Semiparametric Multiplicative GARCH-X Model: Adopting Economic Variables To Explain Volatility*

Heejoon Han† Dennis Kristensen‡

January 2015

Abstract

This paper investigates a multiplicative GARCH-X model, which has a nonparametric long run component induced by an exogenous covariate and a GARCH short run component. Compared to the usual additive GARCH-X model that includes an additional exogenous covariate in the GARCH model, this model contains a nonlinear function of an exogenous covariate that is multiplied to the GARCH model. When the covariate is nonstationary, i.e. integrated, near-integrated or fractionally integrated, the model can explain various stylized facts of financial time series. We suggest a kernel-based estimation procedure for the parametric and nonparametric components and derive related asymptotic properties. The asymptotic analysis is non-standard when the included covariates are non-stationary, and involves novel technique for nonparametric and semiparametric estimation. An empirical application studies the linkage between US and European stock market volatilities using the VIX index as a covariate in our multiplicative GARCH-X model. It is shown that the model outperforms standard models both in terms of in-sample fitting and out-of-sample forecasting.

1 Introduction

This paper considers multiplicative GARCH-X models where the over-all volatility of a given asset return has a semiparametric multiplicative structure. The volatility is assumed to be written as a product of two components: The first component is a standard parametric GARCH component that describes the usual short-run variation in volatility, while the second one is meant to capture long-run movements. We model the long-run volatility component as a nonparametric function of exogeneous economic and financial covariates; these should be informative about long-run factors that influence the financial markets. Our model is related to the so-called spline-GARCH model

---

*We would like to thank the participants of XXXX and seminars at York University for helpful comments and suggestions. Kristensen gratefully acknowledges research support by the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, the Danish National Research Foundation (through a grant to CREATE, University of Aarhus) and the European Research Council (grant no. ERC-2012-StG 312474).

†Sungkyunkwan University. E-mail: heejoonhan@skku.edu.

‡University College London. E-mail: d.kristensen@ucl.ac.uk.
proposed in Engle and Rangel (2008), where a similar structure is imposed but where the long-run component is treated as a deterministic function of time. In contrast, we here model it as a stochastic component.

We show that if the co-variates are persistent, our model can generate the same patterns as the spline-GARCH model, with the component involving additional covariates inducing long-run swings in the volatility. The long-run movements in volatility in our model can be given an economic interpretation (as function of economic variables). This is in contrast to the spline-GARCH model which is silent about the underlying factors generating the long-run swings found in data. These long-run swings can moreover be forecasted using standard methods within our framework; in contrast, the spline-GARCH model only provides an in-sample fit of the observed variation and cannot be used for forecasting purposes.

We develop estimators of the parameters in the GARCH component and the nonparametric function in the long-run component similar to the ones in Hafner and Linton (2010) who considers the case where the long-run component is treated as a deterministic function of time. We extend their asymptotic analysis to allow for random covariates with particular emphasis on the case where the covariate is a non-stationary, unit-root type process. To our knowledge, this is the first study of semiparametric estimators where non-stationary regressors are present in the nonparametric component. Our theory shows that this extension is non-trivial and involve non-standard techniques compared to the case of stationary co-variates, which follows along the lines of Hafner and Linton, 2010). In particular, a second-order expansion of the score function is needed to control for the first-step estimation of the nonparametric component, and we have to invoke non-standard limit results to derive the distributions of this expansion.

In our empirical application, we employ our multiplicative GARCH-X model to study volatility linkages between US and European stock markets, more specifically the British FTSE, French CAC and German DAX index, respectively. We do this by modelling each of the three stock indices using as covariate in our model the VIX index as a proxy for the over-all future volatility of the US markets. Our asymptotic analysis proves relevant in this application since the VIX is highly persistent and is often modelled as an near integrated or fractionally integrated process. For the GARCH component of the model, we adopt the GJR-GARCH specification to accommodate for leverage effects. For comparison, we also fit the GJR-GARCH(1,1) model without covariates and the spline-GARCH model using the estimators of Hafner and Linton (2010).

Compared to these two benchmark models we find the following: First, the fitted version of the standard GJR-GARCH model exhibits very strong persistence as in IGARCH. However, IGARCH disappears in our multiplicative GARCH-X model when including VIX as additional covariate since this captures the long-run movements in volatility. The estimator of the spline-GARCH model proves to be difficult to implement, and we often run into overfitting problems with the nonparametric component absorbing all variation. This is due to the fact that bandwidth selection in the kernel estimation of spline-GARCH models is a very delicate issue. In contrast, no such problems arise in the estimation of our multiplicative GJR-GARCH-X model with VIX as additional covariate.
We conduct within-sample and out-of-sample forecast evaluation of the model. We produce one-step ahead out-of-sample forecast based on the rolling window forecasting procedure with window of 1008 days (4 years). For forecast evaluation, we use realized kernel as the proxy for actual volatility and adopt the QLIKE loss function and the Diebold-Marinao (1995) and West (1996) test. Within-sample and out-of-sample forecast evaluations show that our model performs better than the GJR-GARCH model and the spline-GJR-GARCH model.

Our modelling approach is related to a recent literature on multiplicative volatility models, including SEMIGARCH (Feng, 2004), spline-GARCH model (Engle and Rangel, 2008; Hafner and Linton, 2010), time-varying GARCH model (Amado and Teräsvirta, 2013, 2014; Amado and Laakkonen, 2013) and GARCH-MIDAS (Engle et al., 2013; Conrad and Loch, 2014).

The analysis of the properties of our semiparametric GARCH-X model, when the regressors exhibit unit-root type behaviour, rely on extensions of recent limit results for models with non-stationary regressors as developed in Park and Phillips (2001) and Han and Kristensen (2014), amongst others.

The asymptotic theory of the proposed estimators take as starting point recent results on nonstationary nonparametric regression as found in, amongst others, Wang and Phillips (2009a,b), Wang and Chan (2014). However, we have to extend many of these results in order to employ them in our setting: First, a new uniform rate results for kernel estimators with non-stationary regressors has to be developed. Second, U-statistics results involving non-stationary sequences is needed. Third,

The rest of the paper is organized as follows. Section 2 introduces the model and analyzes its properties. Section 3 describes the estimation method and derives related asymptotic theories. A results of a simulation study is reported in Section 4, while the empirical application is presented in Section 5. Section 7 concludes. All proofs have been relegated to Appendix A, while tables and figures can be found in Appendix B. Before we proceed, a word on notation: Standard terminologies and notations employed in probability and measure theory are used throughout the paper. Notations for various convergences such as \( \rightarrow_{a.s.} \), \( \rightarrow_p \) and \( \rightarrow_d \) frequently appear, where all limits are taken as \( n \to \infty \) except where otherwise indicated.

2 Model and Properties

We observe \( \{y_t, x_t\} \) at time \( t = 1, 2, \cdots, n \) from the following multiplicative volatility model,

\[
y_t = \sigma_t \varepsilon_t \quad \text{where} \quad \sigma_t^2 = h_t f(x_{t-1}),
\]

where \( \{\varepsilon_t\} \) satisfies \( \mathbb{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0, \mathbb{E}[\varepsilon_t^2|\mathcal{F}_{t-1}] = 1 \), and \( \kappa_4 = \mathbb{E}[\varepsilon_t^4|\mathcal{F}_{t-1}] < \infty \), with \( \mathcal{F}_t = \mathcal{F}(\{y_{t-s}, x_{t-s}\}_{s=0}^{\infty}) \) denoting the usual filtration representing the information available at time \( t \). The overall volatility \( \sigma_t^2 = h_t f(x_{t-1}) \) is multiplicative with \( h_t \) capturing the short-run volatility of
the process and is assumed to satisfy a standard linear GARCH specification,

\[ h_t = 1 - \alpha - \beta + \alpha h_{t-1} \varepsilon_{t-1}^2 + \beta h_{t-1} \]  

(2)

for parameters \( \alpha, \beta > 0 \) such that \( \alpha + \beta < 1 \). We have here replaced the usual intercept with \( 1 - \alpha - \beta \) in order to normalize \( h_t \) to satisfy \( E[h_t] = 1 \) so that the long-run mean is determined by \( f(x_{t-1}) \) which represents long-run movements in the volatility: Assuming that \( \{x_t\} \) is strongly exogeneous in the sense that

\[ E[\varepsilon_t^2 | \mathcal{X}] = 1, \text{ where } \mathcal{X} = \{x_t\}, \]

we obtain

\[ E[y_t^2 | \mathcal{X}] = E[f(x_{t-1}) E[h_t | \mathcal{X}]] = f(x_{t-1}), \]

where the second equality follows from the fact that

\[ E[h_t | \mathcal{X}] = 1 - \alpha - \beta + (\alpha + \beta) E[h_{t-1} | \mathcal{X}] = \cdots = (1 - \alpha - \beta) \sum_{k=0}^{\infty} (\alpha + \beta)^k = 1. \]

Thus, conditional on \( \{x_t\} \), the variance of \( \{y_t\} \) is time-varying due to \( f(x_t) \). That is, \( x_t \) can be thought of a set of regressors summarizing the economic environment and other relevant factors affecting the long-run volatility. If \( \{x_t\} \) is stationary and ergodic, if will satisfy \( x_t \xrightarrow{d} x_\infty \) under great generality, where \( x_\infty \) denotes the stationary, long-run distribution of \( x_\infty \). Thus, in this case, the volatility does not exhibit long-run swings. On the other hand, if \( \{x_t\} \) is non-stationary, this will generate stochastic trends in the volatility.

In the existing literature on multiplicative volatility models (Engle and Rangel, 2008; Hafner and Linton, 2010), the regressor \( x_t \) is set to \( x_t = t/n \), and so describes deterministic time-trends in the volatility. We here allow for a broader class of economically relevant regressors, including, but not restricted to, deterministic time trends. This generalized model provides us with a testing ground for determining which (observed) factors that are relevant for explaining long-run movements in financial markets. Moreover, the model can be used for forecasting; this is in contrast to the standard spline-type GARCH models which only provides an in-sample fit of data.

One important type of non-stationary regressors, that will be the main focus of this paper, is the following class of (near) unit-root processes:

**Assumption 1.**

\[ x_t = \left(1 - \frac{c}{n}\right) x_{t-1} + v_t \]

for some \( c \geq 0 \) where \( \{v_t\} \) is independent of \( \{\varepsilon_t\} \).

As is well-known, many economic time series exhibit strong dependence and so Assumption 1 is an empirically relevant case. One of the main goals of this paper is then to examine how persistent covariates, such as the ones in Assumption 1, affect the properties of \( \{y_t\} \) and inference in the above semiparametric volatility model. Importantly, under Assumption 1 (together with
additional assumptions on \( \{ v_t \} \), the particular form of \( f \) will affect the time series properties of \( y_t \) and the asymptotic properties of the estimators.

We will consider two different function classes: The first contains regular integrable (I-regular) functions and the other regular asymptotically homogeneous (H-regular) functions as introduced by Park and Phillips (2001). The I-regular functions have finite integrals and the H-regular functions are asymptotically equivalent to homogeneous functions. In particular, the H-regular function \( f \) can be written as
\[
 f(\lambda x) \approx \kappa(\lambda) \overline{f}(x)
\]
for sufficiently large \( \lambda \) uniformly in \( x \) over any compact interval. \( \kappa \) and \( \overline{f} \) are called the asymptotic order and limit homogeneous function of \( f \), respectively. For example, if \( f(x) = a_0 + a_1 x^2 \), for some \( a_0, a_1 > 0 \), we have \( f(\lambda x) = a_0 + a_1 \lambda^2 x^2 \approx \kappa(\lambda) \overline{f}(x) \), with \( \kappa(\lambda) = \lambda^2 \) and \( \overline{f}(x) = a_1 x^2 \).

Readers are referred to Park (2002) and Park and Phillips (2001) for more details and examples.

We will now demonstrate that under Assumption 1, the multiplicative GARCH-X model can generate long-run stochastic trends in the volatility of \( y_t \). To this end, we consider the sample variance, kurtosis and covariance function, respectively, of \( y_t \) which are defined as
\[
 S_n^2 = \overline{y}_n^2, \quad K_n^4 = \frac{\overline{y}_n^4}{(\overline{y}_n^2)^2}, \quad \text{and} \quad R_{nk}^2 = \frac{\sum_{t=k+1}^{n} (y_t^2 - \overline{y}_n^2) (y_{t-k}^2 - \overline{y}_n^2)}{\sum_{t=1}^{n} (y_t^2 - \overline{y}_n^2)^2},
\]
where \( \overline{y}_n^p := \sum_{t=1}^{n} y_t^p / n, p \geq 2 \). For comparison purposes, recall that for the standard GARCH(1,1) process, which is obtained with \( f(x_{t-1}) = \overline{f} > 0 \) being constant, these sample statistics are known to have the following asymptotic constant limits if \( \mathbb{E} \left[ \left( \beta + \alpha \varepsilon_t^2 \right)^2 \right] < 1 \): \( S_n^2 \to^P S^2 = \omega \),
\[
 K_n^4 \to^P K^4 = \frac{(1 - (\alpha + \beta)^2) K^4}{1 - (\alpha^2 K^4 + 2 \alpha \beta + \beta^2)}, \quad (3)
\]
and
\[
 R_{nk}^2 \to^P R_k^2 = (\alpha + \beta)^{k-1} \frac{\alpha (1 - \alpha \beta - \beta^2)}{1 - 2 \alpha \beta - \beta^2}; \quad (4)
\]
c.f. XXXXX. The result in (3) shows that the sample variance converges to a nonrandom limit unless it is the IGARCH process and also the GARCH(1,1) process is leptokurtic. However, it is well known that the kurtosis implied by the GARCH(1,1) model with normally distributed innovations tends to be far less than the sample kurtosis observed in many financial return series. Moreover, the result in (4) shows that the autocorrelation decreases exponentially and quickly converges to zero. This means that the GARCH(1,1) model cannot explain the long memory property in volatility that is commonly observed in financial return series.

We now consider the case where \( f(x) \) is non-constant and \( \{ x_t \} \) is a unit-root type process. First, we analyze the multiplicative GARCH-X model with an H-regular function \( f \). To this end, we make the following assumptions:
Assumption 2

(i) \( \{v_t\} \) is generated by

\[
v_t = \varphi(L)\eta_t = \sum_{k=0}^{\infty} \varphi_k \eta_{t-k}, \tag{5}
\]

where \( \varphi_0 = 1, \varphi(1) \neq 0 \) with \( \sum_{k=0}^{\infty} k|\varphi_k| < \infty \), and \( \{\eta_t\} \) are i.i.d. random variables with mean zero and \( \mathbb{E}[|\eta_t|^p] < \infty \) for some \( p > 2 \).

(ii) \( \mathbb{E}[|\varepsilon_t|^q] < \infty \) for some \( q \geq 8 \).

(iii) \( \varpi := \mathbb{E}\left[(\beta + \alpha \varepsilon_t^2)^2\right] < 1 \).

Assumption 3

(i) \( \{v_t\} \) are i.i.d.

(ii) \( \mathbb{E}\left[f^2(x + \bar{v}_{k,t})\right] < \infty \) for all \( x \in \mathbb{R} \) and \( k \geq 1 \), where \( \bar{v}_{k,t} = v_{t+1} + \ldots + v_{t+k} \).

(iii) \( \mathbb{E}[|\varepsilon_t|^q] < \infty \) for some \( q \geq 8 \).

(iv) \( \{\varepsilon_t\} \) and \( \{v_t\} \) are mutually independent.

(v) \( v_t \) has distribution absolutely continuous with respect to Lebesgue measure, characteristic function \( \phi(t) \) such that \( t^r\phi(t) \to 0 \) as \( t \to \infty \) for some \( r > 0 \), and \( \mathbb{E}[|v_t|^p] < \infty \) for some \( p > 2 \).

(vi) Assumption 2(iii) holds.

Assumption 2 defines \( \{v_t\} \) as a general linear process. The moment condition in Assumption 2(ii) is for the asymptotic limit of the sample kurtosis to be well-defined. These are used in the analysis of the second and fourth order moments of If we are only interested in the sample variance, it can be relaxed to \( q \geq 4 \). Assumption 3 imposes additional restrictions on the model and is used to analyzeAssumption 3(i)-(v) are identical to Assumption 3W in Han and Park (2008).

Theorem 1 Let \( f \) be \( \mathbb{H} \)-regular with limit homogeneous function \( \tilde{f} \).

(i) Under Assumptions 1 and 2,

\[
\kappa(\sqrt{n})^{-1} S_n^2 \xrightarrow{d} \int_0^1 \tilde{f}(V_c(r))dr.
\]

(ii) Under Assumptions 1 and 2,

\[
K_n^4 \xrightarrow{d} \frac{(1 - (\alpha + \beta)^2)\kappa^4}{1 - \varpi} \frac{\int_0^1 \tilde{f}^2(V_c(r))dr}{\left(\int_0^1 \tilde{f}(V_c(r))dr\right)^2}.
\]
(iii) Under Assumptions 1 and 3, for \( k \geq 1 \),

\[
R_{nk}^2 \overset{d}{\to} R_k^2 = (\alpha + \beta)^{k-1} \frac{\alpha(1 - \alpha \beta - \beta^2)}{1 - 2\alpha \beta - \beta^2} (1 - A_H) + A_H
\]

where

\[
A_H = \frac{\int_0^1 \mathcal{J}(V_c(r))^2dr - \left(\int_0^1 \mathcal{J}(V_c(r))dr\right)^2}{\left(1 - (\alpha + \beta)^2\right)^{k^4} \int_0^1 \mathcal{J}(V_c(r))^2dr - \left(\int_0^1 \mathcal{J}(V_c(r))dr\right)^2}.
\]

Theorem 1 shows that the multiplicative GARCH-X process with \( \mathbb{H} \)-regular function is able to explain some stylized facts of financial time series such as nonstationarity, leptokurtosis and the long memory property in volatility. In Theorem 1(i), the asymptotic limit of the sample variance depends on the asymptotic order of the function \( f \). If \( \kappa(\sqrt{n}) \to \infty \) as \( n \to \infty \) (for example, power functions), the sample variance diverges as the sample size increases. In this case, the unconditional variance is not constant and it is compatible to the time varying unconditional variance of financial time series. On the other hand, if the asymptotic order \( \kappa \) is unity (for example, bounded \( \mathbb{H} \)-regular function), the asymptotic limit is finite.

Theorem 1(ii) shows that the limit of the sample kurtosis is the product of two parts. One part is the same as (3). Unless the limit homogeneous function \( \bar{f} \) is constant, the other part \( \int_0^1 \mathcal{J}^2(V_c(r))dr / (\int_0^1 \mathcal{J}(V_c(r))dr)^2 \) is random and its value is larger than unity due to the Cauchy-Schwarz inequality. This implies that the kurtosis of the multiplicative GARCH-X process is larger than that of the GARCH(1,1) process.

In Theorem 1(iii), the multiplicative GARCH-X process clearly exhibits the long memory property in volatility unless the limit homogeneous function \( \bar{f} \) is constant. As \( k \to \infty \), \( R_k^2 \) decreases exponentially at first and converges to \( A_H \) that does not depend on \( k \). \( A_H \) is random, positive and smaller than unity due to the Cauchy-Schwarz inequality. In this case, the trend of \( R_k^2 \) is quite similar to that commonly observed in financial time series, considering that it decreases quickly at first and converges to a positive random value. If the limit homogeneous function \( \bar{f} \) is constant, \( A_H \) is zero and therefore \( R_k^2 \) is identical to (4).

Theorem 1 shows that nonstationarity, leptokurtosis and the long memory property in volatility are adequately explained by the multiplicative GARCH-X model with an \( \mathbb{H} \)-regular function \( f \) such that its asymptotic order \( \kappa \) increases as sample size increases and its limit homogeneous function \( \bar{f} \) is not constant. These conditions are not restrictive. For example, a simple power function satisfies these conditions.

Next, we consider the multiplicative GARCH-X model with an \( \mathbb{I} \)-regular function:

**Assumption 2’**

(i) \( \{v_t\} \) is generated by (5) where \( \{\eta_t\} \) are iid random variable with distribution absolutely continuous with respect to Lebesgue measure, characteristic function \( \varphi(t) \) satisfying \( t^r \varphi(t) \to 0 \) as \( t \to \infty \) for some \( r > 0 \), and \( \mathbb{E}[|\eta_t|^p] < \infty \) for some \( p > 4 \).
(ii) Assumptions 2(ii) and 2(iii) hold.

**Assumption 3’**

(i) Assumption 3 holds with \( p > 4 \).

Assumptions 2’ and 3’ are very similar to Assumptions 3S-5S in Han and Park (2008). Under Assumption 3’, \( \hat{v}_{k,t} \) has density with respect to Lebesgue measure on \( \mathbb{R} \), and we signify the density by \( p_k \).

**Theorem 2** Let \( f \) be \( I \)-regular.

(i) Under Assumptions 1 and 2’,

\[
\frac{1}{\sqrt{2}} S_n^2 \xrightarrow{d} L_c(1, 0) \int_{-\infty}^{\infty} f(s) ds.
\]

(ii) Under Assumptions 1 and 2’,

\[
\frac{1}{\sqrt{2}} K_n^4 \xrightarrow{d} \frac{(1 - (\alpha + \beta)^2)k^4}{1 - \varpi} \frac{\int_{-\infty}^{\infty} f^2(s) ds}{(\int_{-\infty}^{\infty} f(s) ds)^2},
\]

(iii) Under Assumptions 1 and \( \beta’ \), for \( k \geq 1 \),

\[
R_{nk}^2 \xrightarrow{d} R_k^2 = \frac{\alpha + \beta}{\kappa^2} B_I + (1 - \varpi) \frac{\int_{-\infty}^{\infty} f(x)f(x+y)p_k(y)dy}{\int_{-\infty}^{\infty} f^2(s) ds}
\]

where

\[
B_I = (1 + \alpha + \beta)(\beta + \alpha k^4) - (\alpha + \beta)(1 - \varpi).
\]

Theorem 2 shows that the multiplicative GARCH-X process with an \( I \)-regular function exhibits different time series properties than the process with an \( H \)-regular function. In Theorem 2(i), the sample variance converges in probability to zero as sample size increases. In Theorem 2(ii), \( n^{-1/2}K_n^4 \) has the asymptotic limit that is larger than (3) due to the Cauchy-Schwarz inequality. The sample kurtosis diverges at the rate of \( \sqrt{n} \) as \( n \to \infty \) and therefore is expected to be larger as sample size increases.

Theorem 2(iii) shows that \( R_k^2 \to 0 \) as \( k \to \infty \). This is because

\[
N_k = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(x+y)p_k(y)dydx \to 0
\]

because \( p_k(\cdot) \to 0 \) as \( k \to \infty \) under mild conditions as shown in Park (2002). \( R_k^2 \) is determined by the distribution of \( (v_t) \) as well as the function \( f \). If \( (v_t) \) is normally distributed, \( N_k \) diminishes very slowly at a hyperbolic rate \( k^{-1/2} \), as shown in Corollary 2 in Han and Park (2008), and therefore the long memory is explained.
Additive GARCH-X

The standard way of incorporating exogeneous regressors in GARCH models is to model the volatility in an additive manner,

\[ \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + f(x_{t-1}), \]  

which is typically called the GARCH-X model. As the function \( f(\cdot) \), researchers commonly use simple parametric specifications; For example, \( f(x_{t-1}) = \pi x_{t-1} \) when the covariate is always positive or, more generally, \( f(x_{t-1}) = \pi x_{t-1}^2 \); see Han and Kristensen (2014). More complicated specifications could in principle be allowed for, but estimation of (6) then becomes numerically difficult. If one wants to allow for full flexibility of \( f(\cdot) \) and estimate it nonparametrically, we are not aware of any estimation methods.

Moreover, for the theoretical analysis of the additive GARCH-X with non-stationary \( \{x_t\} \), the function \( f(\cdot) \) has to be restricted further in order to obtain results since \( \sigma_{t-1}^2 \) in (6) contains past values of \( f(x_{t-1}) \) which complicates matters (see, e.g., Hand and Park, 2012).

Assumption 4

(i) Assumption 2 holds with some \( q > 8 \).

(ii) \( \mathbb{E} (\beta + \alpha \varepsilon_t^2)^{q/2} < 1 \) and \( 1/p + 2/q < 1/2 \).

Assumption 4 is generally identical to Assumption 2 in Han and Park (2012). We let \( q > 8 \) instead of \( q > 4 \) for the analysis of \( K_n^4 \) and \( R_n^2 \).

Assumption 5

(i) \( f \) is asymptotically homogeneous with asymptotic order \( \kappa \) such that \( \inf_{x \in \Pi} \kappa(\lambda, \pi) \to \infty \) as \( \lambda \to \infty \).

(ii) \( f \) has an equivalent asymptotically homogeneous function \( \bar{f} \), say, such that

\[ |\bar{f}(x, \pi) - \bar{f}(y, \pi)| \leq \nabla f(z, \pi) |x - y| \]

for any \( x, y \) in a compact subset of \( \mathbb{R} \), where \( x \leq z \leq y \) and \( \nabla f \) is asymptotically homogeneous with asymptotic order \( \nabla \kappa \) such that

\[ \nabla \kappa(\lambda, \pi) = O(\kappa(\lambda, \pi)/\lambda) \]

for all \( \pi \in \Pi \), as \( \lambda \to \infty \).

Power functions satisfy Assumption 5 while the logistic function does not.

Corollary 3 Let \( y_t = \sigma_t \varepsilon_t \) for \( \sigma_t^2 \) defined in (6) with \( \alpha + \beta < 1 \). Then, under Assumptions 1(ii), 4 and 5, the asymptotic limits of \( S_n^2 \), \( R_{nk}^2 \) and \( K_n^4 \) are identical to those in Theorem 1, respectively.
This shows that

**IGARCH**

Han and Park (2014, Theorem 2) analyze the effect of omitting \( f(x_{t-1}) \) in (6). They show that if the relevant covariate \( f(x_{t-1}) \) in (6) is omitted and the usual GARCH(1,1) model is fitted, the model will be estimated approximately as the IGARCH model. If we analyze the effect of omitting \( f(x_{t-1}) \) in (1), we can obtain the same result: Suppose the true data-generating process is the multiplicative GARCH-X process defined in Assumptions 1, 4 and 5, but that we fit the standard GARCH(1,1) model to data,

\[
2_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2,
\]

using the quasi-maximum likelihood estimation method. We denote the QMLE’s of the parameters \( \alpha \) and \( \beta \) in (7) by \( \hat{\alpha}_n \) and \( \hat{\beta}_n \).

**Corollary 4** Let Assumptions 1, 4 and 5 holds with some \( q > 4 \). Then the result of Theorem 2 in Han and Park (2014) holds. More details...

### 3 Estimation

Given a sample \((y_t, x_{t-1}), t = 1, ..., n\), we wish to develop an estimator of the parameters of the model together with the function \( f(x) \). We here impose no parametric specifications on \( f \) and instead treat it as a nonparametric object. We let \( \theta_0 = (\alpha_0, \beta_0) \) and \( f_0 \) denote the data-generating parameters and function respectively, and let \( h_{0,t} \) denote the corresponding true GARCH volatility process.

Next, for estimation purposes, we introduce the normalized process \( w_t(f) \) indexed by \( f \),

\[
w_t(f) = \frac{y_t}{f(x_{t-1})}.
\]

In particular, \( w_{0,t} := w_t(f_0) \) solves a standard parametric GARCH model with normalized second moment,

\[
w_{0,t} = \sqrt{h_{0,t}t}, \quad h_{0,t} = 1 - \alpha_0 - \beta_0 + \alpha_0 w_{0,t-1}^2 + \beta_0 h_{0,t-1}.
\]

To estimate \( \theta \) and \( f(x) \), we adopt the following procedure which is similar to the semiparametric estimator developed in Hafner and Linton (2007): For given \( f \) and \( \theta \), the global log-likelihood takes the form

\[
L_n(\theta, f) = \frac{1}{n} \sum_{t=1}^n \ell_t(\theta, f), \quad \ell_t(\theta, f) := \log(f(x_{t-1}) h_t(\theta)) + \frac{y_t^2}{f(x_{t-1}) h_t(\theta, f)},
\]

where

\[
h_t(\theta, f) = 1 - \alpha - \beta + \alpha w_{t-1}^2(f) + \beta h_{t-1}(\theta, f).
\]

We then propose the following iterative procedure, where \( \theta \) and \( f \) are estimated in turn:
1. Estimate \( f(x) \) as the nonparametric regression function in
\[
y_t^2 = f(x_{t-1}) + u_t, \quad u_t := f(x_{t-1}) \left( h_t \varepsilon_t^2 - 1 \right).
\]
That is,
\[
\hat{f}(x) = \frac{\sum_{t=1}^{n} K_h(x_{t-1} - x) y_t^2}{\sum_{t=1}^{n} K_h(x_{t-1} - x)},
\] (11)
2. Given \( \hat{f}(x) \), estimate \( \theta \) by
\[
\hat{\theta} = \arg \max_{\theta \in \Theta} L_n(\theta, \hat{f}).
\]
3. Given \( \tilde{h}_t = h_t(\hat{\theta}, \hat{f}) \), re-estimate \( f(x) \) as the nonparametric regression function in
\[
y_t^2 \tilde{h}_t = f(x_{t-1}) + \tilde{u}_t,
\]
That is,
\[
\hat{f}(x) = \frac{\sum_{t=1}^{n} K_h(x_{t-1} - x) y_t^2 / \tilde{h}_t}{\sum_{t=1}^{n} K_h(x_{t-1} - x)}.
\]
4. Given \( \hat{f}(x) \), re-estimate \( \hat{\theta} \) by
\[
\hat{\theta} = \arg \max_{\theta \in \Theta} L_n(\theta, \hat{f}).
\]
One can keep iterating until convergence of the algorithm is achieved in a suitable sense. However, as we shall see, this will not lead to any first-order improvements of the estimator; the two-step procedure suffices.

The above kernel regression estimator of \( f(x) \) in Step 1 could be replaced by the following alternative estimator, as proposed in Chen, Cheng and Peng (2009) (see also Yu and Jones, 2004):
\[
\tilde{f}(x) = \exp \left[ \tilde{g}(x) \right] / \tilde{m},
\]
where
\[
\tilde{g}(x) = \frac{\sum_{t=1}^{n} K_h(x_{t-1} - x) \log y_t^2}{\sum_{t=1}^{n} K_h(x_{t-1} - x)}, \quad \tilde{m} = \left( \frac{1}{n} \sum_{t=1}^{n} y_t \exp \left[ -\tilde{g}(x_t) \right] \right)^{-1}
\]
are the estimators of \( g(x) = \log \left[ f(x) \right] m \) and \( m := E \left[ \log \left( h_t \varepsilon_t^2 \right) \right] \) in the following regression,
\[
\log y_t^2 = g(x_{t-1}) + \log \left( h_t \varepsilon_t^2 \right) - m.
\]
Similarly, the kernel regression estimator in Step 3 could be replaced by
\[
\tilde{f}(x) = \exp \left[ \tilde{g}(x) \right] / \tilde{m},
\]
where
\[
\hat{g}(x) = \frac{\sum_{t=1}^{n} K_h(x_{t-1} - x) \log(y_t^2/h_t(\hat{\theta}))}{\sum_{t=1}^{n} K_h(x_{t-1} - x)} , \quad \hat{m} = \left( \frac{1}{n} \sum_{t=1}^{n} \frac{y_t}{h_t(\hat{\theta})} \exp[-\hat{g}(x_t)] \right)^{-1}.
\]

The advantage of this alternative estimator is that it only requires log-moment of \(y_t\), while the standard kernel regression estimator requires 4th moment of \(y_t\).

Also, Step 4 can be replaced by a Newton-Raphson iteration,
\[
\hat{\theta} = \tilde{\theta} - \left( \frac{\partial S^*_n(\tilde{\theta}, \hat{f})}{\partial \theta} \right)^{-1} S^*_n(\tilde{\theta}, \hat{f}),
\]
where \(S^*_n(\theta, f) = \sum_{t=1}^{n} s^*_t(\theta, f) / n\) and
\[
s^*_t(\theta, f) = \left[ \frac{\hat{h}_t(\theta, f)}{\hat{h}_t(\theta, f)} - E \left[ \frac{\hat{h}_t(\theta, f)}{\hat{h}_t(\theta, f)} \right] \right] \left\{ 1 - \frac{y_t^2}{f(x_{t-1}) h_t(\theta, f)} \right\}.
\]

This is the efficient score function, as derived in Hafner and Linton (2010).

### 3.1 Asymptotic Theory

In the first step, we obtain an kernel regression estimator of \(f(x_{t-1})\) as the regression function in eq. (10). The regression error \(u_t = f(x_{t-1})(h_t e_t^2 - 1)\) is not a Martingale difference sequence. This complicates the asymptotic theory of the estimator \(\hat{f}\) given in eq. (11). In particular, we have to take into account the autocorrelation in \(u_t\). The existing literature analyzing the asymptotic properties of kernel estimators with non-stationary regressors have focused on the case of non-autocorrelated errors, and so we have to extend these existing results.

In the theory of the two-step QMLE we wish to allow for both stationary and non-stationary regressors. To this end, we impose the following two high-level conditions inspired by Wang and Chan (2014):

**Assumption 6.** There exists a function \(R(n, h), \gamma > 0\) and \(k_0 > 0\) such that with \(b = O(n^{k_0})\):

\[
\sup_{\|x\| \leq b} \sum_{t=1}^{n} f_t^2 \tilde{K}_{h, t}^2(x) = O_P \left( R^2(n, h) \right),
\]

\[
\sup_{\|x\| \leq b} \left| \sum_{t=1}^{n} f_t \tilde{K}_{h, t}(x) - f(x) \right| = O_P \left( h^\gamma \right),
\]

where \(\tilde{K}_{h, t}(x) = K_h(x_t - x) / \sum_{s=1}^{n} K_h(x_s - x)\).

Taken together, the two conditions of Assumption 6 imply:
Lemma 5. Under Assumptions 1, 2(i) and 6,

$$\sup_{\|x\| \leq b} |\tilde{f}(x) - f(x)| = O_P(h^\gamma) + O_P(R(n,h)).$$

As the following theorem shows, Assumption 6 covers both the case of stationary regressors and the case where $x_t$ is non-stationary $x_t$:

Theorem 6. Suppose that XXXX. Then:

1. If $\{x_t\}$ is stationary and strongly mixing with ..... Then Assumption 6 holds with $\gamma = 2$ and $R^2(n,h) = \log n/(nh^d)$.

2. If $\{x_t\}$ is a unit root process with ..... Then Assumption 6 holds with $\gamma = ?$ and $R^2(n,h) = \log n/(\sqrt{nh})$.

3. If $x_t = (\tilde{x}_t, t/n)$, where $\{\tilde{x}_t\}$ is stationary and strongly mixing..... Then Assumption 6 holds with $\gamma = 2$ and $R^2(n,h) = \log n/(nh^d)$.

If the regression errors in eq. (10) were uncorrelated over time, the second part of the above theorem would follow directly from Wang and Chan (2014). However, this is not the case and so we instead rely on novel uniform rate results for non-stationary kernel regressions with autocorrelated errors.

We now proceed to analyze the first step estimator of the parametric component, $\tilde{\theta}$, where we will employ the above uniform rate result for $\tilde{f}$. However, the uniform rate result is only established on the (expanding) set $\{x : \|x\| \leq b\}$, with $b = O(n^{h_0})$, which complicates the analysis since we cannot control $\tilde{f}$ outside of this set. A popular way of circumventing this problem in the analysis of two-step semiparametric estimator is to either assume that the regressor $x_t$ has bounded support or to introduce fixed trimming of the log-likelihood function. However, an important case that we would like to cover in our analysis is when $x_t$ is a unit-root type process for which it holds that, for any fixed $b > 0$, $P(\|x_t\| \leq b) \rightarrow 0$ as $t \rightarrow \infty$. Thus, these two solutions are not feasible in our set up. Instead, we introduce the following modified estimator of $f$,

$$\tilde{f}_b(x) = \tilde{f}(x)I\{\|x\| \leq b\} + f_1I\{\|x\| > b\},$$

for some constant $f_1 > 0$ and where $I\{\cdot\}$ denotes the indicator function. The inclusion of $f_1 > 0$ is introduced here since $f$ appears in the denominator and the log-term of the Gaussian log-likelihood, and so we need to ensure that its estimator is bounded away from zero with probability approaching 1 (w.p.a.1). We will then let $b \rightarrow \infty$ at a suitable rate consistent with the restrictions in Theorem 6 so that $P(\|x_t\| \leq b) \rightarrow 0$. We then modify the estimator $\tilde{\theta}$ to employ its trimmed version instead of $\tilde{f}$ itself,

$$\tilde{\theta}_b = \arg\max_{\theta \in \Theta} L_n(\theta, \tilde{f}_b).$$

The following theorem shows that the estimator is consistent as long as the trimming vanishes asymptotically:
Theorem 7 Under Assumptions 1, 2(i) and 6, \( \hat{\theta}_n \overset{P}{\to} \theta_0 \), for any \( h, b \to 0 \) such that \( R(n, h) \to 0 \).

Next, we wish to derive the large-sample distribution of the estimator. The derivation of this for two-step semiparametric estimators proceeds as follows: By a Taylor expansion of \( L_n(\theta, \hat{f}_b) \) w.r.t. \( \theta \) we obtain

\[
0 = S_n(\theta, \hat{f}_b) = S_n(\theta_0, \hat{f}_b) + H_n(\theta, \hat{f}_b)(\theta - \theta_0),
\]

where \( \theta \) is on the line connecting \( \theta \) and \( \theta_0 \), \( S_n(\theta, f) = \sum_{t=1}^n s_t(\theta, f) / n \) is the score and \( H_n(\theta, f) = \sum_{t=1}^n h_t(\theta, f) / n \) is the hessian of the log-likelihood \( L_n(\theta, f) \) w.r.t. \( \theta \). Here, we would expect that, by similar arguments as in the proof of Theorem 7, \( H_n(\theta, \hat{f}_b) \to^P H(\theta_0, f) \), where \( H(\theta, f) = E \left[ \frac{\partial^2 L(\theta, f)}{\partial \theta \partial f} \right] \). Next, to take into account the presence of \( \hat{f}_b \) in the score function, we expand this w.r.t. \( f \) and obtain the following first-order functional Taylor expansion,

\[
S_n(\theta_0, \hat{f}_b) = S_n(\theta_0, f) + \nabla S_n[\hat{f}_b - f] + R_n, \tag{13}
\]

where \( \nabla S_n[d f] \) is the pathwise differential of \( S_n(\theta_0, f) \) w.r.t. \( f \) in the direction \( d f \), and \( R_n \) is the remainder term. It is easily shown that \( \sqrt{n} S_n(\theta_0, f) \to^d N(0, \Omega_0(\theta_0, f)) \), where \( \Omega(\theta, f) = E \left[ \frac{\partial^2 L(\theta, f)}{\partial \theta \partial f} \right] \), while \( R_n = O_P \left( ||\hat{f}_b - f||^2_{\infty} \right) \). Thus, for the remainder term to have no impact on the asymptotic distribution, we therefore need \( ||\hat{f}_b - f||^2_{\infty} = O_P (1/\sqrt{n}) \). However, in the case where \( x_t \) is a unit root process this cannot be achieved since here \( \sqrt{n} R^2(n, h) = \log n / h \to +\infty \) as \( h \to 0 \) and \( n \to \infty \). Thus, in order to analyze the non-stationary case, we instead have to use a second-order expansion,

\[
\nabla^2 S_n[d f] \text{ is the second-order pathwise differential of } S_n(\theta_0, f) \text{ w.r.t. } f \text{ in the direction } d f, \text{ while now } R_n = O_P \left( ||\hat{f}_b - f||^3_{\infty} \right) \text{ and so}
\]

\[
\sqrt{n}||\hat{f}_b - f||^3_{\infty} = O_P \left( \sqrt{n} h^{3\gamma} \right) + O_P \left( \sqrt{n} \left\{ \log n / \left( n^{1/2} \sqrt{n} \right) \right\}^3 \right)
\]

\[
= O_P \left( \sqrt{n} h^{3\gamma} \right) + O_P \left( (\log n)^{3/2} / \sqrt{\left( \sqrt{n} h^3 \right) } \right),
\]

and we see that if \( \gamma > 1 \), we can achieve that both the bias and variance component of \( \hat{f}_b \) vanishes fast enough so that the second order expansion suffices. It’s however unclear whether we really need a higher-order expansion - if, for some \( \alpha < 1/2 \),

\[
n^\alpha \nabla S_n[\hat{f}_b - f] \to^d MN(0, V),
\]

where \( MN \) stands for mixed-normal, and \( R_n \) in eq. (13) is of order \( o_P (1/n^\alpha) \), then we obtain

\[
n^\alpha S_n(\theta_0, \hat{f}_b) = n^\alpha \nabla S_n[\hat{f}_b - f] + o_P (1) \to^d MN(0, XXX),
\]

\[
n^\alpha S_n(\theta_0, \hat{f}_b) = n^\alpha \nabla S_n[\hat{f}_b - f] + o_P (1) \to^d MN(0, XXX),
\]
and so the asymptotic distribution of the QMLE is determined alone by the first-step estimator $\tilde{f}$,

$$n^0(\tilde{\theta}_b - \theta_0) \xrightarrow{d} MN \left(0, H_0^{-1}VH_0^{-1}\right).$$

This is in contrast to the standard case where both $S_n(\theta_0, f)$ and $\nabla S_n[\tilde{f}_b - f]$ contribute to the asymptotic distribution with $\sqrt{n}$.

**Theorem 8** Suppose Assumptions 1, 2(i) and 6 hold and $h, b \to 0$ such that XXXX. Then:

1. In the stationary case,

$$\sqrt{n}(\tilde{\theta}_b - \theta_0) \xrightarrow{d} N \left(0, H_0^{-1}\Omega_0 H_0^{-1}\right),$$

where $H_0 = H(\theta_0, f_0)$, $\Omega_0 = \Omega(\theta_0, f_0)$ and

\[
H(\theta, f) = E \left[ \frac{\partial \ell_t(\theta, f)}{\partial \theta} \frac{\partial \ell_t(\theta, f)}{\partial \theta'} \right], \quad \Omega(\theta, f) = E \left[ \frac{\partial \ell_t(\theta, f)}{\partial \theta} \frac{\partial \ell_t(\theta, f)}{\partial \theta'} \right] + \Sigma(\theta, f),
\]

\[
\Sigma(\theta, f) = \sum_{t=-\infty}^{+\infty} \text{Var}(\psi_0(\theta, f), \psi_t(\theta, f)),
\]

\[
\psi_t = \frac{h_t}{h_t^2} \left\{ 1 - \varepsilon_t^2 \right\} + \left\{ E \left[ \frac{h_t}{h_t} \right] - \alpha_0 \sum_{k=0}^{\infty} \beta_k E \left[ w_{t-k}^2 \frac{h_t^2}{h_t^2} \right] \right\} \left\{ w_t^2 - 1 \right\}. \tag{15}
\]

2. In the non-stationary case (under Assumption 1),

$$\sqrt{n}(\tilde{\theta}_b - \theta_0) \xrightarrow{d} N \left(0, XXXXX\right),$$

where XXXXX.

4 A Simulation Study

TBC

5 Empirical Application

5.1 The Data and Models

We consider three daily European stock index return series from 2 January 2004 to 30 December 2013: FTSE (2457 trading days), CAC (2492 trading days), and DAX (2474 trading days). Figure 1 shows graphs of the return series. We demean each return series by subtracting its sample mean which is close to zero. We use the demeaned return series as \{y_t\}.

As the covariate $x_t$ for our semiparametric multiplicative GARCH-X model, we use the VIX index by the Chicago Board Options Exchange. The VIX index is the implied volatility calculated from options on the S&P 500 index. Considering the dominance of the US stock market in the world
and its influence on the European stock markets, it is expected that the VIX index significantly accounts for volatilities of the European stock index return series. The estimated autoregressive coefficient is $0.983$ and the ADF and the KPSS tests indicate that the VIX index can be modeled as a near-integrated process.

Since we study stock return series, we adopt a GJR-GARCH version of our model where

$$h_t(\theta) = XX\cdots X + \alpha h_{t-1} \varepsilon^2_{t-1} + \gamma h_{t-1} \varepsilon^2_{t-1} I(\sqrt{h_{t-1} \varepsilon_{t-1}} < 0) + \beta h_{t-1}. $$

In this way, we can accommodate for the leverage effect. We also estimate the following benchmark models;

$$\sigma^2_t = \omega + \alpha y^2_{t-1} + \gamma y^2_{t-1} I(y_{t-1} < 0) + \beta \sigma^2_{t-1},$$

$$\sigma^2_t = f(x_{t-1}),$$

$$\sigma^2_t = \omega + \alpha y^2_{t-1} + \gamma y^2_{t-1} I(y_{t-1} < 0) + \beta \sigma^2_{t-1} + \pi x_{t-1}. $$

The first benchmark model is the GJR-GARCH(1,1) model, while the second one is a fully non-parametric volatility model as in Han and Zhang (2012). Our model combines these two models in a multiplicative manner and so it is relevant to examine whether such a combination leads to better within-sample fitting and out-of-sample forecasting. The third benchmark model is the additive GJR-GARCH-X model (GJR-X henceforth) in order to investigate whether the multiplicative version performs better than the usual additive GARCH-X model.

For the GJR-GARCH and GJR-X models, we use the Gaussian quasi-maximum likelihood estimation method. For the nonparametric volatility model and the nonparametric component of our model, we use the Nadaraya-Watson kernel estimation method using the Gaussian kernel. For these models, we adopt the cross-validation bandwidth selection method that is designed to minimize the QLIK loss function. For our model, we choose the bandwidth to minimize the following QLIK loss function;

$$h_{CV} = \arg \min_h \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{y_t^2}{h_t(\hat{\theta}) \hat{f}_{t-l}(x_{t-1})} + \log \left( h_t(\hat{\theta}) \hat{f}_{t-l}(x_{t-1}) \right) \right\}$$

where $\hat{f}_{t-l}(x_{t-1})$ is the ‘leave-one-out’ estimator. For the nonparametric volatility model, we choose the bandwidth in the same way by letting $h_t(\hat{\theta}) = 1$ in the above equation.

We compare the within-sample and out-of-sample predictive power of the volatility models. To evaluate volatility forecast, we adopt the following procedure; First, as the proxy for actual volatility, we use the realized kernel, introduced by Barndorff-Nielsen et al. (2008), because it has some robustness to market microstructure effects. The realized kernel of the daily S&P 500 index return series is available at the database ‘Oxford-Man Institute’s realised library’ produced by Heber et al. (2009).
Second, we use the QLIKE loss function defined as

\[
L(\hat{\sigma}_t^2, \sigma_t^2) = \frac{\hat{\sigma}_t^2}{\sigma_t^2} - \log \frac{\hat{\sigma}_t^2}{\sigma_t^2} - 1
\]  

(16)

where \(\sigma_t^2\) is the proxy for actual volatility and \(\hat{\sigma}_t^2\) is the within-sample or out-of-sample forecast. Even if realized measures are known to be better measures, they are imperfect and noisy proxies for actual volatility. There has been research on loss functions that are robust to the use of a noisy volatility proxy. See Hansen and Lunde (2006), Patton (2010) and Patton and Sheppard (2009). Patton (2010) shows that the QLIKE loss function is robust and, in particular, Patton and Sheppard (2009) shows in their simulation study that the QLIKE loss function has the highest power.

Third, the significance of any difference in the QLIKE loss is tested via a Diebold-Marinao and West (henceforth DMW) test (see Diebold-Marinao(1995) and West (1996). A DMW statistic is computed using the difference in the losses of two models

\[
d_t = L(\hat{\sigma}_{t,1}^2, \sigma_t^2) - L(\hat{\sigma}_{t,2}^2, \sigma_t^2)
\]

\[
DMW_T = \frac{\sqrt{T}d_T}{\sqrt{\text{avar}(\sqrt{T}d_T)}}
\]  

(17)

where \(\bar{d}_T\) is the sample mean of \(d_t\) and \(T\) is the number of forecasts. The asymptotic variance of the average is computed using a Newey-West variance estimator with the number of lags set to \([T^{1/3}]\).

5.2 Estimation Results

Table 1 reports the estimation results of the GJR-GARCH, the GJR-X, and the parametric component of our model. We begin by fitting the GJR-GARCH model. The persistence measure \(\hat{\alpha} + \hat{\gamma}/2 + \hat{\beta}\) are 0.990 (FTSE), 0.982 (CAC) and 0.981 (DAX), which are close to unity. This result suggests the IGARCH for the return series. However, the IGARCH disappears in the multiplicative GARCH-X model; \(\hat{\alpha} + \hat{\gamma}/2 + \hat{\beta}\) are 0.930 (FTSE), 0.928 (CAC) and 0.948 (DAX). This corresponds to what we investigate on IGRACH in Section 2; If the relevant covariate \(f(x_{t-1})\) is omitted, the model will be estimated approximately as the IGARCH. The results for the GJR-X model shows that VIX index is significant when it is added in the additive form.

Figure 2 shows the estimated function \(f(x_{t-1})\) for each return series. The VIX index, used as \(x_t\), is ranged from 9.89 to 80.86 in our sample. Figure 2 displays the mapping of \(f(x_{t-1})\) into the grid of values \(\{x_{t-1} = 10 + k; \ k = 0, 1, \cdot \cdot \cdot, 70\}\). Bandwidths are chosen to be 2.572 (FTSE), 3.047 (CAC) and 3.207 (DAX). The estimated functions are different across the return series, which means that each market responds to the VIX index in a different way. In each case, the estimated function is nonlinear and is not a monotonic increasing function. Simple parametric functions can hardly represent the complexities of the estimated functions. This supports the idea of nonparametric
estimation of the function \( f(x_{t-1}) \) in the multiplicative GARCH-X model.

Figure 3 presents sample autocorrelations. In column(a), autocorrelations of squared return series \( y_t^2 \) are given. They clearly show the long memory property in volatility. However, autocorrelations of the rescaled squared return series \( y_t^2 / \hat{f}(x_{t-1}) \) given in column(b) do not exhibit the long memory property. While autocorrelations in column(a) stay positive for large lags, those in column(b) immediately converge to zero. This result corresponds to our asymptotic result in Section 2; The long memory property is generated due to \( f(x_{t-1}) \) in the multiplicative GARCH-X model.

Figure 4 shows the estimated conditional variance of the GJR-GARCH(1,1) model for the original return series \( y_t \) and for the rescaled return series \( y_t / \hat{f}(x_{t-1})^{1/2} \). Graphs in column(b) look more stationary compared to those in column(a). This is because the change of the unconditional variance is absorbed by the nonstationary component \( f(x_{t-1}) \) in the multiplicative GARCH-X model.

5.3 Forecast Evaluation Results

Table 2 contains the within-sample forecast evaluation result based on the QLIKE loss function. In this case, \( \hat{\sigma}_t^2 \) in (16) denotes the fitted values of the volatility models for the entire sample period. Our model shows the smallest QLIKE for all stock return series. Moreover, the DMW test results show that the null hypotheses of equal loss between our model and the rest models are mostly rejected at 1% significance level.

We also evaluate the out-of-sample forecasts. We adopt the rolling window forecast procedure with moving windows of four years (1008 trading days). We obtain one-step ahead forecasts of the models for the period from 4 January 2010 to 30 December 2013. In this case, \( \hat{\sigma}_t^2 \) in (16) now denotes one-period ahead volatility forecasts at time \( t-1 \) and \( T = 982 \) (FTSE), 997 (CAC) or 992 (DAX) in (17). For our model, we use the cross validation bandwidth chosen in the within-sample case.

Table 3 reports QLIKE’s of the models and the DMW test statistics. Similarly as the previous within-sample case, our model shows the smallest QLIKE for all cases. According to the DMW test, the null hypotheses of equal loss between our model and the rest models are mostly rejected at 1% or 5% significance level except for DAX.

Within-sample and out-of-sample forecast evaluations show that our model performs better than the benchmark models. Our model and the GJR-X model outperform the GJR-GARCH model, which means that the inclusion of the VIX index improves within-sample fitting and out-of-sample forecasting. More importantly, our model performs better than the GJR-X model and this indicates that it is worth to let the functional form of the exogenous covariate be flexible to improve volatility fitting and forecasting.
6 Extension to Employ More Economic Variables

The previous sections consider the model using an univariate covariate. Since there can be various relevant economic variables to explain volatility of financial time series, it is desirable to develop a model that can accommodate more than one covariate in the nonparametric component. For this, we adopt a semiparametric single index model as following:

\[ \sigma_t^2 = h_t(\theta) f(x_{t-1}^\prime \beta) \]  

where the vector \( x_t \) is dimension 2 or more and \( \|\beta\| = 1 \), where \( \|\beta\| \) is the Euclidean norm of \( \beta \). This is a semiparametric single index multiplicative GARCH-X model and we call it the single index model henceforth. (better name?) To estimate the single index model, we basically follow the iterative procedure described in Section 4. For semiparametric estimation of \( f(x_{t-1}^\prime \beta) \), the estimate of \( \beta \) and bandwidth are chosen by cross-validation method minimizing the QLIK loss function defined Section 5.1. (adding details on estimations?)

We apply the semiparametric single index model to the data in the previous section. The covariate \( x_t \) now consists of the VIX index, each country’s industrial production index and crude oil price (Europe Brent spot price). Since industrial production indeces are in monthly frequency, they are linearly interpolated. We adopt industiral production and crude oil price because these are are known to be related to stock market in the literature. Engle et al. (2013) and Conrad and Loch (2014) use industiral production as a macroeconomic variable in the GARCH-MIDAS model to explain stock market volatility. For works on the relationship between stock markets and crude oil market, readers are referred to Ciner (2001), Jones and Kaul (2996) and Sadorsky (1999) among others.

The application result shows that there is certainly a room for improvement by employing more relevant economic variables and the single index model is an approach worth adopting. It will be desirable to establish asymptotic distribution of the estimator of the single index model, but asymptotic technique for the semiparametric single index model with nonstationary covariates is not available yet. Therefore, we leave it as future research and we do not conduct statistical inference on the parameter \( \beta \) in the following.

The estimates of \( \beta \) are reported in Table 4. In each case, the coefficient of the VIX index is estimated to be the largest while that of the oil price is the smallest. In Table 4, we also report each estimate divided by the standard deviation of the corresponding covariate. In all three cases, the coefficient of the VIX index has the same sign as that of the oil price. For CAC and DAX, the coefficient of industrial production has an opposite sign from the coefficients of the VIX index and the oil price. This result and the estimated \( \hat{f}(x_{t-1}^\prime \hat{\beta}) \) show that the industrial production is negatively related with stock return volatility while the VIX index and the oil price are positively related. Considering that stock market volatility is in general higher in economic recession, this result corresponds to our expectations. However, for FTSE, the coefficient of industrial production exhibits the same sign as the coefficients of the VIX index and the oil price. This somewhat puzzling
result seems to be due to the different situation of the UK economy. As shown in Figure 5, while the industrial production in France and Germany recovered, to some extent, after recession in 2009, that in UK was even lower in 2012 than 2009. The industrial production index is the lowest in April 2009 in France (93.1) and Germany (85.6), but it is the lowest in October 2012 in UK (94.2). In France and Germany, stock markets became less volatile and industrial production rebounded after 2009. However, in UK, the stock market became less volatile but the industrial production was still low after 2009. This may explain why we obtained a different result for FTSE in terms of the coefficient sign of industrial production.

The persistence measures $\hat{\alpha} + \hat{\gamma}/2 + \hat{\beta}$ for the single index model are estimated to be 0.889 (FTSE), 0.893 (CAC) and 0.936 (DAX), which are lower than those for the multiplicative GARCH-X model using only the VIX index. This implies that the nonparametric component in the single index model (by employing more economic variables) now better captures the persistent aspect of volatility. We examine whether employing more variables leads to better in-sample fitting and out-of-sample forecasting.

Table 5 reports the predictive power of the single index model. Compared to the multiplicative GARCH-X model using only the VIX index in the previous section, the single index model shows improvement in within-sample fitting. The QLIKE losses are lower and the DMW tests show that they are significant. For out-of-sample forecasting, the single index model provides lower QLIKE losses than the model using only the VIX index, while the DMW tests show that it is significant only for FTSE.

7 Conclusion

TBC
A Proofs

A.1 Proofs of Section 2

Proof of Theorem 1. If $\omega := \mathbb{E}[ (\beta + \alpha \varepsilon_t^2)^2 ] < 1$, it is known for the GARCH(1,1) process that

\[
\mathbb{E}(h_t\varepsilon_t^2) = \frac{\omega}{1 - \alpha - \beta},
\]

\[
\mathbb{E}(h_t^2\varepsilon_t^4) = \frac{1 + (\alpha + \beta) \omega^k}{1 - (\alpha + \beta) (1 - \omega)},
\]

and

\[
\mathbb{E}(h_t h_{t-k}\varepsilon_t^2\varepsilon_{t-k}^2) = \omega^2 \left( \frac{1 - (\alpha + \beta)^k}{(1 - (\alpha + \beta))^2} + \frac{1 + \alpha + \beta}{1 - (\alpha + \beta)} \frac{(\alpha + \beta)^{k-1} (\beta + \alpha \kappa^4)}{1 - \omega} \right).
\]

Using the results in Park (2002, proof of Theorem 1) and Han and Park (2008, Lemmas 1-3), we have

\[
(n \kappa(\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^2 = (n \kappa(\sqrt{n}))^{-1} \sum_{t=1}^{n} \mathbb{E}(h_t\varepsilon_t^2) f(x_{t-1}) + o_p(1)
\]

\[
\quad \rightarrow d\mathbb{E}(h_t\varepsilon_t^2) \int_0^1 \bar{f}(V_c(r))dr,
\]

\[
(n \kappa^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^4 = (n \kappa^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} \mathbb{E}(h_t^2\varepsilon_t^4) f(x_{t-1})^2 + o_p(1)
\]

\[
\quad \rightarrow d\mathbb{E}(h_t^2\varepsilon_t^4) \int_0^1 \bar{f}^2(V_c(r))dr,
\]

and

\[
(n \kappa^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^2 y_{t-k}^2 = (n \kappa^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} \mathbb{E}(h_t h_{t-k}\varepsilon_t^2\varepsilon_{t-k}^2) f(x_{t-1})f(x_{t-k-1}) + o_p(1)
\]

\[
\quad \rightarrow d\mathbb{E}(h_t h_{t-k}\varepsilon_t^2\varepsilon_{t-k}^2) \int_0^1 \bar{f}^2(V_c(r))dr.
\]

From these results for the sample moments, we have

\[
(n \kappa^2(\sqrt{n}))^{-1} \sum_{t=k+1}^{n} (y_t^2 - \bar{y}_n^2) (y_{t-k}^2 - \bar{y}_n^2)
\]

\[
= (n \kappa^2(\sqrt{n}))^{-1} \sum_{t=k+1}^{n} y_t^2 y_{t-k}^2 - (\kappa(\sqrt{n})^{-1} \bar{y}_n^2)^2 + o_p(1)
\]

\[
\rightarrow d\mathbb{E}(h_t h_{t-k}\varepsilon_t^2\varepsilon_{t-k}^2) \int_0^1 \bar{f}^2(V_c(r))d - \left( \mathbb{E}(h_t\varepsilon_t^2) \int_0^1 \bar{f}(V_c(r))dr \right)^2.
\]
and
\[
(n \kappa^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} (y_t^2 - \bar{y}_n^2)^2
\]
\[
= (n \kappa^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^4 - (\kappa(\sqrt{n})^{-1} \bar{y}_n^2)^2 + o_p(1)
\]
\[
\rightarrow d\mathbb{E}(h_t^2 \varepsilon_t^4) \int_0^1 \tilde{f}^2(V_c(r))dr - \left( \mathbb{E}(h_t^2 \varepsilon_t^2) \int_0^1 \tilde{f}(V_c(r))dr \right)^2
\]
from which the stated result in part (iii) follows.

The stated result in part (i) is proven above and the stated result in part (ii) is followed by

\[
K_n^4 = (n \kappa^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^4 / \left( (n \kappa(\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^2 \right)^2.
\]

Proof of Theorem 2. We obtain the following results from Park (2002, proof of Theorem 1) and Han and Park (2008, Lemmas 1-3);

\[
n^{-1/2} \sum_{t=1}^{n} y_t^2 = n^{-1/2} \sum_{t=1}^{n} \mathbb{E}(h_t^2 \varepsilon_t^2) f(x_{t-1}) + o_p(1)
\]
\[
\rightarrow d\mathbb{E}(h_t^2 \varepsilon_t^2) L_c(1,0) \int_{-\infty}^{\infty} f(s)ds,
\]

\[
n^{-1/2} \sum_{t=1}^{n} y_t^4 = n^{-1/2} \sum_{t=1}^{n} \mathbb{E}(h_t^2 \varepsilon_t^4) f(x_{t-1})^2 + o_p(1)
\]
\[
\rightarrow d\mathbb{E}(h_t^2 \varepsilon_t^4) L_c(1,0) \int_{-\infty}^{\infty} f^2(s)ds,
\]

and

\[
n^{-1/2} \sum_{t=1}^{n} y_t^2 y_{t-k}^2 = (n \kappa_0^2(\sqrt{n}))^{-1} \sum_{t=1}^{n} \mathbb{E}(h_t h_{t-k} \varepsilon_t^2 \varepsilon_{t-k}^2) f(x_{t-1})f(x_{t-k-1}) + o_p(1)
\]
\[
\rightarrow d\mathbb{E}(h_t h_{t-k} \varepsilon_t^2 \varepsilon_{t-k}^2) L_c(1,0) \int_{-\infty}^{\infty} \mu_k f(s)ds
\]

where \( \mu_k = \int_{-\infty}^{\infty} f(x+y)p_k(y)m(dy) \).
From these results, we deduce that

\[
 n^{-1/2} \sum_{t=k+1}^{n} \left( y_t^2 - \bar{y}_n^2 \right) \left( y_{t-k}^2 - \bar{y}_n^2 \right) = n^{-1/2} \sum_{t=k+1}^{n} y_t^2 y_{t-k}^2 + O_p(n^{-1/2})
\]

\[
\rightarrow \quad d\mathbb{E} \left( h_t h_{t-k} \varepsilon_t^2 \varepsilon_{t-k}^2 \right) L_c (1, 0) \int_{-\infty}^{\infty} \mu_k f(s) ds
\]

and

\[
 n^{-1/2} \sum_{t=1}^{n} \left( y_t^2 - \bar{y}_n^2 \right)^2 = n^{-1/2} \sum_{t=1}^{n} y_t^4 + O_p(n^{-1/2})
\]

\[
\rightarrow \quad d\mathbb{E} \left( h_t^2 \varepsilon_t^4 \right) L_c (1, 0) \int_{-\infty}^{\infty} f^2(s) ds
\]

from which the stated result in part (iii) follows. The stated result in part (i) is proven above and the stated result in part (ii) is followed by

\[
 n^{-1/2} K_n^4 = n^{-1/2} \sum_{t=1}^{n} y_t^4 \left/ \left( n^{-1/2} \sum_{t=1}^{n} y_t^2 \right)^2 \right.
\]

Proof of Corollary 3. We can deduce from Lemma A of Han and Park (2012) that \( \sigma_t^2 (\theta) \) in (6) is asymptotically approximated by \( z_t f(x_{t-1}) \) where

\[
z_t = 1 + \sum_{i=1}^{\infty} \prod_{h=1}^{i} (\beta + \alpha \varepsilon_{t-h}^2).
\]

(19)

It is also known that \( h_t (\theta) \) in (2) is asymptotically approximated by \( \omega z_t \).

From Lemmas A and B in Han and Park (2012), it follow that

\[
(n \kappa (\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^2 = n^{-1} \sum_{t=1}^{n} \left( \kappa (\sqrt{n})^{-1} z_t \varepsilon_t \right)^2 + o_p(1)
\]

\[
\rightarrow \quad d\mathbb{E} \left( z_t^2 \varepsilon_t^2 \right) \int_{0}^{1} \tilde{f}(V_c(r)) dr,
\]

\[
(n \kappa^2 (\sqrt{n}))^{-1} \sum_{t=1}^{n} y_t^4 = n^{-1} \sum_{t=1}^{n} \left( \kappa (\sqrt{n})^{-1} z_t \varepsilon_t \right)^2 \varepsilon_t^4 + o_p(1)
\]

\[
\rightarrow \quad d\mathbb{E} \left( z_t^2 \varepsilon_t^4 \right) \int_{0}^{1} \tilde{f}^2(V_c(r)) dr,
\]
and

\[(n\kappa^2(\sqrt{n}))^{-1} \sum_{t=k+1}^{n} y_t^2 y_{t-k} \]

\[= n^{-1} \sum_{t=k+1}^{n} (\kappa(\sqrt{n})^{-1} z_t f(x_{t-1})) \left( \kappa(\sqrt{n})^{-1} z_{t-k} f(x_{t-k-1}) \right) \varepsilon_t^2 \varepsilon_{t-k}^2 + o_p(1) \]

\[= n^{-1} \sum_{t=k+1}^{n} (z_t z_{t-k} \varepsilon_t^2 \varepsilon_{t-k}^2) \left( \kappa(\sqrt{n})^{-1} f(x_{t-k-1}) \right)^2 + o_p(1) \]

\[\rightarrow \quad d\mathbb{E} \left( z_t z_{t-k} \varepsilon_t^2 \varepsilon_{t-k}^2 \right) \int_0^1 \hat{f}^2(V_c(r)) dr. \]

In particular, the above third line follows because we have from lemma B in Han and Park (2012) that

\[\kappa(\sqrt{n})^{-1} z_t f(x_{t-1}) = \kappa(\sqrt{n})^{-1} z_t f(x_{t-k-1}) + o_p(1) \]

uniformly in \( t = 1, \ldots, n \). This is a simpler approach than the way by Park (2002) and Han and Park (2008). See the proof of theorem 1 in both papers. While they consider more general functions including integrable functions, we consider here only asymptotically homogeneous functions satisfying Assumption 5.

Since \( \mathbb{E} \left( h_t \varepsilon_t^2 \right) = \omega \mathbb{E} \left( z_t \varepsilon_t^2 \right) \), \( \mathbb{E} \left( h_t^2 \varepsilon_t^4 \right) = \omega^2 \mathbb{E} \left( z_t^2 \varepsilon_t^4 \right) \), and \( \mathbb{E} \left( h_t h_{t-k} \varepsilon_t^2 \varepsilon_{t-k}^2 \right) = \omega^2 \mathbb{E} \left( z_t z_{t-k} \varepsilon_t^2 \varepsilon_{t-k}^2 \right) \), the stated results follows. \( \blacksquare \)

**Proof of Corollary 4.** We write the true data-generating process by \( y_t = \sigma_{t,0} \varepsilon_t \) for \( \sigma_{t,0} = h_t (\theta_0) f(x_{t-1}) \) where \( \theta_0 \) and \( \pi_0 \) are true parameters. It follows from Lemma B in Han and Park (2012) that, for all \( i \geq 1 \) uniformly in \( t = 1, \ldots, n \),

\[ \beta^{i-1} \kappa(\sqrt{n})^{-1} \sigma_{t-i,0}^2 = \beta^{i-1} \left\{ \kappa(\sqrt{n})^{-1} f(x_{t-1}) \right\} z_{t-i} + o_p(1) \]

where \( \kappa(\sqrt{n}, \pi_0) \) is the asymptotic order of \( f(x_t, \pi_0) \). Therefore,

\[ \beta^{i-1} \kappa(\sqrt{n}, \pi_0)^{-1} y_{t-i}^2 = \beta^{i-1} \left\{ \kappa(\sqrt{n})^{-1} f(x_{t-1}) \right\} m_{t-i} + o_p(1) \]

for \( m_t = z_t \varepsilon_t^2 \) with \( z_t \) given as (2). Then the equations (24)-(26) in Han and Park (2014) hold and the stated result follows. \( \blacksquare \)

### A.2 Proofs of Section 3

**Proof of Theorem 6.** We verify the conditions of Theorem XXXX of Kanay and Kristensen (2014) from which the result will follow in conjunction with Wang and Chan (2014). \( \blacksquare \)

**Proof of Theorem 7.** We first show that \( \sup_{\theta \in \Theta} | L_n(\theta, \hat{f}_b) - L_{n,b}(\theta, f_0) | \rightarrow \mathcal{L}, \) where \( L_n(\theta, f) \)
is defined in eq. (8). To this end, first note that

\[
|L_n(\theta, \tilde{f}_b) - L_n(\theta, f_0)| \leq \frac{1}{n} \sum_{t=1}^{n} \left| \log \tilde{f}_{b,t} - \log f_{0,t} \right| + \frac{1}{n} \sum_{t=1}^{n} \left| \log \tilde{h}_{b,t}(\theta) - \log h_t(\theta) \right|
\]

\[
+ \frac{1}{n} \sum_{t=1}^{n} y_t^2 \left| \frac{1}{\tilde{h}_{b,t}(\theta) \tilde{f}_{b,t}} - \frac{1}{h_t(\theta) f_{0,t}} \right|
\]

\[
= : A_1 + A_2(\theta) + A_3(\theta),
\]

where \( \tilde{h}_{b,t}(\theta) = h_t(\theta, \tilde{f}_b), h_t(\theta) = h_t(\theta, f_0), \tilde{f}_{b,t} := \tilde{f}_b(x_{t-1}) \) and \( f_{0,t} = f_0(x_{t-1}) \). Also, define \( f_{b,t} = f_b(x_{t-1}) \), where

\[
f_b(x) := f_0(x) \mathbb{I} \{ \| x \| \leq b \} + f_1 \mathbb{I} \{ \| x \| > b \}.
\]

It follows from Lemma 9, that \( A_1 = \mathcal{O}_P(a_n) \), while Lemma 10 together with the mean-value theorem and the fact that \( \tilde{h}_{b,t}(\theta) \) and \( h_t(\theta) \) both are bounded from below by \( \omega_U \) imply that

\[
\sup_{\theta \in \Theta} |A_2(\theta)| \leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} \left| \log \tilde{h}_{b,t}(\theta) - \log h_t(\theta) \right|
\]

\[
\leq \frac{1}{\omega_U n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} \left| \tilde{h}_{b,t}(\theta) - h_t(\theta) \right|
\]

\[
= \mathcal{O}_P(h^\gamma) + \mathcal{O}_P(R(n, h)) + \mathcal{O}_P \left( \frac{1}{n b^{\lambda_0}} \sum_{t=1}^{n} t^{\lambda_0} \right)
\]

\[
= \mathcal{O}_P(h^\gamma) + \mathcal{O}_P(R(n, h)) + \mathcal{O}_P \left( \frac{n^{\lambda_0}}{b^{\lambda_0}} \right),
\]

where we have used that \( \sum_{t=1}^{n} t^\gamma \leq \gamma n^{\gamma+1} \). Finally, by similar arguments,

\[
\sup_{\theta \in \Theta} |A_3(\theta)| \leq \frac{1}{n} \sum_{t=1}^{n} w_t^2 \left| \frac{f_t}{\tilde{h}_{b,t}(\theta) \tilde{f}_{b,t}} - \frac{1}{h_t(\theta)} \right|
\]

\[
\leq \frac{1}{n} \sum_{t=1}^{n} w_t^2 \left\{ \left| \frac{f_t - \tilde{f}_{b,t}}{\tilde{h}_{b,t}(\theta) \tilde{f}_{b,t}} \right| + \left| \frac{1}{\tilde{h}_{b,t}(\theta) \tilde{f}_{b,t}} - \frac{1}{h_t(\theta)} \right| \right\}
\]

\[
\leq \frac{1}{n} \sum_{t=1}^{n} w_t^2 \left\{ ||f_b - \tilde{f}_b||_\infty + f_{0,t} \| \|_T + \frac{1}{\omega_L^2} \left| \tilde{h}_{b,t}(\theta) - h_t(\theta) \right| \right\}
\]

\[
= \mathcal{O}_P(h^\gamma) + \mathcal{O}_P(R(n, h)) + \frac{1}{n} \sum_{t=1}^{n} w_t^2 f_{0,t} \| \|_T + \mathcal{O}_P \left( \frac{n^{\lambda_0}}{b^{\lambda_0}} \right),
\]
Proof of Theorem 8.

where

\[
E \left[ \frac{1}{n} \sum_{t=1}^{n} w_{t}^{2} f_{0,t} \right] = E \left[ w_{t}^{2} \right] \frac{1}{n} \sum_{t=1}^{n} E \left[ f_{0,t} \right] \leq E \left[ w_{t}^{2} \right] \frac{1}{n b_{k_0}} \sum_{t=1}^{n} E \left[ \| x_{t-1} \| \gamma + k_0 \right] 
\]

\[
\leq C \frac{1}{n b_{k_0}} \sum_{t=1}^{n} \epsilon_{h_0} = O \left( \frac{n^{\lambda_0}}{b_{k_0}} \right).
\]

Substituting the above bounds into the right hand side of eq. (20), we obtain

\[
\sup_{\theta \in \Theta} |L_{n,b}(\theta, \tilde{f}) - L_{n,b}(\theta, f_0)| = O_P \left( h_0^{\gamma} \right) + O_P \left( R(n, h) \right) + O_P \left( \frac{n^{\lambda_0}}{b_{k_0}} \right) = o_P(1),
\]

where the last equation follows from the restrictions imposed on \( h \) and \( b \) in the theorem.

Next, write \( L_{n}(\theta, f_0) = L_{n}^{*}(\theta) + \sum_{t=1}^{n} \log f_{0}(x_{t-1}) / n \) where \( L_{n}^{*}(\theta) = \sum_{t=1}^{n} \ell_{t}^{*}(\theta) / n \) and

\[
\ell_{t}^{*}(\theta) := \log h_{t}(\theta) + \frac{w_{t}^{2}}{h_{t}(\theta)}.
\]

First, note that \( \sum_{t=1}^{n} \log f_{0,t} / n \) does not depend on \( \theta \) and so can be ignored in the following. Second, observe that \( L_{n}^{*}(\theta) \) is the negative log-likelihood function for a standard parametric GARCH model and so we can apply existing results for the QMLE of standard GARCH models (see, e.g. Han and Kristensen, 2014, proof of Theorem 1): \( \theta \mapsto \ell_{t}(\theta) \) is continuous; \( L_{n}^{*}(\theta) / n \rightarrow^{P} L^{*}(\theta) := \mathbb{E}[\ell_{t}^{*}(\theta)] \) where the limit exists, \( \forall \theta \in \Theta ; L^{*}(\theta_0) > L^{*}(\theta), \forall \theta \neq \theta_0 \); and \( \mathbb{E}[\sup_{\theta \in \Theta} \ell_{t}^{*}(\theta)] < +\infty \). This combined with eq. (20) completes the proof.

**Proof of Theorem 8.**

with

\[
s_{t}(\theta, f) = \hat{h}_{t}(\theta, f) \left\{ 1 - \frac{y_{t}^{2}}{f(x_{t-1}) h_{t}(\theta, f)} \right\},
\]

\[
h_{t}(\theta, f) = \hat{h}_{t}(\theta, f) \left\{ 1 - \frac{y_{t}^{2}}{f(x_{t-1}) h_{t}(\theta, f)} \right\} + 2 \frac{\hat{h}_{t}(\theta, f) h_{t}(\theta, f)'}{h_{t}^{2}(\theta, f)} f(x_{t-1}) h_{t}(\theta, f).
\]

Here, \( \hat{h}_{t}(\theta, f) = \left( \hat{h}_{\alpha,t}(\theta, f), \hat{h}_{\beta,t}(\theta, f) \right)' \) and

\[
\hat{h}_{t}(\theta, f) = \left[ \begin{array}{c}
\hat{h}_{\alpha,t}(\theta, f) \\
\hat{h}_{\beta,t}(\theta, f)
\end{array} \right]
\]

are the first and second order derivatives of \( h_{t}(\theta, f) \) w.r.t. \( \theta \) with \( \hat{h}_{\alpha,t}(\theta, f) = 0 \) and

\[
\hat{h}_{\alpha,t}(\theta, f) = -1 + w_{t-1}^{2}(f) + \beta \hat{h}_{\alpha,t-1}(\theta, f),
\]

\[
\hat{h}_{\beta,t}(\theta, f) = -1 + h_{t-1}(\theta, f) + \beta \hat{h}_{\beta,t-1}(\theta, f),
\]

\[
\hat{h}_{\alpha,t}(\theta, f) = h_{t-1}(\theta, f) + \beta \hat{h}_{\alpha,t-1}(\theta, f),
\]

\[
\hat{h}_{\beta,t}(\theta, f) = 2 h_{t-1}(\theta, f) + \beta \hat{h}_{\beta,t-1}(\theta, f).
\]

26
In the following, let \( h_t = h_t(\theta_0, f_0) \) and similarly for its derivatives. Next, we expand \( S_n(\theta_0, \tilde{f}_b) \) w.r.t. \( \tilde{f}_b \): For any function \( df(x) \), let \( \nabla S_n[\vec{f}_b] := \sum_{t=1}^n \nabla s_t[\vec{f}_b] / n \) denote the pathwise derivative of \( S_n(\theta_0, f) \) w.r.t. \( f \) in the direction \( df \), where

\[
\nabla s_t[\vec{f}_b] = \frac{\nabla h_t[\vec{f}_b]}{h_t} \left\{ 1 - \frac{y_t^2}{f_{0,t}h_t} \right\} + \frac{\dot{h}_t}{h_t^2} \nabla h_t[\vec{f}_b] \left\{ 2 \frac{y_t^2}{f_{0,t}(x_{t-1})} - 1 \right\} + \frac{\dot{h}_t y_t^2}{f_{0,t}h_t^3} df_t
\]

and \( \nabla h_t[\vec{f}_b] = (\nabla h_{\alpha,t}[\vec{f}_b], \nabla h_{\beta,t}[\vec{f}_b])' \) with

\[
\nabla h_t[\vec{f}_b] = -\alpha_0 w_{t-1}^2 \frac{df_{t-1}}{f_{0,t-1}} + \beta_0 \nabla h_{t-1}[\vec{f}_b],
\]

\[
\nabla h_{\alpha,t}[\vec{f}_b] = -w_{t-1}^2 \frac{df_{t-1}}{f_{0,t-1}} + \beta_0 \nabla h_{\alpha,t-1}[\vec{f}_b],
\]

\[
\nabla h_{\beta,t}[\vec{f}_b] = \nabla h_{t-1}[\vec{f}_b] + \beta_0 \nabla h_{\beta,t-1}[\vec{f}_b].
\]

We claim that, with \( a_t = ... \) and \( b_t = ... \),

\[
TBC : \left\| S_n(\theta_0, \tilde{f}_b) - S_n(\theta_0, f) + \nabla S_n[\vec{f}_b - f] \right\| = O_P \left( \| \vec{f}_b - f \|_\infty \right) + \sum_{t=1}^n a_t f_{0,t}^{-1} h_t^{-1},
\]  

\[
TBC : \left\| H_n(\vec{\theta}, \tilde{f}_b) - H_n(\vec{\theta}, f) \right\| = O_P \left( \| \vec{f}_b - f \|_\infty \right) + \sum_{t=1}^n b_t f_{0,t}^{-1} h_t^{-1}, H_n(\theta_0, f),
\]

where \( H_n(\vec{\theta}, f) \rightarrow^P H(\theta_0, f) \) by the same arguments as in, for example, Francq and Zakoian (2004). Finally, it holds that

\[
\nabla S_n[\vec{f}_b - f] = \left\{ E \left[ \frac{\dot{h}_t}{h_t} - \frac{\alpha_0}{n} E \left[ \frac{w_{t-1}^2}{1 - \beta} \frac{h_t}{h_{t-1}^2} \right] \right] \frac{1}{n} \sum_{s=1}^n \left( w_{s-1}^2 - 1 \right) \right\} + O_P \left( 1/\sqrt{n} \right).
\]

Combining eqs. (23)-(25) with Lemma 5 and the restrictions on \( b \) and \( h \) in the theorem, we obtain

\[
\sqrt{n}(\vec{\theta} - \theta_0) = H_n^{-1}(\theta_0, f) \frac{1}{\sqrt{n}} \sum_{t=1}^n \psi_t + O_P \left( 1/\sqrt{n} \right),
\]

where \( \psi_t \) is the influence function defined in eq. (15), and the theorem now follows by the LLN and CLT for stationary sequences.

To show eq. (25), write

\[
\nabla S_n[\vec{f}_b - f] = \nabla S_n[\vec{f}_b - \tilde{f}_b] + \nabla S_n[\tilde{f}_b - f],
\]

27
where

\[ \bar{f}_b(x_{t-1}) := \frac{\sum_{s=1}^{n} K_h(x_{s-1} - x_{t-1}) f(x_{s-1})}{\sum_{s=1}^{n} K_h(x_{s-1} - x_{t-1})} b_{t-1} + f_1 b_{t-1}. \]

We first establish some results which will prove useful in the analysis: First observe that

\[ ||\bar{f}_b - \bar{f}_b||_\infty = \sup_{||x|| \leq b_n} \left| \frac{\sum_{s=1}^{n} K_h(x_{s-1} - x) u_s}{\sum_{s=1}^{n} K_h(x_{s-1} - x)} \right| = O_P(a_n), \]

and so, for some constant \( M_0 > 0 \), \( ||\bar{f}_b - \bar{f}_b||_\infty \leq M_0 \) w.p.a.1. Next, observe that for any function \( df(x) \),

\[ \nabla h_t[df] = \alpha_0 \sum_{k=0}^{t-1} \beta_0^{k-1} w_{t-1-k}^2 \frac{|df_t|}{f_0, t-1-k} \leq \frac{\alpha_0 ||df||_\infty}{f_L} \sum_{k=0}^{t-1} \beta_0^{k-1} w_{t-1-k}^2 \]

\[ |\nabla h_{\alpha,t}[df]| \leq \frac{\alpha_0 ||df||_\infty}{f_L} \sum_{k=0}^{t-1} \beta_0^{k-1} w_{t-1-k}^2, \]

\[ |\nabla h_{\beta,t}[df]| \leq \frac{\alpha_0 ||df||_\infty}{f_L} \sum_{k=0}^{t-1} \beta_0^{k-1} w_{t-1-k}^2. \]

while

\[ h_{0,t} = (1 - \alpha_0 - \beta_0) \sum_{k=0}^{t-1} \beta_0^{k-1} + \alpha_0 \sum_{k=0}^{t-1} \beta_0^{k-1} w_{t-1-k}^2 + h_{0,0}. \]

Thus, for some constant \( c < \infty \),

\[ \frac{\nabla h_t[df]}{h_t} \leq \frac{||df||_\infty}{f_L} \sum_{k=0}^{t-1} \beta_0^{k-1} w_{t-1-k}^2 = \frac{||df||_\infty}{f_L}, \]

\[ |\nabla h_{\alpha,t}[df]| \leq \frac{||df||_\infty}{\alpha_0 f_L} \sum_{k=0}^{t-1} \beta_0^{k-1} w_{t-1-k}^2 = \frac{||df||_\infty}{\alpha_0 f_L}, \]

where the last bound follows from using arguments similar to the ones in Francq and Zakoïan (2004, p. 619). Now, write

\[ \nabla S_n[\bar{f}_b - \bar{f}_b] = \frac{1}{n} \sum_{t=1}^{n} \frac{1 + \varepsilon_t^2}{h_t} \nabla h_t \left[ \bar{f}_b - \bar{f}_b \right] + \frac{1}{n} \sum_{t=1}^{n} \frac{h_t}{h_t^2} \{2 \varepsilon_t^2 - 1\} \nabla h_t \left[ \bar{f}_b - \bar{f}_b \right] + \frac{1}{n} \sum_{t=1}^{n} \frac{\hat{h}_t}{f_0, t h_t} \varepsilon_t^2 \{\bar{f}_b - \bar{f}_b\} \]

\[ = : B_1 + B_2 + B_3, \]

and, with \( K_{t,s} = \mathbb{I}_{b_{t-1}} K_h(x_{s-1} - x_{t-1}) / \sum_{s=1}^{n} K_h(x_{s-1} - x_{t-1}) \),

\[ \bar{f}_b - \bar{f}_b = \sum_{s=1}^{n} K_{t,s} f_s \{w_s^2 - 1\}. \]
We then have

\[
B_3 = \frac{1}{n} \sum_{t=1}^{n} \frac{h_t}{f_t} \epsilon_t^2 (\tilde{f}_{b,t} - \tilde{f}_{b,t}) = \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{h_t}{f_t} \epsilon_t^2 \tilde{K}_{s,t} f_s \{w_s^2 - 1\}
\]

\[
= \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \{w_s^2 - 1\} f_s \tilde{K}_{s,t} \frac{1}{f_t} \left( \frac{h_t}{h_t} \epsilon_t^2 - E \left[ \frac{h_t}{h_t} \right] \right) + E \left[ \frac{h_t}{h_t} \right] \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} f_s \{w_s^2 - 1\} \tilde{K}_{s,t} \frac{1}{f_t}
\]

\[
= B_{3,1} + B_{3,2},
\]

where \(B_{3,1}\) takes the form of a 2nd order degenerate \(U\)-statistic,

\[
B_{3,1} = \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} U_{1,s} f_s \tilde{K}_{s,t} \frac{1}{f_t} U_{2,t},
\]

with \(U_s := (\frac{h_t}{h_t} \epsilon_t^2 / \frac{h_t}{h_t} - E[\frac{h_t}{h_t}], w_s^2 - 1)\) being stationary and \(\beta\)-mixing. Recall the following result of Yoshihara (1976): Suppose that \(E_X |h(U_s, U_t)|^{2+\delta} < \infty\) and \(E_X |h(U_s^*, U_t^*)|^{2+\delta} < \infty\) for some \(\delta > 0\), where \(U_t^*\) is an i.i.d. sequence with same marginal distribution as \(U_t\). Then, with \(h_2(u, v) := h(u, v) - h_1(u) - h_2(v) - \theta, h_1(u) = E[h(u, U_t)]\) and \(\theta = E[h(U_s, U_t)]\), there exists some constant \(K < \infty\) such that

\[
|E[h_2(U_{t_1}, U_{t_2}) h_2(U_{t_3}, U_{t_4})]| \leq K \beta^{\delta/(2+\delta)} (m),
\]

where \(m := \max\{t_2 - t_1, t_4 - t_3\}\). With \(h(U_s, U_t) = U_{1,s} U_{2,t}, h_1(u) = h_2(v) = 0\) and so

\[
\text{Var}_X(B_{3,1}) \leq \frac{4}{n^2} \sum_{t_{1,2,3,4} = 1}^{n} f_{t_1} \tilde{K}_{t_1,t_2} f_{t_2} \tilde{K}_{t_3,t_4} f_{t_4} \frac{1}{f_{t_4}} \left| E[h_2(U_{t_1}, U_{t_2}) h_2(U_{t_3}, U_{t_4})] \right|
\]

\[
\leq \frac{4K}{n^2} \sum_{t_{1,2,3,4} = 1}^{n} f_{t_1} \tilde{K}_{t_1,t_2} f_{t_2} \tilde{K}_{t_3,t_4} f_{t_4} \frac{1}{f_{t_4}} \beta^{\delta/(2+\delta)} (m).
\]

We now employ arguments similar to those in Yoshihara (1976), see also Dehling and Wendler (2010), to bound the sum on the right hand side by splitting it up into parts according to the relationship between \(t_1, t_2, t_3\) and \(t_4\). Consider, for example, the case \(t_1 < t_2 < t_3 < t_4\) and \(t_2 - t_1 \geq t_4 - t_3\). Then \(m = t_2 - t_1\) and there are at most \(n\) possibilities for \(t_1\) and \(t_3\) and \(m\) possibilities for \(t_4\),

\[
\sum_{t_1 < t_2 < t_3 < t_4 \atop t_2 - t_1 \geq t_4 - t_3} f_{t_1} \tilde{K}_{t_1,t_2} f_{t_2} \tilde{K}_{t_3,t_4} f_{t_4} \frac{1}{f_{t_4}} \beta^{\delta/(2+\delta)} (m)
\]

\[
\leq \sum_{t_1 < t_2 < t_3 < t_4 \atop t_2 - t_1 \geq t_4 - t_3} f_{t_1} \tilde{K}_{t_1,t_2} f_{t_2} \tilde{K}_{t_3,t_4} f_{t_4} \frac{1}{f_{t_4}} (t_2 - t_1)^{1-\rho\delta/(2+\delta)}
\]

\[
= \text{TBC}
\]

(26)
We conclude that \( B_{3,1} = o_P(1/\sqrt{n}) \). On the other hand,

\[
B_{3,2} = E \left[ \frac{\hat{h}_t}{h_t} \right] \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} f_s \{ w_s^2 - 1 \} \left\{ \sum_{t=1}^{n} \bar{K}_{s,t} \frac{1}{f_t} - \frac{1}{f_s} \right\} + E \left[ \frac{\hat{h}_t}{h_t} \right] \frac{1}{n} \sum_{s=1}^{n} \{ w_s^2 - 1 \},
\]

where

\[
\frac{1}{\sqrt{n}} \sum_{s=1}^{n} f_s \{ w_s^2 - 1 \} \rightarrow d?? \text{ - Does this converge with slower than } \sqrt{n}\text{-rate?} \] (27)

while

\[
\frac{1}{\sqrt{n}} \sum_{s=1}^{n} \{ w_s^2 - 1 \} \rightarrow d N \left( 0, \tilde{\sigma}_w^2 \right),
\]

where

\[
\tilde{\sigma}_w^2 = E \left[ \{ w_0^2 - 1 \}^2 \right] + 2 \sum_{s=0}^{\infty} E \left[ \{ w_0^2 - 1 \} \{ w_s^2 - 1 \} \right].
\]

We proceed in a similar manner for the analysis of \( B_1 \) and \( B_2 \). Write

\[
B_2 = \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{\hat{h}_t}{h_t^2} \{ w_s^2 - 1 \} \nabla h_t \left[ \bar{K}_{t,s} \right] f_s \{ w_s^2 - 1 \}
\]

\[
= \frac{2}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{\hat{h}_t}{h_t^2} \{ w_s^2 - 1 \} \nabla h_t \left[ \bar{K}_{t,s} \right] f_s \{ w_s^2 - 1 \} + \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{\hat{h}_t}{h_t^2} \nabla h_t \left[ \bar{K}_{t,s} \right] f_s \{ w_s^2 - 1 \},
\]

where

\[
\nabla h_t \left[ \bar{K}_{t,s} \right] = -\alpha_0 w_{t-1}^2 \frac{\bar{K}_{t-1,s}}{f_{t-1}} + \beta_0 \nabla h_{t-1} \left[ df \right] = -\alpha_0 \sum_{k=0}^{t} \beta_0^k w_{t-1-k}^2 \frac{\bar{K}_{t-1-k,s}}{f_{t-1-k}}.
\]

The first term \( B_{2,1} \) is a degenerate U-statistic which is of order \( o_P(1/\sqrt{n}) \) (**MORE DETAILS NEEDED**), while

\[
B_{2,2} = -\frac{\alpha_0}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{k=0}^{t} \beta_0^k \left\{ \frac{\hat{h}_t}{h_t^2} w_{t-1-k}^2 - E \left[ \frac{w_{t-1-k}^2 \hat{h}_t}{h_t^2} \right] \right\} \frac{\bar{K}_{t-1-k,s}}{f_{t-1-k}} f_s \{ w_s^2 - 1 \}
\]

\[+ \frac{\alpha_0}{n} \sum_{s=1}^{n} \sum_{k=0}^{t} \beta_0^k \frac{E \left[ w_{t-1-k}^2 \hat{h}_t \right]}{h_t^2} \sum_{t=1}^{n} \left\{ \frac{\bar{K}_{t-1-k,s}}{f_{t-1-k}} - \frac{1}{f_s} \right\} f_s \{ w_s^2 - 1 \}
\]

\[+ \frac{\alpha_0}{n} \sum_{k=0}^{t} \beta_0^k \frac{E \left[ w_{t-1-k}^2 \hat{h}_t \right]}{h_t^2} \times \frac{1}{n} \sum_{s=1}^{n} \{ w_s^2 - 1 \}
\]

\[= B_{2,2,1} + B_{2,2,2} + B_{2,2,3}
\]
where
\[
B_{2,2,2} = \frac{a_0}{n} E \left[ w_{t-1-k}^2 \hat{h}_t \right] \sum_{s=1}^{n} \sum_{k=0}^{t} \beta_0^k f_s \left\{ w_s^2 - 1 \right\} \times O_P(h^\gamma),
\]
and
\[
\sqrt{n}B_{2,2,3} \rightarrow^d N \left( 0, \sigma_{B,2}^2 \right),
\]
Finally,
\[
B_1 = \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{1 - \varepsilon_t^2}{h_t} \nabla h_t \left[ \bar{K}_{t,s} \right] f_s \left\{ w_s^2 - 1 \right\} = o_P \left( 1/\sqrt{n} \right)
\]
since it is a degenerate $U$-statistic (MORE DETAILS NEEDED).

To complete the proof, consider
\[
\nabla S_n[\bar{f}_b - f] = \frac{1}{n} \sum_{t=1}^{n} \frac{1 - \varepsilon_t^2}{h_t} \nabla h_t \left[ \bar{f}_b - f \right] + \frac{1}{n} \sum_{t=1}^{n} \frac{\hat{h}_t}{h_t^2} \left\{ 2\varepsilon_t^2 - 1 \right\} \nabla h_t \left[ \bar{f}_b - f \right] + \frac{1}{n} \sum_{t=1}^{n} \frac{\hat{h}_t}{f_{0,t}h_t} \varepsilon_t^2 \left\{ \bar{f}_{b,t} - f_t \right\}
\]
\[
= : C_1 + C_2 + C_3.
\]
Given that $\| \bar{f}_b - f \|_\infty = O_P (h^\gamma)$,
\[
\| C_3 \| \leq \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\hat{h}_t}{f_{0,t}h_t} \varepsilon_t^2 \right\| \left\{ \bar{f}_{b,t} - f_t \right\} = \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\hat{h}_t}{f_{0,t}h_t} \varepsilon_t^2 \right\| \times O_P (h^\gamma),
\]
where
\[
\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\hat{h}_t}{f_{0,t}h_t} \varepsilon_t^2 \right\| \rightarrow \text{??? TBC}.
\]
Similarly, $C_1 = o_P (1/\sqrt{n})$ and $C_2 = o_P (1/\sqrt{n})$ (MORE DETAILS NEEDED).
B Lemmas

Lemma 9 Under Assumptions X-Y, with

\[ \| \log \tilde{f}_b - \log f_b \|_\infty = O_P(h^n) + O_P(R(n,h)), \quad \| f_b/\tilde{f}_b - 1 \|_\infty = O_P(h^n) + O_P(R(n,h)). \]

**Proof.** First note that \( \sup_{\| x \| \leq b_n} | \tilde{f}(x) - f(x) | = o_P(1) \), where \( f(x) \geq f_L > 0 \), and so \( \tilde{f}(x) \geq f_L/2 \) w.p.a.1. Thus, by the mean-value theorem,

\[ \| \log \tilde{f}_b - \log f_b \|_\infty = \sup_{\| x \| \leq b_n} | \log \tilde{f}(x) - \log f(x) | = \sup_{\| x \| \leq b_n} \frac{1}{f(x)} | \tilde{f}(x) - f(x) | \]

where \( \tilde{f}(x) \) lies between \( \tilde{f}(x) \) and \( f(x) \); in particular, \( \tilde{f}(x) \geq f_L/2 \) w.p.a.1. Assumption X now establishes the result. Similarly,

\[ \| f_b/\tilde{f}_b - 1 \|_\infty = \left\| \frac{f_b - \tilde{f}_b}{f_b} \right\|_\infty \overset{w.p.a.1}{=} \frac{2}{f_L} \| f_b - \tilde{f}_b \|_\infty = O_P(a_n). \]

Lemma 10 Under Assumptions X-Y,

\[ \sup_{\theta \in \Theta} | \tilde{h}_{b,t}(\theta) - h_t(\theta) | = O_P(h^n) + O_P(R(n,h)) + O_P \left( \frac{t^{\lambda_0}}{b^{k_0}} \right), \]

where the first two probability bounds hold uniformly in \( t \).

**Proof.** For any given \( (\theta, f) \),

\[ h_t(\theta, f) = 1 - \alpha - \beta + \alpha w_{t-1}^2(f) + \beta h_{t-1}(\theta, f) \]

\[ = \beta^t h_0 + (1 - \alpha - \beta) \sum_{k=0}^{t-1} \beta^k + \alpha \sum_{k=0}^{t-1} \beta^k w_{t-1-k}^2(f). \]

Thus, applying Lemma 9,

\[ | \tilde{h}_{b,t}(\theta) - h_t(\theta) | \leq \alpha \sum_{k=0}^{t-1} \beta^k f_{t-1-k} g_{t-1-k} | f_{t-1-k}/\hat{f}_{b,t-1-k} - 1 | \]

\[ \leq \alpha \sum_{k=0}^{t-1} \beta^k w_{0,t-1-k}^2 \left\| f_b/\tilde{f}_b - 1 \right\|_\infty + | f_{t-1-k}/f_1 - 1 | \right\|_{b,t-1-k} \}

\[ \leq C \sum_{k=0}^{t-1} \beta^k w_{0,t-1-k}^2 \left\{ O_P(a_n) + \sup_{\| x \| > b} | f_1 - f(x) | \right\|_{b,t-1-k} \}

\[ \leq C \left\{ \sum_{k=0}^{t-1} \beta^k w_{0,t-1-k}^2 \right\} \times O_P(a_n) + C \sum_{k=0}^{t-1} \beta^k w_{0,t-1-k}^2 f_{t-1-k} \right\|_{b,t-1-k} \]
where \( E \left[ \sum_{k=0}^{t-1} \beta^k w^2_{0,t-1-k} \right] \leq E \left[ w^2_{0,t} \right] / (1 - \beta) \) and,

\[
E \left[ \sum_{k=0}^{t-1} \beta^k w^2_{0,t-1-k} f_{t-1-k, b_{t-1-k}} \right] = E \left[ w^2_{0,t} \right] \sum_{k=0}^{t-1} \beta^k E \left[ f_{t-1-k, b_{t-1-k}} \right] \leq \frac{E \left[ w^2_{0,t} \right]}{b^{\kappa_0}} \sum_{k=0}^{t-1} \beta^k E \left[ ||x_{t-1-k}||^\gamma_{f+k_0} \right] \leq \frac{E \left[ w^2_{0,t} \right]}{b^{\kappa_0}} \sum_{k=0}^{t-1} \beta^k (t - 1 - k)^{\lambda_0} = O \left( \frac{t^{\lambda_0}}{b^{\kappa_0}} \right).
\]
## C Tables and Figures

### Table 1. Estimation results of the models

<table>
<thead>
<tr>
<th>Series</th>
<th>Parameter</th>
<th>GJR-GARCH</th>
<th>GJR-X</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.021</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>FTSE</td>
<td></td>
<td>(0.017)</td>
<td>(0.014)</td>
<td>(0.009)</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>0.098</td>
<td>0.158</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.028)</td>
<td>(0.021)</td>
<td>(0.027)</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.921</td>
<td>0.811</td>
<td>0.880</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.028)</td>
<td>(0.026)</td>
<td>(0.034)</td>
</tr>
<tr>
<td></td>
<td>$\pi$</td>
<td></td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.007</td>
<td>-0.012</td>
<td>2.08 $\times 10^{-9}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.037)</td>
<td>(0.009)</td>
<td>(5.95 $\times 10^{-9}$)</td>
</tr>
<tr>
<td>CAC</td>
<td>$\gamma$</td>
<td>0.166</td>
<td>0.238</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.045)</td>
<td>(0.029)</td>
<td>(0.027)</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.892</td>
<td>0.763</td>
<td>0.853</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.050)</td>
<td>(0.029)</td>
<td>(0.029)</td>
</tr>
<tr>
<td></td>
<td>$\pi$</td>
<td></td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>DAX</td>
<td>$\alpha$</td>
<td>0.010</td>
<td>0.002</td>
<td>2.08 $\times 10^{-8}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(6.93 $\times 10^{-9}$)</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>0.138</td>
<td>0.170</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.029)</td>
<td>(0.028)</td>
<td>(0.021)</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.901</td>
<td>0.831</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.017)</td>
<td>(0.027)</td>
<td>(0.023)</td>
</tr>
<tr>
<td></td>
<td>$\pi$</td>
<td></td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports estimation results for the GJR-GARCH, GJR-X, and multiplicative GARCH-X models for stock return series. The models are of the form

\[
\begin{align*}
\sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \gamma y_{t-1}^2 I(y_{t-1} < 0) + \beta \sigma_{t-1}^2 : \text{GJR-GARCH} \\
\sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \gamma y_{t-1}^2 I(y_{t-1} < 0) + \beta \sigma_{t-1}^2 + \pi x_{t-1} : \text{GJR-X} \\
\sigma_t^2 &= [\omega + \alpha h_{t-1} \varepsilon_{t-1}^2 + \gamma h_{t-1} \varepsilon_{t-1}^2 I(\sqrt{h_{t-1}} \varepsilon_{t-1} < 0) + \beta h_{t-1}] f(x_{t-1}) : \text{Our model}
\end{align*}
\]

where $(y_t)$ is the demeaned return series and $x_t$ is the VIX index. Standard errors are reported in parentheses.
Table 2. Comparison of within-sample predictive power for the stock return volatilities
(2004.01.02-2013.12.30)

<table>
<thead>
<tr>
<th>Models</th>
<th>FTSE</th>
<th>CAC</th>
<th>DAX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QLIKE</td>
<td>DMW</td>
<td>QLIKE</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.160</td>
<td>4.66***</td>
<td>0.162</td>
</tr>
<tr>
<td>Nonpara-model</td>
<td>0.166</td>
<td>4.09***</td>
<td>0.171</td>
</tr>
<tr>
<td>GJR-X (additive)</td>
<td>0.155</td>
<td>3.82***</td>
<td>0.148</td>
</tr>
<tr>
<td>Our model</td>
<td><strong>0.132</strong></td>
<td><strong>0.140</strong></td>
<td><strong>0.140</strong></td>
</tr>
</tbody>
</table>

Notes: The QLIKE loss is defined in (16) and the DMW test statistic is defined in (17). *, ** and *** signify rejecting the null hypothesis of equal loss for 10%, 5% and 1% tests, respectively.

Table 3. Comparison of out-of-sample predictive power for the stock return volatilities
(2010.01.04-2013.12.30)

<table>
<thead>
<tr>
<th>Models</th>
<th>FTSE</th>
<th>CAC</th>
<th>DAX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QLIKE</td>
<td>DMW</td>
<td>QLIKE</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.152</td>
<td>5.03***</td>
<td>0.160</td>
</tr>
<tr>
<td>Nonpara-mode</td>
<td>0.154</td>
<td>4.97***</td>
<td>0.174</td>
</tr>
<tr>
<td>GJR-X (additive)</td>
<td>0.147</td>
<td>4.59***</td>
<td>0.150</td>
</tr>
<tr>
<td>Our model</td>
<td><strong>0.126</strong></td>
<td><strong>0.135</strong></td>
<td><strong>0.135</strong></td>
</tr>
</tbody>
</table>

Notes: The QLIKE loss is defined in (16) and the DMW test statistic is defined in (17). *, ** and *** signify rejecting the null hypothesis of equal loss for 10%, 5% and 1% tests, respectively.
Table 4. Estimates of the semiparametric single index model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>FTSE</th>
<th>CAC</th>
<th>DAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.970</td>
<td>0.993</td>
<td>0.958</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.236</td>
<td>-0.092</td>
<td>-0.275</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.064</td>
<td>0.075</td>
<td>0.078</td>
</tr>
<tr>
<td>$\hat{\beta}<em>1 / sd(x</em>{1t})$</td>
<td>0.098</td>
<td>0.100</td>
<td>0.097</td>
</tr>
<tr>
<td>$\hat{\beta}<em>2 / sd(x</em>{2t})$</td>
<td>0.037</td>
<td>0.013</td>
<td>0.043</td>
</tr>
<tr>
<td>$\hat{\beta}<em>3 / sd(x</em>{3t})$</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Notes: The table reports estimates of the model given in (18) for $x_t = (x_{1t}, x_{2t}, x_{3t})'$ and $\beta = (\beta_1, \beta_2, \beta_3)'$. $x_{1t}$ is the VIX index, $x_{2t}$ is each country’s industrial production index and $x_{3t}$ is crude oil price. $sd(x_{it})$ denotes the standard deviation of $x_{it}$ for $i = 1, 2, 3$.

Table 5. Predictive power of the single index model

<table>
<thead>
<tr>
<th></th>
<th>FTSE</th>
<th>CAC</th>
<th>DAX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QLIKE</td>
<td>DMW</td>
<td>QLIKE</td>
</tr>
<tr>
<td>Within-sample prediction (2004.01.02-2013.12.30)</td>
<td>0.127</td>
<td>2.44**</td>
<td>0.127</td>
</tr>
<tr>
<td>Out-of-sample prediction (2010.01.04-2013.12.30)</td>
<td>0.118</td>
<td>2.38**</td>
<td>0.133</td>
</tr>
</tbody>
</table>

Notes: The QLIKE loss is defined in (16) and the DMW test statistic is defined in (17). The benchmark model for the DMW test is the multiplicative GARCH-X model using only the VIX index as the covariate. *, ** and *** signify rejecting the null hypothesis of equal loss for 10%, 5% and 1% tests, respectively.
Figure 1. Daily European stock index return series from 2 Jan. 2004 to 30 Dec. 2013.
Figure 2. Estimate of $f(x_{t-1})$ in the multiplicative GARCH-X model for three European stock index return series.
<table>
<thead>
<tr>
<th>Lag</th>
<th>FTSE Correlogram</th>
<th>Lag</th>
<th>FTSE Correlogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lag</th>
<th>CAC Correlogram</th>
<th>Lag</th>
<th>CAC Correlogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lag</th>
<th>DAX Correlogram</th>
<th>Lag</th>
<th>DAX Correlogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

\[(a) \text{ squared return } y_t^2 \]  
\[(b) \text{ rescaled squared return } y_t^2 / \hat{f}(x_{t-1}) \]

Figure 3. Sample autocorrelations of squared return $y_t^2$ and rescaled squared return $y_t^2 / \hat{f}(x_{t-1})$ for three European stock index return series.
(a) estimate of $h_t$ from GJR-GARCH model  
(b) estimate of $h_t$ from our model

Figure 4. Conditional variance of the GJR-GARCH(1,1) model for three European stock index return series and their rescaled return series $y_t/\hat{f}(x_{t-1})^{1/2}$. 
Figure 5. Industrial production and stock index return series for three European countries from 2 Jan. 2004 to 30 Dec. 2013. Thick line represents linearly interpolated industrial production and the other line is stock index return in each country.
References


