A time varying GARCH (p, q) model and related statistical inference

Technical Report No: ISINE/ASD/AOSU/2013-14/001
Report Date: 01.04.2013

Neelabh Rohan
Indian Statistical Institute, North East Centre,
Tezpur University Campus, Tezpur Assam, 784028, INDIA
Email: neelabh@isine.ac.in

Indian Statistical Institute
North-East Centre, Tezpur, Assam-784028

A time varying GARCH (p, q) model and related statistical inference

Neelabh Rohan

Indian Statistical Institute, North East Centre,
Tezpur University Campus, Tezpur Assam, 784028, INDIA
Email: neelabh@isine.ac.in

Abstract

In this paper, we introduce a time varying GARCH (tvGARCH (p, q)) model and consider certain related inferential problems. A two-step local polynomial estimator for the parameter functions of the tvGARCH (p, q) model is proposed. The asymptotic distribution of the suggested estimator depends on the unknown quantities. In order to overcome this issue, a weighted bootstrapped estimator is suggested. We prove that the asymptotic distribution of the bootstrapped estimator coincides with that of the actual local polynomial estimator. The validity of the bootstrapped estimator is also established empirically. Simulation results indicate that the bootstrapped estimator provides a better approximation to normality than the actual estimator. We also suggest a test statistic to test the constancy of the parameter functions of the tvGARCH (p, q) model. The asymptotic distribution of the test statistic is derived. The bootstrapped estimator facilitates in computation of the test statistic. The performance of the test is judged with the help of a simulation study.

Mathematical Subject classification: 62M10, 62G05

Keywords: *Local polynomial estimation, time-varying GARCH, volatility modeling, weighted bootstrap.*

1 Introduction

Modeling financial market volatility using the non-stationary models has received considerable attention in the recent years in the wake of several financial crises and high volatility due to frequent changes in the market. Justification towards the use of such models can be found in Rohan and Ramanathan (2012a, 2012b), Čížek and Spokoiny (2009), Fryzlewicz et al. (2008), Amado and Terasvirta (2008) Dahlhaus and Subba Rao (2006) and Mikosch and Starica (2004) among others. Rohan and Ramanathan (2012a) (RR hereafter) generalized the time varying ARCH (tvARCH) model of Dahlhaus and Subba Rao (2006) to a time varying GARCH (tvGARCH) $(1, 1)$ model, by allowing the parameters of a stationary GARCH model of Bollerslev (1986) to change slowly with time. They also discussed a two-step local polynomial estimation procedure for the estimation of the parameter functions of the model. The superiority of the tvGARCH (1,1) over several other volatility models has been established for various data sets in RR. In this paper, we focus on the general tvGARCH (p, q) model, of which the tvGARCH $(1, 1)$ and tvARCH models of RR and Dahlhaus and Subba Rao (2006) are special cases.

Let $\{\epsilon_t\}$ be a return process with $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$ and $E(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$, where $\mathcal{F}_{t-1}$ denotes the sigma-field generated by the data up to time $t-1$. The tvGARCH (p, q) model is defined as

$$\epsilon_t = \sigma_t v_t,$$
$$\sigma_t^2 = \alpha_0\,(t) + \sum_{i=1}^{p} \alpha_i\,(t)\,\epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j\,(t)\,\sigma_{t-j}^2,$$

where $\{v_t\}$ is a sequence of real valued i.i.d. random variables and $\alpha_0(\cdot)$, $\alpha_i(\cdot)$ and $\beta_j(\cdot)$, $i = 1, 2 \ldots, p$, $j = 1, 2, \ldots, q$ are certain non-negative deterministic functions.

As in the case of tvGARCH (1,1) model (RR), we rescale the domain of parameter functions of the tvGARCH (p, q) model to facilitate the asymptotics. That is, given the sample of size n, we refer to the following as a tvGARCH (p, q) process.

$$\epsilon_t = \sigma_t v_t,$$
$$\sigma_t^2 = \alpha_0\left(\tfrac{t}{n}\right) + \sum_{i=1}^{p} \alpha_i\left(\tfrac{t}{n}\right)\epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j\left(\tfrac{t}{n}\right)\sigma_{t-j}^2,\ t = 1, 2, \ldots, n. \quad (1)$$

We suggest a two-step local polynomial estimator of the parameter functions of the tvGARCH (p, q) model defined in (1) and investigate its asymptotic distributions. It is found that their asymptotic distribution depends on the parameters of a stationary GARCH process, which is unobservable. This limits the scope of asymptotic results. This stationary GARCH process is such that it locally approximates the tvGARCH process (1) at specific time points. More details can be found in Section 2.1. Fryzlewicz et al. (2008) suggested a residual bootstrap algorithm to tackle such problems in the case of the tvARCH model. However, since the least squares as well as local polynomial estimators of the parameter functions are not guaranteed to be non-negative, this method results in some of the bootstrapped residual squares to be negative. To tackle this problem, we propose a two-step weighted bootstrapped local polynomial estimator for the parameter functions of the tvGARCH (p, q) process. A discussion on the weighted bootstrap and its applications in bootstrapping linear estimators of the parameters of a stationary ARCH model can be found in Chatterjee and Bose (2005) and Bose and Mukherjee (2009). It is worth mentioning here that several bootstrap methods such as the Bayesian bootstrap, deleted d-jackknives, classical paired bootstrap and bootstrap clone are the special cases of the weighted bootstrap, see Praestgaard and Wellner (1993) and Chatterjee and Bose (2005). Here, we prove that the distribution of the proposed bootstrap estimator of parameter functions of the tvGARCH (p, q) model asymptotically coincides with that of the actual local polynomial estimator. The validity of the bootstrapped estimator is also investigated using a simulation study. Simulation results reveal that the bootstrapped estimator provides a better approximation to normality than the actual local polynomial estimator.

Various parametric as well as nonparametric tests have been proposed in the literature for detecting structural breaks in the conditional variance dynamics of asset returns. Often, these tests indicate multiple breaks in the volatility over longer period of time, see for example Chu (1995), Andreou and Ghysels (2002), Amado and Terasvirta (2008) and Chen and Hong (2009) among others. Recently, Chen and Hong (2009) constructed a test for detecting changes in the parameters of GARCH models based on the Quasi maximum likelihood (QML). However, QML has been a topic of criticism among researchers for it tends to be shallow about minimum and hence not very reliable for small sample sizes, see for example Shephard (1996), Bose and Mukherjee (2003) and Fryzlewicz et al. (2008).

We suggest a test statistic for testing the constancy of parameter functions of the tvGARCH (p, q) model. The test is based on the supremum of the normalized deviations of the estimated coefficient functions from the true coefficient functions of the tvGARCH (p, q) model. The limiting distribution of the test statistic is derived. Confidence bands for the parameter functions of the tvGARCH (p, q) model are also developed. The proposed bootstrap method facilitates in the easy computation of confidence bands and test statistic. The method is illustrated with the help of simulated data.

The rest of the paper is organized as follows. In Section 2, we develop a bootstrapped local polynomial estimator of the parameter functions of tvGARCH (p, q) model. Here, we also prove the asymptotic normality of the proposed estimators. Section 3 deals with the construction of confidence bands and tests of hypothesis in tvGARCH model. In Section 4, we report the simulations studies. All the proofs are deferred to the Appendix.

2 Local polynomial estimation and bootstrapping

Consider a tvGARCH (1,1) model introduced by RR,

$$\begin{gathered}\epsilon_t = \sigma_t v_t \\ \sigma_t^2 = \alpha_0\left(\tfrac{t}{n}\right) + \alpha\left(\tfrac{t}{n}\right)\epsilon_{t-1}^2 + \beta\left(\tfrac{t}{n}\right)\sigma_{t-1}^2,\end{gathered} \tag{2}$$

By recursive substitution, (2) may be written as

$$\sigma_t^2 = \alpha_0'(\tfrac{t}{n}) + \sum_{k=1}^{t-1}\alpha_k'(\tfrac{t}{n})\epsilon_{t-k}^2 + \sigma_0^2\prod_{i=1}^{t}\beta\left(\tfrac{t-i+1}{n}\right), \tag{3}$$

where

$$\alpha_0'(\tfrac{t}{n}) = \alpha_0\left(\tfrac{t}{n}\right) + \sum_{k=1}^{t-1}\alpha_0\left(\tfrac{t-k}{n}\right)\prod_{i=1}^{k}\beta\left(\tfrac{t-i+1}{n}\right), \alpha_k'(\tfrac{t}{n}) = \alpha\left(\tfrac{t-k+1}{n}\right)\prod_{i=1}^{k-1}\beta\left(\tfrac{t-i+1}{n}\right),$$
$$k = 1, 2, \ldots t-1.$$

Here we take $\prod_{i=1}^{0}\beta\left(\frac{t-i+1}{n}\right) = 1$. Notice that the functions $\alpha_k'(\cdot)$ are geometrically decaying as $k \to \infty$ under the assumption A_1 (Section 2.1). Also, if σ_0^2 is finite with probability one, then $\sigma_0^2\prod_{i=1}^{t}\beta\left(\frac{t-i+1}{n}\right) \xrightarrow{P} 0$ as $t \to \infty,\ n \to \infty$.

Similarly, by recursive substitution for σ_{t-j}^2 in (1), we can write

$$\sigma_t^2 = \mu_0\left(\frac{t}{n}\right) + \sum_{k=1}^{\infty}\mu_k\left(\frac{t}{n}\right)\epsilon_{t-k}^2 \tag{4}$$

where $\mu_k(\cdot), k = 0, 1, \ldots, \infty$ are certain functions of $\alpha_0(\cdot),\ \alpha_i(\cdot)$ and $\beta_j(\cdot)$, $i = 1, 2 \ldots, p$, $j = 1, 2, \ldots, q$. Under the assumption A_1, these functions are non-negative and geometrically decaying. We carry out the estimation of the parameter functions of (1) in two steps. First, we estimate the functions $\mu_k(\cdot), k = 0, 1, \ldots, P$ for a large P and obtain a preliminary estimate of σ_t^2 with the help of the following tvARCH (P) model

$$\epsilon_t^2 = \mu_0(\tfrac{t}{n}) + \mu_1(\tfrac{t}{n})\epsilon_{t-1}^2 \ldots + \mu_P(\tfrac{t}{n})\epsilon_{t-P}^2 + \sigma_t^2(v_t^2 - 1). \tag{5}$$

Here P is such that $P = P_n \to \infty$ as $n \to \infty$. For the derivation of asymptotic properties of the estimators of tvGARCH parameter functions, we require $P_n \to \infty$. However, the suffix n is dropped for notational simplicity. We assume that the parameter functions of (1) possess bounded continuous derivatives upto order $(d+1)$. Given a kernel function $K(\cdot)$, a local polynomial estimate of $\mu_k(u_0),\ u_0 \in (0,1]$, treating $\sigma_t^2(v_t^2-1)$ as error in (5), is defined as

$$\hat{\mu}_k(u_0) = \boldsymbol{e}^\top_{k(d+1)+1,(P+1)(d+1)}(\boldsymbol{X}_1^\top \boldsymbol{W}_1 \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_1 \boldsymbol{Y}_1, \quad k = 0,1,\ldots,P$$

where, $\boldsymbol{X}_1 = [\boldsymbol{Z}_{P+1},\ldots,\boldsymbol{Z}_n]^\top,\ \boldsymbol{Z}_t = [\boldsymbol{U}_t, \epsilon_{t-1}^2\boldsymbol{U}_t,\ldots,\epsilon_{t-P}^2\boldsymbol{U}_t],\ t = 1,2,\ldots,n,$

$$\boldsymbol{U}_t = [1,(u_t-u_0),\ldots,(u_t-u_0)^d]_{1\times(d+1)}, \quad \boldsymbol{Y}_1 = [\epsilon_{P+1}^2,\ldots\epsilon_n^2]^\top\ ,$$

$$\boldsymbol{W}_1 = \ \operatorname{diag}(K_{h_1}(u_{P+1}-u_0),\ldots,K_{h_1}(u_n-u_0)),\ \text{and}\ K_{h_1}(\cdot) = (1/h_1)K(\cdot/h_1).$$

Here and throughout the paper, $u_t = t/n,\ h_1$ denotes the bandwidth of the initial step estimator and $\boldsymbol{e}_{k,m}$ is a column vector of length m with 1 at the k^{th} position and 0 elsewhere. An initial estimate of σ_t^2 is obtained by,

$$\hat{\sigma}_t^2 = \hat{\mu}_0(u_t) + \sum_{k=1}^{P}\hat{\mu}_k(u_t)\epsilon_{t-k}^2. \tag{6}$$

For the practical implementation, set $\epsilon_t^2 = 0,\ \forall\ t \le 0$. Using this initial estimate of $\hat{\sigma}_t^2$, we obtain the estimates of the parameter functions of (1), which can also be written as

$$\epsilon_t^2 = \alpha_0(\tfrac{t}{n}) + \sum_{i=1}^{p}\alpha_i(\tfrac{t}{n})\epsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j(\tfrac{t}{n})\hat{\sigma}_{t-j}^2 - \sum_{j=1}^{q}\beta_j(\tfrac{t}{n})(\hat{\sigma}_{t-j}^2 - \sigma_{t-j}^2) + \sigma_t^2(v_t^2-1). \tag{7}$$

It is shown in Lemmas 1, 2 and 3 (Section 2.1) that for a particular choice of initial step bandwidth $h_1 = o(h_2)$, $E(\hat{\sigma}_{t-j}^2 - \sigma_{t-j}^2)$ is asymptotically negligible, where h_2 denotes the bandwidth in the estimation of the parameter functions of (7). Now, under the $(d+1)^{th}$ order continuous differentiability assumption of the parameter functions, the estimates can be obtained as

$$\begin{aligned}
\hat{\alpha}_0(u_0) &= \boldsymbol{e}^\top_{1,(p+q+1)(d+1)}(\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{Y}_2,\\
\hat{\alpha}_i(u_0) &= \boldsymbol{e}^\top_{i(d+1)+1,(p+q+1)(d+1)}(\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{Y}_2, \quad i = 1,2,\ldots,p\\
\hat{\beta}_j(u_0) &= \boldsymbol{e}^\top_{(j+p)(d+1)+1,(p+q+1)(d+1)}(\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{Y}_2, \quad j = 1,2,\ldots,q.
\end{aligned}$$

where, $\boldsymbol{X}_2 = [\boldsymbol{Z}_{2,r+1},\ldots,\boldsymbol{Z}_{2,n}]^\top,\ r = \max(p,q),$

$$\boldsymbol{Z}_{2,t} = [\boldsymbol{U}_t, \epsilon_{t-1}^2\boldsymbol{U}_t,\ldots,\epsilon_{t-p}^2\boldsymbol{U}_t, \hat{\sigma}_{t-1}^2\boldsymbol{U}_t,\ldots,\hat{\sigma}_{t-q}^2\boldsymbol{U}_t],$$

$$\boldsymbol{W}_2 = \quad \text{diag}(K_{h_2}(u_{r+1}-u_0),\ldots,K_{h_2}(u_n-u_0)), \quad \text{and} \quad \boldsymbol{Y}_2 = [\epsilon^2_{r+1},\ldots,\epsilon^2_n]^\top.$$

In order to construct the bootstrapped local polynomial estimators for the parameter functions of (1), first consider a sequence of exchangeable random variables $\{w_i\}_{i=1}^n$, independent of $\{\epsilon_t\}_{t=1}^n$. Define $W_{B1} = \text{diag}(w_{P+1},\ldots,w_n)$. Then a preliminary bootstrapped estimator of σ_t^2 is given by,

$$\hat{\sigma}^2_{Bt} = \hat{\mu}_{B0}(u_t) + \sum_{k=1}^{P} \hat{\mu}_{Bk}(u_t)\epsilon^2_{t-k}.$$

where $\hat{\mu}_{Bk}(u_0) = \boldsymbol{e}^\top_{k(d+1)+1,(P+1)(d+1)}(\boldsymbol{X}_1^\top \boldsymbol{W}_{B1}\boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_{B1}\boldsymbol{W}_1\boldsymbol{Y}_1, \;\; k=0,1,\ldots,P$ is a bootstrapped local polynomial estimator of the tvARCH (P) model. Hence, the bootstrapped estimators of the parameter functions of tvGARCH (p,q) model can be written as

$$\begin{aligned}
\hat{\alpha}_{B0}(u_0) &= \boldsymbol{e}^\top_{1,(p+q+1)(d+1)}(\boldsymbol{X}^\top_{B2}\boldsymbol{W}_{B2}\boldsymbol{W}_2\boldsymbol{X}_{B2})^{-1}\boldsymbol{X}^\top_{B2}\boldsymbol{W}_{B2}\boldsymbol{W}_2\boldsymbol{Y}_2,\\
\hat{\alpha}_{Bi}(u_0) &= \boldsymbol{e}^\top_{i(d+1)+1,(p+q+1)(d+1)}(\boldsymbol{X}^\top_{B2}\boldsymbol{W}_{B2}\boldsymbol{W}_2\boldsymbol{X}_{B2})^{-1}\boldsymbol{X}^\top_{B2}\boldsymbol{W}_{B2}\boldsymbol{W}_2\boldsymbol{Y}_2 \;\; \text{and}\\
\hat{\beta}_{Bj}(u_0) &= \boldsymbol{e}^\top_{(j+p)(d+1)+1,(p+q+1)(d+1)}(\boldsymbol{X}^\top_{B2}\boldsymbol{W}_{B2}\boldsymbol{W}_2\boldsymbol{X}_{B2})^{-1}\boldsymbol{X}^\top_{B2}\boldsymbol{W}_{B2}\boldsymbol{W}_2\boldsymbol{Y}_2.
\end{aligned}$$

where $\boldsymbol{W}_{B2} = \text{diag}(w_{r+1},\ldots w_n)$ and $\boldsymbol{X}_{B2}$ is same as $\boldsymbol{X}_2$ with $\{\hat{\sigma}_t,\; t=(r-q+1),\ldots,n\}$ replaced by $\hat{\sigma}^2_{Bt}$. In the following section, we show that the bootstrapped estimator has the same asymptotic distribution as that of the actual estimator.

Remark 1. The bandwidth selection for the estimation of tvGARCH (p,q) model can be performed using the the cross validation method of Hart (1994). The detailed procedure is described in RR for the tvGARCH $(1,1)$ model and can be easily extended to the tvGARCH (p,q).

2.1 Asymptotics

We denote the convergence in probability to zero and boundedness in probability by o_P and O_P respectively. Let $P_B,\; E_{P_B},\; V_{P_B},\; o_{P_B}$ and O_{P_B} represent the probability, expectation, variance, convergence in probability to zero and boundedness in probability with respect to the bootstrap distribution conditional on data. Towards deriving the asymptotic distributions of the bootstrapped estimators, we first state the following technical assumptions $(A_1 - A_6)$

$\mathbf{A_1}$. There exists a $\delta > 0$ such that $0 < \sum_{i=1}^{p}\alpha_i(u) + \sum_{j=1}^{q}\beta_j(u) \leq 1-\delta,\; \forall\; u \in (0,1]$ and $\sup_u \alpha_0(u) < \infty$.

$\mathbf{A_2}$. There exist finite constants M_1, M_2 and M_3 such that $\forall\ u_1, u_2 \in (0,1]$,

$$\begin{aligned}|\alpha_0(u_1) - \alpha_0(u_2)| &\le M_1|u_1 - u_2|\\ |\alpha_i(u_1) - \alpha_i(u_2)| &\le M_2|u_1 - u_2|,\ i = 1,2,\ldots,p\\ |\beta_j(u_1) - \beta_j(u_2)| &\le M_3|u_1 - u_2|,\ j = 1,2,\ldots,q.\end{aligned}$$

$\mathbf{A_3}$. The functions $\alpha_0(\cdot), \alpha_i(\cdot)$ and $\beta_j(\cdot)$ (and hence $\mu_k(\cdot)$) have bounded and continuous derivatives up to order $d+1$, in a neighborhood of u_0, $u_0 \in (0,1]$.

$\mathbf{A_4}$. $K(u)$ is a symmetric density function of bounded variation with a compact support.

$\mathbf{A_5}$. The bandwidths h_1 and h_2 are such that $h_1 \to 0, h_2 \to 0$ and $nh_1 \to \infty, nh_2 \to \infty$ as $n \to \infty$.

$\mathbf{A_6}$. The bootstrap weights $\{w_i\}$ are such that $E_{P_B}(w_i) = 1,\ \sigma^2_{wn} = V_{P_B}(w_i) = o(n)$ and $Corr_{P_B}(w_i, w_j) = O(1/n)\ \forall\ i \neq j$.

It can be proved using similar techniques as in RR and Davis and Mikosch (2009) that assumption A_1 is sufficient for the existence of a well defined unique solution to the variance process in (1). Also, it ensures the tvGARCH (p,q) to be a short memory process. The Lipschitz continuity condition A_2 on the parameter functions makes the tvGARCH (p,q) process locally stationary in the sense that it can be approximated by a stationary GARCH process in the neighborhood of a fixed point. Let $\{\tilde{\epsilon}_t(u_0)\},\ u_0 \in (0,1]$ be a process with $E(\tilde{\epsilon}_t(u_0)|\tilde{\mathcal{F}}_{t-1}) = 0$ and $E(\tilde{\epsilon}^2_t(u_0)|\tilde{\mathcal{F}}_{t-1}) = \tilde{\sigma}^2_t(u_0)$ where $\tilde{\mathcal{F}}_{t-1} = \sigma(\tilde{\epsilon}_{t-1}, \tilde{\epsilon}_{t-2}, \ldots)$. Then $\{\tilde{\epsilon}_t(u_0)\}$ is said to follow a stationary GARCH process associated with (1) at time point u_0 if it satisfies,

$$\begin{aligned}\tilde{\epsilon}_t(u_0) &= \tilde{\sigma}_t(u_0)v_t,\\ \tilde{\sigma}^2_t(u_0) = \alpha_0(u_0) + \sum_{i=1}^{p} \alpha_i(u_0)\tilde{\epsilon}^2_{t-i}(u_0) &+ \sum_{j=1}^{q} \beta_j(u_0)\tilde{\sigma}^2_{t-j}(u_0).\end{aligned} \tag{8}$$

It can be shown that tvGARCH (p,q) process can be locally approximated by (8). The result is stated in the Proposition 1. Assumptions A_3 to A_5 are standard assumptions for deriving the asymptotic distributions of local polynomial estimators and are also assumed by RR. The assumption A_6 on the bootstrap weights are the basic conditions assumed by Chatterjee and Bose (2005, Conditions BW). These conditions are required to establish the asymptotic distribution of the bootstrapped estimator and are also assumed by Bose and Mukherjee (2009) for bootstrapping the estimators of stationary ARCH parameters. An example of weights satisfying these conditions is weights following a Multinomial $(n; 1/n, \ldots, 1/n)$ distribution.

Proposition 1. *Let the assumptions A_1 and A_2 be satisfied. Then the process $\{\epsilon_t^2\}$ can be approximated locally by a stationary ergodic process $\{\widetilde{\epsilon}_t^2(u_0)\}$. That is, there exists a well defined stationary ergodic process V_t independent of u_0 and a constant $Q < \infty$ such that*

$$|\epsilon_t^2 - \widetilde{\epsilon}_t^2(u_0)| \leq Q\left(\left|\tfrac{t}{n} - u_0\right| + \tfrac{1}{n}\right) V_t \quad a.s. \tag{9}$$

or equivalently

$$\epsilon_t^2 = \widetilde{\epsilon}_t^2 + O_P\left(\left|\tfrac{t}{n} - u_0\right| + \tfrac{1}{n}\right).$$

In the following lemmas, we state the asymptotic distributions of the local polynomial estimators of the tvARCH (P) and tvGARCH (p, q) processes discussed in the beginning of Section 2. Before going to the main results, we first introduce some notations.

Notations.

$$\begin{aligned}
\tau_k &= \textstyle\int u^k K(u)du, \quad \nu_k = \int u^k K^2(u)du, \\
C_j &= C_j(u_0) = E(\widetilde{\epsilon}_t^2(u_0)\ \widetilde{\epsilon}_{t-j}^2(u_0)), \quad w_j = E(\widetilde{\epsilon}_t^j(u_0)), \\
\boldsymbol{S} &= \boldsymbol{S}(u_0) = E\left([1, \widetilde{\epsilon}_{t-1}^2(u_0), \ldots, \widetilde{\epsilon}_{t-P}^2(u_0)]^\top [1, \widetilde{\epsilon}_{t-1}^2(u_0), \ldots, \widetilde{\epsilon}_{t-P}^2(u_0)]\right), \\
\boldsymbol{\Omega} &= \boldsymbol{\Omega}(u_0) = E\left(\widetilde{\sigma}_t^4(u_0)[1, \widetilde{\epsilon}_{t-1}^2(u_0), \ldots, \widetilde{\epsilon}_{t-P}^2(u_0)]^\top [1, \widetilde{\epsilon}_{t-1}^2(u_0), \ldots, \widetilde{\epsilon}_{t-P}^2(u_0)]\right), \\
\boldsymbol{D}_i &= [\tau_{d+1}, h_i\tau_{d+2}, \ldots, h_i^d\tau_{2d+1}]^\top, \ i = 1, 2, \\
\boldsymbol{e}_m &= \text{a column vector of length } m \text{ with 1 everywhere,}
\end{aligned}$$

$$\boldsymbol{A}_i = \begin{bmatrix} 1 & h_i\tau_1 & \ldots & h_i^d\tau_d \\ h_i\tau_1 & h_i^2\tau_2 & \ldots & h_i^{d+1}\tau_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_i^d\tau_d & h_i^{d+1}\tau_{d+1} & \ldots & h_i^{2d}\tau_{2d} \end{bmatrix}, \quad \boldsymbol{B}_i = \begin{bmatrix} \nu_0 & h_i\nu_1 & \ldots & h_i^d\nu_d \\ h_i\nu_1 & h_i^2\nu_2 & \ldots & h_i^{d+1}\nu_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_i^d\nu_d & h_i^{d+1}\nu_{d+1} & \ldots & h_i^{2d}\nu_{2d} \end{bmatrix}, \quad i = 1, 2.$$

$\boldsymbol{S}_2 = E\left(\boldsymbol{x}_t^\top \boldsymbol{x}_t\right), \boldsymbol{\Omega}_2 = E\left(\widetilde{\sigma}_t^4(u_0)\boldsymbol{x}_t^\top \boldsymbol{x}_t\right)$, where
$\boldsymbol{x}_t = [1, \widetilde{\epsilon}_{t-1}^2(u_0), \ldots \widetilde{\epsilon}_{t-p}^2(u_0), \widetilde{\sigma}_{t-1}^2(u_0), \ldots, \widetilde{\sigma}_{t-q}^2(u_0)]$.

Lemma 1. *Suppose that the assumptions A_1 to A_5 hold and $E|v_t|^8 < \infty$. Then*

$$\begin{aligned}
\sqrt{nh_1}&\left(\hat{\boldsymbol{\mu}}_{tvARCH}(u_0) - \boldsymbol{\mu}_{tvARCH}(u_0) - \tfrac{h_1^{d+1}}{(d+1)!}\boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_1^{-1}\boldsymbol{D}_1\boldsymbol{\mu}_{tvARCH}^{(d+1)}(u_0)\right) \\
&\xrightarrow{D} \boldsymbol{N}_{P+1}\left(0, \boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_1^{-1}\boldsymbol{B}_1\boldsymbol{A}_1^{-1}\boldsymbol{e}_{1,d+1} Var(v_t^2)\boldsymbol{S}^{-1}\boldsymbol{\Omega}\boldsymbol{S}^{-1}\right)
\end{aligned}$$

where $\hat{\boldsymbol{\mu}}_{tvARCH}(u_0)$ and $\boldsymbol{\mu}_{tvARCH}^{(d+1)}(u_0)$ denote the local polynomial estimator and derivative of order $(d+1)$ of $\boldsymbol{\mu}_{tvARCH}(u_0) = [\mu_0(u_0), \mu_1(u_0), \ldots, \mu_P(u_0)]^\top$.

Remark 2. The moment assumption in the Lemma 1, $E|v_t|^8 < \infty$ is slightly strong.

Since, we are dealing with the squared process ϵ_t^2, we are forced to assume $E|v_t|^8 < \infty$ to ensure the existence of asymptotic variances. However, higher moments are not assumed for the return process ϵ_t, but for ϵ_t/σ_t, which may be justifiable as both ϵ_t and σ_t are of the same order in ϵ_t.

Lemma 2. *Let $\hat{\sigma}_t^2$ be as defined in (6). Then under the assumptions of Lemma 1,*

$$bias(\hat{\sigma}_t^2) = E(\hat{\sigma}_t^2 - \sigma_t^2) = O_P(h_1^{d+1}) + O(\rho^{P_n})$$

where $0 < \rho < 1$ and $P_n \to \infty$ as $n \to \infty$.

Lemma 2 shows that the choice of P_n will contribute towards the bias of the conditional variance in the initial step by a term which decays geometrically. Therefore, this term will have negligible effect on final estimators as $P_n \to \infty$. Also, if $h_1 = o(h_2)$, then the term $O_P(h_1^{d+1}) = o_P(h_2^{d+1}) \to 0$ as $n \to \infty$ under assumption A_5. Thus, the bias in the first step of estimation of σ_t^2 is negligible asymptotically.

Now in the following lemma, we show that for a particular choice of the initial step bandwidth, $h_1 = o(h_2)$, the effect of the generated regressors in step 1 vanishes. That is, the local polynomial estimators of the parameter functions of tvGARCH (p, q) in step 2 behave in such a way as if $\sigma_{t-j}^2,\ j = 1, 2, \ldots, q$ were known.

Lemma 3. *Suppose that the assumptions A_1 to A_5 hold and $E|v_t|^8 < \infty$. Further, let the the bandwidth h_2 in the second step of the local polynomial estimation procedure be such that $h_1 = o(h_2)$. Then,*

$$\sqrt{nh_2}\left(\hat{\boldsymbol{\beta}}_{tvGARCH}(u_0) - \boldsymbol{\beta}_{tvGARCH}(u_0) - \frac{h_2^{d+1}}{(d+1)!}\boldsymbol{e}_{1,d+1}^{\top}\boldsymbol{A}_2^{-1}\boldsymbol{D}_2\boldsymbol{\beta}_{tvGARCH}^{(d+1)}(u_0)\right)$$
$$\xrightarrow{D} \boldsymbol{N}_{p+q+1}\left(0, \boldsymbol{e}_{1,d+1}^{\top}\boldsymbol{A}_2^{-1}\boldsymbol{B}_2\boldsymbol{A}_2^{-1}\boldsymbol{e}_{1,d+1}Var(v_t^2)\boldsymbol{S}_2^{-1}\boldsymbol{\Omega}_2\boldsymbol{S}_2^{-1}\right)$$

where $\boldsymbol{\beta}_{tvGARCH}(u_0) = [\alpha_0(u_0), \alpha_1(u_0), \ldots, \alpha_p(u_0), \beta_1(u_0), \ldots, \beta_q(u_0)]^{\top}$.

Thus, if we choose the initial step bandwidth in such a way that $h_1 = o(h_2)$, then the bias and variance expressions for the local polynomial estimators are free from the first step bandwidth. Also, the asymptotic distribution of the parameter functions in step 2 is same as they would have been if $\sigma_{t-j}^2,\ j = 1, 2, \ldots, q$ were known in (1). This means that when the optimal bandwidth is used, then the estimation remains unaffected for a large choice of initial step bandwidth. This makes the estimation procedure relatively easy to implement. The MSE of the final estimator is $O_P(h_2^{2d+2} + (nh_2)^{-1})$, which is independent of the initial step bandwidth. Notice that this MSE achieves the optimal rate of convergence at an order of $O_P(n^{-(2d+2)/(2d+3)})$ for an optimal bandwidth h_2 of order $O(n^{-1/(2d+3)})$ and $h_1 = o(h_2)$.

Evidently, the bias and variance expressions of $\hat{\boldsymbol{\mu}}_{tvARCH}(u_0)$ and $\hat{\boldsymbol{\beta}}_{tvGARCH}(u_0)$ in Lemmas 1 and 3 depend on the parameters of the unobservable stationary GARCH process defined in (8). Therefore, these estimators cannot be directly used for the construction of confidence intervals and testing of hypothesis. To tackle this, we establish the asymptotic distributions of the bootstrapped estimators defined in Section 2. Let $\hat{\boldsymbol{\mu}}_B(u_0) = [\hat{\mu}_{B0}(u_0), \ldots, \hat{\mu}_{BP}(u_0)]^\top$ and $\hat{\boldsymbol{\beta}}_B = [\hat{\alpha}_{B0}(u_0), \hat{\alpha}_{B1}(u_0), \ldots, \hat{\alpha}_{Bp}(u_0), \hat{\beta}_{B1}(u_0), \ldots, \hat{\beta}_{Bq}(u_0)]^\top$ denote the bootstrapped estimators of the parameter functions of the tvARCH (P) and tvGARCH (p, q) respectively. In the following theorems, we establish that the asymptotic distributions of the actual and bootstrapped estimators coincide. This implies that the properties of the actual local polynomial estimator discussed above are also true for the bootstrapped estimator.

Theorem 1. *Let the assumptions A_1 to A_6 hold and $E|v_t|^8 < \infty$. Then,*

$$\sigma_{wn}^{-1}\sqrt{nh_1}\left(\hat{\boldsymbol{\mu}}_B - \hat{\boldsymbol{\mu}}_{tvARCH}(u_0)\right) \xrightarrow{D} \boldsymbol{N}_{P+1}\left(0, \boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_1^{-1}\boldsymbol{B}_1\boldsymbol{A}_1^{-1}\boldsymbol{e}_{1,d+1} Var(v_t^2)\boldsymbol{S}^{-1}\boldsymbol{\Omega}\boldsymbol{S}^{-1}\right).$$

Theorem 2. *Let the assumptions of Lemma 3 hold along with A_6. Then,*

$$\sigma_{wn}^{-1}\sqrt{nh_2}\left(\hat{\boldsymbol{\beta}}_B - \hat{\boldsymbol{\beta}}_{tvGARCH}(u_0)\right) \xrightarrow{D} \boldsymbol{N}_{p+q+1}\left(0, \boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_2^{-1}\boldsymbol{B}_2\boldsymbol{A}_2^{-1}\boldsymbol{e}_{1,d+1} Var(v_t^2)\boldsymbol{S}_2^{-1}\boldsymbol{\Omega}_2\boldsymbol{S}_2^{-1}\right).$$

Comparing the results in Lemma 1 and Theorem 1 and Lemma 3 and Theorem 2, we can see that the asymptotic distributions of $\sigma_{wn}^{-1}\sqrt{nh_1}\left(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu}_{tvARCH}(u_0)\right)$ and $\sqrt{nh_1}\left(\hat{\boldsymbol{\mu}}_{tvARCH} - \boldsymbol{\mu}_{tvARCH}(u_0)\right)$ as well as those of $\sigma_{wn}^{-1}\sqrt{nh_2}\left(\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_{tvGARCH}(u_0)\right)$ and $\sqrt{nh_2}\left(\hat{\boldsymbol{\beta}}_{tvGARCH}(u_0) - \boldsymbol{\beta}_{tvGARCH}(u_0)\right)$ are the same. This implies that the bootstrapped estimator would provide a good approximation to the distribution of the actual tvGARCH estimators. Thus, with the help of repeated bootstrapped iterations, we can obtain the approximate empirical biases and variances of the actual estimators. We use these to construct the confidence bands and the test statistics for testing the constancy of the parameter functions.

3 Confidence bands and testing of hypothesis

It is of interest to test whether all or some of the coefficients of the tvGARCH (p, q) model are constant (possibly equal to zero) or whether they are really time varying. For instance, if $H_0 : \ \beta_j(u) \equiv \ 0 \ \forall \ j$ is not rejected, a tvARCH model is more appropriate. For this purpose, we establish the asymptotic distribution of the supremum of the estimates of parameter functions of the tvARCH and tvGARCH models in Theorems 3 and 4 respectively. These results are also helpful in constructing the confidence bands for the parameter functions of the tvGARCH model.

Theorem 3. *Suppose that the conditions of Lemma 1 are satisfied and* $h_1 = n^{-b_1}$ *for some* $0 < b_1 < 1/2$. *Then*

$$P\left[(-2\log h_1)^{1/2}\left\{\sup_u \left|(Var(\hat{\mu}_k(u)))^{-1/2}\left(\hat{\mu}_k(u) - \mu_k(u) - bias(\hat{\mu}_k(u))\right)\right| - d_n\right\} < z\right]$$

$$\to \exp\{-2\exp(-z)\}$$

where

$$Var(\hat{\mu}_k(u)) = \frac{1}{nh_1}\boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_1^{-1}\boldsymbol{B}_1\boldsymbol{A}_1^{-1}\boldsymbol{e}_{1,d+1}Var(v_t^2)\boldsymbol{e}_{k,P+1}^\top \boldsymbol{S}^{-1}\boldsymbol{\Omega}\boldsymbol{S}^{-1}e_{k,P+1},$$

$$bias(\hat{\mu}_k(u)) = \frac{h_1^{d+1}}{(d+1)!}\boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_1^{-1}\boldsymbol{D}_1\mu_k^{(d+1)}(u_0)$$

and

$$d_n = (-2\log h_1)^{1/2} + \frac{1}{(-2\log h_1)^{1/2}}\left(\frac{1}{4\nu_0\pi}\int (K'(t))^2 dt\right).$$

Let $\beta_{2i}(u)$ denote the i^{th} element of $\boldsymbol{\beta}_{tvGARCH}(u) = [\alpha_0(u), \alpha_1(u), \ldots, \alpha_p(u), \beta_1(u), \ldots, \beta_q(u)]^\top$.

Theorem 4. *Suppose that the conditions of Lemma 3 are satisfied and* $h_2 = n^{-b}$ *for some* $0 < b < 1/2$. *Then*

$$P\left[(-2\log h_2)^{-1/2}\left\{\sup_u \left|\left(Var(\hat{\beta}_{2i}(u))\right)^{-1/2}\left(\hat{\beta}_{2i}(u) - \beta_{2i}(u) - bias(\hat{\beta}_{2i}(u))\right)\right| - d_n\right\} < z\right]$$

$$\to \exp\{-2\exp(-z)\}$$

where

$$Var(\hat{\beta}_{2i}(u)) = (1/nh_2)\boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_2^{-1}\boldsymbol{B}_2\boldsymbol{A}_2^{-1}\boldsymbol{e}_{1,d+1}Var(v_t^2)\boldsymbol{e}_{i,p+q+1}^\top \boldsymbol{S}_2^{-1}\boldsymbol{\Omega}_2\boldsymbol{S}_2^{-1}\boldsymbol{e}_{i,p+q+1},$$

$$bias(\hat{\beta}_{2i}(u)) = \frac{h_2^{d+1}}{(d+1)!}\boldsymbol{e}_{1,d+1}^{\top}\boldsymbol{A}_2^{-1}\boldsymbol{D}_2\beta_{2i}^{(d+1)}(u_0)$$

and d_n is same as that in Theorem 3 with h_1 replaced by h_2.

Notice that the asymptotic distributions of the supremums in Theorems 3 and 4 are similar in spirit to that in the varying coefficient models of Fan and Zhang (2000). However, the tvGARCH model (1) is different from the usual varying coefficient models due to its heteroscedasticity and non-stationary behavior. Also, we adopt a two-step estimation procedure here.

From Theorem 4, we can get $(1-\alpha)$ confidence band for each parameter function of the tvGARCH model (1) as

$$\hat{\beta}_{2i}(u) - bias(\hat{\beta}_{2i}(u)) - z_\alpha,\ \hat{\beta}_{2i}(u) - bias(\hat{\beta}_{2i}(u)) + z_\alpha, \tag{10}$$

where

$$z_\alpha = [d_n + \{\log 2 - \log(-\log(1-\alpha)\}(-2\log h)^{1/2}]\{Var(\hat{\beta}_{2i}(u))\}^{1/2}.$$

Here, $bias(\hat{\beta}_{2i}(u))$ and $Var(\hat{\beta}_{2i}(u))$ are not known in practice. However, we can replace them with the bias and variance of the bootstrapped estimator using Theorem 2.

In order to test the hypothesis that a particular coefficient function is equal to a given constant c, that is, $H_0 : \beta_{2i}(u) \equiv c\ \forall\ u \in (0,1]$, a natural test procedure would be to check whether the $\beta_{2i}(\cdot)$ falls in the confidence band (10) or not. The test statistics that can be used for such a testing problem is

$$(2\log h)^{1/2}\left[\sup_u\left\{\left|\left(Var(\hat{\beta}_{2i}(u))\right)^{-1/2}\left(\hat{\beta}_{2i}(u) - c - bias(\hat{\beta}_{2i}(u))\right)\right|\right\} - d_n\right]. \tag{11}$$

We reject the null hypothesis when the test statistics exceeds the asymptotic critical value $-\log(-0.5\log(1-\alpha))$. If c is unknown and the interest is to test the constancy of the parameter function, then it can be estimated under the null hypothesis. First exploit the fact that $\beta_{2i}(u)$ is a constant and estimate the same using two-step local polynomial procedure with $d = 0$. Then, obtain an estimate of c by averaging over all t,

$$\hat{c} = \frac{1}{n}\sum_{t=1}^{n}\hat{\beta}_{2i}\left(\frac{t}{n}\right).$$

Substitute this estimator of c in (11) and reject the null hypothesis for large values of the test statistics.

Remark 3. Under the conditions of Lemma 3, the bias of the estimator $\hat{c}$ is $O_P(h_2^{d+1})$ and

$$Var(\hat{c}) = \frac{1}{nh_2} Var(v_t^2) \frac{1}{n^2} \sum_{t=1}^{n} \boldsymbol{e}_{i,p+q+1}^\top \boldsymbol{S}_2^{-1}(u_t) \boldsymbol{\Omega}_2(u_t) \boldsymbol{S}_2(u_t)^{-1} \boldsymbol{e}_{i,p+q+1} + o(h_2^2).$$

Notice that the covariance term is of order h_2^{2d+2}.

It may be noted that similar confidence bands and test statistic can be obtained for the parameter functions in a tvARCH model using Theorem 3.

4 Simulation study

We carried out a simulation study to judge the empirical performance of the bootstrapped estimator of the parameter functions of tvGARCH model suggested in Section 2. For computational simplicity, we used the tvGARCH (1,1) model in our simulations. A sample of size $n = 500$ was generated from the following model

$$\begin{aligned} \epsilon_t &= \sigma_t v_t, \\ \sigma_t^2 &= \alpha_0\left(\tfrac{t}{n}\right) + \alpha_1\left(\tfrac{t}{n}\right)\epsilon_{t-1}^2 + \beta\left(\tfrac{t}{n}\right)\sigma_{t-1}^2, \quad t = 1, 2, \ldots, 500, \end{aligned} \tag{12}$$

where $\alpha_0(u) = 2u(1-u^2) + 0.1,\ \alpha_1(u) = 0.2\cos(2\pi u) + 0.25$ and $\beta(u) = 2u(1-u) + 0.2u^3,\ 0 < u \leq 1$. The parameter functions are chosen in such a way that they satisfy the assumptions A_1 to A_3. The local polynomial estimation of the parameter functions is carried out using the Epanechnikov kernel. The value of P in the first step of the estimation is taken as $\log n$. Bandwidth is selected using the cross validation method (as described in RR).

We consider the bootstrap weights to have a multinomial $(n, 1/n, \ldots, 1/n)$ distribution with $n = 500$. The bootstrapped estimator of the parameter functions was obtained based on 1000 bootstrap samples. Figure 1 compares the bootstrapped estimator (blue curve) with the actual estimator (black curve). The red curve in the plot represents the actual function. Notice that both the estimators of $\alpha_0(\cdot)$ are almost similar. However, the performance of the actual estimator is not as good as the bootstrapped one for the functions $\alpha_1(\cdot)$ and $\beta(\cdot)$. Especially at the boundaries, the actual estimator is overestimating while the bootstrapped estimator performs well. Notice that the performance of the estimator at the boundaries of the data is important from the forecasting point of view.

To compare the asymptotic properties of the bootstrapped estimator with those of the actual estimator, 1000 samples were generated, each of size 500 from (12). The local polynomial estimators of all the parameter functions are computed for each of these 1000 datasets and point wise means and variances of the estimators are obtained empirically from these samples of size 1000 each. We present the density of the standardized estimator (blue curve) at three different points ($t = 1, 250, 500$) in Figure 2. The red curve in the figure represents the standard normal density. Similarly, the density of the standardized bootstrapped estimators are plotted along with standard normal density in Figure 3 at the three points. It is clear from the plots that the distribution of the bootstrapped estimator is closer to normal than that of the actual estimator. Specially, at the boundaries ($t = 1, 500$), the performance of the actual estimator is not good and it seems to get overestimated for some samples. Notice that a similar phenomenon is also seen in Figure 1. We also compared the densities of the two estimators at several other randomly selected time points and the second order performance of the estimators was found to be more or less similar to that at point $t = 250$ in Figures 2 and 3.

The 95% confidence bands of the parameter functions based on the methodology described in Section 3 are depicted in Figure 4. We computed the bias and variance of the estimator based on 1000 bootstrap samples.

To examine the empirical performance of the test suggested in Section 3, we generated samples of sizes 500 and 300 from the model (12) with constant parameter functions $\alpha_0(u) = 0.5,\ \alpha_1(u) = 0.2$ and $\beta(u) = 0.3\ \forall\, u$. The test for the constancy of the parameters based on the bootstrapped estimator and test statistics (11) as described in Section 3 is carried out. The empirical probabilities of false rejection based on 1000 samples are given in Table 1 for different nominal levels. Notice that the empirical probabilities are quite close to the nominal levels even for a moderate sample size of 300. However, performance of the test is better in the case of $n = 500$ than in $n = 300$. This is quite natural as our results hold asymptotically.

Acknowledgements

The author would like to thank Dr. T. V. Ramanathan for some helpful discussions.

Appendix: Outline of proofs

Proof of Proposition 1. The tvGARCH (p, q) model can also be written as a tvARCH (∞) as represented in (4). Under assumption A_2, the parameter functions of the tv-GARCH (p, q) model and hence $\mu_k(\cdot),\ k = 1, 2, \ldots$ which are simple functions of $\alpha_0(\cdot),\ \alpha_i(\cdot)$, $i = 1, 2, \ldots, p$ and $\beta_j(\cdot), j = 1, 2, \ldots, q$ are Lipschitz continuous. Therefore, the proposition follows in the same way as Theorem 1 of Dahlhaus and Rao (2006) who prove the local stationarity of a tvARCH (∞) process.

To prove Lemma 1, we first state the following auxiliary lemmas which are proved in RR (Lemmas A.2 and A.4). These lemmas are stated for a general bandwidth h so that they can be used for both h_1 and h_2.

Lemma A.1. *Let the assumptions A_1 to A_5 be satisfied. Then*

$$
\begin{aligned}
(i)\ & \sum_{k=P+1}^{n} \tfrac{1}{nh}(u_k - u_0)^i K\left(\tfrac{u_k-u_0}{h}\right) \epsilon_{k-j_1}^{2l}\epsilon_{k-j_2}^{2m} \xrightarrow{P} h^i \tau_i E(\tilde{\epsilon}_{k-j_1}^{2l}(u_0)\tilde{\epsilon}_{k-j_2}^{2m}(u_0)),\\
& \forall\ l, m \in \{0,1,2\} \text{ and } j_1, j_2 \in \{1, 2, \ldots, P\},\ j_1 \neq j_2\\
(ii)\ & \sum_{k=P+1}^{n} \tfrac{1}{nh}(u_k - u_0)^i K^2\left(\tfrac{u_k-u_0}{h}\right) \sigma_k^4 \epsilon_{k-j_1}^{2l}\epsilon_{k-j_2}^{2m}\\
& \qquad \xrightarrow{P} h^i \nu_i E(\tilde{\sigma}_k^4(u_0)\tilde{\epsilon}_{k-j_1}^{2l}(u_0)\tilde{\epsilon}_{k-j_2}^{2m}(u_0)),\\
& \forall\ l, m \in \{0,1\} \text{ and } j_1, j_2 \in \{1, 2, \ldots, P\},
\end{aligned}
$$

where (ii) is true for $l, m > 0$ only if $E|v_t|^8 < \infty$.

Lemma A.2. *Suppose the assumptions A_1 to A_5 are satisfied. In addition assume that $E|v_t|^8 < \infty$. Then*

$$
\begin{aligned}
Var&\left(\sum_{k=P+1}^{n}(u_k - u_0)^i K_h(u_k - u_0)(v_k^2 - 1)\sigma_k^2[1, \epsilon_{k-1}^2, \ldots, \epsilon_{k-P}^2]^\top\right)\\
&= nh^{2i-1}\nu_{2i} Var(v_t^2)\boldsymbol{\Omega}(1 + o_P(1)), \quad i = 1, 2, \ldots, d.
\end{aligned}
$$

Proof of Lemma 1. To prove the lemma, we first obtain expressions for the asymptotic bias and variance of the first step estimator. Let us denote $\boldsymbol{\beta}_1 = [\mu_{00}, \mu_{01}, \ldots, \mu_{0d}, \ldots, \mu_{P0}, \ldots, \mu_{Pd}]^\top$, where $\mu_{ij} = \mu_i^{(j)}(\cdot)/j!$ and $\mu_i^{(j)}(\cdot)$ denotes the j^{th} derivative of $\mu_i(\cdot)$. Using Taylor's series expansion, we can write,

$$
\boldsymbol{Y}_1 = \boldsymbol{X}_1\left[\mu_0(u_0), \mu_0^{(1)}(u_0), \ldots \tfrac{\mu_0^{(d)}(u_0)}{d!}, \mu_1(u_0), \ldots, \mu_P(u_0), \ldots \tfrac{\mu_P^{(d)}(u_0)}{d!}\right]^\top
$$

$$+\frac{1}{(d+1)!}\begin{bmatrix}\mu_0^{(d+1)}(\zeta_{0(P+1)})(u_{P+1}-u_0)^{d+1}\\ \vdots\\ \mu_0^{(d+1)}(\zeta_{0(n)})(u_n-u_0)^{d+1}\end{bmatrix}$$

$$+\frac{1}{(d+1)!}\sum_{j=1}^{P}\begin{bmatrix}\mu_j^{(d+1)}(\zeta_{j(P+1)})(u_{P+1}-u_0)^{d+1}\epsilon_{P+1-j}^2\\ \vdots\\ \mu_j^{(d+1)}(\zeta_{j(n)})(u_n-u_0)^{d+1}\epsilon_{n-j}^2\end{bmatrix}+\boldsymbol{\sigma}^2*(\boldsymbol{v}^2-e_{n-P})$$

where $\boldsymbol{\sigma}^2=[\sigma_{P+1}^2,\sigma_{P+2}^2,\ldots,\sigma_n^2]^\top$, $\boldsymbol{v}^2=[v_{P+1}^2,v_{P+2}^2,\ldots,v_n^2]^\top$, $*$ denotes the component wise product of vectors and $\zeta_{jk},\ j=0,1,\ldots,P,\ k=P+1,\ldots,n$ are between u_k and u_0. Multiplying both sides by $(\boldsymbol{X}_1^\top\boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\boldsymbol{W}_1$,

$$\hat{\boldsymbol{\beta}}_1(u_0)=\boldsymbol{\beta}_1(u_0)\ +\frac{1}{(d+1)!}(\boldsymbol{X}_1^\top\boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\boldsymbol{W}_1$$

$$\times\begin{bmatrix}\mu_0^{(d+1)}(\zeta_{0(P+1)})(u_{P+1}-u_0)^{d+1}\\ \vdots\\ \mu_0^{(d+1)}(\zeta_{0(n)})(u_n-u_0)^{d+1}\end{bmatrix}+\frac{1}{(d+1)!}\sum_{j=1}^{P}(\boldsymbol{X}_1^\top\boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\boldsymbol{W}_1$$

$$\times\begin{bmatrix}\mu_j^{(d+1)}(\zeta_{j(P+1)})(u_{P+1}-u_0)^{d+1}\epsilon_{P+1-j}^2\\ \vdots\\ \mu_j^{(d+1)}(\zeta_{j(n)})(u_n-u_0)^{d+1}\epsilon_{n-j}^2\end{bmatrix}+(\boldsymbol{X}_1^\top\boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\boldsymbol{W}_1(\boldsymbol{\sigma}^2*(\boldsymbol{v}^2-e_{n-P})).\tag{13}$$

Now it is not difficult to show using Lemma A.1 (i) that

$$\boldsymbol{X}_1^\top\boldsymbol{W}_1\begin{bmatrix}\mu_0^{(d+1)}(\zeta_{0(P+1)})(u_{P+1}-u_0)^{d+1}\\ \vdots\\ \mu_0^{(d+1)}(\zeta_{0(n)})(u_n-u_0)^{d+1}\end{bmatrix}$$

$$=nh_1^{d+1}\mu_0^{(d+1)}(u_0)[1,\boldsymbol{e}_P^\top w_2]^\top(1+o_P(1))\otimes\boldsymbol{D}_1,$$

$$\boldsymbol{X}_1^\top\boldsymbol{W}_1\begin{bmatrix}\mu_j^{(d+1)}(\zeta_{j(P+1)})(u_{P+1}-u_0)^{d+1}\epsilon_{P+1-j}^2\\ \vdots\\ \mu_j^{(d+1)}(\zeta_{j(n)})(u_n-u_0)^{d+1}\epsilon_{n-j}^2\end{bmatrix}$$

$$=nh_1^{d+1}\mu_j^{(d+1)}(u_0)[w_2,C_{j-1},\ldots,C_{j-P}]^\top(1+o_P(1))\otimes\boldsymbol{D}_1,$$

and,

$$(\boldsymbol{X}_1^\top\boldsymbol{W}_1\boldsymbol{X}_1)^{-1}=(1/n)\boldsymbol{S}^{-1}(1+o_P(1))\otimes\boldsymbol{A}_1^{-1}.$$

Hence, the asymptotic bias is given as,

$$\begin{aligned}&E(\hat{\boldsymbol{\beta}}_1(u_0)-\boldsymbol{\beta}_1(u_0))\\ &=\tfrac{h_1^{d+1}}{(d+1)!}\Big(\mu_0^{(d+1)}(u_0)(\boldsymbol{S}^{-1}\otimes\boldsymbol{A}_1^{-1})[(1,w_2\boldsymbol{e}_P^\top]^\top\otimes\boldsymbol{D}_1)\\ &+\sum_{j=1}^{P}\mu_j^{(d+1)}(u_0)(\boldsymbol{S}^{-1}\otimes\boldsymbol{A}_1^{-1})([w_2,C_{j-1},\ldots,C_{j-P}]^\top\otimes\boldsymbol{D}_1)\Big)+o_P(h_1^{d+1}).\end{aligned}$$

Notice that $C_0 = w_4$. Now

$$
\begin{aligned}
&E(\hat{\boldsymbol{\beta}}_1(u_0) - \boldsymbol{\beta}_1(u_0)) \\
&= \tfrac{h_1^{d+1}}{(d+1)!}(\boldsymbol{S}^{-1} \otimes \boldsymbol{A}_1^{-1}) \Big(\big(\mu_0^{(d+1)}(u_0)[1, w_2 \boldsymbol{e}_P^\top]^\top \\
&\quad + \sum_{j=1}^{P} \mu_j^{(d+1)}(u_0)[w_2, C_{j-1}, \ldots, C_{j-P}]^\top \big) \otimes \boldsymbol{D}_1 \Big) + o_P(h_1^{d+1}) \\
&= \tfrac{h_1^{d+1}}{(d+1)!}(\boldsymbol{S}^{-1} \otimes \boldsymbol{A}_1^{-1}) \Big(\boldsymbol{S}[\mu_0^{(d+1)}(u_0), \mu_1^{(d+1)}(u_0), \ldots, \mu_P^{(d+1)}(u_0)]^\top \otimes \boldsymbol{D}_1 \Big) \\
&\quad + o_P(h_1^{d+1}) \\
&= \tfrac{h_1^{d+1}}{(d+1)!} \Big([\mu_0^{(d+1)}(u_0), \mu_1^{(d+1)}(u_0), \ldots, \mu_P^{(d+1)}(u_0)]^\top \otimes \boldsymbol{A}_1^{-1}\boldsymbol{D}_1 \Big) + o_P(h_1^{d+1})
\end{aligned}
$$

Notice that Bias $(\hat{\mu}_j(u_0))= e_{j(d+1)+1,(P+1)(d+1)}^\top$ Bias $(\hat{\boldsymbol{\beta}}_1(u_0))$. Hence the bias expression is obtained.

Now the asymptotic variance is

$$
\begin{aligned}
&Var(\hat{\boldsymbol{\beta}}_1(u_0)) \\
&= (1/n)(\boldsymbol{S}^{-1}(1 + o_P(1)) \otimes \boldsymbol{A}_1^{-1}) Var(\boldsymbol{X}_1^\top \boldsymbol{W}_1(\boldsymbol{\sigma}^2 * (\boldsymbol{v}^2 - \boldsymbol{e}_{n-P}))) \\
&\quad \times (1/n)(\boldsymbol{S}^{-1}(1 + o_P(1)) \otimes \boldsymbol{A}_1^{-1}). \\
&= (1/n)(\boldsymbol{S}^{-1}(1 + o_P(1)) \otimes \boldsymbol{A}_1^{-1})((n/h_1) Var(v_t^2)\boldsymbol{\Omega}(1 + o_P(1)) \otimes \boldsymbol{B}_1) \\
&\quad \times (1/n)(\boldsymbol{S}^{-1}(1 + o_P(1)) \otimes \boldsymbol{A}_1^{-1}).
\end{aligned}
$$

using Lemma A.2. The desired expression for the variance can be obtained after some simplification using the properties of Kronecker product. The asymptotic normality follows using the Martingale central limit theorem.

Proof of Lemma 2. It is clear from Lemma 1 that the bias of $\hat{\mu}_k(u_0)$ is $O_P(h_1^{d+1})$. Also, the parameter functions $\mu_k(\cdot)$ are geometrically decaying. Hence using the expressions (4) and (6), the lemma follows.

To prove Lemma 3, we state the following lemma which is similar to Lemma A.2.

Lemma A.3. *Under the same assumptions as in Lemma 3,*

$$
\begin{aligned}
&Var\left(\sum_{k=p+1}^{n} (u_k - u_0)^i K_{h_2}(u_k - u_0)(v_k^2 - 1)\sigma_k^2 [1, \epsilon_{k-1}^2, \ldots, \epsilon_{k-p}^2 \hat{\sigma}_{k-1}^2, \ldots, \hat{\sigma}_{k-q}^2] \right) \\
&\qquad = n h_2^{2i-1} \nu_{2i} Var(v_t^2) \boldsymbol{\Omega}_2 (1 + o_P(1)), \quad i = 1, 2, \ldots, d.
\end{aligned}
$$

Proof of Lemma 3. Denote

$\boldsymbol{\beta}_2 = (\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0d}, \alpha_{10}, \ldots, \alpha_{1d}, \alpha_{p0}, \ldots, \alpha_{pd}, \beta_{10}, \ldots, \beta_{1d}, \beta_{q0}, \ldots, \beta_{qd})$, where $\alpha_{00}, \ldots, \beta_{qd}$ are constants. Using Taylor's series expansion in (7),

$$
\hat{\boldsymbol{\beta}}_2(u_0) = \boldsymbol{\beta}_2(u_0) \; + \frac{1}{(d+1)!}(\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{W}_2 \begin{bmatrix} \alpha_0^{(d+1)}(\xi_{0,r+1})(u_{r+1} - u_0)^{d+1} \\ \vdots \\ \alpha_0^{(d+1)}(\xi_{0,n})(u_n - u_0)^{d+1} \end{bmatrix}
$$

$$+\frac{1}{(d+1)!}(\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{W}_2 \sum_{i=1}^{p}\begin{bmatrix} \alpha_i^{(d+1)}(\xi_{i,r+1})(u_{r+1}-u_0)^{d+1}\epsilon_{r+1-i}^2 \\ \vdots \\ \alpha_i^{(d+1)}(\xi_{i,n})(u_n-u_0)^{d+1}\epsilon_{n-i}^2 \end{bmatrix}$$

$$+\frac{1}{(d+1)!}(\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{W}_2 \sum_{j=1}^{q}\begin{bmatrix} \beta_j^{(d+1)}(\xi_{j,r+1})(u_{r+1}-u_0)^{d+1}\hat{\sigma}_{r+1-j}^2 \\ \vdots \\ \beta_j^{(d+1)}(\xi_{j,n})(u_n-u_0)^{d+1}\hat{\sigma}_{n-j}^2 \end{bmatrix}$$

$$+o_P(h_2^{d+1}) + (\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{W}_2(\boldsymbol{\sigma}_2^2 * (\boldsymbol{v}_2^2 - \boldsymbol{e}_{n-r})),$$

where $\xi_{0,t}, \xi_{i,t}$ and $\xi_{j,t}$ are between u_t and u_0. Here $\boldsymbol{v}_2^2 = [v_{r+1}^2, \ldots, v_n^2]^\top$ and $\boldsymbol{\sigma}_2^2 = [\sigma_{r+1}^2, \ldots, \sigma_n^2]^\top$. Notice that the bias of $\hat{\sigma}_{t-j}^2$ in (7) is $o_P(h_2^{d+1})$ using Lemma 2 and the assumption $h_1 = o(h_2)$. We ignore the term $O(\rho^{p_n})$ as it is negligible asymptotically. In a similar way as in the proof of Lemma 1, it can be seen that

$$\boldsymbol{X}_2^\top \boldsymbol{W}_2 \begin{bmatrix} \alpha_0^{(d+1)}(\xi_{0,r+1})(u_{r+1}-u_0)^{d+1} \\ \vdots \\ \alpha_0^{(d+1)}(\xi_{0,n})(u_n-u_0)^{d+1} \end{bmatrix}$$

$$= nh_2^{d+1}\alpha_0^{(d+1)}(u_0)E[\boldsymbol{x}_t]^\top(1+o_P(1)) \otimes \boldsymbol{D}_2,$$

$$\boldsymbol{X}_2^\top \boldsymbol{W}_2 \begin{bmatrix} \alpha_i^{(d+1)}(\xi_{i,r+1})(u_{r+1}-u_0)^{d+1}\epsilon_{r+1-i}^2 \\ \vdots \\ \alpha_i^{(d+1)}(\xi_{i,n})(u_n-u_0)^{d+1}\epsilon_{n-i}^2 \end{bmatrix}$$

$$= nh_2^{d+1}\alpha_i^{(d+1)}(u_0)E[\tilde{\epsilon}_{t-i}^2 \boldsymbol{x}_t]^\top(1+o_P(1)) \otimes \boldsymbol{D}_2,$$

$$\boldsymbol{X}_2^\top \boldsymbol{W}_2 \begin{bmatrix} \beta_j^{(d+1)}(\xi_{j,r+1})(u_{r+1}-u_0)^{d+1}\hat{\sigma}_{r+1-j}^2 \\ \vdots \\ \beta_j^{(d+1)}(\xi_{j,n})(u_n-u_0)^{d+1}\hat{\sigma}_{n-j}^2 \end{bmatrix}$$

$$= nh_2^{d+1}\beta_j^{(d+1)}(u_0)E[\tilde{\sigma}_{t-j}^2 \boldsymbol{x}_t]^\top(1+o_P(1)) \otimes \boldsymbol{D}_2$$

and

$$(\boldsymbol{X}_2^\top \boldsymbol{W}_2 \boldsymbol{X}_2)^{-1} = (1/n)\boldsymbol{S}_2^{-1}(1+o_P(1)) \otimes \boldsymbol{A}_2^{-1}.$$

Therefore,

$$\begin{aligned} &\mathrm{Bias}(\hat{\boldsymbol{\beta}}_2(u_0)) \\ &= \tfrac{h_2^{d+1}}{(d+1)!}(\boldsymbol{S}_2^{-1}(1+o_P(1)) \otimes \boldsymbol{A}_2^{-1})\Big(\Big(\alpha_0^{(d+1)}(u_0)E[\boldsymbol{x}_t]^\top \\ &\quad + \sum_{i=1}^{p}\alpha_i^{(d+1)}(u_0)E[\tilde{\epsilon}_{t-i}^2 \boldsymbol{x}_t]^\top + \sum_{j=1}^{q}\beta_j^{(d+1)}(u_0)[\tilde{\sigma}_{t-j}^2 \boldsymbol{x}_t]^\top\Big)(1+o_P(1)) \otimes \boldsymbol{A}_2^{-1}\boldsymbol{D}_2\Big) + o_P(h_2^{d+1}) \\ &= \tfrac{h_2^{d+1}}{(d+1)!}(\boldsymbol{S}_2^{-1} \otimes \boldsymbol{A}_2^{-1})\Big((\boldsymbol{S}_2[\alpha_0^{(d+1)}(u_0), \alpha_1^{(d+1)}(u_0), \ldots, \alpha_p^{d+1}(u_0), \beta_1^{(d+1)}(u_0), \ldots, \beta_q^{d+1}(u_0)]^\top) \otimes \boldsymbol{D}_2\Big) \\ &\quad + o_P(h_2^{d+1}) \\ &= \tfrac{h_2^{d+1}}{(d+1)!}[\alpha_0^{(d+1)}(u_0), \alpha_1^{(d+1)}(u_0), \ldots, \alpha_p^{d+1}(u_0), \beta_1^{(d+1)}(u_0), \ldots, \beta_q^{d+1}(u_0)]^\top \otimes \boldsymbol{A}_2^{-1}\boldsymbol{D}_2 + o_P(h_2^{d+1}). \end{aligned}$$

The bias expressions can be obtained by using

$\text{Bias}(\hat{\alpha}_0(u_0)) = e^\top_{1,(p+q+1)(d+1)}\text{Bias}(\hat{\boldsymbol{\beta}}_2(u_0)),\ \text{Bias}(\hat{\alpha}_i(u_0)) = e^\top_{i(d+1)+1,(p+q+1)(d+1)}\text{Bias}(\hat{\boldsymbol{\beta}}_2(u_0))$ and $\text{Bias}(\hat{\beta}_j(u_0)) = e^\top_{(i+j)(d+1)+1,(p+q+1)(d+1)}\text{Bias}(\hat{\boldsymbol{\beta}}_2(u_0))$.

Now using Lemma A.3

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}_2(u_0)) &= (1/n)\boldsymbol{S}_2^{-1}(1+o_P(1)) \otimes \boldsymbol{A}_2^{-1} Var(\boldsymbol{X}_2^\top \boldsymbol{W}_2(\boldsymbol{\sigma}_2^2 * (\boldsymbol{v}_2^2 - \boldsymbol{e}_{n-r}))) \\ &\quad \times (1/n)\boldsymbol{S}_2^{-1}(1+o_P(1)) \otimes \boldsymbol{A}_2^{-1} \\ &= \tfrac{1}{nh_2} Var(v_t^2)(\boldsymbol{S}_2^{-1} \otimes \boldsymbol{A}_2^{-1})(\boldsymbol{\Omega}_2 \otimes \boldsymbol{B}_2)(\boldsymbol{S}_2^{-1} \otimes \boldsymbol{A}_2^{-1})(1+o_P(1)). \end{aligned}$$

The variance expression given in Lemma 3 can be arrived at after some simplification. The asymptotic normality follows using the martingale central limit theorem.

To prove Theorems 1 and 2, we state the following lemma, which can be proved using (9) in a similar way as Lemma A.1.

Lemma A.4. *Suppose that the assumptions A_1 to A_4 and A_6 hold. Let h denote a bandwidth such that $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Then*

$$\sum_{k=P+1}^{n} \tfrac{1}{nh}(u_k - u_0)^i K\left(\tfrac{u_k-u_0}{h}\right) w_k \epsilon^{2l}_{k-j_1}\epsilon^{2m}_{k-j_2} \overset{P_B}{\to} h^i \tau_i E(\tilde{\epsilon}^{2l}_{k-j_1}(u_0)\tilde{\epsilon}^{2m}_{k-j_2}(u_0)),$$
$$\forall\, l, m \in \{0,1,2\} \text{ and } \forall\ j_1, j_2,\ j_1 \neq j_2,\ i = 0,1,2\dots,2d.$$

Proof of Theorem 1. Let

$$\boldsymbol{e}_{P+1} = \begin{bmatrix} \boldsymbol{e}^\top_{1,(P+1)(d+1)} \\ \boldsymbol{e}^\top_{(d+1)+1,(P+1)(d+1)} \\ \vdots \\ \boldsymbol{e}^\top_{P(d+1)+1,(P+1)(d+1)} \end{bmatrix}_{(P+1)\times(P+1)(d+1)}.$$

Then

$$\begin{aligned} &\sigma_{wn}^{-1}\sqrt{nh_1}\,(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu}_{tvARCH}(u_0)) \\ &\qquad = \boldsymbol{e}_{P+1}\sigma_{wn}^{-1}\sqrt{nh_1}(\boldsymbol{X}_1^\top \boldsymbol{W}_{B1}\boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_{B1}\boldsymbol{W}_1(\boldsymbol{Y}_1 - \boldsymbol{X}_1\boldsymbol{\beta}_{B1}(u_0)), \end{aligned}$$

where $\boldsymbol{\beta}_{B1}(u_0) = [\mu_0(u_0), \mu_0^{(1)}(u_0), \dots, \frac{1}{d!}\mu_0^{(d)}(u_0), \dots, \frac{1}{d!}\mu_P^{(d)}(u_0)]^\top$. Now, using Lemma A.4, it can be shown that

$$\frac{1}{n}\boldsymbol{X}_1^\top \boldsymbol{W}_{B1}\boldsymbol{W}_1\boldsymbol{X}_1 = \frac{1}{nh_1}\sum_{k=P+1}^{n} K\left(\frac{u_k-u_0}{h_1}\right) w_k \boldsymbol{Z}_k\boldsymbol{Z}_k^\top = \boldsymbol{S}\otimes\boldsymbol{A}_1\left(1+o_{P_B}(1)\right).$$

Using the Taylor's series expansion of order $d+1$ for each parameter function in (5), we can write

$$\begin{aligned} &\sigma_{wn}^{-1}\sqrt{h_1/n}\boldsymbol{X}_1^\top \boldsymbol{W}_{B1}\boldsymbol{W}_1(\boldsymbol{Y}_1 - \boldsymbol{X}_1\boldsymbol{\beta}_{B1}(u_0)) \\ &\qquad = \sigma_{wn}^{-1}\sqrt{nh_1}\sum_{k=P+1}^{n}\tfrac{1}{nh_1}K\left(\tfrac{u_k-u_0}{h_1}\right) w_k \boldsymbol{Z}_k^\top \sigma_k^2(v_k^2-1) + \sigma_{wn}^{-1}\sqrt{nh_1}O_{P_B}(h_1^{d+1}), \end{aligned}$$

where the second term in the above expression is due to the remainder in the Taylor's series expansion. It can be shown in a similar way as Lemma A.2 that the bootstrap variance of each element of the first term goes to zero. Therefore

$$\sigma_{wn}^{-1}\sqrt{h_1/n}\boldsymbol{X}_1^\top \boldsymbol{W}_{B1}\boldsymbol{W}_1(\boldsymbol{Y}_1 - \boldsymbol{X}_1\boldsymbol{\beta}_{B1}(u_0)) = o_{P_B}(1) + \sigma_{wn}^{-1}\sqrt{nh_1}O_{P_B}(h_1^{d+1}).$$

Notice that under the assumption A_6, $\sigma_{wn}^2 = o(n)$. Therefore

$$\sigma_{wn}^{-1}\sqrt{nh_1}\left(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu}_{tvARCH}(u_0)\right) = \sigma_{wn}^{-1}\sqrt{nh_1}O_{P_B}(h_1^{d+1}) + o_{P_B}(1).$$

Now,

$$\begin{aligned}
&\sigma_{wn}^{-1}\sqrt{nh_1}\left(\hat{\boldsymbol{\mu}}_B - \hat{\boldsymbol{\mu}}_{tvARCH}(u_0)\right) \\
&\qquad = \sigma_{wn}^{-1}\sqrt{nh_1}\left(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu}_{tvARCH}(u_0)\right) - \sigma_{wn}^{-1}\sqrt{nh_1}\left(\hat{\boldsymbol{\mu}}_{tvARCH}(u_0) - \boldsymbol{\mu}_{tvARCH}(u_0)\right) \\
&\qquad = \sigma_{wn}^{-1}\sqrt{nh_1}(O_{P_B}(h_1^{d+1}) - O_P(h_1^{d+1})) + o_{P_B}(1),
\end{aligned}$$

using Lemma 1. Rest of the theorem can be easily proved.

Proof of Theorem 2. Let $\boldsymbol{y}_t = [1, \epsilon_{t-1}^2, \ldots, \epsilon_{t-P}^2]^\top$. Then we can write as $n \to \infty$

$$\hat{\sigma}_{Bt}^2 = \hat{\boldsymbol{\mu}}_B^\top \boldsymbol{y}_t = \left(\boldsymbol{\mu}_{tvARCH} + \frac{h_1^{d+1}}{(d+1)!}\boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_1^{-1}\boldsymbol{D}_1\boldsymbol{\mu}_{tvARCH}^{(d+1)}(u_0)\right)^\top \boldsymbol{y}_t.$$

Since the parameter functions $\mu_k(\cdot)$ are geometrically decaying, we can write using (4), for some $0 < \rho < 1$,

$$\hat{\sigma}_{Bt}^2 - \sigma_t^2 = O_{P_B}(h_1^{d+1}) + O_P(\rho^{P_n}).$$

Using this and Lemma A.4, it can be shown that

$$\frac{1}{n}\boldsymbol{X}_{B2}^\top \boldsymbol{W}_{B2}\boldsymbol{W}_2\boldsymbol{X}_{B2} = \boldsymbol{S}_2 \otimes \boldsymbol{A}_2\left(1 + o_{P_B}(1)\right).$$

Let $\boldsymbol{\beta}_{B2}(u_0) = [\alpha_0(u_0), \alpha_0^{(1)}(u_0), \ldots, \frac{1}{d!}\alpha_0^{(d)}(u_0), \alpha_1(u_0), \ldots, \frac{1}{d!}\beta_q^{(d)}(u_0)]^\top$. Then using Taylor series expansion of all the parameter functions in (7), we can write

$$\begin{aligned}
&\sigma_{wn}^{-1}\sqrt{h_2/n}\boldsymbol{X}_{B2}^\top \boldsymbol{W}_{B2}\boldsymbol{W}_2(\boldsymbol{Y}_2 - \boldsymbol{X}_{B2}\boldsymbol{\beta}_{B2}(u_0)) \\
&\qquad = \sigma_{wn}^{-1}\sqrt{nh_2}\sum_{k=P+1}^{n}\tfrac{1}{nh_2}K\left(\tfrac{u_k-u_0}{h_2}\right)w_k\boldsymbol{Z}_{Bk}^\top\sigma_k^2(v_k^2-1) + \sigma_{wn}^{-1}\sqrt{nh_2}O_{P_B}(h_2^{d+1}) \\
&\qquad\quad - \sum_{j=1}^{q}\beta_j\left(\tfrac{t}{n}\right)\left(O_{P_B}(h_1^{d+1}) + O_P(\rho^{P_n})\right),
\end{aligned}$$

where $\boldsymbol{Z}_{Bk} = [\boldsymbol{U}_k, \epsilon_{k-1}^2\boldsymbol{U}_k, \ldots, \epsilon_{k-p}^2\boldsymbol{U}_k, \hat{\sigma}_{Bk-1}^2\boldsymbol{U}_k, \ldots, \hat{\sigma}_{Bk-q}^2\boldsymbol{U}_k]^\top$. Here the first term can be shown to converge to zero in probability, second term is due to the remainder term in

Taylor series expansion and the third term is due to the bias of initial estimation of σ_t^2. If the initial step bandwidth is chosen in such a way that $h_1 = o(h_2)$, then the last term can be ignored and

$$\sigma_{wn}^{-1}\sqrt{h_2/n}\boldsymbol{X}_{B2}^\top \boldsymbol{W}_{B2}\boldsymbol{W}_2(\boldsymbol{Y}_2 - \boldsymbol{X}_{B2}\boldsymbol{\beta}_{B2}(u_0)) = \sigma_{wn}^{-1}\sqrt{nh_2}O_{P_B}(h_2^{d+1}) + o_{P_B}(1).$$

The rest of the theorem can be proved in a similar way as in Theorem 1.

Proof of Theorem 3. Denote $\boldsymbol{\sigma}^2 = [\sigma_{P+1}^2, \ldots, \sigma_n^2]$ and similarly $\boldsymbol{v}^2 = [v_{P+1}^2, \ldots, v_n^2]$. Notice that

$$(\hat{\mu}_k(u) - \mu_k(u) - bias(\hat{\mu}_k(u))) = \boldsymbol{e}_{k(d+1)+1,(P+1)(d+1)}^\top (\boldsymbol{X}_1^\top \boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_1\boldsymbol{\sigma}^2 * (\boldsymbol{v}^2 - 1),$$

where * denotes the element-wise product of vectors. It can be shown that

$$\frac{1}{n}\boldsymbol{X}_1^\top \boldsymbol{W}_1\boldsymbol{X}_1 \xrightarrow{P} \boldsymbol{S} \otimes \boldsymbol{A}_1.$$

Therefore,

$$\begin{aligned}&(\boldsymbol{X}_1^\top \boldsymbol{W}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_1\boldsymbol{\sigma}^2(\boldsymbol{v}^2 - 1)\\ &\qquad = (1/n)(\boldsymbol{S}^{-1}(1 + o_P(1)) \otimes \boldsymbol{A}_1^{-1}) \sum_{i=P+1}^{n} \boldsymbol{y}_i \otimes \boldsymbol{U}_i\sigma_i^2(v_i^2 - 1)K_h(u_i - u_0),\end{aligned}$$

where $\boldsymbol{y}_i = [1, \epsilon_{i-1}^2, \ldots, \epsilon_{i-P}^2]^\top$. Using (9) and ignoring the terms of order $(1/n^2)$, we can write

$$\begin{aligned}(1/n)\sum_{i=P+1}^{n} \boldsymbol{y}_i \otimes \boldsymbol{U}_i\sigma_i^2(v_i^2 - 1)K_h(u_i - u_0) &= (1/n)\sum_{i=P+1}^{n} \widetilde{\boldsymbol{y}}_i \otimes \boldsymbol{U}_i\widetilde{\sigma}_i^2(v_i^2 - 1)K_h(u_i - u_0)\\ &\quad + (1/n)\sum_{i=P+1}^{n} O_P\left(\left|\tfrac{i}{n} - u_0\right| + \tfrac{1}{n}\right)(v_i^2 - 1)K_h(u_i - u_0).\end{aligned}$$

Here $\widetilde{\boldsymbol{y}}_i = [1, \widetilde{\epsilon}_{i-1}^2, \ldots, \widetilde{\epsilon}_{i-P}^2]^\top$. We drop (u_0) from the notations of stationary processes for simplicity. The second term converges to zero in probability. Now, suppose that $F_{n,v^2,\boldsymbol{z}}(v^2, \boldsymbol{z})$ denotes the empirical distribution function of $\{\epsilon_t^2/\hat{\sigma}_t^2, \boldsymbol{z}_t\}$ where $\boldsymbol{z}_t = [\epsilon_{t-1}^2, \ldots \epsilon_{t-P}^2]^\top$. Let $F(v^2, \widetilde{\boldsymbol{z}})$ denote the joint distribution function of $\{v_t^2, \widetilde{\boldsymbol{z}}_t\}$ where $\widetilde{\boldsymbol{z}}_t = [\widetilde{\epsilon}_{t-1}^2, \widetilde{\epsilon}_{t-2}^2, \ldots, \widetilde{\epsilon}_{t-P}^2]$. Then using (9), it can be easily shown that $F_{n,v^2,\boldsymbol{z}}(v^2, \boldsymbol{z}) \xrightarrow{P} F(v^2, \widetilde{\boldsymbol{z}})$. We can write

$$n^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_1\boldsymbol{\sigma}^2(\boldsymbol{v}^2 - 1) = \int_{v^2,\widetilde{\boldsymbol{z}},u} \widetilde{\boldsymbol{y}} \otimes \boldsymbol{U}\widetilde{\sigma}^2(v^2 - 1)K_h(u - u_0)dF_{n,v^2,\boldsymbol{z}}(v^2, \widetilde{\boldsymbol{z}})du.$$

Let $Z_n = \sqrt{n}(F_{n,v^2,\boldsymbol{z}}(v^2, \boldsymbol{z}) - F(v^2, \widetilde{\boldsymbol{z}}))$ denote the empirical process of $\{v_t^2, \widetilde{\boldsymbol{z}}_t\}$. Then,

$$n^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_1\boldsymbol{\sigma}^2(\boldsymbol{v}^2 - 1) = (n^{-3/2})\int_{v^2,\widetilde{\boldsymbol{z}},u} \widetilde{\boldsymbol{y}} \otimes \boldsymbol{U}\widetilde{\sigma}^2(v^2 - 1)K_h(u - u_0)dZ_n(v^2, \widetilde{\boldsymbol{z}})du.$$

Here, the remaining term vanishes as $E(v_t^2 - 1) = 0$. Let $B(v^2, \widetilde{\boldsymbol{z}})$ denote the Brownian bridge based on the uniform measure on $[0,1]^{P+1}$. Then,

$$n^{-1}\boldsymbol{X}_1^\top \boldsymbol{W}_1 \boldsymbol{\sigma}^2(\boldsymbol{v}^2 - 1) = (n^{-3/2}) \int\limits_{v^2, \widetilde{\boldsymbol{z}}, u} \widetilde{\boldsymbol{y}} \otimes \boldsymbol{U} \widetilde{\sigma}^2 (v^2 - 1) K_h(u - u_0) dB(v^2, \widetilde{\boldsymbol{z}}) du + \eta_n,$$

where

$$\eta_n = (n^{-3/2}) \int\limits_{v^2, \widetilde{\boldsymbol{z}}, u} \widetilde{\boldsymbol{y}} \otimes \boldsymbol{U} \widetilde{\sigma}^2 (v^2 - 1) K_h(u - u_0) \left[dZ_n(v^2, \widetilde{\boldsymbol{z}}) - dB(v^2, \widetilde{\boldsymbol{z}}) \right] du.$$

Therefore,

$$(\boldsymbol{X}_1^\top \boldsymbol{W}_1 \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top \boldsymbol{W}_1 \boldsymbol{\sigma}^2 (\boldsymbol{v}^2 - 1) = (n^{-3/2})(\boldsymbol{S}^{-1}(1 + o_P(1)) \otimes \boldsymbol{A}_1^{-1})$$

$$\times \int\limits_{v^2, \widetilde{\boldsymbol{z}}, u} \widetilde{\boldsymbol{y}} \otimes \boldsymbol{U} \widetilde{\sigma}^2 (v^2 - 1) K_h(u - u_0) dB(v^2, \widetilde{\boldsymbol{z}}) du + n^{-1}(\boldsymbol{S}^{-1}(1 + o_P(1)) \otimes \boldsymbol{A}_1^{-1})\eta_n.$$

Now, using a similar arguments as in Lemma 3 of Gruet (1996), it can be shown that the second term is negligible in the above expression. Considering $(k(d+1)+1)^{th}$ element of this vector, we have

$$\begin{aligned} &\boldsymbol{e}_{k(d+1)+1,(P+1)(d+1)}^\top (\boldsymbol{X}_1^\top \boldsymbol{W}_1 \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top \boldsymbol{W}_1 \boldsymbol{\sigma}^2 * (\boldsymbol{v}^2 - 1) \\ &= n^{-3/2} \int\limits_{v^2, \widetilde{\boldsymbol{z}}, u} \left(\boldsymbol{e}_{k,P+1}^\top \boldsymbol{S}^{-1}(1 + o_P(1)) \widetilde{\boldsymbol{y}} \right) \left(e_{1,d+1}^\top \boldsymbol{A}_1^{-1} \boldsymbol{U} \right) \widetilde{\sigma}^2 (v^2 - 1) K_h(u - u_0) dB(v^2, \widetilde{\boldsymbol{z}}) du \end{aligned}.$$

Rest of the theorem can be argued in a similar way as in Gruet (1996, Lemma 3), Hardle (1989) and Bickel and Rosenblatt (1973) using some algebraic adjustments and the fact that

$$Var(\hat{\mu}_k(u)) = \frac{1}{nh_1} \boldsymbol{e}_{1,d+1}^\top \boldsymbol{A}_1^{-1} \boldsymbol{B}_1 \boldsymbol{A}_1^{-1} \boldsymbol{e}_{1,d+1} Var(v_t^2) \boldsymbol{e}_{k,P+1}^\top \boldsymbol{S}^{-1} \boldsymbol{\alpha}_0 \boldsymbol{S}^{-1} e_{k,P+1}.$$

Proof of Theorem 4. This proof can be formulated in a similar way as in Theorem 3. The bias in the estimation of σ_t^2 in the first step is negligible under the assumption $h_1 = o(h_2)$. Details are omitted.

References

Amado, C., Terasvirta, T., 2008. Modeling conditional and unconditional heteroskedasticity with smoothly time-varying structure. Working paper series in Economics and Finance NIPE WP 3 / 2008, Universidade do Minho.

Andreou, E., Ghysels, E., 2002. Detecting multiple breaks in financial market volatility dynamics. Journal of Applied Econometrics 17, 579-600.

Bickel, P., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Annals of Statistics 1, 1071-1095.

Bose, A., Mukherjee, K., 2003. Estimating the ARCH parameters by solving linear equations. Journal of Time Series Analysis 24, 127-136.

Bose, A., Mukherjee, K., 2009. Bootstrapping a weighted linear estimator of the ARCH parameters. Journal of Time Series Analysis 30, 315-331.

Chatterjee, S., Bose, A., 2005. Generalized bootstrap for estimating equations. Annals of Statistics 33:414-436.

Chen, B., Hong, Y., 2009. Detecting for smooth structural changes in GARCH models. Working paper.

Chu, C. J., 1995. Detecting parameter shift in GARCH models. Econometric Reviews 14, 241-266.

Čížek, P., Spokoiny, V., 2009. Varying coefficient GARCH models. In: Handbook of Financial Time Series, 169-186, (eds T. G. Andersen, R. A. Davis, J. P. Kreib, T. Mikosch), Springer-Verlag, Berlin Heidelberg.

Dahlhaus, R., Subba Rao, S., 2006. Statistical inference for time-varying ARCH processes. Annals of Statistics 34, 1075-1114.

Davis, R. A., Mikosch, T., 2009. Extreme value theory for GARCH processes. In Handbook of Financial Time Series, 187-200, (eds T. G. Andersen, R. A. Davis, J. P. Kreib, T. Mikosch), Springer-Verlag, Berlin Heidelberg.

Fan, J., Zhang, W., 2000. Simultaneous confidence bands and hypothesis testing in varying coefficient models. Scandinavian Journal of Statistics 27, 715-731.

Fryzlewicz, P., Sapatinas, T., Subba Rao, S., 2008. Normalized least-squares estimation in time-varying ARCH models. Annals of Statistics 36, 742-786.

Gruet, M.A., 1996. A nonparametric calibration analysis. Annals of Statistics 24, 1474-1492.

Hardle, W., 1989. Asymptotic maximal deviation of m-smoothers. Journal of Multivariate Analysis 29, 163-179.

Hart, J. D., 1994. Automated kernel smoothing of dependent data by using time series cross- validation. Journal of Royal Statistical Society Series B 56, 529-542.

Mikosch, T., Starica, C., 2004. Nonstationarities in financial time series, the long-range dependence and the IGARCH effects. Review of Economics and Statistics 86, 378-390.

Praestgaard, J., Wellner, J., 1993. Exchangeably weighted bootstrap of the general empirical process. Annals of Probability 21, 2053-2086.

Rohan, N., Ramanathan, T. V., 2012a. Non-parametric estimation of a time varying GARCH model. To appear: Journal of Nonparametric Statistics.
Available from URL: http://stats.unipune.ernet.in/NR-TVR-3.pdf.

Rohan, N., Ramanathan, T. V., 2012b. Asymmetric volatility models with structural breaks. Communications in Statistics- Simulations and Computations. 41, 1519-1543.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Time Series Models in Econometrics, Finance and Other Fields (eds D. R. Cox, D. V. Hinkley, O. E. Barndorff-Nielsen), Chapman and Hall, London.

Table 1. Empirical probabilities of false rejection

	$n = 500$			$n = 300$		
Confidence level	0.01	0.05	0.1	0.01	0.05	0.1
$H_0 : \alpha_0(\cdot) = c$	0.009	0.064	0.120	0.006	0.076	0.156
$H_0 : \alpha(\cdot) = c$	0.007	0.061	0.114	0.012	0.036	0.051
$H_0 : \beta(\cdot) = c$	0.003	0.056	0.105	0.002	0.052	0.114

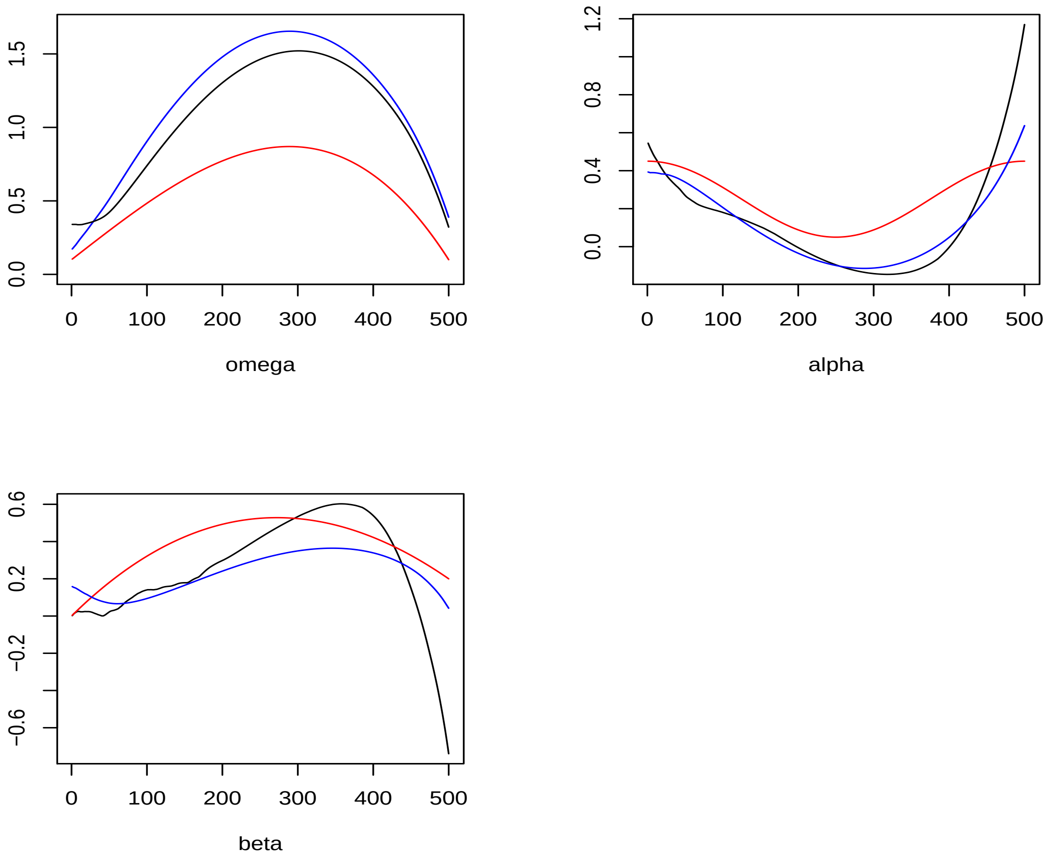

Figure 1. Plot of bootstrapped (blue) and actual (black) estimators along with the parametric function (red)

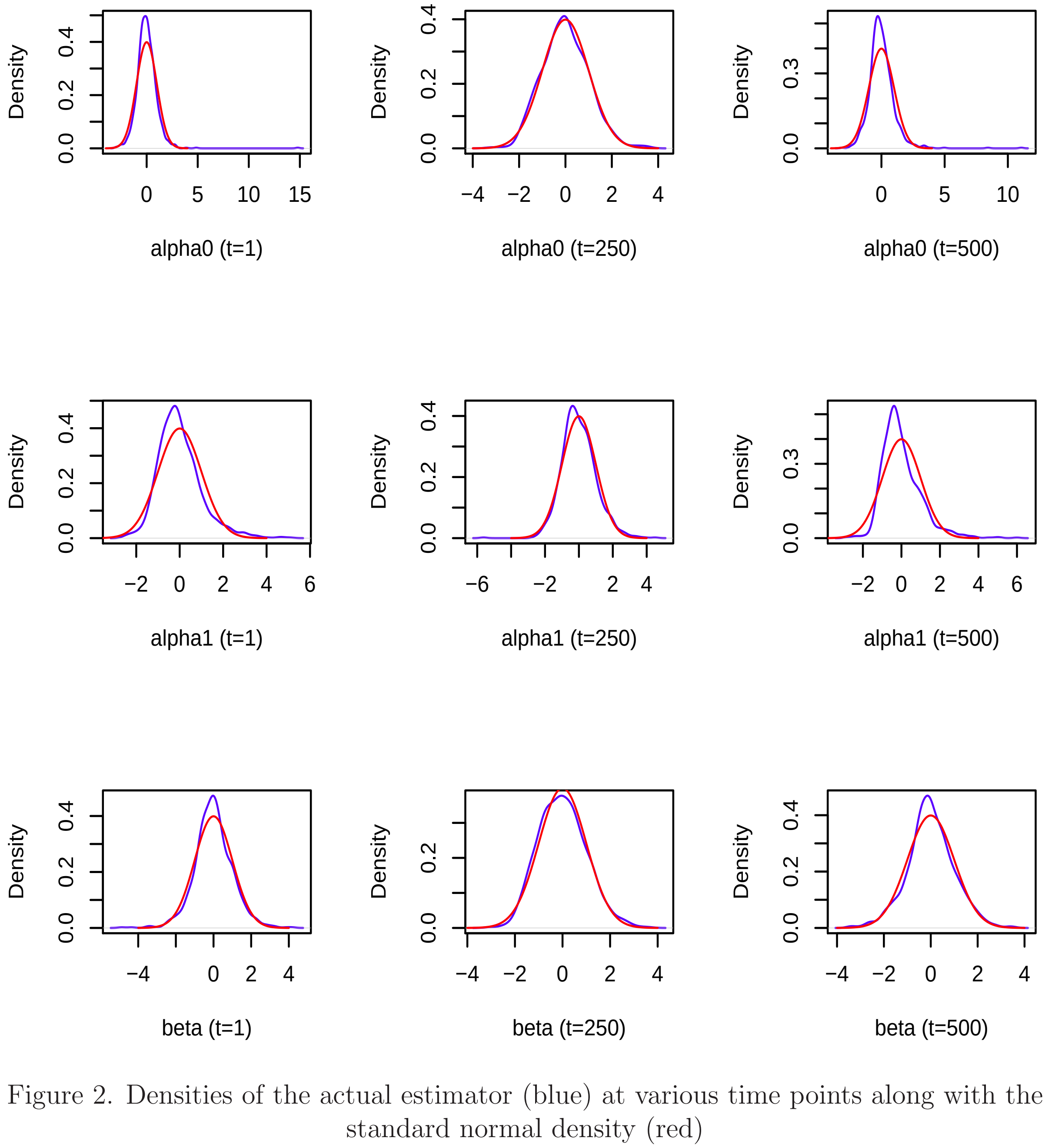

Figure 2. Densities of the actual estimator (blue) at various time points along with the standard normal density (red)

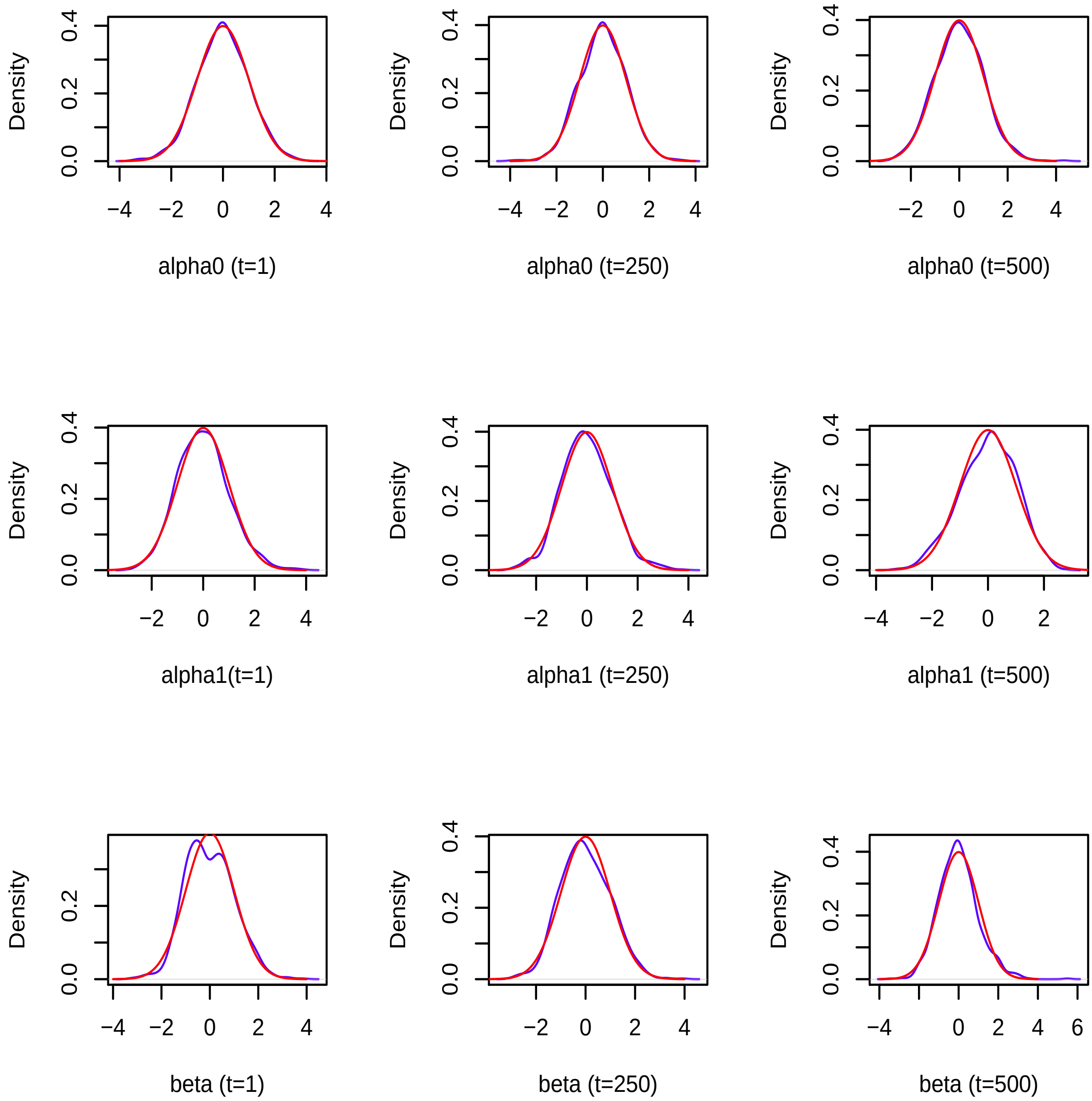

Figure 3. Densities of the bootstrapped estimator (blue) at various time points along with the standard normal density (red)

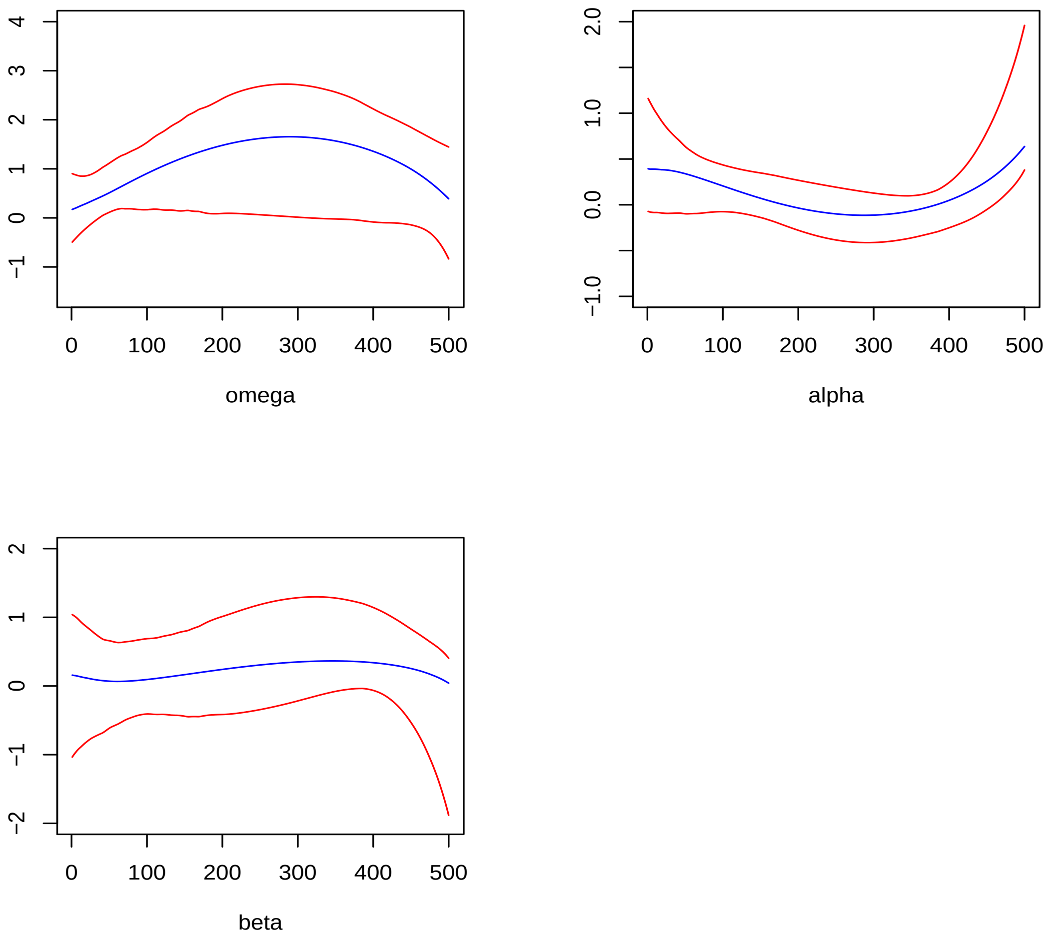

Figure 4. 95% confidence bands for the parameter functions