-C''\left(1-\kappa^{-1}-\gamma^{-1}\right)} =: C\frac{TL(T)}{h^2}\, T^{-C'''},
\end{aligned} \tag{81}
$$

where $C'' > 0$. Hence, by setting $L(T) = O\big(Th^{-3}\ln T\big)$, we obtain $E_{1,t,1}^{(\alpha,j)} = CT^{2-C'''}h^{-5}\ln T = o(1)$ as we can simply choose C'', and thus C''', to be sufficiently large. Recalling (77), we thus have shown that $E_{1,t}^{(\alpha,j)} = o(1)$ and, consequently, $E_1^{(\alpha,j)} = o(1)$, which together with (74) and completely analogous calculations for $\Delta\mathscr{G}_{T,\beta}^{(3,1,j)}$ proves (71).

In a similar (and slightly simplified) fashion, we can show that the second term in (69) satisfies

$$
\mathscr{G}_T^{(3,2)} = -\sum_{j=1}^{J}\frac{1}{T}\sum_{s=1}^{T}\varepsilon_s\,\mathbb{E}_{(s)}\Big[f_s^{(j)}\Big(x_{t-1}^{(j)}\Big)q_t\Big] + o_p\big(T^{-1/2}\big). \tag{82}
$$

Finally, by application of Lemma A.10 and following the above proof strategy, we can show for the remaining two terms in (68) that

$$
\mathscr{G}_T^{(1)} = o_p\big(T^{-1/2}\big), \tag{83}
$$

$$
\begin{aligned}
\mathscr{G}_T^{(2)} = -\frac{\alpha_0}{T}\sum_{j=1}^{J}\sum_{s=1}^{T}\sum_{l=0}^{\infty}\varepsilon_s\,\mathbb{E}_{(s)}\Bigg[&\frac{K_h\Big(x_{t-(l+1)}^{(j)}, x_{s-1}^{(j)}\Big)\frac{v_t'}{v_t^2}\frac{u_{t-(l+1)}^2}{g_{t-(l+1)}}}{\int_0^1 K_h\Big(x_{t-(l+1)}^{(j)}, z\Big)\,dz\,p^{(j)}\Big(x_{t-(l+1)}^{(j)}\Big)} \\
&+ f_s^{(j)}\Big(x_{t-(l+1)}^{(j)}\Big)\frac{v_t'}{v_t^2}\frac{u_{t-(l+1)}^2}{g_{t-(l+1)}}\Bigg] + o_p\big(T^{-1/2}\big),
\end{aligned} \tag{84}
$$

where, for (84) to hold, we need analogously to the arguments in (78)-(80) that $\mathbb{E}\big[|u_t|^{8+\epsilon}\big] < \infty$, $\epsilon > 0$. The latter is guaranteed by Assumption 2(e), which completes the proof. $\square$

Lemma A.6. *Let Assumptions 1 as well as 2(b) and 2(e) be fulfilled. Further, assume that* $h = O\big(T^{-(1/4+\epsilon)}\big)$ *for some* $\epsilon > 0$. *Then,*

$$\begin{aligned}\nabla_{\hat{g}^B - B}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) = -\frac{1}{T}\sum_{t=1}^{T}\Bigg\{\alpha_0 \sum_{l=0}^{\infty}\beta_0^l\, \mathbb{E}\Bigg[\frac{u^2_{t-(l+1)}}{g_{t-(l+1)}}\frac{v'_t}{v_t^2}\Bigg] \\ -\mathbb{E}\Bigg[\frac{v'_t}{v_t g_t}\Bigg]\Bigg\}\sum_{j=1}^{J} g^{(j)}\big(x^{(j)}_{t-1}\big) + o_p\big(T^{-1/2}\big), \\ =: \frac{1}{T}\sum_{t=1}^{T} y_t^{(g,B)} + o_p\big(T^{-1/2}\big).\end{aligned} \tag{85}$$

Proof. $\nabla_{\hat{g}^B - B}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) := \frac{\partial}{\partial \nu}\mathscr{G}_T\big(\theta_0, g + \nu\left(\hat{g}^B - g\right), \mathscr{S}_N\big)|_{\nu=0}$ satisfies, analogous to the proof of Lemma A.5,

$$\begin{aligned}\nabla_{\hat{g}^B - B}\, \mathscr{G}_T(\theta_0, g, \mathscr{S}_N) = \frac{1}{T} - \sum_{j=1}^{J}\sum_{t=1}^{T}\frac{\nabla_{\hat{g}^{(j),B}-g^{(j)}}\, v'_t}{v_t}\big(1-\eta_t^2\big) \\ -\frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\nabla_{\hat{g}^{(j),B}-g^{(j)}}\, v_t \frac{v'_t}{v_t^2}\big(2\eta_t^2 - 1\big) \\ -\frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{v'_t}{v_t}\frac{\eta_t^2}{g_t}\Big[\hat{g}^{(j),B}\big(x^{(j)}_{t-1}\big) - g^{(j)}\big(x^{(j)}_{t-1}\big)\Big], \\ := \nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(1)} + \nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(2)} + \nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(3)}.\end{aligned} \tag{86}$$

Again, we first cover the third term. The latter can be decomposed as

$$\begin{aligned}\nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(3)} = -\frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{v'_t}{v_t}\frac{\eta_t^2}{g_t}\Big[\hat{g}^{(j),B}\big(x^{(j)}_{t-1}\big) - g^{(j)}\big(x^{(j)}_{t-1}\big) + \overline{g^{(j)}}\Big] \\ +\frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{v'_t}{v_t}\frac{\eta_t^2}{g_t}\overline{g^{(j)}}, \\ =: \nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(3,1)} + \nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(3,2)},\end{aligned} \tag{87}$$

where $\overline{g^{(j)}} := T^{-1}\sum_{t=1}^{T} g^{(j)}\big(x^{(j)}_{t-1}\big)$, $j = 1, \ldots, J$. For $\nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(3,1)}$, we distinguish between situations with $x^{(j)}_{t-1}$, $t = 1, \ldots, T$, $j = 1, \ldots, J$, being in the interior and in the boundary region of the support, respectively, i.e.,

$$\nabla_{\hat{g}^B - g}\, \mathscr{G}_T^{(3,1)} \leq \sum_{j=1}^{J}\sup_{x^{(j)}\in I_h}\Big|\hat{g}^{(j),B}\big(x^{(j)}\big) - g^{(j)}\big(x^{(j)}\big) + \overline{g^{(j)}}\Big|\,\frac{1}{T}\sum_{t=1}^{T}\frac{\|v'_t\|}{v_t}\frac{\eta_t^2}{g_t}\mathbb{I}_{\big\{x^{(j)}_{t-1}\in I_h\big\}} \tag{88}$$

$$
\begin{aligned}
&+\sum_{j=1}^{J}\sup_{x^{(j)}\in[0,1]}\left|\hat{g}^{(j),B}\big(x^{(j)}\big)-g^{(j)}\big(x^{(j)}\big)+\overline{g^{(j)}}\right|\frac{1}{T}\sum_{t=1}^{T}\frac{\left\|v_t'\right\|}{v_t}\frac{\eta_t^2}{g_t}\mathbb{I}_{\left\{x_{t-1}^{(j)}\notin I_h\right\}},\\
&\leq O_p\big(h^2\big)+O_p(h)\left\{\mathbb{E}\left[\frac{\left\|v_t'\right\|}{v_t}\frac{\eta_t^2}{g_t}\mathbb{I}_{\left\{x_{t-1}^{(j)}\notin I_h\right\}}\right]+O_p\big(T^{-1/2}\big)\right\},\\
&\leq O_p(h)\,\mathbb{E}\Big[\mathbb{E}\Big[\left\|v_t'\right\|\eta_t^2\,|x_{t-1}^{(j)}\Big]\,\mathbb{I}_{\left\{x_{t-1}^{(j)}\notin I_h\right\}}\Big]+o_p\big(T^{-1/2}\big),\\
&\leq O_p(h)\,\mathbb{E}\Big[\mathbb{I}_{\left\{x_{t-1}^{(j)}\notin I_h\right\}}\Big]+o_p\big(T^{-1/2}\big),\\
&\leq O_p(h)\int_{x^{(j)}\notin I_h}p^{(j)}\big(x^{(j)}\big)\,dx^{(j)}+o_p\big(T^{-1/2}\big),\\
&\leq O_p\big(h^2\big)+o_p\big(T^{-1/2}\big)=o_p\big(T^{-1/2}\big),
\end{aligned}
$$

where for the second inequality, we have applied Lemma A.2, while the third inequality follows from the undersmoothing condition, the law of iterated expectations, as well as Assumption 1(i) and 2(b).[14] Further, for the fourth inequality, we note that, by another application of the law of iterated expectations,

$$
\mathbb{E}\Big[\left\|v_t'\right\|\eta_t^2\,|x_{t-1}^{(j)}\Big]=\mathbb{E}\Big[\mathbb{E}\Big[\left\|v_t'\right\||\mathscr{H}_{t-1}^{(x)}\Big]|x_{t-1}^{(j)}\Big]\leq\mathbb{E}\Big[C_1+C_2\mathbb{E}\Big[v_t|\mathscr{H}_{t-1}^{(x)}\Big]|x_{t-1}^{(j)}\Big]\leq O(1),
$$

by Lemma A.9 and $\mathbb{E}\Big[v_t|\mathscr{H}_{t-1}^{(x)}\Big]=1$ with $\mathscr{H}_t^{(x)}$ being the filtration generated by $\{x_t,\ldots,x_{-\infty}\}$. Finally, the boundedness of the density $p^{(j)}$ according to Assumption 1(b) yields the last inequality in (88). We can now tackle the second term in (87), which satisfies

$$
\begin{aligned}
\nabla_{\hat{g}^B-g}\,\mathscr{G}_T^{(3,2)}&=\sum_{j=1}^{J}\overline{g^{(j)}}\left\{\mathbb{E}\left[\frac{v_t'}{g_tv_t}\right]+O_p\big(T^{-1/2}\big)\right\}, &(89)\\
&=\mathbb{E}\left[\frac{v_t'}{g_tv_t}\right]\sum_{j=1}^{J}\overline{g^{(j)}}+o_p\big(T^{-1/2}\big),
\end{aligned}
$$

by Theorem 18.5.3 in Ibragimov and Linnik (1971), the fact that $\{\eta_t\}$ is i.i.d. with unit mean and $\overline{g^{(j)}}=O_p\big(T^{-1/2}\big)=o_p(1)$. The remaining two terms in (86) can be treated analogously after applying Lemma A.10. This yields

$$
\nabla_{\hat{g}^B-g}\,\mathscr{G}_T^{(1)}=o_p\big(T^{-1/2}\big), \tag{90}
$$

$$
\nabla_{\hat{g}^B-g}\,\mathscr{G}_T^{(2)}=-\alpha_0\sum_{l=0}^{\infty}\beta_0^l\,\mathbb{E}\left[\frac{u_{t-(l+1)}^2}{g_{t-(l+1)}}\frac{v_t'}{v_t^2}\right]\sum_{j=1}^{J}\overline{g^{(j)}}+o_p\big(T^{-1/2}\big), \tag{91}
$$

[14]For the remainder of this appendix, $\|\cdot\|$ denotes the Euclidean norm.

where, for the proof of (91), we additionally need that $\mathbb{E}\left[\left\|v_t'\right\| u_{t-1-k}^2 | x_{t-2-k}^{(j)}\right] < \infty$. By Lemma A.9 and repeated application of the Cauchy-Schwarz inequality, it is straightforward to show that a necessary and sufficient condition for the latter is that $\mathbb{E}\left[u_t^4 | \mathscr{H}^{(x,j)}\right] < \infty$, $j = 1, \ldots, J$, where $\mathscr{H}^{(x,j)}$ is the sigma algebra generated by $\left\{x_{\infty}^{(j)}, \ldots, x_{-\infty}^{(j)}\right\}$. This condition is fulfilled by means of Assumption 2(e), which completes the proof. □

The final Lemma in this section shows the consistency of the sample Hessian given pre-estimated seasonality and semiparametric component.

Lemma A.7. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that $h = O\left(T^{-(1/4+\epsilon)}\right)$ for some $\epsilon > 0$. Then,*

$$\mathscr{H}_T\left(\bar{\theta}, \hat{g}, \hat{\mathscr{S}}_N\right) = \mathscr{H}(\theta_0, g, \mathscr{S}_N) + o_p(1), \tag{92}$$

where $\mathscr{H}_T\left(\bar{\theta}, \hat{g}, \hat{\mathscr{S}}_N\right) := T^{-1} \sum_{t=1}^{T} \partial^2 / \left(\partial \theta \partial \theta^T\right) l_t\left(\theta, \hat{g}, \hat{\mathscr{S}}_N\right)|_{\theta=\bar{\theta}}$, $\mathscr{H}(\theta_0, g, \mathscr{S}_N) := \mathbb{E}\left[\partial^2 / \left(\partial \theta \partial \theta^T\right) l_t(\theta, g, \mathscr{S}_N)|_{\theta=\theta_0}\right]$ *and* $\bar{\theta} := \lambda \hat{\theta} + (1-\lambda)\theta_0$ *for* $\lambda \in [0,1]$.

Proof. We can bound the estimation error of the Hessian by noting that

$$\begin{aligned}
\left\|\mathscr{H}_T\left(\bar{\theta}, \hat{g}, \hat{\mathscr{S}}_N\right) - \mathscr{H}(\theta_0, g, \mathscr{S}_N)\right\| &\leq \sup_{\theta \in \Theta} \left\|\mathscr{H}_T\left(\theta, \hat{g}, \hat{\mathscr{S}}_N\right) - \mathscr{H}_T(\theta, \hat{g}, \mathscr{S}_N)\right\| \\
&+ \sup_{\theta \in \Theta} \|\mathscr{H}_T(\theta, \hat{g}, \mathscr{S}_N) - \mathscr{H}_T(\theta, g, \mathscr{S}_N)\| \\
&+ \left\|\mathscr{H}_T\left(\bar{\theta}, g, \mathscr{S}_N\right) - \mathscr{H}(\theta_0, g, \mathscr{S}_N)\right\|, \\
&\leq o_p(1) + o_p(1) + o_p(1) = o_p(1),
\end{aligned} \tag{93}$$

where, for the second inequality, the bounds for the first and second term can be straightforwardly derived using Theorem 1 and 2, respectively, while the result for the third term follows, e.g., from part (vi) of the proof of Theorem 2.2 in Francq and Zakoian (2004). □

Basic Lemmas

The first lemma summarises important results on uniform convergence rates of uni- and bivariate kernel density estimators as well as the stochastic part of the univariate Nadaraya-Watson estimator. These are based on Masry (1996) and Mammen et al. (1999).

Lemma A.8. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that $h = O\left(T^{-(1/4+\epsilon)}\right)$ for some $\epsilon > 0$. Then, for $j, k = 1, \ldots, J$, $j \neq k$,*

$$\sup_{x^{(j)} \in I_h, x^{(k)} \in I_h} \left|\hat{p}^{(j,k)}\left(x^{(j)}, x^{(k)}\right) - p^{(j,k)}\left(x^{(j)}, x^{(k)}\right)\right| = O_p\left(\sqrt{\frac{\ln T}{T h^2}}\right) + O_p\left(h^2\right), \tag{94}$$

$$\sup_{x^{(j)}\in I_h}\left|\hat{p}^{(j)}\big(x^{(j)}\big)-p^{(j)}\big(x^{(j)}\big)\right|=O_p\left(\sqrt{\frac{\ln T}{Th}}\right)+O_p\big(h^2\big), \tag{95}$$

$$\sup_{x^{(j)}\in[0,1],x^{(k)}\in[0,1]}\left|\hat{p}^{(j,k)}\big(x^{(j)},x^{(k)}\big)-\tilde{p}^{(j,k)}\big(x^{(j)},x^{(k)}\big)\right|=O_p\left(\sqrt{\frac{\ln T}{Th^2}}\right)+O_p(h), \tag{96}$$

$$\sup_{x^{(j)}\in[0,1]}\left|\hat{p}^{(j)}\big(x^{(j)}\big)-\tilde{p}^{(j)}\big(x^{(j)}\big)\right|=O_p\left(\sqrt{\frac{\ln T}{Th}}\right)+O_p(h), \tag{97}$$

$$\sup_{x^{(j)}\in[0,1]}\left|\hat{g}^{(j),A}_{NW}\big(x^{(j)}\big)\right|=O_p\left(\sqrt{\frac{\ln T}{Th}}\right), \tag{98}$$

where $\tilde{p}^{(j,k)}\big(x^{(j)},x^{(k)}\big) := p^{(j,k)}\big(x^{(j)},x^{(k)}\big)\Big(\int_0^1 K_h\big(x^{(j)},v\big)\,dv\Big)\Big(\int_0^1 K_h\big(x^{(k)},v\big)\,dv\Big)$ *and* $\tilde{p}^{(j)}\big(x^{(j)}\big):=p^{(j)}\big(x^{(j)}\big)\int_0^1 K_h\big(x^{(j)},v\big)\,dv.$

The next lemma provides the gradient of observations of the unit-GARCH dynamics (4).

Lemma A.9. *For* $t=1,\dots,T$*,*

$$v'_t(\theta,g,\mathscr{S}_N):=\frac{\partial}{\partial\theta}v_t(\theta,g,\mathscr{S}_N)=\Big(v'_{\alpha,t}(\theta,g,\mathscr{S}_N),v'_{\beta,t}(\theta,g,\mathscr{S}_N)\Big)^T, \tag{99}$$

$$\begin{aligned} v'_{\alpha,t}(\theta,g,\mathscr{S}_N) &:= -1+\frac{r^2_{t-1}}{s_{t-1}g_{t-1}}+\beta\, v'_{\alpha,t-1}(\theta,g,\mathscr{S}_N), \qquad (100)\\ &= -\sum_{k=0}^{t-1}\beta^k+\sum_{k=0}^{t-1}\beta^k\frac{r^2_{t-1-k}}{s_{t-1-k}g_{t-1-k}},\\ v'_{\beta,t}(\theta,g,\mathscr{S}_N) &:= -1+v_{t-1}(\theta,g,\mathscr{S}_N)+\beta\, v'_{\beta,t-1}(\theta,g,\mathscr{S}_N), \qquad (101)\\ &= -\sum_{k=0}^{t-1}\beta^k+(1-\alpha-\beta)\sum_{k=0}^{t-1}\sum_{l=0}^{t-2-k}\beta^{k+l}\\ &\quad +\alpha\sum_{k=0}^{t-1}\sum_{l=0}^{t-2-k}\beta^{k+l}\frac{r^2_{t-2-(k+l)}}{s_{t-2-(k+l)}g_{t-2-(k+l)}}+\sum_{k=0}^{t-1}\beta^{t-1}v_0. \end{aligned}$$

Proof. (100) is trivial. For (101), use that $v_t(\theta,g,\mathscr{S}_N)=(1-\alpha-\beta)\sum_{k=0}^{t-1}\beta^k+\alpha\sum_{k=0}^{t-1}\beta^k\frac{r^2_{t-1-k}}{s_{t-1-k}g_{t-1-k}}+\beta^t v_0.$ □

The lemma below provides directional derivatives of observations of the unit-GARCH dynamics (4) and their gradient with respect to the semiparametric component g.

Lemma A.10. *For* $t=1,\dots,T$*,*

$$\begin{aligned} \nabla_{\Delta g}\,v_t(\theta,g,\mathscr{S}_N) &:= \frac{\partial}{\partial v}\,v_t(\theta_0,g+v\,\Delta g,\mathscr{S}_N)|_{v=0}, \qquad (102)\\ &= -\alpha\frac{r^2_{t-1}}{s_{t-1}g_{t-1}}\frac{\Delta g_{t-1}}{g_{t-1}}+\nabla_{\Delta g}\,v_{t-1}(\theta_0,g,\mathscr{S}_N), \end{aligned}$$

$$
\begin{aligned}
&= -\alpha \sum_{k=0}^{t-1} \beta^k \frac{r_{t-1-k}^2}{s_{t-1-k}\, g_{t-1-k}} \frac{\Delta g_{t-1-k}}{g_{t-1-k}}, \\
\nabla_{\Delta g}\, v_t'(\theta, g, \mathscr{S}_N) &:= \frac{\partial}{\partial \nu} v_t'(\theta_0, g + \nu \Delta g, \mathscr{S}_N)|_{\nu=0}, \qquad (103)\\
&= \left(\nabla_{\Delta g}\, v_{\alpha,t}'(\theta, g, \mathscr{S}_N), \nabla_{\Delta g}\, v_{\beta,t}'(\theta, g, \mathscr{S}_N)\right)^T, \\
\nabla_{\Delta g}\, v_{\alpha,t}'(\theta, g, \mathscr{S}_N) &= -\frac{r_{t-1}^2}{s_{t-1} g_{t-1}} \frac{\Delta g_{t-1}}{g_{t-1}} + \nabla_{\Delta g}\, v_{\alpha,t-1}'(\theta_0, g, \mathscr{S}_N), \qquad (104)\\
&= -\sum_{k=0}^{t-1} \beta^k \frac{r_{t-1-k}^2}{s_{t-1-k}\, g_{t-1-k}} \frac{\Delta g_{t-1-k}}{g_{t-1-k}}, \\
\nabla_{\Delta g}\, v_{\beta,t}'(\theta, g, \mathscr{S}_N) &:= \nabla_{\Delta g}\, v_{t-1}(\theta_0, g, \mathscr{S}_N) + \beta\, \nabla_{\Delta g}\, v_{\beta,t-1}'(\theta, g, \mathscr{S}_N), \qquad (105)\\
&= \alpha \sum_{k=0}^{t-1} \sum_{l=0}^{t-2-k} \beta^{k+l} \frac{r_{t-2-(k+l)}^2}{s_{t-2-(k+l)} g_{t-2-(k+l)}} \frac{\Delta g_{t-2-(k+l)}}{g_{t-2-(k+l)}},
\end{aligned}
$$

Proof. (102) is trivial, while (104) and (105) directly follow from Lemma A.9. □

The following lemma establishes uniform convergence rates for ratios involving the component estimators.

Lemma A.11. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that* $h = O\big(T^{-(1/4+\epsilon)}\big)$ *for some* $\epsilon > 0$. *Then,*

$$
\sup_{t=1,\dots,N} \left| \frac{\hat{s}_t - s_t}{\hat{s}_t} \right| = O_p\big(T^{-1/2}\big), \qquad (106)
$$

$$
\sup_{x \in [0,1]} \left| \frac{\hat{g}(x) - g(x)}{\hat{g}(x)} \right| = O_p(h), \qquad (107)
$$

$$
\sup_{\theta \in \Theta} \left| \frac{v_t\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big) - v_t(\theta, g, \mathscr{S}_N)}{v_t\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big)} \right| = O_p(h). \qquad (108)
$$

Proof. Use Theorems 1 and 2, Lemma A.12, as well as the fact that s_t, $g(x)$ and v_t are bounded from below uniformly in $t = 1, \dots, N$, $x \in [0,1]$ and $\theta \in \Theta$ according to Assumption 1(h), 1(i) and 2(b), respectively. □

The final lemma provides uniform rates for the observations of the unit GARCH process and its gradient based on pre-estimated seasonality and semiparametric component.

Lemma A.12. *Let Assumptions 1 and 2(b) be fulfilled. Further, assume that* $h = O\big(T^{-(1/4+\epsilon)}\big)$ *for some* $\epsilon > 0$. *Then, for* $t = 1, \dots, T$,

$$
\sup_{\theta \in \Theta} \left| v_t\big(\theta, \hat{g}, \hat{\mathscr{S}}_N\big) - v_t(\theta, g, \mathscr{S}_N) \right| = O_p(h), \qquad (109)
$$

$$\sup_{\theta\in\Theta}\left|v_t\big(\theta,\hat{g},\hat{\mathcal{S}}_N\big)-v_t(\theta,\hat{g},\mathcal{S}_N)\right|=O_p\big(T^{-1/2}\big), \tag{110}$$

$$\sup_{\theta\in\Theta}\left|v_t\big(\theta,\hat{g},\hat{\mathcal{S}}_N\big)-v_t\big(\theta,g,\hat{\mathcal{S}}_N\big)\right|=O_p(h), \tag{111}$$

$$\sup_{\theta\in\Theta}\left\|v_t'\big(\theta,\hat{g},\hat{\mathcal{S}}_N\big)-v_t'(\theta,\hat{g},\mathcal{S}_N)\right\|=O_p\big(T^{-1/2}\big), \tag{112}$$

$$\sup_{\theta\in\Theta}\left\|v_t'(\theta,\hat{g},\mathcal{S}_N)-v_t'(\theta,g,\mathcal{S}_N)\right\|=O_p(h). \tag{113}$$

Proof. For (109), we use

$$v_t(\theta,g,\mathcal{S}_N)=\beta^t v_0+(1-\alpha-\beta)\sum_{k=0}^{t-1}\beta^k+\alpha\sum_{k=0}^{t-1}\beta^k u_{t-1-k}^2(g,s_{t-1-k}), \tag{114}$$

to obtain

$$\begin{aligned}
\left|v_t\big(\theta,\hat{g},\hat{\mathcal{S}}_N\big)-v_t(\theta,g,\mathcal{S}_N)\right| &\le \alpha\sum_{k=0}^{\infty}\beta^k u_{t-1-k}^2(g,s_{t-1-k})\left|1-\frac{s_{t-1-k}g_{t-1-k}}{\hat{s}_{t-1-k}\hat{g}_{t-1-k}}\right|, \\
&\le \alpha\sum_{k=0}^{\infty}\beta^k u_{t-1-k}^2\left(\sup_{x\in[0,1]}\left|1-\frac{g(x_{t-1-k})}{\hat{g}(x_{t-1-k})}\right|\right. \\
&\left.+\sup_{x\in[0,1]}\left|\frac{g(x_{t-1-k})}{\hat{g}(x_{t-1-k})}\right|\sup_{t=1,\ldots,N}\left|1-\frac{s_{t-1-k}}{\hat{s}_{t-1-k}}\right|\right), \\
&=\alpha\sum_{k=0}^{\infty}\beta^k u_{t-1-k}^2\left(O_p(h)+O_p\big(T^{-1/2}\big)\right), \\
&=O_p(1)\,O_p(h)=O_p(h),
\end{aligned} \tag{115}$$

where, for the third inequality, we have used Lemma A.11 and Assumption 1(i), while for the last inequality, we have exploited that $\mathbb{E}\big[\sum_{k=0}^{\infty}\beta^k u_{t-1-k}^2\big]=\mathbb{E}[u_t^2]/(1-\beta)<\infty$, such that $\sum_{k=0}^{\infty}\beta^k u_{t-1-k}^2=O_p(1)$ by the Markov inequality. To prove (110) and (111), we proceed analogously, additionally using Assumption 1(h) in the latter case. For the proof of (112), note that by Lemma A.9

$$\begin{aligned}
\left|v_{\alpha,t}'\big(\theta,\hat{g},\hat{\mathcal{S}}_N\big)-v_{\alpha,t}'(\theta,\hat{g},\mathcal{S}_N)\right| &\le \sum_{k=0}^{\infty}\beta^k\frac{u_{t-1-k}^2}{\hat{g}_{t-1-k}}\left|\frac{s_{t-1-k}-\hat{s}_{t-1-k}}{\hat{s}_{t-1-k}}\right|, \\
&=O_p\big(T^{-1/2}\big),
\end{aligned} \tag{116}$$

and

$$\left|v_{\beta,t}'\big(\theta,\hat{g},\hat{\mathcal{S}}_N\big)-v_{\beta,t}'(\theta,\hat{g},\mathcal{S}_N)\right|\le\alpha\sum_{k=0}^{t-1}\sum_{l=0}^{t-2-k}\beta^{k+l}\frac{u_{t-2-(k+l)}^2}{\hat{g}_{t-2-(k+l)}} \tag{117}$$

$$
\begin{aligned}
&\times \left| \frac{s_{t-2-(k+l)} - \hat{s}_{t-2-(k+l)}}{\hat{s}_{t-2-(k+l)}} \right|, \\
&= O_p\left(T^{-1/2}\right),
\end{aligned}
$$

using Lemma A.11 and Assumption 1(i). Finally, (113) can be proven analogously, exploiting Assumption 1(h). □

B Additional Tables

Table 6: Ljung-Box Statistics. LB_l: Ljung-Box statistic associated with l lags. 5% (1%) critical values associated with lag lengths 30, 100 and 200: 43.78 (50.89), 124.34 (135.81) and 233.99 (249.45), respectively. DEP_{1-5} and DEP_{16-20}: average relative depth on levels 1 to 5 and 16 to 20, respectively. Results for (average) relative depth on levels 6 to 10 and 11 to 15 omitted due to space constraints, but available upon request.

	AAPL			GOOG		
	$LRET^2$	DEP_{1-5}	DEP_{16-20}	$LRET^2$	DEP_{1-5}	DEP_{16-20}
LB_{30}	1913.739	4091.917	1722.043	496.781	243.527	943.531
LB_{100}	3095.095	10093.756	3176.169	570.901	310.212	1815.279
LB_{200}	4040.890	16085.051	4125.625	706.499	480.058	2560.643
	FB			MSFT		
	$LRET^2$	DEP_{1-5}	DEP_{16-20}	$LRET^2$	DEP_{1-5}	DEP_{16-20}
LB_{30}	2606.008	616.182	1342.833	15462.293	4636.652	863.461
LB_{100}	3217.624	1051.275	2095.142	18160.858	5393.931	1575.020
LB_{200}	3711.725	1540.223	2842.626	23848.078	5667.031	2241.236

Table 7: Results of Augmented Dickey-Fuller Unit-Root-Test for Relative Depths. Lag length selected by SIC.

	$RDEPTH^{(1-5)}$		$RDEPTH^{(6-10)}$		$RDEPTH^{(11-15)}$		$RDEPTH^{(16-20)}$	
	DF	P_{DF}	DF	P_{DF}	DF	P_{DF}	DF	P_{DF}
AAPL	-14.150	0.000	-13.131	0.000	-16.364	0.000	-17.385	0.000
GOOG	-16.046	0.000	-17.490	0.000	-17.893	0.000	-17.517	0.000
FB	-14.907	0.000	-17.859	0.000	-18.701	0.000	-18.477	0.000
MSFT	-13.840	0.000	-13.924	0.000	-16.588	0.000	-15.061	0.000