cemmap
centre for microdata methods and practice

Let’s get LADE: robust estimation of semiparametric multiplicative volatility models

Bonsoo Koo
Oliver Linton

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP11/13

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Let's Get LADE: Robust Estimation of Semiparametric Multiplicative Volatility Models

Bonsoo Koo* Monash University

Oliver Linton† University of Cambridge

March 19, 2013

Abstract

We investigate a model in which we connect slowly time varying unconditional long-run volatility with short-run conditional volatility whose representation is given as a semi-strong GARCH (1,1) process with heavy tailed errors. We focus on robust estimation of both long-run and short-run volatilities. Our estimation is semiparametric since the long-run volatility is totally unspecified whereas the short-run conditional volatility is a parametric semi-strong GARCH (1,1) process. We propose different robust estimation methods for nonstationary and strictly stationary GARCH parameters with nonparametric long run volatility function. Our estimation is based on a two-step LAD procedure. We establish the relevant asymptotic theory of the proposed estimators. Numerical results lend support to our theoretical results.

Key words: semiparametric, heavy-tailed errors, time varying, nonstationary multiplicative GARCH

Journal of Economic Literature Classification: C13, C14, C22

*Department of Econometrics and Business Statistics, Monash University, PO Box 11E, Clayton Campus, VIC 3800, Australia; e-mail: bonsoo.koo@monash.edu

†Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk. Thanks to the ERC for financial support.

1 Introduction

There is a lot of empirical evidence that is consistent with time varying dependent and heavy-tailed innovations in economic and financial time series. To some extent the autoregressive conditional heteroskedastic (ARCH) or generalized ARCH (GARCH) type of models address these features. However, there remain some issues. For instance, one salient empirical feature often found in financial time series data is that return volatility is higher during recessions. See Black (1976), Schwert (1989), and Bollerslev *et al.* (1992). This is not compatible with many widely used models whose assumption is based on constant unconditional return volatility. In addition, the estimation methods of many of those models are based on assumptions (such as i.i.d. innovations and moment conditions) that are violated often in financial time series data, which renders their estimators unreliable.

This paper proposes robust estimation of a GARCH type model that incorporates time varying aspects of long-run volatility as in Engle and Rangel (2008) and heavy-tailed innovation processes as in Jensen and Rahbek (2004a, 2004b) and Linton, Pan and Wang (2010). We investigate a model in which we connect slowly time varying long-run volatility with short-run volatility whose representation is given as a semi-strong GARCH (1,1) model with heavy tailed errors. Our model accommodates the idea that there are two different types of volatilities embedded in volatility processes we observed in financial markets. One is the short-run volatility that reflects market risks. The other one is long-run volatility that reflects the risks of real economic activity and is slowly time varying. We focus on robust estimation of both volatilities given that our approach does not require moment conditions of innovations usually required in the other ARCH/GARCH type literature.

Our model can be considered as generalisation of Engle and Rangel (2008). However, while Engle and Rangel's model captures both time varying characteristics of the state of the economy and many stylised facts of financial volatility, their approach neglects the possibility that even short-run dynamics might not be a weakly stationary stochastic process. In truth, high kurtosis and persistence of financial volatility are often found in many datasets and this is incompatible with weakly stationary ARCH/GARCH models. Moreover, their model is confined to the case of normal i.i.d. errors or at least errors that have a large number of moments. One of the salient features in financial data is fat-tailed distributions. When errors are leptokurtic, the existence of certain moments of errors is not guaranteed. Therefore, weak stationarity assumption is unlikely to hold in many cases. This paper incorporates heavy tailed errors in innovation process and possible nonstationarity in short-run dynamics into their model.

Our relaxation of these conditions is important since it is well documented that the residuals after GARCH filtering are heavy-tailed and far from normal. (Mittnik and Rachev (2000), Rachev (2003)). Even after spline-GARCH filtering, heavy tailed innovations of financial time series data remain. In Figure 1, we show a time series of daily log returns of the Russian RTS index (RTS) along with the corresponding conditional variance. It can be seen that there is time trending in the time series of log returns of RTS. Figure 2 shows QQ-plot of log returns of RTS in the left panel and QQ-plot of residuals after AR(2)-GARCH(1,1) filtering. This shows that even after AR(2)-GARCH(1,1) filtering, the residuals are far from normal. Heavy-tailed errors and nonstationarity poses a significant challenge to both estimation of parameters of ARCH/GARCH type models and the asymptotic theories of their proposed estimators. Heavy-tailed innovations are too important to neglect!

FIGURE 1,2 ABOUT HERE

Consequently, both nonstationarity and heavy tailed innovations should be allowed for. This necessitates the development of robust estimation methods for such models. We investigate a model under which the essential structure of Engle and Rangel (2008) is generalized to allow for both these features and we develop estimation technology to handle this case.

Whereas Engle and Rangel (2008) estimated the long-run volatility by a spline methodology, we employ a kernel estimation methodology to make our asymptotic analysis tractable. Moreover, our focus is on robust estimation of both the long-run and short-run volatilities, which can be particularly useful for risk management such as Value at Risk and expected shortfalls.

Our estimation approach is based on the least absolute deviation estimation (LADE) in Hall, Peng and Yao (2002) and Peng and Yao (2003). The LADE is known to have several advantages compared with least squares. Among those, robustness and milder moment conditions are worth mentioning. The robustness feature becomes more important for seemingly nonstationary financial data we consider in this paper. We approach a strictly stationary and a nonstationary semi-strong GARCH (1,1) processes separately since an assumption of stationarity makes our estimation procedure different. We propose different robust estimation methods for nonstationary and strictly stationary GARCH parameters with nonparametric long run volatility function. We establish the relevant asymptotic theory of the proposed estimators.

The remainder of this paper is organized as follows. Section 2 briefly reviews relevant literature. Section 3 introduces the model and the related framework. Section 4 suggests our estimation procedure for the unknown parameter function which captures the long-run volatility and unknown parameters which capture the short-run dynamics. Section 5 develops distribution theories in relation to our proposed estimators. Simulation studies of our estimation procedure are explored in Section 6. Section 7 concludes. The mathematical proofs are provided in the Appendix.

The following notations are used. The integral $\int$ is taken over $(-\infty, \infty)$ unless specified otherwise. $||\cdot||$ denotes any norm over the relevant space. Let g be any function from $\mathbb{R}^d \to \mathbb{R}$. $||g||_\infty = \sup_{x \in \mathbb{R}^d} |g(x)|$, $||g||_q = (\int |g(x)|^q dx)^{1/q}$, $||g||_q^p = (||g||_q)^p$ and $g^{(m)}(u)$ denotes the mth derivative with respect to u. $\mathbb{C}_2(b)$ denotes the space of twice continuously differentiable real valued functions with first and second partial derivatives of all of their arguments bounded by b and $||g||_\infty < b$. $K_h(\cdot)$ denotes $K(\cdot/h)$ with $\int K(u)\,du = 1$ and the corresponding bandwidth $h(T) \to 0$ as $T \to \infty$. Subscript 0 implies true values or function of unknown parameters or function. $\mathbb{E}_{t-1}$ is conditional expectation on an information set $\mathcal{F}_{t-1}$ including past information up to $t-1$. $1(A)$ is the indicator function for the set A. C is a generic constant which may be different at different places.

2 Literature Review

The long-run economic environment is known to be closely related to short-run movements of financial markets. In addition, this long-run economic environment changes over time, which will manifest itself in time varying long-run volatility. Nevertheless, financial practitioners often rely on models whose concern focuses on only short-run volatility by neglecting the time varying characteristic of this long-run volatility. In fact, whereas there is plethora of literature whose unconditional volatility is assumed to be constant, there has not been much literature which attempts to capture this time varying characteristic of unconditional volatility with time varying conditional short-run volatility under an unified framework. However, quite often, this apparently unclear relationship between the long-run volatility and short-run volatility embedded in financial data is too informative to neglect given that financial risk and ever changing environment of financial markets play a crucial role in the contemporary economics and finance literature.

Quite recently, however, there have been a few attempts which incorporate changing unconditional long-run volatility. Veronesi (1999), Engle and Rangel (2008) and Bikbov and Chernov (2010). For example, Engle and Rangel (2008) proposed a model which related high frequency financial risks to the low frequency macroeconomic risks based on the assumption that the long-run volatility captures the macroeconomic environment. They adopted semiparametric approach to capture both short- and long-run volatility. Their model is designed to separate out the long run patterns of volatility detected in the financial data. The estimated long run volatility is then used to empirically investigate its causes. Later, Hafner and Linton (2010) extended their univariate multiplicative volatility model to a multivariate one and provide the asymptotic properties of their proposed estimators. In addition, Van Bellegem (2011) provided comparison study among many locally stationary time series including a multiplicative model.

The long-run volatility in this paper is time varying. This time varying characteristic of stochastic processes have gained a momentum. In particular, locally stationary processes have lain at the centre of active investigation. See, for example, Dahlhaus (1997), Giurcanu and Spokoiny (2004), and Koo and Linton (2012). However, quantile regression analysis for time varying processes has emerged quite recently. See Zhou and Wu (2010).

The short-run volatility in this paper is represented by a semi-strong GARCH (1,1) process. This semi-strong GARCH process is not unknown elsewhere in the financial econometrics literature. Drost and Nijman (1993) and Lee and Hansen (1994) investigated this process. Quite recently, Linton, Pan, and Wang (2010) extended their results. Linton *et al.* (2010) studied the estimation of a semi-strong GARCH (1,1) with $\mathbb{E}\varepsilon_t^4 = \infty$. They proposed that the semi-strong GARCH (1,1) process be estimated consistently by least absolute deviations estimator (LADE) and quasi-maximum-likelihood estimator (QMLE) under suitable regularity conditions. Moreover, asymptotic properties of both estimators for the semi-strong GARCH (1,1) process are also provided. Specifically, in their paper, LADE is preferred to QMLE since the former is shown to be asymptotically normal if $\mathbb{E}|\varepsilon_t|^{2+\delta} < \infty$ and the conditional densities of $\log \varepsilon_t^2$ given $\mathcal{F}_t (= \sigma(..., \varepsilon_{i-1}, \varepsilon_i))$ satisfy some regularity conditions where $\mathcal{F}_t$ denotes an information set including the history of returns up to time t. Also, our semi-strong GARCH (1,1) process could be nonstationary. Jensen and Rahbek (2004a, 2004b) developed the distribution theory of the QMLE for a nonstationary GARCH (1,1) model with the assumption of $\mathbb{E}\log(\alpha_0\varepsilon_t^2 + \beta_0) \geq 0$ and $\mathbb{E}\varepsilon_t^4 < \infty$. Jensen and Rahbek (2004a, 2004b) explored Nonstationary GARCH models. In addition, there are an array of robust estimation methods regarding ARCH/GARCH type models. For example, Hall and Yao (2003) and Peng and Yao (2003). After Hall and Yao (2003) pointed out that for heavy tailed errors whose fourth moment is infinite, the asymptotic distribution of QMLE be nonnormal and be difficult to obtain directly from standard methods, Peng and Yao (2003) suggested the LADE estimation method. Huang, Wang and Yao (2008) proposed that the tail heaviness of the innovation distribution plays an important role in determining the relative performance of those two estimation methods. Most recently, Linton *et al.* (2010) used the LADE to estimate the more general nonstationary GARCH (1,1) model under the condition that the fourth moment of errors is infinite.

3 The Model

Suppose that we observe a time series $\{y_{t,T}\}$ for $t = 1, \ldots, T$, $T = 1, 2, \ldots$. The process $\{y_{t,T}\}_{t=1}^{T}$ is assumed to follow the multiplicative volatility model given by

$$y_{t,T} = \sqrt{\tau_{t,T}} h_{t,T} = \sqrt{\tau_{t,T} g_{t,T}} \varepsilon_t, \tag{1}$$

where $\tau_{t,T}(= \tau(t/T))$ is a positive deterministic time-varying long-run component of volatility, $g_{t,T}$ is a short-run dynamic process, which represents high frequency short-run volatility, and ε_t is a strictly stationary and ergodic sequence of random variables such that the (conditional) median of $\log \varepsilon_t^2$ is zero.

There are some notable features of our model. To begin with, $\{y_{t,T}\}$ defines a triangular array of observations. By its nature, $\{y_{t,T}\}$ is dependent and heterogeneous. The process $\{y_{t,T}\}$ is assumed to consist of a long-run component, which we interpret as unconditional "volatility", and a short-run component which has temporal dependence and which we interpret to be conditional "volatility". The long-run volatility is represented by a slowly time varying deterministic function.[1] See Rodriguez-Poo and Linton (2001) and Hafner and Linton (2010). In particular, we shall suppose that the long-run volatility process $\tau_{t,T}$ is a totally unspecified smooth function of time.[2]

To justify the asymptotic theory for our estimators, we use the following rescaling method. Let $\tau(.)$ be a function on $[0, 1]$ and let

$$\tau_{t,T} = \tau\left(t/T\right), \quad t = 1, \ldots, T.$$

[1] Unlike Engle and Rangel (2008), we let the long-run volatility be a totally unspecified function of time without specifying factors that could contribute to this long-run volatility. This is because our paper focuses on the separate estmation of those two distinct volatilities instead of finding the underlying factors for changing long-run volatility.

[2] The function $\tau(.)$ can be either smooth or have a finite number of structural breaks in time. Even though we focus on everywhere-continuous $\tau(.)$ in this paper, $\tau(.)$ that allows for a finite number of breaks in time can be considered via some extension of Delgado and Hidalgo (2000) and Koo (2012).

Note that $\tau_{t,T}$ depends on the sample size T and the domain of $\tau(.)$ becomes more dense in t/T as T tends to infinity; this renders many asymptotic results available. See Härdle and Tuan (1986), Robinson (1989), and Dahlhaus (1997). We drop the subscript T for simplicity in the following wherever possible.

For the dynamics of short-run volatility, we model $g_{t,T}$ parametrically by considering the first order semi-strong GARCH model given as

$$h_t = \sqrt{g_t}\varepsilon_t, \qquad g_t(\phi) = \omega + \alpha h_{t-1}^2 + \beta g_{t-1}(\phi), \tag{2}$$

where $\phi = (\eta, \omega, \alpha, \beta)^\top$, with $\eta = g_0(\phi), \omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ are unknown parameters, and $\{\varepsilon_t\}$ is strictly stationary and ergodic. For nonstationary g_t, we need to consider η as one of unknown parameters since the analysis is conditional on the initial observed value, but for strict stationary g_t, we can drop the parameter η from ϕ.

We do not assume that $\{\varepsilon_t\}$ is an i.i.d. sequence of random variables. Instead, we assume that $\{\varepsilon_t\}$ is stationary and ergodic, which makes g_t follow a semi-strong GARCH model. The process g_t does not necessarily have a weakly stationary solution since we do not require $\alpha_0 + \beta_0 < 1$. Instead, we investigate g_t according to whether $\mathbb{E}\log(\alpha_0\varepsilon_t^2 + \beta_0) < 0$ or $\mathbb{E}\log(\alpha_0\varepsilon_t^2 + \beta_0) \geq 0$. It is worth noting that Theorem 1 in Linton *et al.* (2010) states that (2) defines a unique, strictly stationary and ergodic solution if and only if $\mathbb{E}\log(\alpha_0\varepsilon_t^2 + \beta_0) < 0$ (under some regularity conditions). In addition, we do not assume that $\mathbb{E}\varepsilon_t^4 < \infty$ (the QMLE is well-behaved when $\mathbb{E}\varepsilon_t^4 < \infty$, see Hall and Yao (2003)). In sum, we focus on two distinct cases in which (2) is either strictly stationary but not weakly stationary or nonstationary. However, this non-weak stationarity feature of g_t in this paper incurs an undesirable consequence in our set-up, since the weak stationarity condition can not be imposed on the coefficients of g_t any longer.[3] Due to a multiplicative relationship between unobservable long and short run volatility components, without a further restriction, we cannot identify those two different volatilities separately. Therefore, a serious identification problem arises naturally. To avoid this difficulty, we impose a restriction on the long-run component rather than on the GARCH coefficients. Also, instead of imposing moment restrictions

[3] $E\log\left(\alpha_0\varepsilon_t^2 + \beta_0\right) < 0$ is implied by $\alpha_0 + \beta_0 < 1$ when ε_t is i.i.d. with $E\varepsilon_t^2 = 1$ since

$$E\left[\log\left(\alpha_0\varepsilon_t^2 + \beta_0\right)\right] < \ln E\left[\alpha_0\varepsilon_t^2 + \beta_0\right] = \ln(\alpha_0 + \beta_0)$$

from Jensen's inequality. This implies that the condition for strict stationarity is weaker than that of weak stationarity. However, when $E\varepsilon_t^2 = \infty$ as in our case, it is much more complex.

on $\{\varepsilon_t\}$, a median restriction is imposed on $\{\varepsilon_t\}$. Specifically, we assume that

$$\int_0^1 \tau(u)\,du = 1, \tag{3}$$

$$\mathbf{med}(\log \varepsilon_t^2 | \mathcal{F}_{t-1}) = 0, \tag{4}$$

where $\mathbf{med}(X|\mathcal{F}_t)$ denotes the median of X conditional on $\mathcal{F}_t$.

Note that a median restriction (4) is equivalent to $\mathbf{med}(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$. Due to one of the quantile characteristics, say the quantile does not change under any monotonic transformation, logarithm and median operators can be switched. Therefore, it can be said that the standard mean restriction, $\mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$ is simply replaced by (4) in our model. Furthermore, this condition implies that the unconditional median takes the same values, i.e., $\mathbf{med}(\varepsilon_t^2) = 1$ and $\mathbf{med}(\log \varepsilon_t^2) = 0$.[4]

Volatility here is not associated with moments such as the variance. Instead, it is a more general "scale" measure that is defined in the absence of such moments and which would be equal to a constant times variance if the required moments were to exist.

As we will see shortly, the estimation procedures are different according to the assumption on the $\{g_t\}$ process. We will investigate this in more detail in the following section.

4 Estimation Procedure

In this section, procedures of estimation for (1) are proposed. Recall that we have two components to estimate in order to obtain the multiplicative volatility. Since h_t^2 is equivalent to y_t^2/τ_t, the equations we are to estimate can be given by (1) and

$$g_t(\phi) = \omega + \alpha \frac{y_{t-1}^2}{\tau_{t-1}} + \beta g_{t-1}(\phi), \tag{5}$$

subject to (3) and (4). The process (1) can be written as

$$\log y_t^2 = \log \tau_t + \log g_t + \log \varepsilon_t^2. \tag{6}$$

[4]This follows because

$$E\left[1\left(\log \varepsilon_t^2 > 0\right)\right] = E\left[E\left[1\left(\log \varepsilon_t^2 > 0\right)|\mathcal{F}_{t-1}\right]\right] = \frac{1}{2}.$$

We consider two different cases. First, we assume that g_t is a strictly stationary but not necessarily a weakly stationary process. One salient example of this case is the integrated GARCH model (IGARCH). The IGARCH model is strictly stationary even though $\alpha+\beta=1$. Since we do not assume any moments for g_t, the usual approach is not feasible. Secondly, we allow g_t to be neither weakly stationary nor strictly stationary. This requires a different approach.

4.1 Strict stationary $\{g_t\}$

4.1.1 Step 1 : Estimation of Long-run Volatility

Our estimation procedure consists of two main steps. The first step is to estimate $\tau(.)$ nonparametrically based on the kernel estimation method. The second step involves the estimation of parameters of a semi-strong GARCH (1,1) model based on the estimate from the first step. When $\{g_t\}$ is strictly stationary but not weakly stationary, we cannot apply various results based on the classical framework of second order characteristics. However, the implications of strict stationarity can be used. For instance, when $\{g_t\}$ and $\{\varepsilon_t\}$ are stationary, the median of $\{g_t\varepsilon_t^2\}$ is the same for all observations. Based on this realisation, if we take the median of both sides of (6) with the restriction (4), we have

$$\begin{aligned}\mathbf{med}\left(\log y_t^2\right) &= \log\tau\left(t/T\right)+\mathbf{med}\left(\log g_t\varepsilon_t^2\right)\\ &= \log\tau\left(t/T\right)+C=\log\tau^*(t/T),\end{aligned} \tag{7}$$

where $C=\mathbf{med}\left(\log g_t\varepsilon_t^2\right)$. Since we have assumed that $\int_0^1\tau\left(u\right)du=1$, for τ^* specified in (7), then

$$\tau\left(u\right)=\frac{\tau^*\left(u\right)}{\int_0^1\tau^*\left(u\right)du}, \tag{8}$$

since, from (7)

$$\frac{\tau^*\left(u\right)}{\int_0^1\tau^*\left(u\right)du}=\frac{\tau\left(u\right)\exp\left(C\right)}{\exp\left(C\right)}=\tau\left(u\right).$$

We estimate $\tau(.)$ nonparametrically by the LAD method. It is worth noting that the LAD method requires less moment conditions than the QMLE or regression based estimation does. That is why our estimation is robust to the presence of heavy-tailed innovations. From (7), our kernel estimator for the long run volatility, $\check{\tau}\left(u\right)$ can be obtained as

$$\log \check{\tau}(u) = \arg\min_{\tau \in \mathbb{R}_+} \sum_{t=1}^{T} \left| \log y_t^2 - \log \tau^* \right| K_h(u - t/T), \tag{9}$$

where $K_h(\cdot) = K(\cdot/h)/h$ with a kernel function $K(\cdot)$ and a bandwidth $h > 0$. Once we obtain $\check{\tau}(u)$, due to (3) and (8), we renormalize $\check{\tau}(u)$ by calculating

$$\hat{\tau}(u) = \frac{\check{\tau}(u)}{\int_0^1 \check{\tau}(u)\, du}, \tag{10}$$

which yields our estimator $\hat{\tau}(u)$ for $\tau(u)$.

4.1.2 Step 2 : Estimation of Short-run Volatility

Since our focus is robust estimation and we do not assume $\mathbb{E}\varepsilon_t^4 < \infty$, we restrict ourselves to the LAD estimation method.[5] The objective function we minimize is given by

$$S_T(\phi) = \sum_{t=v+1}^{T} \left| \log \hat{h}_t^2 - \log g_t(\phi) \right| = \sum_{t=v+1}^{T} \left| \log \frac{y_t^2}{\hat{\tau}_t} - \log g_t(\phi) \right|,$$

where $\phi = (\omega, \alpha, \beta)^\top$ and $v = v(T)$ is a non-negative integer. Therefore, the least absolute deviations estimator for ϕ is as follows

$$\hat{\phi}_{LAD} \equiv \arg\min_{\phi} S_T(\phi). \tag{12}$$

This can be motivated by the regression relationship

$$\log \frac{y_t^2}{\tau_t} = \log \{g_t(\phi)\} + \log \varepsilon_t^2,$$

[5] If the data satisfy certain regularity conditions (i.e., moment conditions), we could use the following QMLE method. Note that h_t is unobservable and can be written as $h_t^2 = y_t^2/\tau_t$. The log likelihood takes the form given by

$$\mathcal{L}(\phi) = \frac{1}{T} \sum_{t=1}^{T} l_t(\phi) \text{ where } l_t(\phi) = -\left(\log g_t(\phi) + \frac{\tilde{h}_t^2}{g_t(\phi)} \right) \tag{11}$$

where $\tilde{h}_t^2 = y_t^2/\tilde{\tau}_t$, $\tilde{\tau}_t$ is the estimator from the first step and g_t is defined as (5). The initial QMLE is the maximizer of (11) given by

$$\tilde{\phi}_{ML} \equiv \arg\min_{\phi} \frac{1}{T} \sum_{t=1}^{T} \left(\log g_t(\phi) + \frac{y_t^2}{\tilde{\tau}_t g_t(\phi)} \right).$$

where the (conditional) median of $\log(\varepsilon_t^2)$ is zero under the restriction (4). The LAD estimator is known to be consistent and asymptotically normal under very mild conditions. We extend this work to allow for estimated τ_t.

4.2 Nonstationary $\{g_{t,T}\}$

We now turn to the case where $\{g_{t,T}\}$ is not even strictly stationary. In this case, the previous estimation method breaks down. Nevertheless, the following estimation method can be used. Unlike the stationary case, we estimate parameters for the semi-strong GARCH (1,1) process first. Then, based on the estimates for the GARCH parameters, we estimate the long-run volatility.

4.2.1 Step 1 : Estimation of short-run Volatility

Recall that the parameter vector of the nonstationary $\{g_{t,T}\}$ is $\phi = (\eta, \omega, \alpha, \beta)$ with true value $\phi_0 = (\eta_0, \omega_0, \alpha_0, \beta_0)$. We have

$$\log y_t^2 = \log \tau_t + \log g_t + \log \varepsilon_t^2 = \log \tau_t + \log(\omega + \alpha h_{t-1}^2 + \beta g_{t-1}(\phi)) + \log \varepsilon_t^2,$$

where $h_{t-1}^2 = y_{t-1}^2/\tau_{t-1}$. In the neighbourhood of a time point u, $\tau(u)$ can be considered as a constant, and therefore, in the vicinity of each time point $u \in [0,1]$, we have

$$\begin{aligned} \log y_t^2 &= \log\left[\tau(u)(\omega(u) + \alpha(u)h_{t-1}^2 + \beta(u)g_{t-1}(\phi))\right] + \log \varepsilon_t^2 + o(1) \qquad (13) \\ &= \log[\tau(u)\,\omega(u) + \alpha(u)y_{t-1}^2 + \tau(u)\beta(u)g_{t-1}(\phi)] + \log \varepsilon_t^2 + o(1), \end{aligned}$$

since $\tau_t - \tau_{t-1} = O(1/T)$ under our smoothness assumptions below. As seen from (13), we estimate the parameters of short-run volatility up to some constant scale. However, (13) can be reparameterised as, for each sub-sample in the neighbourhood of each time point $u \in [0,1]$,

$$\begin{aligned} y_t &= \sqrt{\tilde{g}_t(\tilde{\phi})}\varepsilon_t \qquad (14) \\ \tilde{g}_t(\tilde{\phi}) &= \tilde{\omega} + \tilde{\alpha} y_{t-1}^2 + \tilde{\beta}\tilde{g}_{t-1}(\tilde{\phi}) \end{aligned}$$

where $\tilde{\omega}(= \tau(u)\omega)$, $\tilde{\alpha}(= \alpha)$, and $\tilde{\beta}(= \tau(u)\beta)$ with the initial $\tilde{g}_0$. Note that (14) is nothing but another nonstationary semi-strong GARCH process whose parameters are $\tilde{\phi} = (\tilde{g}_0, \tilde{\omega}, \tilde{\alpha}, \tilde{\beta})^\top$ for each time point. Therefore, for each time point $u \in [0,1]$, this family of nonstationary semi-strong GARCH processes can be estimated consistently via the method of Linton *et al.* (2010) using the corresponding sub-sample to which u corresponds. Following the method in Linton *et al.* (2010), we estimate $\tilde{\phi}$ for each time point u. More specifically, the local LADE is defined as a local minimiser of the following objective function

$$S_T\left(\tilde{\phi}\right) = \sum_{t=v+1}^{T} \left|\log y_t^2 - \log \tilde{g}_t(\tilde{\phi})\right| K_h\left(u - t/T\right),$$

where v is a non-negative integer. That is, let

$$\hat{\tilde{\phi}}_{LAD}(u) \equiv \arg\min_{\tilde{\phi}} S_T(\tilde{\phi}). \tag{15}$$

Integrate the estimator for $\tilde{\phi}$ over u, we obtain an estimator for ϕ, since, with (3), we have: $\int_0^1 \tau(u)\omega(u)du = \omega \int_0^1 \tau(u)du = \omega$, $\int_0^1 \alpha(u)du = \alpha$, and $\int_0^1 \tau(u)\beta(u)du = \beta \int_0^1 \tau(u)du = \beta$. Therefore, we let

$$\hat{\phi} = \int_0^1 \hat{\tilde{\phi}}_{LAD}(u)du. \tag{16}$$

Specifically, we obtain $\hat{\alpha} = \int_0^1 \hat{\tilde{\alpha}}_{LAD}(u)du$ and $\hat{\beta} = \int_0^1 \hat{\tilde{\beta}}_{LAD}(u)du$. It is worth noting that as explained in Theorem 2.(ii) and Remark 5 of Linton *et al.* (2010), it is known that by taking any fixed value of $(\tilde{g}_0, \tilde{\omega})$, $(\tilde{\alpha}, \tilde{\beta})$ can be consistently estimated and hence we can estimate (α, β) consistently.

4.2.2 Step 2 : Estimation of long run Volatility

We then use the relationship $\tilde{\beta} = \tau(u)\beta$. We let

$$\hat{\tau}(u) = \frac{\hat{\tilde{\beta}}(u)}{\hat{\beta}}.$$

We could plug this long run volatility estimator back into the objective function to estimate the short run volatility estimator for the GARCH parameters. We may iterate between these two estimation problems, and the convergent values for τ_t and ϕ are conjectured to be more efficient. However, since this is beyond the scope of this paper, we do not pursue this.

5 Asymptotics

5.1 Distribution theory for multiplicative model with strict stationary $\{g_t\}$

To begin with we derive the asymptotic properties of the proposed estimators following the procedure under the assumption of strict stationary $\{g_t\}$. For the presentation of the asymptotic analysis, the following notation is introduced. When $\{y_t\}$ would be strictly stationary if it were not for time index $\tau = t/T$, then we can define a measurable function $\Phi : [0,1] \times \mathbb{R}^\infty \mapsto \mathbb{R}$ such that

$$y_{t,T} = \Phi\left(t/T, \mathcal{Z}_t\right) \tag{17}$$

where $\mathcal{Z}_t = (\cdots, \varepsilon_{t-2}, \varepsilon_{t-1}, \varepsilon_t)$ with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ a stationary ergodic process. This representation is possible for the following reasons. In (17), $\{y_{t,T}\}$ depends on the time index and all the past history of stationary ergodic process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. From (1) and (2), the underlying data generating process $\{y_{t,T}\}$ is a measurable function of all the past history of $\{\varepsilon_t\}$ and time index. Moreover, $\{y_{t,T}\}$ changes so smoothly over time that local stationarity of $\{y_{t,T}\}$ can be ensured for each time index. See also Zhou and Wu (2009).

Let $\xi_t(u)$ be a measurable random variable for all $u \in [0,1]$ such that $\xi_t(u) := \Phi(u, \mathcal{Z}_t)$ with a cumulative distribution function $F(u,x,\mathcal{Z}_t)(= \Pr(\xi_t(u) \leq x))$ for $x \in \mathbb{R}$ and $u \in [0,1]$ and a continuous density function $f(u,x,\mathcal{Z}_t)$. Note that this implies that $\xi_t(u)$ is the local version of $y_{t,T}$. For fixed time index u, $\{\xi_t(u)\}$ is a stationary process. We drop $\mathcal{Z}_t$ for $f(u,x,\mathcal{Z}_t)$ in the following.

Assumption 1 *(a) $y_{t,T}$ is β-mixing with exponentially decaying mixing coefficients and $\sup_T \mathbb{E}|y_{t,T}|^\delta < \infty$ for some $\delta \geq 1$. (b) $f(u,x)$ is uniformly away from zero and uniformly continuous. (c) There exists C and $\epsilon > 0$ such that $\|\xi_t(u_1) - \xi_t(u_2)\|^\epsilon \leq C|u_1 - u_2|$ holds for all u_1, $u_2 \in [0,1]$.*

Assumption 2 *(a) ε_t is strictly stationary and ergodic, ε_t^2 is non-degenerate and for some $\varrho > 0$, there exists a $G < \infty$ such that $E\left(|\varepsilon_t|^{2+\varrho}|\mathcal{F}_{t-1}\right) \leq G$ a.s. (b) Conditional on $\mathcal{F}_{t-1}$, $\log\left(\varepsilon_t^2\right)$ has zero median and a differentiable density function $\mathcal{E}_t(x)$ satisfying $\mathcal{E}_t(0) \equiv \mathcal{E}(0) > 0$, and $\sup_{x\in\mathbb{R},t\geq 1}|\mathcal{E}_t'(x)| < C < \infty$.*

Assumption 3 *(a) $\tau(u)$ is uniformly positive and twice continuously differentiable on $[0,1]$; (b) $\tau(u)$ is Lipschitz continuous, i.e. $|\tau(u_1) - \tau(u_2)| \leq C|u_1 - u_2|$ for $\forall u_1, u_2 \in [0,1]$; (c) $\int_0^1 \tau(u)\,du = 1$.*

Assumption 4 *(i) For* $\phi = (\omega, \alpha, \beta)^\top \in \Theta$ *in strictly stationary* $\{g_t\}$*, the parameter space* Θ *is compact and* $\phi \in int\,(\Theta)$ *where* $int\,(\cdot)$ *denotes an interior point of the space of interest; (ii) For* $\phi = (\eta, \omega, \alpha, \beta)^\top \in \Theta$ *in nonstationary* $\{g_{t,T}\}$*, the parameter space* Θ *is compact and* $\phi \in int\,(\Theta)$ *where* $int\,(\cdot)$ *denotes an interior point of the space of interest.*

Assumption 5 $v \to \infty$ *and* $v/T \to 0$*, as* $T \to \infty$*.*

Assumption 6 *The kernel* $K(\cdot)$ *is a bounded symmetric around zero function such that: (i) it is continuously differentiable up to order* r *on* $\mathbb{R}$ *with* $2 \le r$*; (ii) it belongs to* L^2*,* $\int |K(x)|dx < \infty$ *,* $\int K(x)dx = 1$*,* $||K||_2^2 = \int K^2(x)dx < \infty$ *and the support of* K *is contained in* $[-1, 1]$*; (iii)* $\mu_i(K) = \int x^i K(x)dx = 0$*,* $i = 1, \ldots, r-1$*, and:* $\int x^r K(x)dx \neq 0$*,* $\int |x|^r |K(x)|dx < \infty$*,* $\lim_{||u|| \to \infty} ||u|| K(u) = 0$*; (iv)* $K(\cdot)$ *is Lipschitz continuous, i.e.* $|K(u) - K(u')| \le C|u - u'|$ *for all* $u, u' \in \mathbb{R}^2$*.*

Assumption 7 *As* $T \to \infty$*,* $h \to 0$*,* $Th^4 \to 0$*,* $Th/\log T \to \infty$ *and* $\liminf_{T \to \infty} Th^{1+2\delta} > 0$ *for* δ *defined in Assumption 1.(a).*

Assumption 1.(a) is required to apply the invariance principle in Doukhan, Massart and Rio (1995) for our semiparametric estimation and to obtain the Bahadur representation for the limit distribution of our long-run volatility estimator as well as our semiparametric estimation. For the Bahadur representation, we only need α-mixing in order for the central limit theorem to hold. See Doukhan, Massart and Rio (1994) for more details. However, β-mixing implies α-mixing. Assumption 1.(b) is standard. In particular, we require $\inf_{u \in [0,1], x \in \mathbb{R}} f(u, x) > 0$ for the uniform Bahadur representation. Assumption 1.(c) implies that the process of interest changes smoothly over time, which is consistent with stochastic Lipschitz continuity in Zhou and Wu (2009). This ensures that the underlying data generating process changes smoothly in time. Also, this implies that we do not consider the long memory dependence. Assumption 2 applies to $\{\varepsilon_t\}$. This implies that our short-run volatility is represented by a semi-strong GARCH model with heavy-tailed errors. Assumption 3 is required to ensure that our long-run volatility is slowly time varying and ensures that our nonparametric LADE method and asymptotic arguments based on local stationarity go through along with Assumption 1. Assumptions 4 and 5 are standard for the LADE of a semi-strong GARCH (1,1) from Linton *et al.* (2010) and Peng and Yao (2003). Assumptions 6 and 7 are standard for nonparametric estimation except for $\liminf_{T \to \infty} Th^{1+2\delta} > 0$. The conditions that $\liminf_{T\to\infty} Th^{1+2\delta} > 0$ and $Th^4 \to 0$ are required for Bahadur representation and Theorem 3. See Hall, Peng and Yao (2002).

Assumption 1.(a) further merits our attention. If $\{\varepsilon_t\}$ were a sequence of i.i.d. random variables with $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\varepsilon_t^2 = 1$, the results in Carrasco and Chen (2002) could be applied in order to verify Assumption 1.(a). Our model is much more complicated and hence the results in the paper are not directly applicable. Nevertheless, if our focus is restricted to a sequence of i.i.d random variables $\{\varepsilon_t\}$ without any moment restriction, we can provide some valid argument based on the results in Francq and Zakoïan (2006) whose contribution is to show β-mixing with exponential decay only under strictly stationarity without any moment condition on a general class of GARCH(1,1) processes. Note that our model can be cast into the following framework

$$\begin{aligned} h_t &= \sqrt{g_t}\varepsilon_t \\ g_t &= A(\varepsilon_{t-1})\, g_{t-1} + B(\varepsilon_{t-1}) \end{aligned}$$

where $h_t = y_t/\sqrt{\tau_t}$, $A(\varepsilon_{t-1}) = (\alpha\varepsilon_t^2 + \beta)$ and $B(\varepsilon_{t-1}) = \omega$. Suppose ε_t is i.i.d. without any moment restriction. Due to Theorem 3 in Francq and Zakoïan (2006), $\{h_t\}$ in (1) is shown to be β-mixing with exponential decay given $\mathbb{E}\log(\beta + \alpha\varepsilon_t^2) < 0$ and Assumption A in Francq and Zakoïan (2006), both of which are met under our setting. Since $\{y_t\}$ is obtained by $h_t/\sqrt{\tau_t}$ where τ_t is smoothly time varying, $\{y_t\}$ is also β-mixing with exponential decay by the definition of mixing. Since our model assumes $\{\varepsilon_t\}$ is strictly stationary and ergodic instead of a sequence of i.i.d. $\{\varepsilon_t\}$, our situation is much more complicated and more investigation is required. We leave this as a future research topic.

5.1.1 Long-run Volatility

Our asymptotic theory for the estimator $\hat{\tau}(u)$ is based on Bahadur representation. It is worth noting that asymptotic theories with respect to usual quantile regression involve estimators whose representation are nonlinear, which makes usual asymptotic arguments challenging. It is known that we could approximate these nonlinear quantile estimators by linear forms via the Bahadur representation. See Koenker (2005) and references therein for more details with respect to Bahadur representations. Also, since we obtain $\hat{\tau}(u)$ from (10), the usual Δ-method comes into play and the asymptotic distribution changes accordingly.

Let $\xi_t(u)$ be as defined in the introduction of Section 5 and let:

$$b_\tau(u) = \frac{\mu_2(K)\Gamma(u)\left[\tau^*(u)\right]\left[\log\tau^*(u)\right]^{(2)}}{2} \quad ; \quad V_\tau(u) = \frac{||K||_2^2\left[\Gamma(u)\right]^2\left[\tau^*(u)\right]^2}{4f^2(u, \mathbf{med}(\log\xi_t^2(u))}$$

$$\Gamma(u) = \frac{\int_0^1 \tau^*(u)du - \tau^*(u)}{[\int_0^1 \tau^*(u)du]^2},$$

where $[\log\tau^*(u)]^{(2)}$ is the second derivative of $\log\tau^*(u)$ with respect to u and $\tau^*(u)$ is specified in (7).

Theorem 1 *Suppose Assumptions 1 - 3, 6 and 7 hold. Then, for* $u \in (0,1)$

$$\sqrt{Th}\left(\hat{\tau}(u) - \tau(u) - h^2 b_\tau(u)\right) \xrightarrow{d} N\left(0, V_\tau(u)\right).$$

Theorem 1 shows that for $u \in (0,1)$, our estimator for slowly time varying long-run volatility is asymptotically normal. $\Gamma(u)$ is the first derivative of the continuous function, $\tau^*(u)/\int_0^1 \tau^*(u)du$ and comes from the usual Δ-method. We next discuss how to conduct inference about the functions of interest. Let V_τ denote the asymptotic variance for $\hat{\tau}(u)$. The consistent estimator of $V_\tau(u)$ can be constructed as

$$\hat{V}_\tau(u) = \|K\|_2^2\left[\hat{\Gamma}(u)\right]^2\left[\check{\tau}(u)\right]^2 / 4\hat{f}^2(u, \hat{\zeta}(u)) \tag{18}$$

where $\zeta = \mathbf{med}(\log\xi_t^2(u))$ and with $\check{\tau}(u)$ is specified in (9):

$$\begin{aligned}
\hat{\zeta}(u) &= \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{u - t/T}{h}\right)|\log y_t^2|, \\
\hat{f}(u, \hat{\zeta}(u)) &= \frac{1}{Th_1h_2}\sum_{t=1}^{T} K\left(\frac{u - t/T}{h_1}\right) K\left(\frac{y_t - \hat{\zeta}(u)}{h_2}\right), \\
\hat{\Gamma}(u) &= \frac{\int_0^1 \check{\tau}(u)du - \check{\tau}(u)}{\left[\int_0^1 \check{\tau}(u)du\right]^2}.
\end{aligned}$$

Remark. It would be of interest to test whether τ_t is time varying or a constant. Härdle and Mammen (1993) proposed test statistics based on the L_2-distance between a parametric estimate as a null and a nonparametric estimate as an alternative. This could be tested using the following approach.

$$\begin{aligned} H_0 &: \tau(u) = 1 \text{ for } \forall u \in (0,1) \\ H_1 &: \tau(u) \neq 1 \text{ for some } u \in (0,1) \end{aligned}$$

The test statistic for this hypothesis test could be

$$\mathcal{T} = Th \int_0^1 \left[\hat{\tau}(u) - 1\right]^2 \pi(u) du$$

where $\pi(u)$ is some weight function. It can be shown that under H_0, suitably scaled T has a standard normal distribution. This test statistic is based on the weighted L_2-distance between a parametric estimate as a null and a nonparametric estimate as an alternative.

5.1.2 Short-run Volatility

In the second step, we estimate the short-run volatility g_t via using our estimate $\hat{\tau}(u)$ from the first step due to $\hat{h}_t^2 = y_t^2/\hat{\tau}_t$. Therefore, it is indispensable to show that $\hat{\tau}(u)$ is so close to $\tau_0(u)$ that we can use the estimated $\hat{\tau}(u)$ instead of the true functional form of $\tau(u)$ for parametric estimation of short-run volatility. This is shown in the proof of Theorem 2. Once we show this, the remainder of the relevant asymptotic theory is analogous to that of Peng and Yao (2003).

Let us introduce the following notation. Let $\mathcal{A}_t = (\mathcal{A}_{0t}(\phi), \mathcal{A}_{1t}(\phi), \mathcal{A}_{2t}(\phi))^\top$, where

$$\begin{aligned} \mathcal{A}_{0t}(\phi) &= \frac{\partial g_t^2(\phi)}{\partial \omega} \frac{1}{g_t^2(\phi)} = \frac{1}{1-\beta} \frac{1}{g_t^2(\phi)} \\ \mathcal{A}_{1t}(\phi) &= \frac{\partial g_t^2(\phi)}{\partial \alpha} \frac{1}{g_t^2(\phi)} = \sum_{j=1}^{t} \beta^{j-1} \frac{\hat{h}_{t-j}^2}{g_t^2(\phi)} \\ \mathcal{A}_{2t}(\phi) &= \frac{\partial g_t^2(\phi)}{\partial \beta} \frac{1}{g_t^2(\phi)} = \sum_{j=1}^{t} \beta^{j-1} \frac{g_{t-j}^2(\phi)}{g_t^2(\phi)}. \end{aligned}$$

Let $\mathcal{E}(\cdot)$ be the density function of $\log \varepsilon_t^2$, $\mathcal{E}(0)$ be the density function evaluated at the median of $\log \varepsilon_t^2$, and let $\Sigma = E\left(\mathcal{A}_t \mathcal{A}_t^\top\right)$.

Theorem 2 *Suppose Assumptions 1 - 7 hold. Then, there exists a local minimizer*

$\hat{\phi} = (\hat{\omega}, \hat{\alpha}, \hat{\beta})^{\top}$ *of* $S_T(\phi)$ *as in (12) such that*

$$\sqrt{T}\left(\hat{\phi} - \phi_0\right) \xrightarrow{d} N\left(0, \frac{1}{4\mathcal{E}^2(0)}\Sigma^{-1}\right).$$

We can conduct inference about the functions of interest by obtaining the consistent estimator of the asymptotic variance for $\hat{\phi}(u)$. The consistent estimator of V_ϕ can be constructed as

$$\hat{V}_\phi = \frac{1}{4\hat{\mathcal{E}}^2(0)}\hat{\Sigma}^{-1}$$

where (with $\hat{\varepsilon}_t^2 = y_t^2/\hat{\tau}(u)g_t(\hat{\phi})$):

$$\begin{aligned} \hat{\mathcal{E}}(0) &= \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\log \hat{\varepsilon}_t^2}{h}\right) \\ \hat{\Sigma} &= \frac{1}{T}\sum_{t=1}^{T}[\mathcal{A}_t(\hat{\phi})\mathcal{A}_t^{\top}(\hat{\phi})]. \end{aligned}$$

5.2 Distribution theory for multiplicative model with nonstationary $\{g_{t,T}\}$

To begin with, we state assumptions for nonstationary $\{g_{t,T}\}$. The following assumptions replace any assumption related to strict stationary $\{g_t\}$.

Assumption 8 *Suppose we consider the model (1) and (2) with* $E\log(\alpha_0\varepsilon_t^2 + \beta_0) \geq 0$ *and* $E\varepsilon_t^4 = \infty$. *Let* $S(\phi, \tau(u)) = E|\log y_t^2 - \log\tau(u) - \log g_t(\phi)| < \infty$ *where* $\phi = (\eta, \omega, \alpha, \beta)^{\top}$. *Then, there exists a unique pair* $(\phi_0, \tau_0(u))$ *which minimizes* $S(\phi, \tau(u))$ *uniquely.*

Assumption 9 *In the case of* $E\log(\alpha_0\varepsilon_t^2 + \beta_0) = 0$, ε_t^2 *is* φ*-mixing with* $\sum_{j=1}^{\infty}\varphi_j^{1/2} < \infty$ *where*

$$\varphi_j = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_j^{\infty}, \Pr(A)>0} |\Pr(B) - \Pr(B|A)|$$

with $\mathcal{F}_i^j = \sigma(\varepsilon_t, i \leq t \leq j)$.

Assumption 8 is a minimal high level assumption which ensures identification of both the unknown parameter function of long-run volatility and the unknown parameters of our nonstationary GARCH process. The primary reason behind this is that we are only able to estimate a local minimizer of (α, β) given the local value of (η, ω) due to nonstationarity of our GARCH process. See Jensen and Rahbek (2004b) and Linton *et al.* (2010). Assumption 8 ensures our estimation for $(\phi_0, \tau_0(u))$ leads to the identification of both long-run and short-run volatilities due to the uniqueness of the minimizer of unknown parameters of the short-run GARCH process and the long run volatility function. Assumption 9 ensures that if and only if $E \log (\alpha_0 \varepsilon_t^2 + \beta_0) < 0$, the semi strong GARCH (1,1) defines a unique, strictly stationary and ergodic solution, which implies that $\{g_{t,T}\}$ in this section is a nonstationary semi-strong GARCH (1,1) process.

5.2.1 Short-run Volatility

To begin with, we divide $\tilde{\phi}$ into $\tilde{\theta} = (\tilde{\alpha}, \tilde{\beta})^\top$ and $\tilde{\psi} = (\tilde{\eta}, \tilde{\omega})^\top$. Likewise, we set $\theta = (\alpha, \beta)^\top$ and $\psi = (\eta, \omega)^\top$. This is because we are not able to obtain the asymptotic properties of the estimated ψ due to nonstationarity. Rather, we can estimate $\theta = (\alpha, \beta)^\top$ consistently by taking any value of ψ. See Linton *et al.* (2010) for more details.

We analyse the asymptotics for the short-run volatility as follows. To begin with, we derive the limiting distribution of our local estimator for the short-run volatility parameter for each time point $u \in [0, 1]$ along the line of Linton *et al.* (2010) and Jensen and Rahbek (2004). Based on the limiting distribution of the local estimator, we estimate the global estimator for the short-run dynamics.

The following notation is introduced for the local estimator for the short-run volatility parameter. Note that this is the local estimator and hence each parameter is the function of u. However, unless any confusion is expected, we dispense with u for the simplicity of notation, that is, $\tilde{\phi} = \tilde{\phi}(u)$. With respect to the model (14), define $\tilde{\mathcal{A}}_t(\tilde{\phi}) = (\tilde{\mathcal{A}}_{1t}(\tilde{\phi}), \tilde{\mathcal{A}}_{2t}(\tilde{\phi}))^\top$ where

$$\begin{aligned} \tilde{\mathcal{A}}_{1t}(\tilde{\phi}) &= \frac{\partial \tilde{g}_t^2(\tilde{\phi})}{\partial \tilde{\alpha}} \frac{1}{\tilde{g}_t^2(\tilde{\phi})} = \sum_{j=1}^{t} \beta^{j-1} \frac{y_{t-j}^2}{\tilde{g}_t^2(\tilde{\phi})} \\ \tilde{\mathcal{A}}_{2t}(\tilde{\phi}) &= \frac{\partial \tilde{g}_t^2(\tilde{\phi})}{\partial \tilde{\beta}} \frac{1}{\tilde{g}_t^2(\tilde{\phi})} = \sum_{j=1}^{t} \beta^{j-1} \frac{\tilde{g}_{t-j}^2(\tilde{\phi})}{\tilde{g}_t^2(\tilde{\phi})}. \end{aligned}$$

Also, define the following two corresponding stationary and ergodic processes, $\mathcal{D}_t(\tilde{\alpha}, \tilde{\beta}) =$

$(\mathcal{D}_{1t}(\tilde{\alpha}, \tilde{\beta}), \mathcal{D}_{2t}(\tilde{\alpha}, \tilde{\beta}))$ where

$$\begin{aligned}
\mathcal{D}_{1t}(\tilde{\alpha}, \tilde{\beta}) &= \sum_{j=1}^{\infty} \tilde{\beta}_0^{j-1}(u)\varepsilon_{t-j}^2 \prod_{k=1}^{j} \frac{1}{\tilde{\alpha}_0 \varepsilon_{t-k}^2 + \tilde{\beta}}, \\
\mathcal{D}_{2t}(\tilde{\alpha}, \tilde{\beta}) &= \sum_{j=1}^{\infty} \tilde{\beta}_0^{j-1}(u) \prod_{k=1}^{j} \frac{1}{\tilde{\alpha}_0 \varepsilon_{t-k}^2 + \tilde{\beta}}.
\end{aligned}$$

Dealing with $\tilde{\mathcal{A}}_t(\tilde{\phi})$ directly is tricky since it might be nonstationary. Therefore, we approximate $\tilde{\mathcal{A}}_t(\tilde{\phi})$ by tractable stationary ergodic processes, $\mathcal{D}_t(\cdot)$ for our asymptotic analysis. As is shown in Lemma 3 in Linton *et al.* (2010) and Lemmas 3 and 4 in Jensen and Rahbek (2004b), $\mathcal{D}_{it}(\cdot)$ approximates $\tilde{\mathcal{A}}_{it}(\phi)$ in L_p sense for $i = 1, 2$.

For U being a uniform random variable on $[0, 1]$, let

$$V_{\hat{\theta}} = \frac{\mathbb{E}\left[\left[\mathcal{D}_t(\tilde{\theta}(U))\mathcal{D}_t^{\top}(\tilde{\theta}(U))\right]^{-1}\right]}{4\mathcal{E}^2(0)}.$$

Theorem 3 *Suppose Assumptions 1 - 9 hold. Let ψ_* be any fixed value of ψ. Then, there exists a local minimizer $\hat{\theta} = (\hat{\alpha}, \hat{\beta})^{\top}$ such that*

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, V_{\hat{\theta}}\right).$$

Note that $\hat{\theta}$ is derived from $\hat{\tilde{\theta}}$ by integrating $\hat{\tilde{\theta}}$ over u. In addition, according to Remark 5 in Linton *et al.* (2010), $(\alpha, \tilde{\beta})$ can be estimated by taking any value of $(\tilde{g}_0, \tilde{\omega})$. One may estimate $(\tilde{g}_0, \bar{\omega})$, but the asymptotic properties of the estimated $(\tilde{g}_0, \bar{\omega})$ have not been obtained. In addition, it is worth mentioning that as is shown in the proof of Theorem 3, the rate of convergence is $O_p(T^{-1/2})$ instead of $O_p((Th)^{-1/2})$ when the bandwidth h is properly chosen due to integrating the nonparametric kernel estimator out over u. Moreover, the bias term does not come into play either. These are along the lines of Zhang, Fan and Sun (2009). While they are concerned with local least squares estimation method, we are concerned with local absolute deviation approach with the Bahadur representation.

5.2.2 Long-run Volatility

Since we estimate $\tau(u)$ by using

$$\hat{\tau}(u) = \hat{\hat{\beta}}(u)/\hat{\beta},$$

the limit distribution of $\hat{\tau}(u)$ follows the limit distribution of $\hat{\hat{\beta}}(u)$ very closely.

With $\mathcal{E}(0)$ as defined in Theorem 2, let:

$$b_\tau^*(u) = \frac{\mu_2(K)\frac{\partial^2 \mathbb{E}\mathcal{D}_{2t}(\tilde{\theta}(u))}{\partial u^2}}{2\beta} \quad ; \quad V_\tau^*(u) = \frac{||K||_2^2}{4\mathcal{E}^2(0)\beta^2}\Omega_{22,1}^{-1},$$

$$\Omega_{22,1}^{-1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$$

$$\mathbb{E}\mathcal{D}_t(\tilde{\theta}(u))\mathcal{D}_t^\top(\tilde{\theta}(u)) = \Omega = \begin{bmatrix} \Omega_1 & \Omega_2 \\ \Omega_2^\top & \Omega_3 \end{bmatrix}.$$

Corollary 4 *Suppose that the Assumptions in Theorem 3 hold. Then,*

$$\sqrt{Th}(\hat{\tau}(u) - \tau(u) - h^2 b_\tau^*(u)) \xrightarrow{d} N(0, V_\tau^*(u)).$$

6 Simulation Studies

This section provides our simulation results to examine the finite-sample performance of our estimators for long-run and short-run volatilities. The data are generated from the model (1) with our GARCH (1,1) specification (2). To check robustness of our method, we use two different parameter values for GARCH (1,1) and two different distributions for the innovation. One is for the case where GARCH parameter implies non-weakly stationarity but strong stationarity. The other one is for the case where $\{\varepsilon_t\}$ has a very thick tail. More specifically, for the first simulation study, we use the specification of IGARCH (1,1) with student-t distribution with the degrees of freedom 5. That is, the parameter vector for financial volatility is $(\omega, \eta, \alpha, \beta)^\top = (0.0001, 0.0001, 0.1, 0.9)^\top$ and $\{\varepsilon_t\} \sim t(5)$. For the second simulation study, we use the specification of GARCH (1,1) whose parameter vector is $(\omega, \eta, \alpha, \beta)^\top = (0.0001, 0.0001, 0.1, 0.7)^\top$ along with student-t distribution with the degrees of freedom 2, i.e. $\{\varepsilon_t\} \sim t(2)$.[6] These are consistent with

[6] When the IGARCH parameter values as in the first simulation, (i.e. $(\omega, \eta, \alpha, \beta)^\top = (0.0001, 0.0001, 0.1, 0.9)^\top$) are combined with student-t distribution with the degrees of freedom 2, the condition $\mathbb{E}\log(\alpha_0\varepsilon_t^2 + \beta_0) < 0$ is not met and hence g_t is nonstationary. Note that we do not pursue a simulation study for nonstationary g_t since limit variances of limiting distribution for $\hat{\tau}$ and $\hat{\theta}$ are

our model. Since we want $\tau(u)$ to display expansions and contractions of a business cycle, we set τ_t as

$$\tau(u) = 0.001(0.5 \sin(4\pi u)) + 0.004.$$

We consider two different number of data to provide an asymptotic validity of our approach. We create 600 and 2000 observations with 1000 iterations respectively. We employ an Epanechnikov Kernel for the estimation of long-run volatility. Our bandwidth is given by Silverman's rule of thumb, i.e. $h = std(u)T^{-1/5}$. All the results are found in Appendix B.

For each simulation study, our estimates for parameters of the short-run GARCH(1,1) models are provided in Table 1. Table 1 confirms that our estimates for GARCH parameters come closer to their corresponding true values as the number of observations gets larger, which confirms that our estimates for GARCH parameters are consistent. Table 1 also provides the mean squared error associated with each estimate for both simulations and with 2000 observations, the estimates are close to the true parameter value.

On the other hand, our estimates for long run volatilities of both simulation studies are plotted in Figures 3 and 4 respectively. We truncate boundaries due to boundary issues associated with the local constant nonparametric estimation. This boundary issue is less problematic when you have more data as can be confirmed by Figures 3 and 4. Moreover, this issue can be well addressed when the local linear nonparametric estimation method is adopted. These figures lend clear credence to our estimation procedure. Figures 3 and 4 show that our estimates for the long run volatility functions in our simulation studies are consistent because the estimates get closer as the number of observations gets larger.

In sum, we conduct two simulation studies whereby two different specifications are considered. The results from these simulation studies confirm that our proposed method is quite robust to the change of parameter and distributional specifications.

7 Conclusion

This paper studies the robust semiparametric estimation of a multiplicative model which combines the long-run volatility and the short-run volatility under a unified framework. We allow for heavy tailed errors and time varying unconditional long-run volatility. In addition, this has practical significance. Risk management such as Value at Risk

difficult to estimate due to several nuisance parameters in those quantities.

and expected shortfall should be different between the recession and boom. Moreover, robustness is essential for risk management.

Our estimation strategy primarily involves the separation of the long-run and short-run volatilities. The proposed estimator for the long-run and short-run volatilities are based on well-established nonparametric quantile regression and GARCH estimation. Therefore, consistency results follow directly from those well-established estimation methods given our estimation strategy.

It could be useful to derive the efficiency bound for our semiparametric model and see whether our proposed estimator achieves this bound. However, it is non-trivial given that our model contains a complicated data structure. The existing literature on semiparametric efficiency bounds used a different approach in order to obtain efficiency bounds for different models, i.e. a case-by-case basis. To the best of our knowledge, there is no known semiparametric efficiency bound for trending dependent data without strict moment restrictions. Although our model is too complicated to be covered by the most recent literature, for instance, Ai and Chen (2012). Thus, we leave this intriguing question as a future research topic.

Appendix A

Proof of Theorem 1. Let $S_T(u)$ be defined as

$$S_T(u) = \sum_{t=1}^{T} \{ | \log y_t^2 - \log \tau^*(u) - (Th)^{-1/2}\phi(u) | \} K_h(u - t/T)$$

where $\hat{\phi}(u) = \sqrt{Th}(\log \check{\tau}(u) - \log \tau^*(u))$ such that

$$\hat{\phi}(u) = \underset{\phi}{\operatorname{argmin}} \frac{1}{T} S_T(u). \tag{19}$$

Then, it can be verified that (19) is equivalent to (9) with the relationship $\hat{\phi}(u) = (Th)^{1/2}[\log \check{\tau} - \log \tau^*]$. For simplicity of exposition, we introduce the following notation. $\log y_t^2 = Y_t$, $\log \tau^*(u) = m(u)$, $\tilde{Y}_t = \log y_t^2 - \log \tau^*(u) = Y_t - m(u)$. By construction, the conditional median of $\tilde{Y}_t$ is zero since, for $t/T = u$,

$$\mathbf{med}[\tilde{Y}_t(u)] = \mathbf{med}[Y_t(u) - m(u)] = 0$$

To begin with, $S_T(u)$ can be rewritten as

$$S_T(u) = \sum_{t=1}^{T} \{|\tilde{Y}_t - (Th)^{-1/2}\phi(u)|\} K_h(u - t/T). \tag{20}$$

For the asymptotic distribution of $\hat{\phi}(u)$, we need to find a linear approximation, say Bahadur representation. In this end, we consider the sign function, $sgn(x) = 1 - 2 \cdot \mathbf{1}(x \leq 0)$. Note that $sgn(x)$ is the left derivative of $|x|$. It is worth mentioning that the left derivative of a check function in the quantile regression is used for the asymptotic analysis of the quantile estimator. As is the case in the quantile regression, we define the following sequence,

$$\mathcal{W}_T(\phi) = \sum_{t=1}^{T} W_t(\phi)$$

where $W_t(\phi) = \{|\tilde{Y}_t - (Th)^{-1/2}\phi(u)| - |\tilde{Y}_t| + (Th)^{-1/2}\phi(u) sgn(\tilde{Y}_t)\} K_h(u - t/T)$. It can be verified that

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}}\, S_T(u) = \underset{\phi}{\operatorname{argmin}}[\mathcal{W}_T(\phi) - \sum_{t=1}^{T} \frac{\phi(u)}{\sqrt{Th}} sgn(\tilde{Y}_t) K_h(u - t/T)]. \tag{21}$$

Since $sgn(x) = \mathbf{1}(y > 0) - \mathbf{1}(y < 0)$, $|x| = x \cdot sgn(x)$, and $|x - y| - |x| = -y \cdot sgn(x) + 2(y - x)(\mathbf{1}(0 < x < y) - \mathbf{1}(y < x < 0))$,

$$\begin{aligned} & T\mathbb{E}[\frac{1}{T}\mathcal{W}_T(\phi)] \\ = \ & 2T\mathbb{E}[((Th)^{-1/2}\phi(u) - \tilde{Y}_t)(\mathbf{1}(\mathfrak{a}) - \mathbf{1}(\mathfrak{b}))K_h(u - t/T)] \end{aligned}$$

where $\mathfrak{a} := 0 < \tilde{Y}_t < (Th)^{-1/2}\phi(u)$ and $\mathfrak{b} := (Th)^{-1/2}\phi(u) < \tilde{Y}_t < 0$ respectively.

$$\begin{aligned} & \mathbb{E}[((Th)^{-1/2}\phi(u) - \tilde{Y}_t)(\mathbf{1}(\mathfrak{a}) - \mathbf{1}(\mathfrak{b}))K_h(u - t/T)] \\ = \ & \int_0^1 K_h(u - v) \int_{\mathbb{R}} (\frac{\phi(u)}{\sqrt{Th}} - \tilde{y})(\mathbf{1}_{(0 < \tilde{y} < \frac{\phi(u)}{\sqrt{Th}})} - \mathbf{1}_{(\frac{\phi(u)}{\sqrt{Th}} < \tilde{y} < 0)}) f(u, \tilde{y} + m(u))\, d\tilde{y} dv \\ = \ & \int_0^1 K_h(u - v) \underbrace{\int_{\mathbb{R}} (\frac{\phi(u)}{\sqrt{Th}} - \tilde{y}) \mathbf{1}_{(0 < \tilde{y} < \frac{\phi(u)}{\sqrt{Th}})} f(u, \tilde{y} + m(u))\, d\tilde{y}}_{A} \, dv \\ & - \int_0^1 K_h(u - v) \underbrace{\int_{\mathbb{R}} (\frac{\phi(u)}{\sqrt{Th}} - \tilde{y}) \mathbf{1}_{(\frac{\phi(u)}{\sqrt{Th}} < \tilde{y} < 0)} f(u, \tilde{y} + m(u))\, d\tilde{y}}_{B} \, dv \end{aligned}$$

We can focus on A since the part B is analogous. Define $\epsilon = \frac{\phi(u)}{\sqrt{Th}}$ and then $\epsilon \to 0$ as $T \to \infty$.

$$\begin{aligned} A &= \int_{\mathbb{R}} (\frac{\phi(u)}{\sqrt{Th}} - \tilde{y}) \mathbf{1}_{(0<\tilde{y}<\frac{\phi(u)}{\sqrt{Th}})} f\left(u, \tilde{y} + m(u)\right) d\tilde{y} \\ &= \int_{\mathbb{R}} (\epsilon - \tilde{y}) \mathbf{1}_{(0<\tilde{y}<\epsilon)} f\left(u, \tilde{y} + m(u)\right) d\tilde{y} \\ &= \int_0^{\epsilon} (\epsilon - \tilde{y}) f\left(u, \tilde{y} + m(u)\right) d\tilde{y} \\ &= \int_{m(u)}^{\epsilon + m(u)} (\epsilon - y + m(u)) f(u, y) dy \\ &= \int_{m(u)}^{\epsilon + m(u)} (\epsilon + m(u)) f(u, y) dy - \int_{m(u)}^{\epsilon + m(u)} y f(u, y) dy \end{aligned}$$

Following usual standard kernel estimation method and the mean value theorem, the first term of the above quantity becomes

$$\begin{aligned} & (\epsilon + m(u))[F(u, \epsilon + m(u)) - F(u, m(u))] \\ = \; & (\epsilon + m(u))[\epsilon f(u, m(u)) + \frac{1}{2}\epsilon^2 f^{(1)}(u, m(u)) + o(\epsilon^2)]. \end{aligned}$$

The same argument applies to the second term and we can get

$$A = \frac{1}{2}\epsilon^2 [f(u, m(u))] + o(\delta^2)$$

where $\epsilon^2 = \frac{\phi(u)^2}{Th}$. Therefore,

$$\begin{aligned} & T\mathbb{E}[\frac{1}{T}\mathcal{W}_T(\phi)] \\ = \; & 2T \int_0^1 K_h(u - v)(\frac{1}{2}\epsilon^2 [f(u, m(u))] + o_p(\epsilon^2)) dv \\ = \; & Th \int_0^1 [K(x) \frac{\phi(x + hu)^2}{Th} [f(x + hu, m(x + hu)]] dx + o_p(1) \\ = \; & f(u, m(u))\mu^2(u) + o_p(1) \\ = \; & f(u, \mathbf{med}(\log \xi_t^2(u)))\phi^2(x) + o_p(1) \end{aligned} \tag{22}$$

From (22),

$$\mathbb{E}R_T(\phi) = o_p(1) \tag{23}$$

where $R_T(\phi) = \mathcal{W}_T(\phi) - f(u, \mathbf{med}(\log \xi_t^2(u)))\phi^2(u)$. Moreover, using the similar method in the proof of Theorem 1 of Hall, Peng and Yao (2002),

$$\begin{aligned} \Pr(|\sum_{t=1}^{T} W_t(\phi) - \mathbb{E}W_t(\phi)| &> \varepsilon) \leq \frac{1}{\varepsilon^2}\mathrm{var}(\sum_{t=1}^{T} W_t(\phi)) \\ &\leq \frac{1}{\varepsilon^2}[\sum_{t=1}^{T} \mathbb{E}(W_t^2(\phi)) + 2\sum_{t=1}^{T}(T-t)Cov(W_1, W_{t+1})] \\ &= o_p((Th)^{-1/2}) \end{aligned} \tag{24}$$

since the first term of the second inequality is $o_p((Th)^{-1/2})$ using the standard nonparametric method and the second term of the second inequality is also $o_p((Th)^{-1/2})$ due to covariance inequality in Doukhan (1994). This is expected since strong mixing ensures the asymptotic independence. This implies $R_T(\mu) \xrightarrow{L_p} 0$ with (23). Combining (22) and (24),

$$\mathcal{W}_T(\phi) \xrightarrow{p} f(u, \mathbf{med}(\log \xi_t^2(u)))\phi^2(x).$$

This implies that from (21),

$$\hat{\phi}(u) = \underset{\phi}{\mathrm{argmin}}[f(u, \mathbf{med}(\log \xi_t^2(u)))\phi^2(x) - \sum_{t=1}^{T} \frac{\phi(u)}{\sqrt{Th}} sgn(\tilde{Y}_t)K_h(u - t/T)] + o_p(1)$$

Let $\mathfrak{B}_T$ be defined as

$$\mathfrak{B}_T(u) = \frac{1}{Th}\sum_{t=1}^{T} sgn(\tilde{Y}_t)K_h(u - t/T). \tag{25}$$

Then, by the usual first order condition,

$$2f(u, \mathbf{med}(\log \xi_t^2(u)))\hat{\phi}(u) = (\sqrt{Th})^{-1}\mathfrak{B}_T(u) + o_p(1). \tag{26}$$

This convergence is actually uniform on compact sets Θ for each fixed u due to convexity Lemma in Pollard (1991, p187).

Recall that our aim is to obtain the asymptotic distribution of $\hat{\phi}(u)(= \sqrt{Th}(\log \check{\tau}(u) - \log \tau^*(u)))$. This can be achieved by deriving the limit distribution of the right hand side of (26). From Theorem 1 of Doukhan *et al.* (1994), the central limit theorem can be applied to the right hand side of (26) due to the assumptions of Theorem 1. Moreover, using the standard nonparametric estimation method,

$$\begin{aligned}\mathbb{E}\mathfrak{B}_T(u) &= \frac{1}{h}\int sgn(Y_t - m(u))K(\frac{u-v}{h})dv \\ &= \int K(x)sgn(Y_t - m(x+hu))f(u+hx)dx \\ &= (h^2\mu_2(K)m^{(2)}(u))/2.\end{aligned}$$

In addition,

$$\mathbb{E}[sgn(\tilde{Y}_t)K_h(u-t/T)]^2 = \mathbb{E}K_h(u-t/T)^2 = ||K||_2^2$$

Hence the following result holds.

$$(Th)^{1/2}[\mathfrak{B}_T(u) - \frac{h^2\mu_2(K)m^{(2)}(u)}{2}] \xrightarrow{d} N(0, ||K||_2^2).$$

where $m^{(2)}(u)$ is the second derivative of $\log \tau^*(u)$ with respect to u. Since $\hat{\phi}(u) = \frac{(\sqrt{Th})^{-1}\mathfrak{B}_T(u)}{2f(u,\mathbf{med}(\log \xi_t^2(u)))} + o_p(1)$, we can conclude

$$\sqrt{Th}\left(\log \check{\tau} - \log \tau^* - \frac{h^2\mu_2(K)[\log \tau^*(u)]^{(2)}}{2}\right) \xrightarrow{d} N\left(0, \frac{||K||_2^2}{4f^2(u, \mathbf{med}(\log \xi_t^2\,(u))}\right),$$

where $[\log \tau^*(u)]^{(2)}$ is the second derivative of $\log \tau^*(u)$ with respect to u. Since $\tau^* = \exp(\log \tau^*)$ and $\frac{\partial \exp(\log \tau^*)}{\partial \log \tau^*} = \tau^*$, due to Δ-method,

$$\sqrt{Th}\left(\check{\tau}(u) - \tau^*(u) - \frac{h^2\mu_2(K)\,[\tau^*(u)]\,[\log \tau^*(u)]^{(2)}}{2}\right) \xrightarrow{d} N\left(0, \frac{||K||_2^2\,[\tau^*(u)]^2}{4f^2(u, \mathbf{med}(\log \xi_t^2\,(u))}\right)$$

Due to (8) and another Δ-method,

$$\sqrt{Th}\left(\hat{\tau}\,(u) - \tau\,(u) - \frac{h^2\mu_2(K)\Gamma(u)\,[\tau^*(u)]\,[\log \tau^*(u)]^{(2)}}{2}\right) \xrightarrow{d} N\left(0, \frac{||K||_2^2\,[\tau^*(u)]^2\,[\Gamma(u)]^2}{4f^2(u, \mathbf{med}(\log \xi_t^2\,(u))}\right),$$

where

$$\Gamma(u) = \frac{\int_0^1 \tau^*(u)du - \tau^*(u)}{[\int_0^1 \tau^*(u)du]^2}.$$

This completes the proof. ∎

Proof of Theorem 2. It is worth noting that our two step estimation of long-run volatility and short-run dynamics is based on the usual semiparametric estimation.

Notice that the objective function of LAD estimation, (12) takes a form of $M(\phi, \tau(\cdot))$ where $M(\cdot)_{\mathbb{R}^d \times \mathcal{T} \mapsto \mathbb{R}}$ is a nonrandom measurable function such that $M(\phi_0, \tau_0(\cdot)) = 0$ with $\phi_0 \in \Theta \subset \mathbb{R}^d$ and $\tau(\cdot) \in \mathcal{T}$. That is, an estimator of a finite-dimensional parameter ϕ is obtained by using an estimator of an infinite-dimensional nuisance parameter τ. In order for the infinite-dimensional estimator $\hat{\tau}$ to suffice instead of using the true functional form $\tau_0(\cdot)$, there are two issues to show. These are the uniform Bahadur representation and functional invariance principle for stochastic equicontinuity.

Let us start with the former. To begin with, let $B_h(u) = [b_T, 1 - b_T]$ where $b_T \to 0$. Under the assumptions of Theorem 1, then from straightforward extension of Kong *et al.* (2009) and Zhou and Wu (2009), the following holds.

$$\sup_{u \in B_h} \left| f(u, \mathbf{med}(\log \xi_t^2(u)))(\hat{m}(u) - m(u)) - \mathfrak{B}_T(u) \right| = O_p\left(\frac{\log T}{Th}\right) \tag{27}$$

where $\mathfrak{B}_T(u)$ is defined as in (25) and $m(u) = \log \tau^*(u)$. (27) implies that the true function and its associated estimator are uniformly close.

For the latter, unlike Andrews (1994), and Chen, Linton and Van Keilegom (2003) whose focus is this type of estimation with i.i.d. data, the theory involved in our setting uses characteristics of β-mixing with the exponentially decaying mixing coefficients due to Douhkan, Massart, and Rio (1995). For this, define a random process ε_t. Let $\hat{\varepsilon}_t = \frac{Y_t}{\hat{m}(u)}$ and $\mathcal{U} = [u - h, u + h]$. Then, it suffices to show that under the assumptions of Theorem 2,

$$\begin{aligned} \sup_x |\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(\hat{\varepsilon}_t &\leq x)\mathbf{1}(u \in \mathcal{U}) - \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(\varepsilon_t \leq x)\mathbf{1}(u \in \mathcal{U}) \\ -\{\Pr(\hat{\varepsilon} &\leq x)\Pr(u \in \mathcal{U}) - \Pr(\varepsilon \leq x)\Pr(u \in \mathcal{U})\}| = o_p(T^{-1/2}) \end{aligned} \tag{28}$$

where $\Pr(\hat{\varepsilon} \leq x)$ is the distribution of $\hat{\varepsilon}_t = \frac{Y_t}{\hat{m}(u)}$ in the neighborhood of a time point u.

Let $\mathcal{R}$ be defined such that

$$\begin{aligned} \mathcal{R} &= \mathbf{1}(\hat{\varepsilon} \leq x)\mathbf{1}(u \in \mathcal{U}) - \mathbf{1}(\varepsilon \leq x)\mathbf{1}(u \in \mathcal{U}) \\ -\Pr(\hat{\varepsilon} &\leq x)\Pr(u \in \mathcal{U}) + \Pr(\varepsilon \leq x)\Pr(u \in \mathcal{U}). \end{aligned}$$

Then, $\mathcal{R}$ can be rewritten as

$$\mathcal{R} = \mathbf{1}(\varepsilon \leq x + \delta x)\mathbf{1}(u \in \mathcal{U}) - \mathbf{1}(\varepsilon \leq x)\mathbf{1}(u \in \mathcal{U}) - \Pr(\varepsilon \leq x + \delta x)\Pr(u \in \mathcal{U}) + \Pr(\varepsilon \leq x)\Pr(u \in \mathcal{U})\}, \tag{29}$$

where $\delta = \frac{\hat{m}(u)-m(u)}{m(u)}$. (28) is concerned with stochastic equicontinuity. To show (28), we need the functional invariance principle in the sense of Donsker for β-mixing processes. Due to Theorem 1 and eq. (2.11) in Doukhan *et al.* (1995) with the assumption 1 and the boundedness of the indicator function, this can be proven if the entropy with bracketing with respect to $\mathcal{R}$ satisfies the following integrability condition,

$$\mathcal{J}_{[\,]}(\mathcal{R}, L_{\varkappa,2}(P)) = \int_0^1 \sqrt{\log \mathcal{N}_{[\,]}(u, \mathcal{R}, L_{\varkappa,2}(P))}du < \infty \tag{30}$$

where $\mathcal{N}_{[\,]}(\cdot)$ is the covering number, $L_{\varkappa,2}(P)$ denotes functional space equipped with the usual Orlicz norm associated with the function $x \to \varkappa(x^2)$ and $\mathcal{J}_{[\,]}(\cdot)$ is the bracketing integral. It is known that if $\varkappa(x) = x^r$, $L_{\varkappa,2}(P) = L_{2r}(P)$, the Orlicz norm is the usual norm such that $||g||_{2r} = [\int |g(x)|^{2r} dP(x)]^{1/2r}$ for any function $g : \mathcal{X} \to \mathbb{R} \in L_{2r}(P)$. See Doukhan *et al.* (1995, pp 401-404) for more details. Let me begin with $\mathbf{1}(\varepsilon \leq x + \delta x)\mathbf{1}(u \in \mathcal{U})$ in (29). In order to show that (30) holds for $\mathbf{1}(\varepsilon \leq x + \delta x)\mathbf{1}(u \in \mathcal{U})$, it suffices to show that $\mathbf{1}(\varepsilon \leq x + \delta x)\mathbf{1}(u \in \mathcal{U})$ is locally uniformly $L_{2r}(P)$ continuous with respect to x and k. For k_1 and k_2 such that $|k_1 - k_2| < \delta$ and x_1 and x_2 such that $|x_1 - x_2| < \delta$, the following holds

$$\begin{aligned}
\mathbf{1}(\varepsilon_t &\leq x_1 + k_1 x_1)\mathbf{1}(u \in \mathcal{U}) - \mathbf{1}(\varepsilon_t \leq x_2 + k_2 x_2)\mathbf{1}(u \in \mathcal{U}) \\
&\leq \mathbf{1}(|\varepsilon_t - x_1 - k_1 x_1| \leq |x_1 - x_2 + k_1 x_1 - k_2 x_2|)\mathbf{1}(u \in \mathcal{U}) \\
&\leq \mathbf{1}(|\varepsilon_t - x_1 - k_1 x_1| \leq \sup_{\substack{|k_1-k_2|<\delta, \\ |x_1-x_2|<\delta}} |x_1 - x_2 + k_1 x_1 - k_2 x_2|)\mathbf{1}(u \in \mathcal{U}) \\
&\leq \mathbf{1}(|\varepsilon_t - x_1 - k_1 x_1| \leq |\delta + k_1\delta + \delta x_2|)
\end{aligned} \tag{31}$$

Using (31), the following holds. For $r > 1$,

$$\begin{aligned}
\mathbb{E}[\sup_{|k_1-k_2|<\delta, |x_1-x_2|<\delta} [[\mathbf{1}(\varepsilon_t &\leq x_1 + k_1 x_1) - \mathbf{1}(\varepsilon_t \leq x_2 + k_2 x_2)]\mathbf{1}(u \in \mathcal{U})]^{2r}] \\
&\leq \Pr(|\varepsilon_t - x_1 - k_1 x_1| < |\delta + k_1\delta + \delta x_2|) \\
&\leq F(\varepsilon < x_1 + k_1 x_1 + \delta + k_1\delta + \delta x_2) \\
-F(\varepsilon &< x_1 + k_1 x_1 - (\delta + k_1\delta + \delta x_2)) \leq C\delta \to 0
\end{aligned}$$

The first equality comes from (31) and the second equality comes from the definition of the CDF. Therefore, the Donsker theorem holds for $\mathbf{1}(\varepsilon \leq x + \delta x)\mathbf{1}(u \in \mathcal{U})$, which yields the desired result. For the other terms of (29), the similar argument applies due to the straightforward extension of the result in Theorem 6 of Andrews (1994). Then,

(28) follows. From (27) and (28), we can proceed to estimate the parameters of financial volatility as if we knew the true functional form, $\tau(u)$. That is, we can use $\hat{h}_t$ instead of h_t. Then, the problem can be dealt with as if it were a parametric situation just like Theorem 1 in Peng and Yao (2003). As a result, the remainder is analogous to the proof of Theorem 1 in Peng and Yao (2003). ■

Proof of Theorem 3. This proof is the local version of Theorem 2 in Linton *et al.* (2010) along the lines of kernel nonparametric estimation. Just as Peng and Yao (2003) and Linton *et al.* (2010), we define $Z_t(\tilde{\phi}) = \log y_t^2 - \log \tilde{g}_t(\tilde{\phi})$ and $Z_t(\tilde{\theta}) = Z_t(\tilde{\phi})\Big|_{\phi=(\tilde{\theta}^\top,\tilde{\psi}_*)}$ where $\tilde{\theta} = \tilde{\theta}_0 + \frac{1}{\sqrt{Th}}\nu$, $\nu = (\nu_1, \nu_2)^\top \in \mathbb{R}^2$. From the usual nonparametric estimation, it can be shown that $\hat{\tilde{\theta}}(u) = \tilde{\theta}_0 + \frac{1}{\sqrt{Th}}\hat{\nu}$ where $\hat{\nu}$ is the minimizer of

$$\mathcal{T}(\nu) = \sum_{t=v+1}^{T} \left(|Z_t(\tilde{\theta}_0 + (Th)^{-1/2}\nu)| - \left|Z_t(\tilde{\theta}_0)\right| \right) K_h(u - t/T) \tag{32}$$

As is shown in Peng and Yao (2003) and Linton *et al.* (2010), since (32) has the same limit distribution as

$$\mathcal{T}^+(\nu) = \sum_{t=v+1}^{T} \left(|Z_t(\tilde{\theta}_0) - (Th)^{-1/2}\nu^\top \tilde{\mathcal{A}}_t| - \left|Z_t(\tilde{\theta}_0)\right| \right),$$

where $\tilde{\mathcal{A}}_t = (\tilde{\mathcal{A}}_{1t}, \tilde{\mathcal{A}}_{2t})^\top$ with $\tilde{\mathcal{A}}_{it}$ defined in section 5.2.1, we focus on $\mathcal{T}^+(\nu)$. In the following, we use $t = 1$ instead of $t = v + 1$ since the results are identical under the assumption 5 due to Theorem 2 of Peng and Yao (2003). Since

$$|x - y| - |x| = -y \cdot sgn(x) + 2(y - x)(\mathbf{1}(0 < x < y) - \mathbf{1}(y < x < 0)),$$

$$\begin{aligned}
\mathcal{T}^+(\nu) &= -(Th)^{-1/2} \sum_{t=1}^{T} \nu^\top \tilde{\mathcal{A}}_t sgn(\log \varepsilon_t^2) K_h(u - t/T) + 2 \sum_{t=1}^{T} \left[(Th)^{-1/2}\nu^\top \tilde{\mathcal{A}}_t - \log \varepsilon_t^2 \right] \\
&\quad \times \left[\mathbf{1}(0 < \log \varepsilon_t < (Th)^{-1/2}\nu^\top \tilde{A}_t) - \mathbf{1}((Th)^{-1/2}\nu^\top \tilde{\mathcal{A}}_t < \log \varepsilon_t^2 < 0) \right] K_h(u - t/T) \\
&= \mathcal{J}_{1T} + \mathcal{J}_{2T}
\end{aligned}$$

where

$$\mathcal{J}_{1T} = -(Th)^{-1/2} \sum_{t=1}^{T} \nu^\top \tilde{\mathcal{A}}_t sgn(\log \varepsilon_t^2) K_h(u - t/T)$$

and $\mathcal{J}_{2T}$ is the remainder. For $\mathcal{J}_{1T}$, note that $\{\nu^\top \tilde{\mathcal{A}}_t sgn(\log \varepsilon_t^2)\}$ is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$ under the Assumption 2. Therefore, from the usual nonparametric estimation and Lemmas 3, 4 and 5 in Linton *et al.* (2010),

$$\begin{aligned}
\mathbb{E}\mathcal{J}_{1T} &= \nu^\top \frac{1}{h}\int_0^1 \tilde{\mathcal{A}}_t sgn(\log\varepsilon_t^2)K(\frac{u-x}{h})dx \\
&= \nu^\top \left[\int_0^1 \left[\mathbb{E}\left[\tilde{\mathcal{A}}_t sgn(\log\varepsilon_t^2)\right] + \frac{h^2 v^2}{2}\frac{\partial^2 \mathbb{E}\left[\tilde{\mathcal{A}}_t(\tilde{\theta}(u))\right]}{\partial u^2}\right] K(v)dv\right] \\
&= \nu^\top \frac{h^2\mu_2(K)\frac{\partial^2 \mathbb{E}\left[\tilde{\mathcal{A}}_t(\tilde{\theta}(u))\right]}{\partial u^2}}{2} \xrightarrow{p} \nu^\top \frac{h^2\mu_2(K)\frac{\partial^2 \mathbb{E}\mathcal{D}_t(\tilde{\theta}(u))}{\partial u^2}}{2},
\end{aligned}$$

where the stationary and ergodic process $\mathcal{D}_t(\tilde{\theta})$ approximates $\tilde{\mathcal{A}}_t(\tilde{\theta})$ in L^p sense. Moreover,

$$\mathbb{E}\left[\mathcal{J}_{1T}\right]^2 = \nu^\top \varphi\Omega\nu$$

Therefore,

$$(Th)^{1/2}\left[\mathfrak{W}_T(u) - \nu^\top \frac{h^2\mu_2(K)\frac{\partial^2 \mathbb{E}\mathcal{D}_t(\tilde{\theta}(u))}{\partial u^2}}{2}\right] \xrightarrow{d} N(0, ||K||_2^2 \nu^\top \Omega\nu) \tag{33}$$

where

$$\mathfrak{W}_T(u) = (Th)^{-1}\sum_{t=1}^T \nu^\top \tilde{\mathcal{A}}_t sgn(\log\varepsilon_t^2)K_h(u-t/T), \tag{34}$$

$\Omega = \mathbb{E}(\mathcal{D}_t\mathcal{D}_t^\top)$ with $\mathcal{D}_{it}$ defined in section 5.2.1. For $\mathcal{J}_{2T}$, define

$$\begin{aligned}
\mathcal{J}_{2T}^{(1)} &= \left[(Th)^{-1/2}\nu^\top\tilde{\mathcal{A}}_t - \log\varepsilon_t^2\right]\mathbf{1}(0 < \log\varepsilon_t < (Th)^{-1/2}\nu^\top\tilde{A}_t)K_h(u-t/T), \\
\mathcal{J}_{2T}^{(2)} &= \left[(Th)^{-1/2}\nu^\top\tilde{\mathcal{A}}_t - \log\varepsilon_t^2\right]\mathbf{1}((Th)^{-1/2}\nu^\top\tilde{\mathcal{A}}_t < \log\varepsilon_t^2 < 0)K_h(u-t/T).
\end{aligned}$$

Note that it can be shown that from (A.6) in Linton *et al.* (2010),

$$\begin{aligned}
\sum_{t=1}^T \mathbb{E}[(\mathcal{J}_{2T}^{(1)} - \mathcal{J}_{2T}^{(2)})|\mathcal{F}_{t-1}] &= A_1 + A_2 + B_1 + B_2 \\
&= A_1 + B_1 + o_p(1),
\end{aligned}$$

where

$$
\begin{aligned}
A_1 &= \sum_{t=1}^{T} \mathbf{1}(\nu^\top \tilde{A}_t > 0) \int_0^1 \int_0^{\frac{\nu^\top \tilde{A}_t}{\sqrt{Th}}} \left[(Th)^{-1/2}\nu^\top \tilde{\mathcal{A}}_t - z\right] \mathcal{E}\left(0\right) K_h(u-x)dzdx, \\
A_2 &= \sum_{t=1}^{T} \mathbf{1}(\nu^\top \tilde{A}_t > 0) \int_0^1 \int_0^{\frac{\nu^\top \tilde{A}_t}{\sqrt{Th}}} \left[(Th)^{-1/2}\nu^\top \tilde{\mathcal{A}}_t - z\right] \left[\mathcal{E}_t\left(z\right) - \mathcal{E}\left(0\right)\right] K_h(u-x)dzdx, \\
B_1 &= \sum_{t=1}^{T} \mathbf{1}(\nu^\top \tilde{A}_t \leq 0) \int_0^1 \int_{\frac{\nu^\top \tilde{A}_t}{\sqrt{Th}}}^{0} \left[(Th)^{-1/2}\nu^\top \tilde{\mathcal{A}}_t - z\right] \mathcal{E}\left(0\right) K_h(u-x)dzdx, \\
B_2 &= \sum_{t=1}^{T} \mathbf{1}(\nu^\top \tilde{A}_t \leq 0) \int_0^1 \int_{\frac{\nu^\top \tilde{A}_t}{\sqrt{Th}}}^{0} \left[(Th)^{-1/2}\nu^\top \tilde{\mathcal{A}}_t - z\right] \left[\mathcal{E}_t\left(z\right) - \mathcal{E}\left(0\right)\right] K_h(u-x)dzdx.
\end{aligned}
$$

$$
\begin{aligned}
A_1 + B_1 &= \frac{\mathcal{E}\left(0\right)}{2} \frac{1}{Th} \sum_{t=1}^{T} \int_0^1 \left[\nu^\top \tilde{\mathcal{A}}_t \tilde{\mathcal{A}}_t^\top \nu\right] K_h(u-x)dx \\
&\xrightarrow{p} \frac{\mathcal{E}\left(0\right)}{2} \mathbb{E}(\nu^\top \mathcal{D}_t \mathcal{D}_t^\top \nu) + o_p(1).
\end{aligned}
$$

As a result,

$$
\sum_{t=v+1}^{T} \mathbb{E}[(\mathcal{J}_{2T}^{(1)} - \mathcal{J}_{2T}^{(2)})|\mathcal{F}_{t-1}] \xrightarrow{p} \frac{\mathcal{E}\left(0\right)}{2} \mathbb{E}(\nu^\top \mathcal{D}_t \mathcal{D}_t^\top \nu) + o_p(1), \tag{35}
$$

Define $\hat{\mu}(u) = \sqrt{Th}(\hat{\tilde{\theta}}(u) - \tilde{\theta}(u))$. From (33) and (35),

$$
2\mathcal{E}\left(0\right) \mathbb{E}(\mathcal{D}_t \mathcal{D}_t^\top) \hat{\mu}(u) = (\sqrt{Th})^{-1} \mathfrak{B}_T(u) + o_p(1).
$$

where $\mathfrak{B}_T(u) = (Th)^{-1} \sum_{t=1}^{T} \tilde{\mathcal{A}}_t sgn(\log \varepsilon_t^2) K_h(u - t/T)$, which is the Bahadur representation of $\hat{\mu}(u)$. Note that $\mathfrak{W}_T(u)$ from (34) has the relationship with $\mathfrak{B}_T(u)$ such that $\mathfrak{W}_T(u) = \nu^\top \mathfrak{B}_T(u)$. For the remainder, the similar argument as in Linton *et al.* (2010) applies for each time point u by construction along with the proof of Theorem 1 in this paper. This means that convexity is ensured for $\mathcal{T}^+(\nu)$ and $\mathcal{T}(\nu)$. Following

$$
\sqrt{Th}\left(\hat{\tilde{\theta}}(u) - \tilde{\theta}(u) - \frac{h^2 \mu_2(K) \frac{\partial^2 \mathbb{E}\mathcal{D}_t(\tilde{\theta}(u))}{\partial u^2}}{2}\right) \xrightarrow{d} N\left(0, \frac{||K||_2^2}{4\mathcal{E}^2\left(0\right)} \Omega^{-1}\right) \tag{36}
$$

where $\mathcal{E}\left(\cdot\right)$ is defined in Theorem 2, and $\Omega = E\left(\mathcal{D}_t(\tilde{\theta}(u)) \mathcal{D}_t^\top(\tilde{\theta}(u))\right)$.

From (36) and (16), we can derive the limit distribution of $\hat{\theta}$ in a similar way to the proof of Theorem 5 in Zhang *et al.* (2009) due to asymptotic independence of kernel estimator shown in Lemma 7.1 of Robinson (1983) under our β-mixing assumption 1. ■

Proof of Corollary 4. Recall that

$$\hat{\tau}(u) = \hat{\tilde{\beta}}(u)/\hat{\beta}.$$

Note that

$$\begin{aligned}
\sqrt{Th}(\hat{\tau}(u) - \tau_0(u)) &= \sqrt{Th}\left(\frac{\hat{\tilde{\beta}}(u)}{\hat{\beta}} - \frac{\tilde{\beta}(u)}{\beta}\right) \\
&= \sqrt{Th}\left(\frac{\hat{\tilde{\beta}}(u)}{\hat{\beta}} - \frac{\hat{\tilde{\beta}}(u)}{\beta} + \frac{\hat{\tilde{\beta}}(u)}{\beta} - \frac{\tilde{\beta}(u)}{\beta}\right) \\
&= \sqrt{Th}\hat{\tilde{\beta}}(u)\left(\frac{1}{\hat{\beta}} - \frac{1}{\beta}\right) + \frac{\sqrt{Th}}{\beta}(\hat{\tilde{\beta}}(u) - \tilde{\beta}(u)) \\
&= \frac{\sqrt{Th}}{\beta}(\hat{\tilde{\beta}}(u) - \tilde{\beta}(u)) + o_p(1)
\end{aligned}$$

where $\hat{\beta} = \beta_0 + O_p(T^{-1/2})$. Then, due to Theorem 3, the result follows. ■

Appendix B: Tables and Figures

Table 1. Estimates for the short run GARCH parameter (ω, α, β)

Simulation I	T=600			T=2000		
	ω	α	β	ω	α	β
estimate	0.0008	0.088	0.889	0.0009	0.097	0.896
Bias	0.0002	0.012	0.011	6.4e-5	0.003	0.004
MSE	0.0001	0.009	0.012	2.2e-5	0.003	0.005
Simulation II	T=600			T=2000		
estimate	0.0007	0.081	0.677	0.0009	0.092	0.689
Bias	0.0003	0.019	0.023	0.0001	0.008	0.011
MSE	0.0003	0.018	0.056	0.0001	0.008	0.021

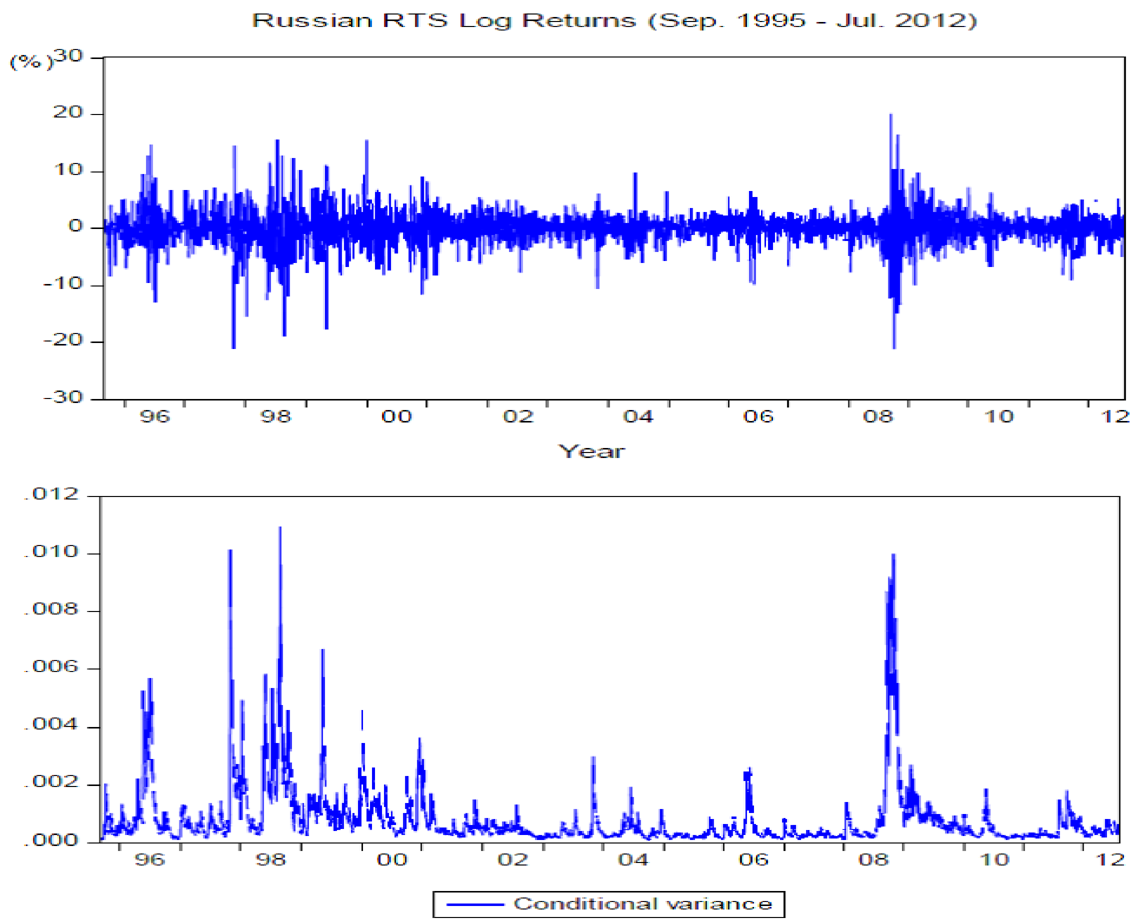


Figure 1: 1) Daily Log returns of Russian RTS index (Sep. 1975 - Jul. 2012) (upper panel); (2) Conditional Variance based on the AR(2)-GARCH(1,1) model (lower panel)

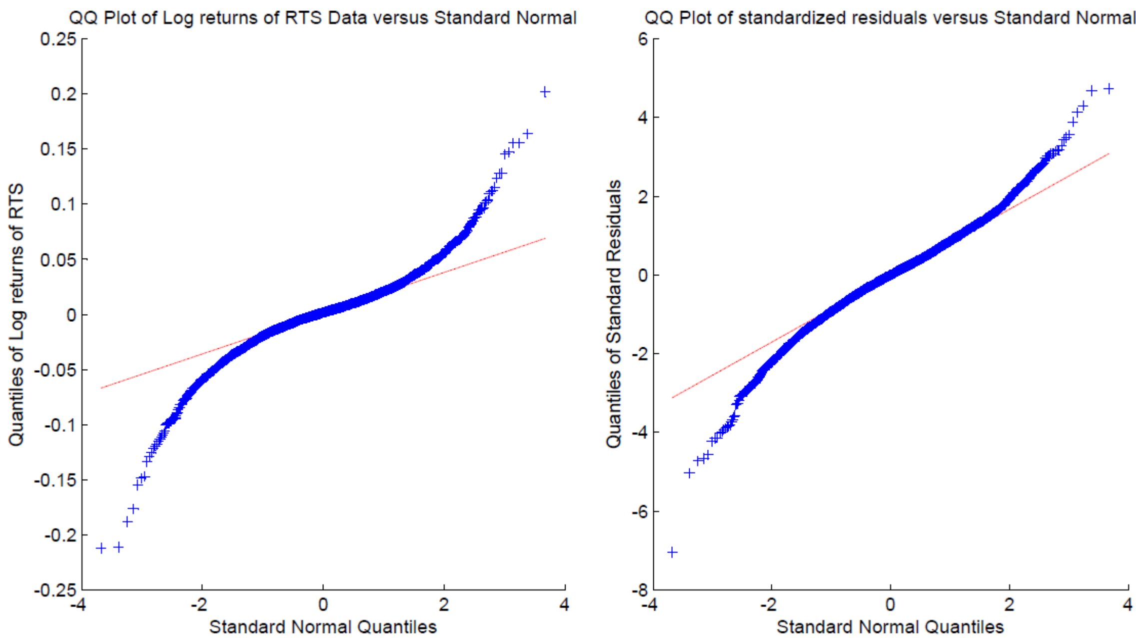


Figure 2: 1) QQ-plot of log return of daily RTS index (Sep. 1995 - Jul. 2012) (left panel); 2) QQ-plot of standardized residuals after the AR(2)-GARCH(1,1) filtering based on the same data as in the left panel. (right panel)

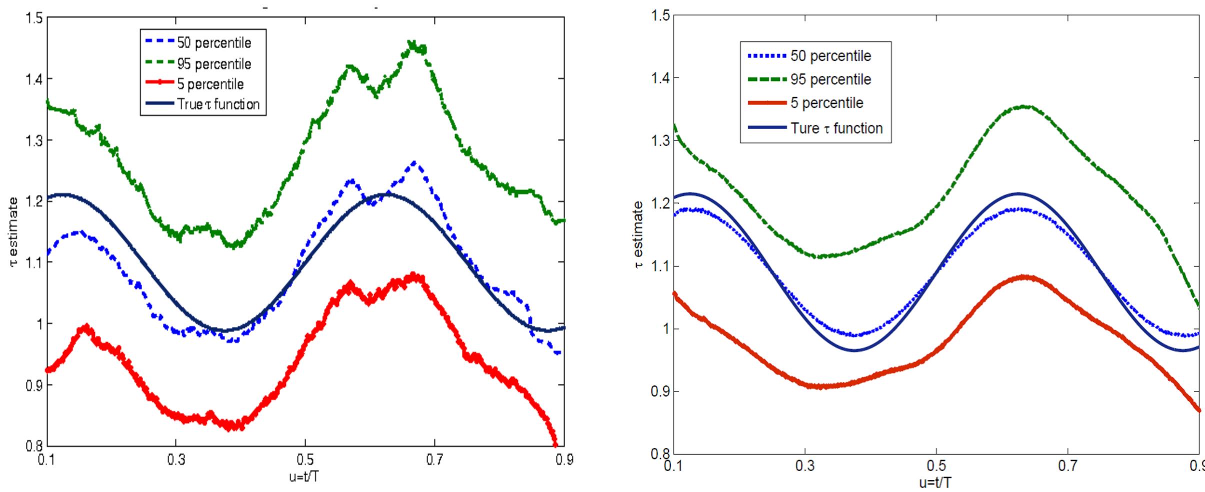
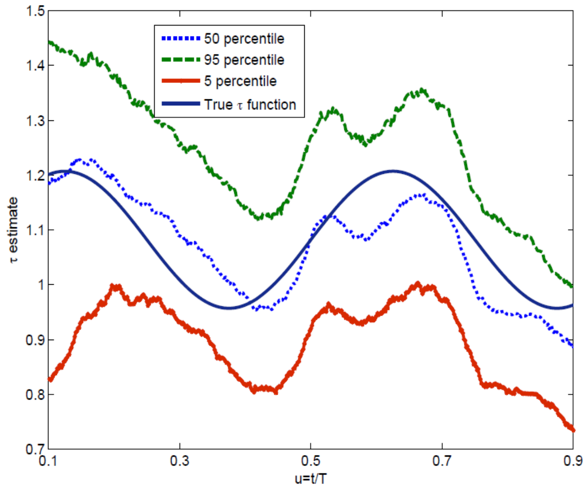


Figure 3:

(a) Simulation I: $(\omega, \eta, \alpha, \beta)^\top = (0.0001, 0.0001, 0.1, 0.9)^\top$, $\{\varepsilon_t\} \sim t(5)$.

(b) Truncation occurs at $u = 0.1$ and $u = 0.9$ to avoid usual boundary issue of nonparametric estimation.

(c) For the left and right panels, the number of generated data is 600 and 2000 respectively.



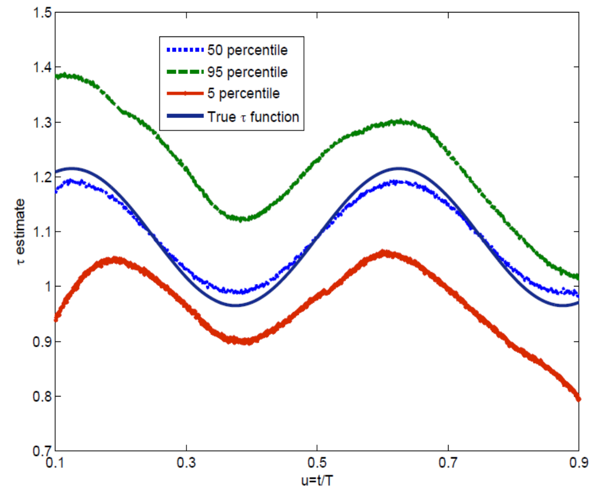


Figure 4:

(a) Simulation II: $(\omega, \eta, \alpha, \beta)^\top = (0.0001, 0.0001, 0.1, 0.7)^\top$, $\{\varepsilon_t\} \sim t(2)$.

(b) Truncation occurs at $u = 0.1$ and $u = 0.9$ to avoid usual boundary issue of nonparametric estimation.

(c) For the left and right panels, the number of generated data is 600 and 2000 respectively.

References

Ai, C., Chen, X. (2012) The Semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions.

Andrews, D. W. K. (1994) Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity. *Econometrica* 62, 43-72.

Bikbov, R., Chernov, M. (2010) No-Arbitrage Macroeconomic Determinants of the Yield Curve. *Journal of Econometrics* 159, 166-182.

Black, F. (1976) Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section*, American Statistical Association, 171-181

Bollerslev, T., Chou, R., Kroner, K. (1992) ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics* 52, 5-59.

Carrasco, M., Chen, X. (2002) Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory* 18, 17-39.

Chen, X., Linton, O., Van Keilegom, I. (2003) Estimation of Semiparametric Models when the Criterion Function is not Smooth, *Econometrica* 71, 1591-1608.

Doukhan, P. (1994) Mixing : properties and examples. Springer, New York

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

Delgado, M.A., Hidalgo, J. (2000) Nonparametric inference on Structural Breaks. *Journal of Econometrics* 96, 113-144.

Doukhan, P., Massart, P., Rio, E. (1994) The Central Limit Theorem for Strongly Mixing Processes. *Annales de l'Institut Henri Poincare, Series B, Probability and Statistics* 30, 63-82.

Doukhan, P., Massart, P., Rio, E. (1995) Invariance Principles for Absolutely Regular Processes. *Annales de l'Institut Henri Poincare, Series B, Probability and Statistics* 31, 393-427.

Drost, F.C., Nijman, T.E. (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909-927.

Engle, R., Rangel, J. (2008) The spline-GARCH Model for Low Frequency Volatility and Its Global Macroeconomic Causes. *Review of Financial Studies* 21, 1187-1222.

Francq, C., Zakoïan, J. (2006) Mixing Properties of a General Class of GARCH (1,1) Models without Moment Assumptions on the Observed Process. *Econometric Theory* 22, 815-834.

Giurcanu, M., Spokoiny, V. (2004) Confidence estimation of the covariance function of stationary and locally stationary processes. *Statistical Decisions* 22, 283-300.

Hall, P., Peng, L., Yao, Q. (2002) Prediction and Nonparametric Estimation for Time Series with Heavy Tails. *Journal of Time Series Analysis* 23, 313-331.

Hall, P., Yao, Q. (2003) Inference in ARCH and GARCH Models with Heavy-Tailed Errors. *Econometrica* 71, 285-317.

Hafner, C., Linton, O. (2010) Efficient Estimation of a Multivariate Multiplicative Volatility Model. *Journal of Econometrics* 159, 55-73.

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

Härdle, W., Tuan, P.-D. (1986) Some Theory on M-Smoothing of Time Series. *Journal of Time Series Analysis* 7, 191–204

Huang, D., Wang, H., Yao, Q. (2008) Estimating GARCH models: when to use what? *Econometrics Journal* 11. 27-38.

Jensen, S., Rahbek, A. (2004a) Asymptotic Normality of the QMLE Estimator of ARCH in the Nonstationary Case. *Econometrica* 72, 641-646

Jensen, S., Rahbek, A. (2004b) Asymptotic Inference for Nonstationary GARCH. *Econometric Theory* 20, 1203-1226.

Koenker, R. (2005) Quantile regression, Cambridge University Press, Cambridge.

Kong, E., Linton, O., Xia, Y. (2010) Uniform Bahadur Representation for Local Polynomial Estimates of M-Regression and Its Application to the Additive Model. *Econometric Theory* 26, 1529-1564.

Koo, B., Linton, O. (2012) Estimation of Semiparametric Locally Stationary Diffusion Models. *Journal of Econometrics* 170, 210-233.

Koo, B. (2012) Locally Stationary Diffusion Processes with Structural Breaks, Manuscript, Monash University.

Lee, S., Hansen, B. (1994) Asymptotic Theory for the Garch (1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory* 10, 29-52.

Linton, O., Pan, J., Wang, H. (2010) Estimation for a non-stationary semi-strong GARCH (1,1) model with heavy-tailed errors. *Econometric Theory* 26, 1-28.

Mittnik, S., Rachev, S.T. (2000) Stable Paretian Models in Finance. Wiley, New York.

Peng, L., Yao, Q. (2003) Least Absolute Deviations Estimations for ARCH and GARCH models. *Biometrika* 90. 967-975.

Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186-198

Rachev, S.T. (2003) Handbook of Heavy Tailed Distributions in Finance. North Holland

Robinson, P., (1983) Nonparametric Estimators For Time Series. *Journal of Time Series Analysis* 4, 185-207.

Robinson, P. (1989) Nonparametric Estimation of Time-Varying Parameters. Statistical Analysis and Forecasting of Economic Structural Change. Springer-Verlag, 253-264.

Rodriguez-Poo, J., Linton, O. (2001) Nonparametric Factor Analysis for Residual Time Series. *Test* 10. 161-182.

Schwert, W. (1989) Why Does Stock Market Volatility Change Over Time? *Journal of Finance* 44, 1115-1153.

van Bellegem, S. (2011) Locally Stationary Volatility Modelling. Preprint.

Veronesi, P. (1999) Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model. *Review of Financial Studies* 12, 975-1007.

Zhang, W., Fan, J., Sun, Y. (2009) A Semiparametric Model for Cluster Data. *The Annals of Statistics 37*, 2377-2408.

Zhou, Z., Wu, W. (2009) Local linear Qantile Estimation For Non-stationary Time Series. *The Annals of Statistics 37*, 2696-2729.