Econometric Theory, **32**, 2016, 1023–1054.
doi:10.1017/S0266466615000055

SPLINE ESTIMATION OF A SEMIPARAMETRIC GARCH MODEL

RONG LIU
Soochow University and University of Toledo

LIJIAN YANG
Soochow University

The semiparametric GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) model of Yang (2006, *Journal of Econometrics* 130, 365–384) has combined the flexibility of a nonparametric link function with the dependence on infinitely many past observations of the classic GARCH model. We propose a cubic spline procedure to estimate the unknown quantities in the semiparametric GARCH model that is intuitively appealing due to its simplicity. The theoretical properties of the procedure are the same as the kernel procedure, while simulated and real data examples show that the numerical performance is either better than or comparable to the kernel method. The new method is computationally much more efficient than the kernel method and very useful for analyzing large financial time series data.

1. INTRODUCTION

Volatility forecasting is of special interest for risk management and portfolio choice that involve many financial time series such as stock and foreign exchange returns. Empirical evidence had led to the understanding that for such series, the volatility often depends on infinitely many past returns with diminishing weights. The GARCH(p, q) model of Bollerslev (1986), for example, allows the volatility function to depend on all past observations, with geometrically decaying rate.

As a special case, the GARCH(1, 1) model describes a process $\{Y_t\}_{t=-\infty}^{\infty}$ of the form $Y_t = \sigma_t \xi_t, t \in \mathbf{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ where the innovations $\{\xi_t\}_{t\in\mathbf{Z}}$ are i.i.d random variables satisfying $E(\xi_t) = 0$, $E(\xi_t^2) = 1$, and $\{\sigma_t^2\}_{t=-\infty}^{\infty}$ denotes the conditional volatility series $\sigma_t^2 = \mathrm{var}(Y_t | Y_{t-1}, Y_{t-2}, \ldots)$ i.e.,

$$\sigma_t^2 = w + \beta_0 \sum_{j=1}^{\infty} \theta_0^{j-1} Y_{t-j}^2, t \in \mathbf{Z}, w, \beta_0, \theta_0 > 0, \theta_0 + \beta_0 < 1. \quad (1)$$

The comments from editor Peter Phillips, co-editor Oliver Linton and two anonymous referees have resulted in substantial improvement of the work. This research has been supported in part by NSF awards DMS 0706518, 1007594, Jiangsu Specially-Appointed Professor Program SR10700111, Jiangsu Province Key-Discipline (Statistics) Program ZY107002, ZY107992, National Natural Science Foundation of China Award NSFC 11371272, Research Fund for the Doctoral Program of Higher Education of China Award 20133201110002. Address correspondence to Lijian Yang, Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou 215006, China; e-mail: yanglijian@suda.edu.cn.

Nelson (1990) established necessary and sufficient conditions for the stationarity and ergodicity of the GARCH(1, 1) process. Engle and Ng (1993), Glosten, Jaganathan, and Runkle (1993), Hentschel (1995), Duan (1997), Hafner and Herwartz (2006), and Hafner (2008) examined various useful extensions of model (1), mostly providing empirical evidence without establishing asymptotic results. For related theoretical works on GARCH model, see Peng and Yao (2003), Sun and Stengos (2006), Chan, Deng, Peng, and Xia (2007), Giraitis, Leipus, and Surgailis (2010), Meitz and Saikkonen (2011), and Andrews (2012). Linton, Pan, and Wang (2010) and Zhang and Ling (2014) established asymptotic results for heavy-tailed noises.

In recent years, there has been a surge of interest in applying nonparametric smoothing theory to volatility estimation, as in Yang, Härdle, and Nielsen (1999), Dahl and Levine (2006), Levine (2006), Brown and Levine (2007), and Kim and Linton (2011). In particular, Linton and Mammen (2005) proposed an iterative algorithm for nonparametric GARCH model of the form

$$\sigma_t^2 = \sum_{j=1}^{\infty} \theta_0^{j-1} m\left(Y_{t-j}\right), t \in \mathbf{Z}, 0 < \theta_0 < 1$$

with unknown parameter θ_0 and unknown smooth news impact function m, without asymptotic theory. A truncated version of the above nonparametric model was studied in Yang (2000), Yang (2002), and Wang, Feng, Song, and Yang (2012) with asymptotic results, yet it failed to capture the dependence of σ_t^2 on infinitely many past Y_{t-j}.

As an alternative, Yang (2006) proposed a class of semiparametric GARCH model

$$\sigma_t^2 = m\left\{\sum_{j=1}^{\infty} \theta_0^{j-1} \upsilon\left(Y_{t-j}; \eta_0\right)\right\}, t \in \mathbf{Z}, \theta_0 \in (0,1), \eta_0 \in [\eta_1, \eta_2] \subset [0, \infty), \quad \text{(2)}$$

where $\upsilon(y; \eta) = y^2 + \eta_0 y^2 1_{(y<0)}$, with unknown parameter vector $\gamma_0 = (\theta_0, \eta_0)$ and unknown smooth link function m. The unknown γ_0 and m were estimated by kernel estimation method with satisfactory theoretical properties and numerical accuracy in simulations and applications. Like all the aforementioned works based on kernel smoothing, the algorithm in Yang (2006) is extremely slow due to the intensive computation of solving as many least squares problems as the sample size. The average computing time for the local linear based algorithm in Yang (2006) is contained in Table 3 for sample sizes n from 1,000 to 4,000, and one can see that it grows at the rate of n^2. At $n = 4{,}000$, which is a moderate sample size for financial time series, the estimation of unknown parameter vector γ_0 takes 5 hours. The method of Yang (2006) is therefore not appealing for practical use.

Model (1) has been extended to multivariate GARCH by Bauwens, Laurent, and Rombouts (2006), Silvennoinen and Terasvirta (2009), Linton (2009), Conrad and Karanasos (2010), Hafner and Linton (2010), and Francq and Zakoïan (2012) which take into account conditional correlations in addition to conditional volatility. Extending the semiparametric model (2) to multivariate time series would bring much progress to an active field and this paper serves as an important first step in this direction.

It is widely recognized that global smoothing methods such as those by spline or wavelet are much more computationally efficient than local kernel smoothing, see for example the comparison of computing time in Xue and Yang (2006) and Wang and Yang (2007). Recent development of regression spline smoothing in terms of local asymptotics (Huang, 2003) and high dimensional and weakly dependent data (Huang and Yang, 2004; Xue and Yang, 2006; Wang and Yang, 2007) has presented convincing incentives for applying spline smoothing to solve challenging problems in time series analysis. We apply cubic spline smoothing to the semiparametric GARCH model (2), which resulted in a procedure which is much faster but shares the same theoretical and numerical properties of the kernel smoothing procedure in Yang (2006). Table 3 shows the computing time comparison between the proposed cubic spline method versus the local linear method in estimating γ_0. Clearly, the cubic spline method is superior for large samples as its computing time is proportional to n^{-1} of the corresponding time of the local linear method. The advantage of spline method had already been recognized by Engle and Ng (1993), who proposed spline estimation for the news impact curve for extensions of model (1), without developing justifications by asymptotic theory. Theoretical justifications may be also rather difficult to establish for wavelet or other basis, when applied in the time series context. Some comparisons can be found in Baraud, Comte, and Viennet (2001).

The paper is organized as follows. In Section 2 we discuss the assumptions of the model (2), the spline estimation of the unknown parameter γ_0 and asymptotic properties including oracle efficiency. In Section 3 we describe the implementation of the estimator. In Sections 4 and 5 we apply the method to simulated and empirical examples. All technical proofs are given in the Appendix.

2. ESTIMATION METHOD

The statistical inference of the semiparametric GARCH model (2) consists of estimating both parameter γ_0 and link function m. In this paper we focus on estimating the parameter and one can estimate the link function by using $\widehat{\gamma}$ as the true value of γ_0, but the theoretical properties of such plug-in estimation require further research.

For convenience, define

$$X_t = \sum_{j=1}^{\infty} \theta_0^{j-1} \upsilon\left(Y_{t-j}; \eta_0\right), t \in \mathbf{Z},$$

which simplifies model (2) to $Y_t = m^{1/2}(X_t)\xi_t, \sigma_t^2 = m(X_t), t \in \mathbf{Z}$ while the process $\{X_t\}_{t=-\infty}^{\infty}$ satisfies the Markovian equation $X_t = \theta_0 X_{t-1} + \upsilon\left\{m(X_{t-1})\xi_{t-1}^2; \eta_0\right\}, t \in \mathbf{Z}$.

The following assumptions on the data generating process are used.

A1: The process $\{Y_t\}_{t=-\infty}^{\infty}$ is strictly stationary, and the innovations $\{\xi_t\}_{t\in\mathbf{Z}}$ have finite 6-th absolute moments $E|\xi_t|^6 < \infty$.

A2: The link function $m(\cdot)$ is positive everywhere on R_+ and has Lipschitz continuous 4-th derivative. There exist constants $0 < \delta, c < \infty$ such that $EX_t^{\delta} < \infty$ and

$$\lim_{x\to\infty} \sup m^{2/\delta}(x)/x = m_0 \in (0, c).$$

Since γ_0 is an unknown parameter vector in $(0, 1) \times [\eta_1, \eta_2]$, to make numerical optimization feasible, we assume that θ_0 lies in the interior of $[\theta_1, \theta_2]$, where $0 < \theta_1 < \theta_2 < 1$, are boundary values known a priori. In practice, one takes sufficiently small θ_1 and large θ_2 based on prior knowledge of the data, and denotes $\Gamma = [\theta_1, \theta_2] \times [\eta_1, \eta_2]$. Define next $X_{\gamma,t}$ as a series analogous to X_t but with any candidate value of $\gamma \in \Gamma$

$$X_{\gamma,t} = \sum_{j=1}^{\infty} \theta^{j-1} \upsilon\left(Y_{t-j};\eta\right), t \in \mathbf{Z}. \tag{3}$$

We need the following assumptions on the processes $\left\{X_{\gamma,t}\right\}_{t=-\infty}^{\infty}, \gamma \in \Gamma$.

A3: The processes $\left\{X_{\gamma,t}\right\}_{t=-\infty}^{\infty}, \gamma \in \Gamma$ are jointly strictly stationary and geometrically α-mixing, i.e., the α -mixing coefficient $\alpha(k) \le c\rho^k$, for constants $c > 0$, $0 < \rho < 1$, where

$$\alpha(k) = \sup_{A\in\sigma\left(X_{\gamma,t},t\le 0,\gamma\in\Gamma\right),B\in\sigma\left(X_{\gamma,t},t\ge k,\gamma\in\Gamma\right)} |P(A)P(B) - P(A\cap B)|.$$

The ergodicity and mixing properties of $X_{\gamma,t}$ were discussed in Carrasco and Chen (2002) and Yang (2006). Mixing conditions similar to Assumption (A3) are standard in the time series literatures, see Linton and Mammen (2005) and Wang and Yang (2007), although primitive conditions that ensure Assumption (A3) remain unavailable. From Assumption (A3) and the fact that the innovations $\{\xi_t\}_{t=-\infty}^{\infty}$ are iid, the joint distribution of $\left(Y_t, \xi_t, X_{\gamma,t}, \gamma \in \Gamma\right)$ is strictly stationary. Since the range of each $X_{\gamma,t}$ is $(0, +\infty)$, one first transforms all $X_{\gamma,t}$'s by a common transformation to make their range $[0, 1]$, so B spline regression can be applied to the transformed variables. For each $\gamma \in \Gamma$, define the transformed variables for the $X_{\gamma,t}$ as,

$$U_{\gamma,t} = F\left(X_{\gamma,t}\right) = \frac{F_{\gamma_1}\left(X_{\gamma,t}\right) + F_{\gamma_2}\left(X_{\gamma,t}\right)}{2}, 1 \le t \le n$$

in which F_{γ_1} and F_{γ_2} are cdfs of $X_{\gamma_1,t}$ and $X_{\gamma_2,t}$ respectively, where $\gamma_1 = (\theta_1, \eta_1)$ and $\gamma_2 = (\theta_2, \eta_2)$. Since the $X_{\gamma,t}$'s are increasing in both θ and η, one has $F_{\gamma_1} \le F_{\gamma} \le F_{\gamma_2}$ for any $\gamma \in \Gamma$, thus the common transformation function F assigns sufficient probability mass to the whole range of $[0, 1]$. In particular, we denote $U_t = U_{\gamma_0,t} = F\left(X_{\gamma_0,t}\right) = F(X_t)$. With previous transformation, one assumes

A4: The pdf associated with F is $f(x) > 0, \forall x \in (0, +\infty)$ and $U_{\gamma,t}$ has a pdf $\varphi_\gamma(\cdot)$ which is Lipschitz continuous and there exist constants c_φ, C_φ such that $\inf_{\gamma \in \Gamma, 0 \le u \le 1} \varphi_\gamma(u) \ge c_\varphi$ and $\sup_{\gamma \in \Gamma, 0 \le u \le 1} \varphi_\gamma(u) \le C_\varphi$.

For any $\gamma \in \Gamma$ define the predictor of Y_t^2 based on $U_{\gamma,t}$ as $g_\gamma(u) = E(Y_t^2 | U_{\gamma,t} = u), 0 < u < 1$. In particular, denote $g(U_t) = g_{\gamma 0}(U_{\gamma 0,t}) = E(Y_t^2 | U_{\gamma 0,t}) = m(X_t)$. Define the risk function of γ as $R(\gamma) = E\left\{Y_t^2 - g_\gamma(U_{\gamma,t})\right\}^2$. Apparently Y_t has finite 4-th moment due to Assumptions (A1) and (A2). So $R(\gamma)$ allows the usual bias-variance decomposition $R(\gamma) = E\left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}^2 + (E|\xi_t|^4 - 1)$ $Eg^2(U_t)$, together with $g(U_t) \equiv g_{\gamma 0}(U_{\gamma 0,t})$, imply that

$$R(\gamma) = E\left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}^2 + R(\gamma_0) \ge R(\gamma_0), \forall \gamma \in \Gamma.$$

We need the following assumption on the function $R(\gamma)$,

A5: The function $R(\gamma)$ has positive definite Hessian matrix at γ_0, and consequently $R(\gamma)$ is locally convex at γ_0, i.e., for any $\varepsilon > 0$, there exists $\delta > 0$ such that $R(\gamma) - R(\gamma_0) < \delta$ implies $\|\gamma - \gamma_0\| < \varepsilon$, where $\|\cdot\|$ is the Euclidean norm.

Thus by minimizing the prediction error of Y_t^2 on $U_{\gamma,t}$, one should be able to locate the true parameter γ consistently via polynomial spline smoothing. To introduce the space of splines, we divide [0, 1] into $(N+1)$ subintervals $J_j = \left[t_j, t_{j+1}\right), j = 0, \ldots, N-1, J_N = [t_N, 1]$, where $T := \left\{t_j\right\}_{j=1}^N$ is a sequence of equally-spaced points, called interior knots, given as

$$t_{1-k} = \cdots = t_{-1} = t_0 = 0 < t_1 < \cdots < t_N < 1 = t_{N+1} = \cdots = t_{N+k},$$

in which $t_j = jh, j = 0, 1, \ldots, N+1, h = 1/(N+1)$ is the distance between neighboring knots. The j-th B-spline of order k for the knot sequence T denoted by $B_{j,k}$ is recursively defined by de Boor (2001) as

$$B_{j,k}(u) = \frac{\left(u - t_j\right) B_{j,k-1}(u)}{t_{j+k-1} - t_j} - \frac{\left(u - t_{j+k}\right) B_{j+1,k-1}(u)}{t_{j+k} - t_{j+1}}, 1 - k \le j \le N$$

for $k > 1$, with

$$B_{j,1}(u) = I_{\{u \in J_j\}} = \begin{cases} 1 & t_j \le u < t_{j+1} \\ 0 & \text{otherwise} \end{cases}.$$

Define the spaces of linear, quadratic, and cubic spline functions on [0, 1] as

$$\begin{aligned} G^{(k-2)} &= G^{(k-2)}[0,1] \\ &= \left\{\psi : \psi(u) \equiv \sum\nolimits_{J=1-k}^{N+1} \lambda_J B_{J,k}(u), u \in [0,1]\right\}, \quad k = 2, 3, 4. \end{aligned}$$

Given a realization $\{Y_t\}_{t=1}^n$, define for $\forall\gamma \in \Gamma$ the cubic spline estimator of $g_\gamma(\cdot)$

$$\hat{g}_\gamma(\cdot) = \text{argmin}_{\psi \in G^{(2)}} \frac{1}{n''} \sum\nolimits_{t=n'+1}^{n} \left\{Y_t^2 - \psi(U_{\gamma,t})\right\}^2$$

with $n'' = n - n'$, where the first n' data points are not used in the above estimator for implementation reasons given in Section 3. Define next the empirical risk function

$$\hat{R}(\gamma) = \frac{1}{n''} \sum\nolimits_{t=n'+1}^{n} \left\{Y_t^2 - \hat{g}_\gamma(U_{\gamma,t})\right\}^2$$

and let γ be the minimizer of $\hat{R}(\gamma)$, i.e.

$$\widehat{\gamma} = \text{argmin}_{\gamma \in \Gamma} \hat{R}(\gamma). \tag{4}$$

A6: The number of interior knots N satisfies: $n^{1/6} \ll N = N_n \ll n^{1/5} (\log n)^{-2/5}$, $h = (N+3)^{-1}$.

The proofs of the following proposition and theorems use complex spline smoothing arguments and are given in the Appendix.

PROPOSITION 1. *Under Assumptions (A1)–(A6), as $n \to \infty$*

$$\sup_{\gamma \in \Gamma} \left|\nabla^{(k)} \hat{R}(\gamma) - \nabla^{(k)} R(\gamma)\right| \to 0, a.s., k = 0, 1, 2.$$

The next theorem establishes the strong consistency of $\hat{\gamma}$.

THEOREM 1. *Under assumptions (A1)–(A6), as $n \to \infty$,* $\left|\hat{\gamma} - \gamma_0\right| \to 0, a.s.$.

Denote the asymptotic variance of $\widehat{\gamma}$ by the following formula

$$\Sigma(\gamma_0) = \left\{\nabla^2 R(\gamma_0)\right\}^{-1} \Psi(\gamma_0) \left\{\nabla^2 R(\gamma_0)\right\}^{-1} \tag{5}$$

with

$$\Psi(\gamma_0) = 4E\left[\left\{g_\gamma(U_{\gamma_0,t}) - Y_t^2\right\}^2 \left\{\nabla g_\gamma(U_{\gamma,t})\right\}\left\{\nabla g_\gamma(U_{\gamma,t})\right\}^T |_{\gamma=\gamma_0}\right]^2 \tag{6}$$

and

$$\nabla^2 R(\gamma_0) = 2E\left[\left\{g_{\gamma_0}(U_{\gamma_0,t}) - Y_t^2\right\} \nabla^2 g_\gamma(U_{\gamma,t})|_{\gamma=\gamma_0} + \left\{\nabla g_\gamma(U_{\gamma,t})\right\}\left\{\nabla g_\gamma(U_{\gamma,t})\right\}^T |_{\gamma=\gamma_0}\right]. \tag{7}$$

Here $\nabla^2 g_\gamma(U_{\gamma,t})|_{\gamma=\gamma_0}$ is not the same as $\nabla^2 g_\gamma(U_{\gamma_0,t})$ since both g_γ and $U_{\gamma,t}$ depend on γ. The next theorem establishes $\hat{\gamma}$'s $\sqrt{n}$-asymptotic normality.

THEOREM 2. *Under assumptions (A1)–(A6), as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\gamma} - \gamma_0\right) \to_d N\left(\mathbf{0}, \Sigma\left(\gamma_0\right)\right). \tag{8}$$

According to Theorem 2, the true parameter vector γ_0 can be estimated by $\widehat{\gamma}$ at $\sqrt{n}$-rate. One can then use the estimated $\widehat{\gamma}$ in place of the unknown γ_0 for the estimation of function m. We define next the "would-be oracle" estimator of γ_0 as $\widetilde{\gamma} = \operatorname{argmin}_{\gamma \in A} \widetilde{R}(\gamma_0)$ under the oracle assumption that the link function g is known, where the oracle empirical risk is $\widetilde{R}(\gamma) = \left(n''\right)^{-1} \sum_{t=n'+1}^{n} \left\{Y_t^2 - g(U_{\gamma,t})\right\}^2$. So $\widetilde{\gamma}$ serves as a benchmark of oracle optimality and the following theorem states the asymptotic oracle efficiency of the estimator $\widetilde{\gamma}$.

THEOREM 3. *Under assumptions (A1)–(A6), as $n \to \infty$, the estimator $\widehat{\gamma}$ is asymptotically oracally efficient, i.e., it is asymptotically as efficient as $\widetilde{\gamma}$. Specifically, $\sqrt{n}\left(\widetilde{\gamma} - \gamma_0\right) \to_d N\left(\mathbf{0}, \Sigma\left(\gamma_0\right)\right)$, where the variance $\Sigma\left(\gamma_0\right)$ is the same as in (5) and (8).*

The proof of Theorem 3 consists of routine arguments in parametric inference, thus it is omitted. If γ_0 is known, asymptotic convergence rate of the estimation of functions g and m are given by Stone (1985, Thm. 1) and Huang and Yang (2004, Lemma 4).

3. IMPLEMENTATION

For a given realization $\{Y_t\}_{t=1}^{n}$, denote in the following two integers

$$n' = \left[2\log n / \log\left(\theta_2^{-1}\right)\right] + 1, n'' = n - n'.$$

It is easily verified that

$$\sup_{\theta \in [\theta_1, \theta_2]} \theta^{n'} = \theta_2^{n'} < n^{-2},$$

which is the magnitude of error one would incur if the infinite series in (2) were truncated at n'. In practice, one always has to replace the infinite series of $X_{\gamma,t}$ in (3) by a finite truncated $\sum_{j=1}^{n'} \theta^{j-1} \upsilon\left(Y_{t-j}; \eta\right)$ for $t \in \mathbf{Z}$, the difference between the two being

$$\begin{aligned}\sum_{j=n'+1}^{\infty} \theta^{j-1} \upsilon\left(Y_{t-j}; \eta\right) &\le \sum_{j=n'+1}^{\infty} \theta_2^{j-1} \upsilon\left(Y_{t-j}; \eta\right) \\ &= \theta_2^{n'} \sum_{j=1}^{\infty} \theta_2^{j-1} \upsilon\left(Y_{t-n'-j}; \eta\right) \\ &= \theta_2^{n'} X_{\gamma_2, t-n'} < n^{-2} X_{\gamma_2, t-n'},\end{aligned}$$

which is bounded by n^{-2} times a stationary process with finite variance according to Assumption (A1). Thus instead of computing the infinite sum $\sum_{j=1}^{\infty} \theta^{j-1} v\left(Y_{t-j}; \eta\right)$, slowly growing truncation $\sum_{j=1}^{n'} \theta^{j-1} v\left(Y_{t-j}; \eta\right)$ is used for implementing the algorithm due to practicality. Also due to practicality, the empirical cdfs $\hat{F}_{\gamma_1}$ and $\hat{F}_{\gamma_2}$ of $X_{\gamma_1,t}$ and $X_{\gamma_2,t}$ are used in place of F_{γ_1} and F_{γ_2} respectively to compute the transformation function F. The range $[\gamma_1, \gamma_2]$ can be chosen as wide as possible. In practice, one can start with a narrow range, and expand the range in the case of the estimator $\hat{\gamma}$ being too close to γ_1 or γ_2, and re-estimate with the new parameter range. Lastly, the number of interior knots $N = N_n$ is computed according to the formula $N = \min\left(\left[n^{2/11}\right] + 1, n/4 - 1\right)$, which satisfies the Assumption (A6).

One computes the value of $\hat{R}(\gamma)$ over an equally spaced grid of points from γ_1 to γ_2, and takes the one with smallest $\hat{R}(\gamma)$ value as $\widehat{\gamma}$ according to (4). In the next two sections, numerical evidence is presented on how the proposed procedures work for both simulated and real time series data.

4. SIMULATION

To investigate the finite-sample precision of the proposed estimator, the procedure is applied to time series data generated according to (2) with $\theta_0 = 0.85, \eta_0 = 0.5, \Gamma = [\theta_1, \theta_2] \times [\eta_1, \eta_2] = [0.75, 0.95] \times [0.5, 0.6]$, and function

$$m(x) = 0.01\,(2x + 1 + \sin(x/5))\,/\,(1 - \theta_0) \tag{9}$$

and ξ_t has either the standard double exponential distribution or the standard normal distribution. The data generating process actually follows the standard GARCH model possessing all the known theoretical properties presented in Engle and Ng (1993) and Glosten et al. (1993).

For sample sizes $n = 1{,}000, 2{,}000, 4{,}000, 8{,}000$, a total of $K = 100$ realizations of length $n + 400$ are generated according to model (2), with functions $m(x)$ as (9). For each realization, the last n observations are kept as our data for inference. Truncating the first 400 observations off the series ensures that the remaining series behaves like a stationary one. Estimator $\widehat{\gamma}$ of the parameter γ_0 is computed according to the setups described in Section 3, using cubic spline. For comparison, we also compute the infeasible oracle estimator $\widetilde{\gamma}$ and the maximum likelihood estimator $\check{\gamma}$ with the correct link function and treating the ξ_t as standard normal.

For a parameter λ and its estimate $\hat{\lambda}$, define $\mathrm{Var}\left(\hat{\lambda}\right) = K^{-1}\sum_{k=1}^{K} \left(\hat{\lambda}_k - K^{-1}\sum_{k=1}^{K}\hat{\lambda}_k\right)^2$, $\mathrm{Bias}\left(\hat{\lambda}\right) = K^{-1}\sum_{k=1}^{K}\left(\hat{\lambda}_k - \lambda\right)$, $\mathrm{MSE}\left(\hat{\lambda}\right) = K^{-1}\sum_{k=1}^{K} \left(\hat{\lambda}_k - \lambda\right)^2$, where $\hat{\lambda}_k$ is obtained from k-th replication. Table 1 consists of the average sum of squared error, bias, variance for $\hat{\theta}, \tilde{\theta}$, and $\check{\theta}$, with $n = 1{,}000, 2{,}000, 4{,}000, 8{,}000$, and efficiencies $\mathrm{EFF}\left(\hat{\theta}, \tilde{\theta}\right) = \mathrm{MSE}\left(\tilde{\theta}\right)/\mathrm{MSE}\left(\hat{\theta}\right)$, $\mathrm{EFF}\left(\hat{\theta}, \check{\theta}\right) = \mathrm{MSE}\left(\check{\theta}\right)/\mathrm{MSE}\left(\hat{\theta}\right)$. Table 2 contains the same measures for $\hat{\eta}$, $\tilde{\eta}$ and $\check{\eta}$.

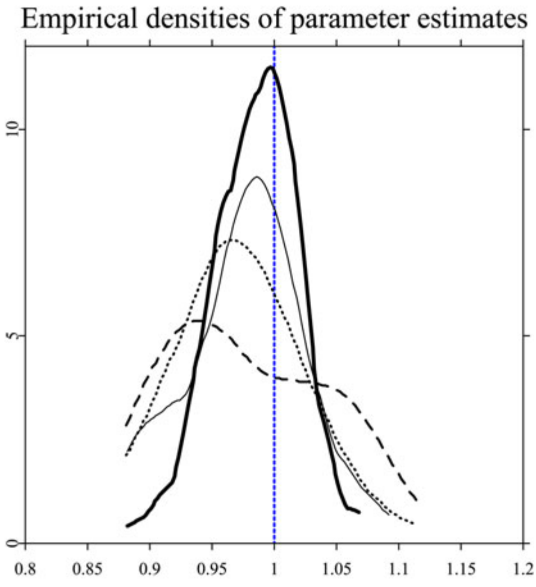
TABLE 1. Simulated example: The MSE, Bias, Var for $\hat{\theta}$, $\tilde{\theta}$, $\check{\theta}$ and $\mathrm{EFF}(\hat{\theta},\tilde{\theta})$, $\mathrm{EFF}(\hat{\theta},\check{\theta})$ for $n = 1{,}000, 2{,}000, 4{,}000, 8{,}000$ with 100 replications. Numbers outside/inside parenthesis correspond to ξ_t having standard double exponential/standard normal distributions

		$n = 1{,}000$	$n = 2{,}000$	$n = 4{,}000$	$n = 8{,}000$
MSE	$\hat{\theta}$	0.00290 (0.00298)	0.00221 (0.00244)	0.00158 (0.00120)	0.00076 (0.00086)
	$\tilde{\theta}$	0.00227 (0.00250)	0.00191 (0.00206)	0.00137 (0.00105)	0.00073 (0.00080)
	$\check{\theta}$	0.00368 (0.00246)	0.00271 (0.00188)	0.00230 (0.00093)	0.00134 (0.00068)
Bias	$\hat{\theta}$	−0.0110 (−0.0081)	−0.0104 (−0.0204)	0.0004 (0.00051)	0.0070 (−0.00134)
	$\tilde{\theta}$	−0.0048 (0.0050	−0.0034 (0.0026)	−0.0040 (−0.0062)	−0.0020 (0.0064)
	$\check{\theta}$	−0.0078 (−0.0042)	−0.0052 (−0.0074)	0.0044 (−0.0136)	0.0150 (−0.0040)
Var	$\hat{\theta}$	0.00277 (0.00283)	0.00210 (0.00202)	0.00158 (0.0011)	0.00071 (0.00068)
	$\hat{\theta}$	0.00197 (0.00241)	0.00188 (0.00192)	0.00135 (0.0009)	0.00067 (0.00072)
	$\check{\theta}$	0.00366 (0.00239)	0.00271 (0.00182)	0.00228 (0.0008)	0.00105 (0.00067)
EFF $(\hat{\theta},\tilde{\theta})$		0.783 (0.839)	0.864 (0.844)	0.867 (0.875)	0.961 (0.930)
EFF$(\hat{\theta},\check{\theta})$		1.269 (0.825)	1.226 (0.770)	1.456 (0.775)	1.684 (0.79)

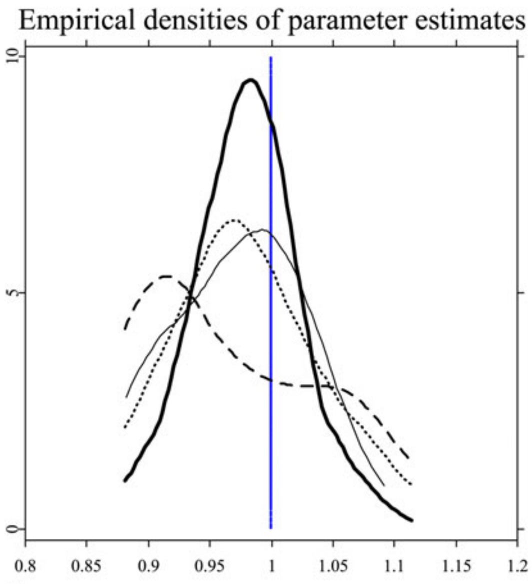
TABLE 2. Simulated example: The MSE, Bias, Var $\hat{\eta}$, $\tilde{\eta}$, $\check{\eta}$ and $\mathrm{EFF}(\hat{\eta},\tilde{\eta})$, for $n = 1{,}000, 2{,}000, 4{,}000, 8{,}000$ with 100 replications

		$n = 1{,}000$	$n = 2{,}000$	$n = 4{,}000$	$n = 8{,}000$
MSE	$\hat{\eta}$	0.00256 (0.00265)	0.00210 (0.00244)	0.00138 (0.00114)	0.00073 (0.00071)
	$\tilde{\eta}$	0.00184 (0.00235)	0.00163 (0.00206)	0.00145 (0.00105)	0.00068 (0.00062)
	$\check{\eta}$	0.00334 (0.00226)	0.00236 (0.00163)	0.00205 (0.00078)	0.00128 (0.00057)
Bias	$\hat{\eta}$	0.0067 (−0.0064)	−0.0174 (−0.0105)	−0.0024 (0.0036)	0.0066 (−0.0076)
	$\tilde{\eta}$	−0.0035 (0.0038)	−0.0022 (−0.0038)	−0.0045 (−0.0045)	−0.0018 (0.0024)
	$\check{\eta}$	−0.0064 (−0.0029)	−0.0052 (−0.0068)	0.0032 (−0.0027)	0.0078 (−0.0034)
Var	$\hat{\eta}$	0.00249 (0.00254)	0.00181 (0.00215)	0.00138 (0.0009)	0.00081 (0.00068)
	$\tilde{\eta}$	0.00173 (0.00226)	0.00162 (0.00184)	0.00142 (0.0009)	0.00065 (0.00059)
	$\check{\eta}$	0.00366 (0.00204)	0.00271 (0.00156)	0.00228 (0.0007)	0.00105 (0.00054)
EFF $(\hat{\eta},\tilde{\eta})$		0.719 (0.839)	0.776 (0.844)	0.917 (0.875)	0.939 (0.930)
EFF$(\hat{\eta},\check{\eta})$		1.437 (0.853)	1.123 (0.668)	1.485 (0.684)	1.753 (0.803)

Tables 1 and 2 show that the estimated $\hat{\theta}$ and $\hat{\eta}$ converge to the true parameters θ_0 and η_0 as the sample size increases, corroborating the asymptotics in Theorem 2. In Figure 1, the probability density functions of $\hat{\theta}/\theta_0$ and $\hat{\eta}/\eta_0$ are estimated by kernel smoothing based on the 100 replications, which also confirm



A



B

FIGURE 1. Plot of densities of (a) $\hat{\theta}/\theta_0$ and (b) $\hat{\eta}/\eta_0$ of $n = 1{,}000$ - dashed line, $n = 2{,}000$ - dotted line, $n = 4{,}000$ - thin solid line, $n = 8{,}000$ - thick solid line.

TABLE 3. Computing time (in seconds) of cubic spline estimation and local linear estimation of parameter θ_0 for one replication with $n = 1{,}000, 2{,}000, 4{,}000, 8{,}000$ (For $n = 8{,}000$, the time is omitted for local linear estimation as it is excessive.)

n	1,000	2,000	4,000	8,000
Spline estimation	12	35	92	252
Local linear estimation	650	3600	18000	–
Time ratio	1 : 57	1 : 103	1 : 196	–

the numerical convergence. The efficiency results in Tables 1 and 2 show that both $\text{EFF}\left(\hat{\theta}, \tilde{\theta}\right)$ and $\text{EFF}\left(\hat{\eta}, \tilde{\eta}\right)$ converge to 1 as Theorem 3 states; on the other hand, both $\text{EFF}\left(\hat{\theta}, \check{\theta}\right)$ and $\text{EFF}\left(\hat{\eta}, \check{\eta}\right)$ are much less than 1 for normal innovations and much greater than 1 for double exponential innovations. The latter phenomena are clearly caused by $\check{\gamma}$ being optimally efficient when the correct parametric model (in this case Gaussian model when actual innovations are normal) is specified, and completely wrong when an incorrect one is used (in this case Gaussian model when actual innovations are double exponential).

We have experimented with knot numbers ranging from 3 to 10, and have not seen significant changes in the simulation study.

As discussed in the introduction, Table 3 shows the computing time comparison between the proposed cubic spline method and the local linear method of Yang (2006) in estimating parameter γ_0. Since for each candidate parameter vector γ, the cubic spline method needs to solve one linear least squares problem in order to compute the empirical risk while the local linear has to solve n, one for each data point, so the ratio of their computing times is inversely proportional to n. As a matter of fact, the computing times are of order n and n^2 respectively for the cubic spline and the local linear methods. Since the theoretical properties and numerical performance of the two are similar, the cubic spline method is the one we would recommend for the estimation of parameter γ_0. Once the parameter γ_0 has been efficiently estimated, the estimation of functions g and m can be done via either kernel type or spline type method, using the estimated $\widehat{\gamma}$ in place of γ_0.

Given the above empirical observations, $\widehat{\gamma}$ is a very competitive estimator for γ_0 in terms of robustness, efficiency and computing time. Since the sample sizes we have used are common for high frequency financial time series such as the data set in the next section, the satisfactory numerical performance provides the assurance one needs to apply the procedure to real data.

5. APPLICATIONS

In this section, we apply the semiparametric GARCH model on stock daily percentage returns of the BMW share price from 1 June 1986 to 30 January 1994,

TABLE 4. Fitting the BMW stock returns, the GARCH(1, 1) model has $\hat{\theta} = 0.87$, the GJR(1, 1) model has $\hat{\theta} = 0.87$, $\hat{\eta} = 0.66$, and the semiparametric GARCH(Spline) has $\hat{\theta} = 0.87, \hat{\eta} = 0.66$

Fitted model	Log-likelihood	Volatility prediction error
GARCH(1,1)	−0.8528	15.65
GJR	−0.8234	15.04
Semi. GARCH(Spline)	−0.7952	14.15

a total of 2,000 observations. We truncated Y_t by its 0.01 and 0.99 quantiles. For more details, see Wang et al. (2012). In analyzing the data set, a process $\{X_{\gamma,t}\}_{t=1}^{2000}$ is generated for every parameter value γ. The parameter estimate $\hat{\gamma}$ is first obtained according to Section 3. In the second step, we use the estimated $\hat{\gamma}$ in place of the unknown γ_0 for the Nadaraya–Watson estimation of function g. The volatility forecasts are $\hat{\sigma}_t^2 = \hat{g}_{\hat{\gamma}}\left(U_{\hat{\gamma},t}\right)$, while the residuals are $\hat{\xi}_t = Y_t/\hat{\sigma}_t, t = 1, \ldots, 2000$.

In Table 4, we have compared the goodness-of-fit of our model with GARCH(1,1) and GJR(1,1) in terms of volatility prediction error $\sum_{t=51}^{2000}\left(Y_t^2 - \hat{\sigma}_t^2\right)^2/1950$ and loglikelihood $-(1/2)\sum_{t=51}^{2000}\left\{Y_t^2/\hat{\sigma}_t^2 + \ln\left(\hat{\sigma}_t^2\right)\right\}/1950$ with $n' = 50$. The semiparametric GARCH model (2) with spline estimation method has the best log-likelihood and prediction error. In Figure 2, we show the standardized residuals and estimated autocorrelation function (ACF) in the daily return series and there is very little if any dependence left in the residuals. Further evidence of the residuals' randomness is provided in Table 5, where p-values are listed for the Ljung–Box and Box-Pierce tests of the semiparametric GARCH residuals. All p-values are large, and hence there is no evidence of any serial dependence in the residuals.

Figure 3 (a) shows the graph of the estimation of function g. Then we can estimate the unknown smooth link function m by using $\hat{m}\left(X_{\hat{\gamma},t}\right) = \hat{g}_{\hat{\gamma}}\left(U_{\hat{\gamma},t}\right)$, which is shown in Figure 3 (b).

TABLE 5. Significance probabilities of Portmanteau tests on the residuals of the semiparametric GARCH model

Lag	LB	BP
20	0.7569	0.7633
30	0.7268	0.7308
40	0.5538	0.5771

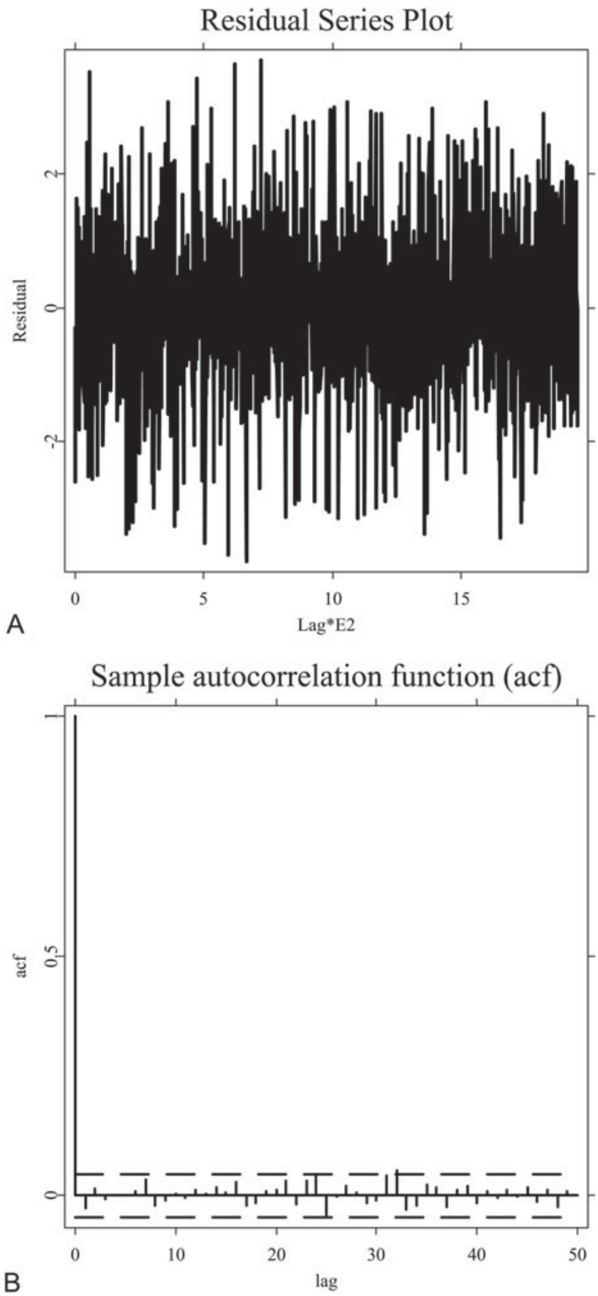
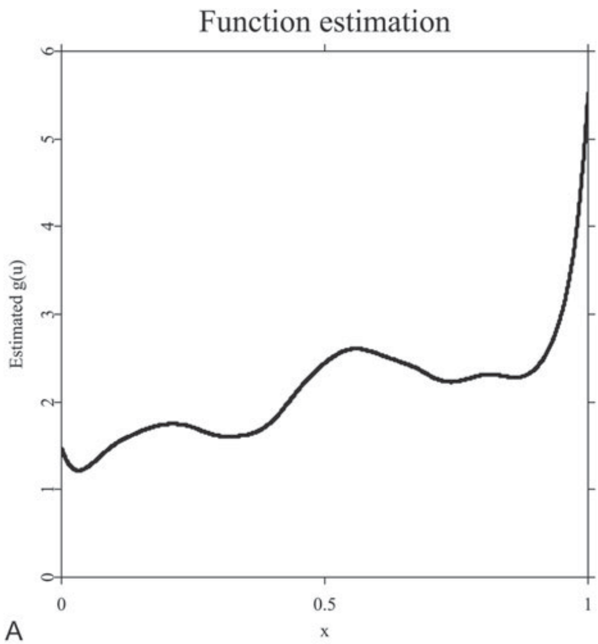
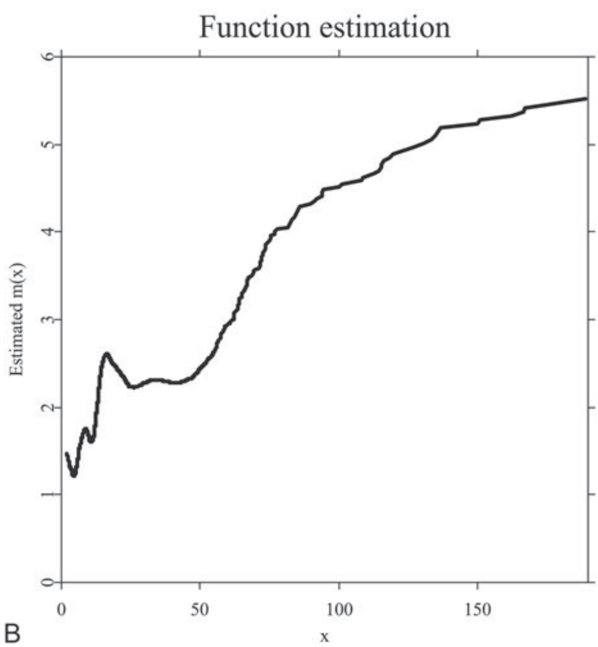


FIGURE 2. Semiparametric GARCH modeling of BMW stock returns: (a) Standardized residuals (b) the estimated ACF with 95% Bartlett intervals.



A



B

FIGURE 3. Plots of estimated function (a) $\hat{g}$ and (b) $\hat{m}$ by $\hat{m}\left(X_{\hat{\gamma},t}\right) = \hat{g}_{\hat{\gamma}}\left(U_{\hat{\gamma},t}\right)$ for the semiparametric GARCH model.

REFERENCES

Andrews, B. (2012) Rank-based estimation for GARCH processes. *Econometric Theory* 28, 1037–1064.

Baraud, Y., F. Comte, & G. Viennet (2001) Model selection for (auto-)regression with dependent data. *ESAIM: Probability and Statistics* 5, 33–49.

Bauwens, L., B. Laurent, & J. Rombouts (2006) Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21, 79–109.

Bollerslev, T.P. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes*. Springer-Verlag.

Brown, L.D. & M. Levine (2007) Variance estimation in nonparametric regression via the difference sequence method. *Annals of Statistics* 35, 2219–2232.

Carrasco, M. & X. Chen (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Chan, N., S. Deng, L. Peng, & Z. Xia (2007) Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics* 137, 556–576.

Conrad, C. & M. Karanasos (2010) Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory* 26, 838–862.

Dahl, C.M. & M. Levine (2006) Nonparametric estimation of volatility models with serially dependent innovations. *Statistics and Probability Letters* 76, 2007–2016.

de Boor, C. (2001) *A Practical Guide to Splines*. Springer-Verlag.

DeVore, R.A. & G.G. Lorentz (1993) *Constructive Approximation: Polynomials and Splines Approximation*. Springer-Verlag.

Duan, J.C. (1997) Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics* 79, 97–127.

Engle, R.F. & V. Ng (1993) Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749–1778.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag.

Francq, C. & J.M. Zakoïan (2012) QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* 28, 179–206.

Giraitis, L., R. Leipus, & R. Surgailis (2010) Aggregation of the random coefficient GLARCH (1,1) process. *Econometric Theory* 26, 406–425.

Glosten, L.R., R. Jaganathan, & D.E. Runkle (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Hafner, C.M. (2008) Temporal aggregation of multivariate GARCH processes. *Journal of Econometrics* 142, 467–483.

Hafner, C.M. & H. Herwartz (2006) Volatility impulse responses for multivariate GARCH models: An exchange rate illustration. *Journal of International Money and Finance* 25, 719–740.

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

Hentschel, L. (1995) All in the family: Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39, 71–104.

Huang, J.Z. (2003) Local asymptotics for polynomial spline regression. *Annals of Statistics* 31, 1600–1635.

Huang, J.Z. & L. Yang (2004) Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society Series B* 66, 463–477.

Kim, W. & O. Linton (2011) Estimation of a semiparametric IGARCH(1,1) model. *Econometric Theory* 27, 639–661.

Levine, M. (2006) Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: A possible approach. *Computational Statistics and Data Analysis* 50, 3405–3431.

Linton, O. (2009) Semiparametric and nonparametric ARCH modeling. In T.G. Andersen, R.A. Davis, J.P. Kreiss and Th. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 157–167. Springer-Verlag.

Linton, O. & E. Mammen (2005) Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica* 73, 771–836.

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

Meitz, M. & P. Saikkonen (2011) Parameter estimation in nonlinear AR–GARCH models. *Econometric Theory* 27, 1236–1278.

Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1, 1) Model. *Econometric Theory* 6, 318–334.

Peng, L. & Q. Yao (2003) Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* 90, 967–975.

Silvennoinen, A. & T. Teräsvirta (2009) Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373–411.

Stone, C. (1985) Additive regression and other nonparametric models. *Annals of Statistics* 13, 689–705.

Sun, Y. & T. Stengos (2006) Semiparametric efficient adaptive estimation of asymmetric GARCH models. *Journal of Econometrics* 133, 373–386.

Wang, L., C. Feng, Q. Song, & L. Yang (2012) Efficient semiparametric GARCH modelling of financial volatility. *Statistica Sinica* 22, 249–270.

Wang, L. & L. Yang (2007) Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics* 35, 2474–2503.

Xue, L. & L. Yang (2006) Additive coefficient modeling via polynomial spline. *Statistica Sinica* 16, 1423–1446.

Yang, L. (2000) Finite nonparametric GARCH model for foreign exchange volatility. *Communications in Statistics Theory and Methods* 5 & 6, 1347–1365.

Yang, L. (2002) Direct estimation in an additive model when the components are proportional. *Statistica Sinica* 12, 801–821.

Yang, L. (2006) A semiparametric GARCH model for foreign exchange volatility. *Journal of Econometrics* 130, 365–384.

Yang, L., W. Härdle, & J.P. Nielsen (1999) Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis* 20, 597–604.

Zhang, R. & S. Ling (2014) Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory*. Available on CJO2014. doi:10.1017/S0266466614000632.

APPENDIX

A.1. Preliminaries

In the whole section, we use $\|g\|_\infty$ denote $\sup_x |g(x)|$. For any functions $g_1, g_2 \in L_2[0,1]$, define for $\forall \gamma \in \Gamma$ the theoretical inner product and norm as

$$\langle g_1, g_2 \rangle_\gamma = \int_0^1 g_1(u) g_2(u) \varphi_\gamma(u) du, \ \|g_1\|_{2,\gamma}^2 = \langle g_1, g_1 \rangle_\gamma .$$

For any vector $\lambda = (\lambda_1, \dots, \lambda_p)$ and $0 < r < \infty$, $\|\lambda\|_r = \left(\sum_{i=1}^p |\lambda_i|^r\right)^{1/r}$ and $\|\lambda\|_\infty = \max(|\lambda_1|, \dots, |\lambda_p|)$. In particular, denote $\|\lambda\| = \|\lambda\|_2$.

LEMMA A.1 (Bernstein's inequality, Bosq, 1998, Thm. 1.4). *Let $\{\xi_t\}$ be a zero mean real valued process, $S_n = \sum_{i=1}^n \xi_i$. Suppose that there exists $c > 0$ such that for $i = 1, \cdots, n$, $k \geq 3, E|\xi_i|^k \leq c^{k-2} k! E\xi_i^2 < +\infty, m_r = \max_{1 \leq i \leq N} \|\xi_i\|_r, r \geq 2$. Then for each $n > 1$, integer $q = q_n \in [1, n/2]$, each $\varepsilon_n > 0$ and $k \geq 3$*

$$P\left\{\left|\sum_{i=1}^n \xi_i\right| > n\varepsilon_n\right\} \leq a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) + a_2(k)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{\frac{2k}{2k+1}},$$

where

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right), \qquad a_2(k) = 11n\left(1 + \frac{5m_k^{2k/(2k+1)}}{\varepsilon_n}\right).$$

Next, we introduce some properties of the B-spline. We denote by $Q_T(g)$ the 4-th order quasi-interpolant of g corresponding to the knots T, see DeVore and Lorentz (1993, eqn. 4.12, p. 146). According to DeVore and Lorentz (1993, Thm. 7.7.4, p. 225), the following lemma holds.

LEMMA A.2. *There exists a constant $C_\infty > 0$ such that for any $g \in C^{(r)}[0,1]$ and $0 \leq k \leq 2$, $\left\|(Q_T(g) - g)^{(k)}\right\|_\infty \leq C_\infty \left\|g^{(k)}\right\|_\infty h^{r-k}$.*

LEMMA A.3 (B-spline Property).

(i) *Partition of Unity. (de Boor, 2001, p. 96) The sequence $\{B_{j,k}\}_{j=-k+1}^N$ provides a positive and local partition of unity, i.e., each $B_{j,k}$ is positive on (t_j, t_{j+k}), is zero off $[t_j, t_{j+k}]$, $\sum_{j=-k+1}^N B_{j,k} = 1$.*

(ii) *Differentiation. (de Boor, 2001, p. 116)*

$$\frac{d}{du} B_{j,k}(u) = (k-1)\left\{\frac{B_{j,k-1}(u)}{t_{j+k-1} - t_j} - \frac{B_{j+1,k-1}(u)}{t_{j+k} - t_{j+1}}\right\}, 1 - k \leq j \leq N.$$

(iii) *Good Condition. (DeVore and Lorentz, 1993, Thm. 5.4.2, p. 145) There is a constant $D_k > 0$ such that for each spline $S = \sum_{j=-k+1}^N c_j B_{j,k}$ of order k and each $0 < r \leq \infty$,*

$$\begin{cases} D_k \|\mathbf{c}'\|_r \leq \|S\|_r \leq \|\mathbf{c}'\|_r, & 1 \leq r \leq \infty, \\ D_k \|\mathbf{c}'\|_r \leq \|S\|_r \leq k^{1/r} \|\mathbf{c}'\|_r, & 0 < r < 1. \end{cases}$$

LEMMA A.4. *There exist constants $c > 0$ such that for any $\lambda := (\lambda_{-1,2}, \lambda_{0,2}, \dots, \lambda_{N,2}, \dots, \lambda_{N,4}) \in R^{3N+6}$.*

$$\begin{cases} ch^{1/r}\|\lambda\|_r \leq \left\|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}\right\|_r \leq \left(3^{r-1}kh\right)^{1/r} \|\lambda\|_r, & 1 \leq r \leq \infty, \\ ch^{1/r}\|\lambda\|_r \leq \left\|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}\right\|_r \leq (3kh)^{1/r} \|\lambda\|_r, & 0 < r < 1. \end{cases}$$

In particular, under Assumption (A4), $\exists$ *constants* $c, C \in (0, +\infty)$ *such that*

$$ch^{1/2}\|\lambda\|_2 \le \left\|\sum\nolimits_{k=2}^{4}\sum\nolimits_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}\right\|_{2,\gamma} \le Ch^{1/2}\|\lambda\|_2, \forall\gamma\in\Gamma.$$

Proof. It follows from Lemma A.3 (i) that, $\sum_{k=2}^{4}\sum_{j=-k+1}^{N}B_{j,k}\equiv 3$ on $[0,1]$. So the right inequality follows immediately for $r=\infty$. When $1\le r<\infty$, Hölder's inequality implies that

$$\left|\sum\nolimits_{k=2}^{4}\sum\nolimits_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}\right| \le 3^{1-1/r}\left(\sum\nolimits_{k=2}^{4}\sum\nolimits_{j=-k+1}^{N}\left|\lambda_{j,k}\right|^r B_{j,k}\right)^{1/r}.$$

Since all the knots are equally spaced, Lemma A.3 (i) ensures that $\int_{-\infty}^{\infty}B_{j,k}(u)\,du\le kh$, the right inequality follows from $\int_0^1\left|\sum_{k=2}^{4}\sum_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}(u)\right|^r du\le 3^{r-1}kh\|\lambda\|_r^r$. When $r<1$, we have $\left|\sum_{k=2}^{4}\sum_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}\right|^r \le \sum_{k=2}^{4}\sum_{j=-k+1}^{N}\left|\lambda_{j,k}\right|^r B_{j,k}^r$. Since $\int_{-\infty}^{\infty}B_{j,k}^r(u)\,du\le t_{j+k}-t_j=kh$ and

$$\int_0^1\left|\sum\nolimits_{k=2}^{4}\sum\nolimits_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}(u)\right|^r du \le \|\lambda\|_r^r\int_{-\infty}^{\infty}B_{j,k}^r(u)\,du\le 3kh\|\lambda\|_r^r,$$

the right inequality follows in this case as well. For the left inequalities, we derive from Lemma A.3 (iii), for any $0<r\le\infty$

$$\left|\lambda_{j,k}\right|^r \le C_1^r h^{-1}\int_{t_j}^{t_{j+1}}\left|\sum\nolimits_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}(u)\right|^r du.$$

Since each $u\in[0,1]$ appears in at most k intervals (t_j, t_{j+k}), adding up these inequalities, we obtain that

$$\begin{aligned}\|\lambda\|_r^r &\le C_1 h^{-1}\sum_{k=1}^{4}\int_{t_j}^{t_{j+k}}\left|\sum\nolimits_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}(u)\right|^r du\\ &\le 3Ch^{-1}\left\|\sum\nolimits_{j=-k+1}^{N}\lambda_{j,k}B_{j,k}\right\|_r^r.\end{aligned}$$

The left inequality follows. ■

Define for any functions $g_1, g_2\in L_2[0,1]$ and any $\gamma\in\Gamma$ the empirical inner product and norm as

$$\langle g_1, g_2\rangle_{n,\gamma} = \left(n''\right)^{-1}\sum_{t=n'+1}^{n} g_1\left(U_{\gamma,t}\right)g_2\left(U_{\gamma,t}\right),\ \|g_1\|_{2,n,\gamma}^2 = \langle g_1, g_1\rangle_{n,\gamma}.$$

LEMMA A.5. *Under Assumptions (A3), (A4), and (A6), as* $n\to\infty$, *with probability* 1

$$\sup_{\gamma\in\Gamma}\max_{\substack{k,k'=2,3,4\\1-k\le j\le N,1-k'\le j'\le N}}\left|\left\langle B_{j,k}, B_{j',k'}\right\rangle_{n,\gamma} - \left\langle B_{j,k}, B_{j',k'}\right\rangle_{\gamma}\right| = O\left\{(nN)^{-1/2}\log n\right\}.$$

Proof. We only prove the case $k = k' = 4$, all other cases are similar. Let

$$\zeta_{\gamma,j,j',t} = B_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t}) - EB_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})$$

with the second moment

$$E\zeta^2_{\gamma,j,j',t} = E\left[B^2_{j,4}(U_{\gamma,t}) B^2_{j',4}(U_{\gamma,t})\right] - \{EB_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})\}^2,$$

where $E\left[B^2_{j,4}(U_{\gamma,t}) B^2_{j',4}(U_{\gamma,t})\right] \sim N^{-1}$, $\{EB_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})\}^2 \sim N^{-2}$ uniformly for all $-3 \le j, j' \le N$ by Assumption (A4) and Lemma A.4. Hence, $E\zeta^2_{\gamma,j,j',t} \sim N^{-1}$ uniformly for all $-3 \le j, j' \le N$. The $k-$th moment is

$$\begin{aligned} E\left|\zeta_{\gamma,j,j',t}\right|^k &= E\left|B_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t}) - EB_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})\right|^k \\ &\le 2^{k-1}\left\{E\left|B_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})\right|^k + \left|EB_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})\right|^k\right\}, \end{aligned}$$

where $E\left|B_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})\right|^k \sim N^{-1}$, $\left|EB_{j,4}(U_{\gamma,t}) B_{j',4}(U_{\gamma,t})\right|^k \sim N^{-k}$ uniformly for all $-3 \le j, j' \le N$. Thus, there exists a constant $C > 0$ such that $E\left|\zeta_{\gamma,j,j',t}\right|^k \le C2^{k-1}k!E\zeta^2_{\gamma,j,j',t}$ for all $-3 \le j, j' \le N$. So Cramér's condition in Lemma A.1 is satisfied and one has for $\delta_n = \delta \log n/\sqrt{nN}$, for some constant c such that $cn/\log n \le q = q_n < n/2$, and fixed γ

$$P\left\{\frac{1}{n''}\left|\sum\nolimits_{t=n'+1}^{n} \zeta_{\gamma,j,j',t}\right| > \delta_n\right\} \le n^{-10}. \tag{A.1}$$

We divide interval $[\theta_1, \theta_2]$ and $[\eta_1, \eta_2]$ into $M_n = n^3$ equally spaced intervals with disjoint endpoints $\theta_1 = \theta_1 < \cdots < \theta_{M_n} = \theta_2$ and $\eta_1 = \eta_1 < \cdots < \eta_{M_n} = \eta_2$ (Discretization). Then $\sup_{\gamma\in\Gamma} \max_{-3\le j,j'\le N} \left|\zeta_{\gamma,j,j',t}\right|$ is bounded by

$$\sup_{1\le r\le M_n} \max_{-3\le j,j'\le N} \left|\zeta_{a_r,j,j',t}\right| + \max_{-3\le j,j'\le N} \sup_{1\le r\le M_n} \max_{\gamma\in[\theta_r,\theta_{r+1}]\times[\eta_s,\eta_{s+1}]} \left|\zeta_{\gamma,j,j',t} - \zeta_{\gamma_{rs},j,j',t}\right|. \tag{A.2}$$

with $\gamma_{rs} = (\theta_r, \eta_s)$. While (A.1) implies that

$$\sup_{1\le r,s\le M_n} \max_{-3\le j,j'\le N} \left(n''\right)^{-1}\left|\sum\nolimits_{t=n'+1}^{n} \zeta_{\gamma_{rs},j,j',t}\right| = O\left\{(nN)^{-1/2}\log n\right\}, a.s. \tag{A.3}$$

by Borel–Cantelli lemma. Employing Lipschitz continuity of the cubic B-spline, one has with probability 1

$$\begin{aligned} &\max_{-3\le j,j'\le N} \sup_{1\le r\le M_n} \max_{\gamma\in[\theta_r,\theta_{r+1}]\times[\eta_s,\eta_{s+1}]} \left|\left(n''\right)^{-1}\sum\nolimits_{t=n'+1}^{n}\left(\zeta_{\gamma,j,j',t} - \zeta_{\gamma_{rs},j,j',t}\right)\right| \\ &= O\left(M_n^{-1}h^{-6}\right). \end{aligned} \tag{A.4}$$

Therefore Assumption A4, (A.2), (A.3), and (A.4) lead to the result. ■

Denote by $G = G^{(0)} \cup G^{(1)} \cup G^{(2)}$ the space of all linear, quadratic, and cubic spline functions on $[0, 1]$. We establish the uniform rate at which the empirical inner product approximates the theoretical inner product for all B-splines $B_{j,k}$ with $k = 2, 3, 4$.

LEMMA A.6. *Under Assumptions (A3), (A4), and (A6), as $n \to \infty$, one has*

$$A_n = \sup_{\gamma \in \Gamma} \sup_{\phi_1, \phi_2 \in G} \left| \frac{\langle \phi_1, \phi_2 \rangle_{n,\gamma} - \langle \phi_1, \phi_2 \rangle_\gamma}{\|\phi_1\|_{2,\gamma} \, \|\phi_2\|_{2,\gamma}} \right| = O\left\{ (nh)^{-1/2} \log n \right\}, a.s.. \tag{A.5}$$

Proof. Denote $\phi_a = \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \phi_{a,jk} B_{j,k}, a = 1, 2$, without loss of generality. Then

$$\langle \phi_1, \phi_2 \rangle_{n,\gamma} = \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \sum_{k'=2}^{4} \sum_{j'=-k+1}^{N} \phi_{1,j,k} \phi_{2,j'k'} \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\gamma},$$

$$\|\phi_1\|_{2,\gamma}^2 = \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \sum_{k'=2}^{4} \sum_{j'=-k+1}^{N} \phi_{1,j,k} \phi_{1,j'k'} \left\langle B_{j,k}, B_{j',k'} \right\rangle_{\gamma},$$

$$\|\phi_2\|_{2,\gamma}^2 = \sum_{k=2}^{4} \sum_{j=-k+1}^{N} \sum_{k'=2}^{4} \sum_{j'=-k+1}^{N} \phi_{2,j,k} \phi_{2,j'k'} \left\langle B_{j,k}, B_{j',k'} \right\rangle_{\gamma}.$$

Let $\phi_1 = (\phi_{1,-1,2}, \phi_{1,0,2}, \ldots, \phi_{1,N,2}, \ldots, \phi_{1,N,4})$, $\phi_2 = (\phi_{2,-1,2}, \phi_{2,0,2}, \ldots, \phi_{2,N,2}, \ldots, \phi_{2,N,4})$. According to Lemma A.4, one has for any $\gamma \in \Gamma$,

$$ch \|\phi_1\|_2^2 \le \|\phi_1\|_{2,\gamma}^2 \le Ch \|\phi_1\|_2^2, ch \|\phi_2\|_2^2 \le \|\phi_2\|_{2,\gamma}^2 \le Ch \|\phi_2\|_2^2,$$

$$ch \|\phi_1\|_2 \|\phi_2\|_2 \le \|\phi_1\|_{2,\gamma} \|\phi_2\|_{2,\gamma} \le Ch \|\phi_1\|_2 \|\phi_2\|_2 .$$

Hence

$$\begin{aligned} A_n &= \sup_{\gamma \in \Gamma} \sup_{\phi_1 \in \phi, \phi_2 \in G} \left| \frac{\langle \phi_1, \phi_2 \rangle_{n,\gamma} - \langle \phi_1, \phi_2 \rangle_\gamma}{\|\phi_1\|_{2,\gamma} \, \|\phi_2\|_{2,\gamma}} \right| \\ &\le \frac{\|\phi_1\|_\infty \|\phi_2\|_\infty}{c_1 h \|\phi_1\|_2 \|\phi_2\|_2} \sup_{\gamma \in \Gamma} \max_{\substack{k,k'=2,3,4 \\ 1-k \le j \le N, 1-k' \le j' \le N}} \left| \left\{ \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\gamma} - \left\langle B_{j,k}, B_{j',k'} \right\rangle_\gamma \right\} \right| \\ &\le c_0 h^{-1} \sup_{\gamma \in \Gamma} \max_{\substack{k,k'=2,3,4 \\ 1-k \le j \le N, 1-k' \le j' \le N}} \left| \left\{ \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\gamma} - \left\langle B_{j,k}, B_{j',k'} \right\rangle_\gamma \right\} \right|, \end{aligned}$$

which, together with Lemma A.5, imply (A.5). ■

For any fixed γ, one has $\mathbf{Y}^2 = \mathbf{g}_\gamma + \mathbf{g} - \mathbf{g}_\gamma + \mathbf{E} = \mathbf{g}_\gamma + \mathbf{E}_\gamma + \mathbf{E}$, where $\mathbf{E}^T = \left\{ g(U_t) (\xi_t^2 - 1) \right\}_{t=n'+1}^{n}, \mathbf{E}_\gamma = \left\{ g(U_t) - g_\gamma(U_{\gamma,t}) \right\}_{t=n'+1}^{n}$. Then one can break the cubic spline estimation error as

$$\hat{g}_\gamma(u) - g_\gamma(u) = \tilde{g}_\gamma(u) - g_\gamma(u) + \tilde{\varepsilon}_\gamma(u) + \hat{\varepsilon}_\gamma(u), \tag{A.6}$$

where

$$\tilde{g}_\gamma(u) = \{B_{j,4}(u)\}^T_{-3\le j\le N} \mathbf{V}^{-1}_{n,\gamma} \left\{\langle \mathbf{g}_\gamma, B_{j,4}\rangle_{n,\gamma}\right\}^N_{j=-3},$$

$$\tilde{\varepsilon}_\gamma(u) = \{B_{j,4}(u)\}^T_{-3\le j\le N} \mathbf{V}^{-1}_{n,\gamma} \left\{\langle \mathbf{E}_\gamma, B_{j,4}\rangle_{n,\gamma}\right\}^N_{j=-3},$$

$$\hat{\varepsilon}_\gamma(u) = \{B_{j,4}(u)\}^T_{-3\le j\le N} \mathbf{V}^{-1}_{n,\gamma} \left\{\langle \mathbf{E}, B_{j,4}\rangle_{n,\gamma}\right\}^N_{j=-3},$$

$$\mathbf{V}_{n,\gamma} = \left\{\langle B_{j,4}, B_{j',4}\rangle_{n,\gamma}\right\}^N_{j,j'=-3}, \mathbf{V}_\gamma = \left\{\langle B_{j,4}, B_{j',4}\rangle_\gamma\right\}^N_{j,j'=-3}. \quad \textbf{(A.7)}$$

The next proposition is used in proving Proposition 1.

PROPOSITION 2. *Under Assumptions (A1)–(A4), (A6), as* $n \to \infty$

$$\sup_{\gamma\in\Gamma}\sup_{u\in[0,1]} \left|\hat{g}_\gamma(u) - g_\gamma(u)\right| = O\left\{(nh)^{-1/2}\log n + h^4\right\}, a.s., \quad \textbf{(A.8)}$$

$$\sup_{\gamma\in\Gamma}\max_{n'+1\le t\le n} \left|\nabla\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right) - g_\gamma\left(U_{\gamma,t}\right)\right\}\right| = O\left\{n^{-1/2}h^{-3/2}\log n + h^3\right\}, a.s., \quad \textbf{(A.9)}$$

$$\sup_{\gamma\in\Gamma} \left|\nabla^2\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right) - g_\gamma\left(U_{\gamma,t}\right)\right\}\right| = O\left\{n^{-1/2}h^{-5/2}\log n + h^2\right\}, a.s.. \quad \textbf{(A.10)}$$

In order to prove the above proposition, we need several technical lemmas. The following is a special case in DeVore and Lorentz (1993, Thm. 13.4.3). We denote for square positive definite symmetric matrix $\mathbf{B} = \left(b_{i,j}\right)$, $\|\mathbf{B}\|_2 = \sup\{\|\mathbf{Bx}\|_2 / \|\mathbf{x}\|_2 : \mathbf{x}\neq\mathbf{0}\} = \sup\left\{\mathbf{x}^T\mathbf{B}\mathbf{x}_2/\|\mathbf{x}\|_2^2 : \mathbf{x}\neq\mathbf{0}\right\}$, which is the largest eigenvalue of $\mathbf{B}$, and $\|\mathbf{B}\|_\infty = \max_i \sum_j \left|b_{i,j}\right|$.

LEMMA A.7. *If a bi-infinite matrix with bandwidth r has a bounded inverse* $\mathbf{A}^{-1}$ *on* l_2 *(defined in DeVore and Lorentz, 1993, p. 19) and* $\kappa = \kappa(\mathbf{A}) = \|\mathbf{A}\|_2 \left\|\mathbf{A}^{-1}\right\|_2$ *is the condition number of* $\mathbf{A}$, *then* $\left\|\mathbf{A}^{-1}\right\|_\infty \le 2c_0(1-\upsilon)^{-1}$, *with* $c_0 = \upsilon^{-2r}\|\mathbf{A}\|_2$, $\upsilon = \left(\kappa^2-1\right)^{1/4r}\left(\kappa^2+1\right)^{-1/4r}$.

LEMMA A.8. *Under Assumptions (A3), (A4), and (A6), there exist constants* $0 < c_{\mathbf{V}} < C_{\mathbf{V}}$ *such that for any vector* $\mathbf{w}\neq\mathbf{0}$,

$$c_{\mathbf{V}} N^{-1}\|\mathbf{w}\|_2^2 \le \mathbf{w}^T\mathbf{V}_\gamma\mathbf{w} \le C_{\mathbf{V}} N^{-1}\|\mathbf{w}\|_2^2 \quad \textbf{(A.11)}$$

$$c_{\mathbf{V}} N^{-1}\|\mathbf{w}\|_2^2 \le \mathbf{w}^T\mathbf{V}_{n,\gamma}\mathbf{w} \le C_{\mathbf{V}} N^{-1}\|\mathbf{w}\|_2^2 \quad \textbf{(A.12)}$$

with matrices $\mathbf{V}_\gamma$ *and* $\mathbf{V}_{n,\gamma}$ *defined in (A.7). In addition, there exists a constant* $C > 0$ *such that*

$$\sup_{\gamma\in\Gamma}\left\|\mathbf{V}^{-1}_{n,\gamma}\right\|_\infty \le CN, a.s., \sup_{\gamma\in\Gamma}\left\|\mathbf{V}^{-1}_\gamma\right\|_\infty \le CN. \quad \textbf{(A.13)}$$

Proof. Let $\mathbf{w}$ be any $(N+4)$-vector and $\phi_{\mathbf{w}}(u) = \sum_{j=-3}^{N} w_J B_{j,4}(u)$, then $\mathbf{B}_\gamma \mathbf{w} = \left\{\phi_{\mathbf{w}}\left(U_{\gamma,n'}\right), \ldots, \phi_{\mathbf{w}}\left(U_{\gamma,n-1}\right)\right\}$ and A_n in (A.5) entails that

$$\|\phi_{\mathbf{w}}\|_{2,\gamma}^2 (1 - A_n) \leq \mathbf{w}^T \mathbf{V}_{n,\gamma} \mathbf{w} \leq \|\phi_{\mathbf{w}}\|_{2,\gamma}^2 (1 + A_n). \tag{A.14}$$

By DeVore and Lorentz (1993, Thm. 5.4.2) and Assumption (A4), one has

$$c_\varphi \frac{C}{N} \|\mathbf{w}\|_2^2 \leq \mathbf{w}^T \mathbf{V}_\gamma \mathbf{w} \leq C_\varphi \frac{C}{N} \|\mathbf{w}\|_2^2 \tag{A.15}$$

which, together with (A.14), yield

$$c_\varphi \frac{C}{N} \|\mathbf{w}\|_2^2 (1 - A_n) \leq \mathbf{w}^T \mathbf{V}_{n,\gamma} \mathbf{w} \leq C_\varphi \frac{C}{N} \|\mathbf{w}\|_2^2 (1 + A_n).$$

Then one has (A.11) and (A.12) by (A.15), (A.14), and (A.5). Next, denote by $\lambda_{\max}\left(\mathbf{V}_{n,\gamma}\right)$ and $\lambda_{\min}\left(\mathbf{V}_{n,\gamma}\right)$ the maximum and minimum eigenvalue of $\mathbf{V}_{n,\gamma}$, by the definition of the $\| \|_2$, one has $c_\varphi \frac{C}{N}\left(1 - A_n\right) \leq \dfrac{\mathbf{w}^T \mathbf{V}_{n,\gamma} \mathbf{w}}{\|\mathbf{w}\|_2^2} \leq \|\mathbf{V}_{n,\gamma}\|_2 = \lambda_{\max}\left(\mathbf{V}_{n,\gamma}\right)$, then there exists constant $c_{\mathbf{V}} > 0$, $c_V N^{-1} \leq \|\mathbf{V}_{n,\gamma}\|_2 = \lambda_{\max}\left(\mathbf{V}_{n,\gamma}\right)$, $a.s.$. Similarly, $\lambda_{\min}^{-1}\left(\mathbf{V}_{n,\gamma}\right) = \|\mathbf{V}_{n,\gamma}^{-1}\|_2 \leq C_V N^{-1}$, $a.s.$, thus $\kappa = \|\mathbf{V}_{n,\gamma}\|_2 \|\mathbf{V}_{n,\gamma}^{-1}\|_2 = \lambda_{\max}\left(\mathbf{V}_{n,\gamma}\right) \lambda_{\min}^{-1}\left(\mathbf{V}_{n,\gamma}\right) = C_V / c_V < \infty, a.s.$. One can also show that $\kappa \geq C > 1, a.s.$. Combining the above and Lemma A.7 with $\upsilon = \left(\kappa^2 - 1\right)^{1/16}\left(\kappa^2 + 1\right)^{-1/16}$, one gets $\|\mathbf{V}_{n,\gamma}^{-1}\|_\infty \leq 2\upsilon^{-8} N (1 - \upsilon)^{-1} = CN, a.s.$, which is part one of (A.13). Part two of (A.13) can be proved similarly. ∎

A.2. Proof of Proposition 2

We only illustrate the first element in the vector $\nabla\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right) - g_\gamma\left(U_{\gamma,t}\right)\right\}$ and matrix $\nabla^2 \hat{g}_\gamma\left(U_{\gamma,t}\right)$, i.e., $\partial\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right) - g_\gamma\left(U_{\gamma,t}\right)\right\}/\partial\theta$ and $\partial^2\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right) - g_\gamma\left(U_{\gamma,t}\right)\right\}/\partial\theta^2$. The proofs for other elements are similar.

LEMMA A.9. *Under Assumptions (A2)–(A4) and (A6), as $n \to \infty$*

$$\sup_{\gamma \in \Gamma} \left\| \left(\tilde{g}_\gamma - g_\gamma\right)^{(k)} \right\|_\infty \leq C \left\| m^{(4)} \right\|_\infty h^{4-k}, a.s., 0 \leq k \leq 2. \tag{A.16}$$

Proof. According to Huang (2003, Lemma 5.1), there exists an absolute constant $C > 0$, such that

$$\sup_{\gamma \in \Gamma} \|\tilde{g}_\gamma - g_\gamma\|_\infty \leq C \sup_{\gamma \in \Gamma} \inf_{\phi \in G^{(2)}} \|\phi - g_\gamma\|_\infty \leq C \left\| m^{(4)} \right\|_\infty h^4, a.s.,$$

which proves for the case $k = 0$. Applying Lemma A.2, one has for $0 \leq k \leq 2$

$$\sup_{\gamma \in \Gamma} \left\| \left(Q_T\left(g_\gamma\right) - g_\gamma\right)^{(k)} \right\|_\infty \leq C \sup_{\gamma \in \Gamma} \left\| g_\gamma^{(4)} \right\|_\infty h^{4-k} \leq C \left\| m^{(4)} \right\|_\infty h^{4-k}, a.s.. \tag{A.17}$$

So $\sup_{\gamma\in\Gamma}\left\| Q_T(g_\gamma)-\tilde g_\gamma\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^4$ a.s., which entails that

$$\sup_{\gamma\in\Gamma}\left\|\left(Q_T(g_\gamma)-\tilde g_\gamma\right)^{(k)}\right\|_\infty \le C\left\|m^{(4)}\right\|_\infty h^{4-k}, a.s., 0\le k\le 2. \quad \textbf{(A.18)}$$

Then the lemma is proved by combining (A.17) and (A.18). ∎

Denote $\mathbf{B}_\gamma = \left\{B_{j,4}(U_{\gamma,t})\right\}_{t=n'+1,j=-3}^{n,N}$ and

$$\mathbf{P}_\gamma = \mathbf{B}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T \quad \textbf{(A.19)}$$

as the projection matrix onto the cubic spline space spanned by $G^{(2)}$, and $\dot{\mathbf{B}}_\gamma = \left\{\partial B_{j,4}(U_{\gamma,t})/\partial\theta\right\}_{t=n'+1,j=-3}^{n,N}$, $\dot{\mathbf{P}}_\gamma = \partial\mathbf{P}_\gamma/\partial\theta$.

LEMMA A.10. *Under Assumptions (A4), one has*

$$\dot{\mathbf{B}}_\gamma = \Big[\left\{B_{j,3}(U_{\gamma,t})-B_{j+1,3}(U_{\gamma,t})\right\} f(X_{\gamma,t})h^{-1} \times \sum_{j=1}^{\infty}(j-1)\theta^{j-2}Y_{t-j}^2\left\{1+\eta 1_{(Y_{t-j}<0)}\right\}\Big]_{t=n'+1,j=-3}^{n,N}, \quad \textbf{(A.20)}$$

$$\dot{\mathbf{P}}_\gamma = \left(\mathbf{I}-\mathbf{P}_\gamma\right)\dot{\mathbf{B}}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T + \mathbf{B}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\dot{\mathbf{B}}_\gamma^T\left(\mathbf{I}-\mathbf{P}_\gamma\right). \quad \textbf{(A.21)}$$

Proof. Property (ii) in Lemma A.3 implies that

$$\begin{aligned}
\dot{\mathbf{B}}_\gamma &= \left\{\nabla B_{j,4}(U_{\gamma,t})\right\}_{t=n'+1,j=-3}^{n,N} = \left\{B'_{j,4}(U_{\gamma,t})\nabla U_{\gamma,t}\right\}_{t=n'+1,j=-3}^{n,N} \\
&= \Big[3\left\{\frac{B_{j,3}(U_{\gamma,t})}{t_{j+3}-t_j} - \frac{B_{j+1,3}(U_{\gamma,t})}{t_{j+4}-t_{j+1}}\right\} f(X_{\gamma,t}) \\
&\qquad \times \sum_{j=1}^{\infty}(j-1)\theta^{j-2}Y_{t-j}^2\left\{1+\eta 1_{(Y_{t-j}<0)}\right\}\Big]_{t=n'+1,j=-3}^{n,N} \\
&= \Big[\left\{B_{j,3}(U_{\gamma,t})-B_{j+1,3}(U_{\gamma,t})\right\} f(X_{\gamma,t})h^{-1} \\
&\qquad \times \sum_{j=1}^{\infty}(j-1)\theta^{j-2}Y_{t-j}^2\left\{1+\eta 1_{(Y_{t-j}<0)}\right\}\Big]_{t=n'+1,j=-3}^{n,N}.
\end{aligned}$$

Next, note that

$$\dot{\mathbf{P}}_\gamma = \dot{\mathbf{B}}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T + \mathbf{B}_\gamma\frac{\partial}{\partial\theta}\left\{\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\right\}\mathbf{B}_\gamma^T + \mathbf{B}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\dot{\mathbf{B}}_\gamma$$

and

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left\{\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\right\} &= -(\mathbf{B}_\gamma^T\mathbf{B}_\gamma)^{-1}\frac{\partial}{\partial\theta}\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1} \\
&= -\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\left(\dot{\mathbf{B}}_\gamma^T\mathbf{B}_\gamma + \mathbf{B}_\gamma^T\dot{\mathbf{B}}_\gamma\right)\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}. \quad \textbf{(A.22)}
\end{aligned}$$

Hence $\dot{\mathbf{P}}_\gamma$ is

$$\dot{\mathbf{B}}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \mathbf{B}_\gamma^T - \mathbf{B}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \mathbf{B}_\gamma^T \dot{\mathbf{B}}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \mathbf{B}_\gamma^T$$
$$-\mathbf{B}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \dot{\mathbf{B}}_\gamma^T \mathbf{B}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \mathbf{B}_\gamma^T + \mathbf{B}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \dot{\mathbf{B}}_\gamma^T$$
$$= \left(\mathbf{I} - \mathbf{P}_\gamma\right) \dot{\mathbf{B}}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \mathbf{B}_\gamma^T + \mathbf{B}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \dot{\mathbf{B}}_\gamma^T \left(\mathbf{I} - \mathbf{P}_\gamma\right). \quad \textbf{(A.23)}$$

■

LEMMA A.11. *Under Assumptions (A3), (A4), and (A6), as $n \to \infty$*

$$\sup_{\gamma \in \Gamma} \left\| (n'')^{-1} \mathbf{B}_\gamma^T \right\|_\infty \le Ch, \ \sup_{\gamma \in \Gamma} \left\| (n'')^{-1} \dot{\mathbf{B}}_\gamma^T \right\|_\infty \le C, a.s. \quad \textbf{(A.24)}$$

$$\sup_{\gamma \in \Gamma} \left\| \mathbf{P}_\gamma \right\|_\infty \le C, \ \sup_{\gamma \in \Gamma} \left\| \dot{\mathbf{P}}_\gamma \right\|_\infty \le Ch, \ \sup_{\gamma \in \Gamma} \left\| \frac{\partial}{\partial \theta} \left\{ \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma\right)^{-1} \right\} \right\|_\infty = O(N), a.s. \quad \textbf{(A.25)}$$

Proof. For any vector $\mathbf{a} \in R^{n''}$, one has

$$\left\| (n'')^{-1} \mathbf{B}_\gamma^T \mathbf{a} \right\|_\infty \le \|\mathbf{a}\|_\infty \max_{-3 \le j \le N} \left| (n'')^{-1} \sum_{t=n'+1}^{n} B_{j,4}\left(U_{\gamma,t}\right) \right| \le Ch \|\mathbf{a}\|_\infty, a.s.$$

and using equation (A.20), $\left\| (n'')^{-1} \dot{\mathbf{B}}_\gamma^T \mathbf{a} \right\|_\infty$ is bounded with probability 1 by

$$\|\mathbf{a}\|_\infty \max_{-3 \le j \le N} \left| (n''h)^{-1} \sum_{t=n'+1}^{n} \left\{ \left(B_{j,3} - B_{j+1,3}\right)\left(U_{\gamma,t}\right) \right\} f\left(X_{\gamma,t}\right) \right.$$
$$\left. \times \sum_{j=1}^{\infty} (j-1)\theta^{j-2} Y_{t-j}^2 \left\{ 1 + \eta 1_{(Y_{t-j} < 0)} \right\} \right| \le C \|\mathbf{a}\|_\infty.$$

Then one has (A.25) by (A.19), (A.13), (A.24), (A.23), and (A.22). Equations (A.20) and (A.21) are needed for proving the rest of the inequalities. ■

LEMMA A.12. *Under Assumptions (A2)–(A4) and (A6),*

$$\sup_{\gamma \in \Gamma} \left| \frac{\partial^k}{\partial \theta^k} \left\{ \tilde{g}_\gamma\left(U_{\gamma,t}\right) - g_\gamma\left(U_{\gamma,t}\right) \right\} \right| \le C \left\| m^{(4)} \right\|_\infty h^{4-k}, a.s., k = 1, 2. \quad \textbf{(A.26)}$$

Proof. According to the definition of $\tilde{g}_\gamma$, one has

$$\frac{\partial}{\partial \theta} \left[\left\{ Q_T\left(g_\gamma\right) - \tilde{g}_\gamma \right\} \left(U_{\gamma,t}\right) \right] = \frac{\partial}{\partial \theta} \mathbf{P}_\gamma \left[\left\{ Q_T\left(g_\gamma\right) - g_\gamma \right\} \left(U_{\gamma,t}\right) \right]$$
$$= \dot{\mathbf{P}}_\gamma \left[\left\{ Q_T\left(g_\gamma\right) - g_\gamma \right\} \left(U_{\gamma,t}\right) \right] + \mathbf{P}_\gamma \frac{\partial}{\partial \theta} \left[\left\{ Q_T\left(g_\gamma\right) - g_\gamma \right\} \left(U_{\gamma,t}\right) \right],$$

$$\frac{\partial}{\partial \theta} \left[\left\{ Q_T\left(g_\gamma\right) - g_\gamma \right\} \left(U_{\gamma,t}\right) \right]$$
$$= \left[\left\{ Q_T\left(\frac{\partial}{\partial \theta} g_\gamma\right) - \frac{\partial}{\partial \theta} g_\gamma \right\} \left(U_{\gamma,t}\right) \right] + \left[\left\{ Q_T\left(g_\gamma\right) - g_\gamma \right\}^{(1)} \left(U_{\gamma,t}\right) \right]$$
$$\times f\left(X_{\gamma,t}\right) h^{-1} \sum_{j=1}^{\infty} (j-1)\theta^{j-2} Y_{t-j}^2,$$

which yield (A.26) for $k = 1$ by (A.17), (A.25), and Lemma A.2. The proof for $k = 2$ is similar. ■

LEMMA A.13. *Under Assumptions (A2)–(A4) and (A6), as $n \to \infty$, one has with probability* 1

$$\sup_{\gamma\in\Gamma}\left\|\frac{\mathbf{B}_\gamma^T\mathbf{E}}{n''}\right\|_\infty = O\left(\frac{\log n}{\sqrt{nN}}\right), \qquad \sup_{\gamma\in\Gamma}\left\|\frac{\mathbf{B}_\gamma^T\mathbf{E}_\gamma}{n''}\right\|_\infty = O\left(\frac{\log n}{\sqrt{nN}}\right), \tag{A.27}$$

$$\sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\theta}\left(\frac{\mathbf{B}_\gamma^T\mathbf{E}}{n''}\right)\right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}}\right), \qquad \sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\theta}\left(\frac{\mathbf{B}_\gamma^T\mathbf{E}_\gamma}{n''}\right)\right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}}\right), \tag{A.28}$$

$$\sup_{\gamma\in\Gamma}\left\|\frac{\dot{\mathbf{B}}_\gamma^T\mathbf{E}}{n''}\right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}}\right), \qquad \sup_{\gamma\in\Gamma}\left\|\frac{\dot{\mathbf{B}}_\gamma^T\mathbf{E}_\gamma}{n''}\right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}}\right). \tag{A.29}$$

Proof. we prove only the first equation in (A.27) and the second equation of (A.28), other equations can be proved similarly. One has

$$\frac{\mathbf{B}_\gamma^T\mathbf{E}}{n''} = \left[(n'')^{-1}\sum\nolimits_{t=n'}^{n} B_{j,4}(U_{\gamma,t})\,g(U_t)\left(\xi_t^2-1\right)\right]_{j=-3}^{N}.$$

Denote $Z_t = g(U_t)\left(\xi_t^2-1\right) = Z_{t,1}^{D_n} + Z_{t,2}^{D_n} + Z_{t,3}^{D_n}$, where $D_n = n^\eta$ $(1/3 < \eta < 2/5)$,

$$Z_{t,1}^{D_n} = g(U_t)\left(\xi_t^2-1\right) I\left\{\left|g(U_t)\left(\xi_t^2-1\right)\right| > D_n\right\},$$
$$Z_{t,2}^{D_n} = g(U_t)\left(\xi_t^2-1\right) I\left\{\left|g(U_t)\left(\xi_t^2-1\right)\right| \le D_n\right\} - Z_{t,3}^{D_n},$$
$$Z_{t,3}^{D_n} = E\left[g(U_t)\left(\xi_t^2-1\right) I\left\{\left|g(U_t)\left(\xi_t^2-1\right)\right| \le D_n\right\}\right].$$

Note that the B-spline basis is bounded and $Eg(U_t)\left(\xi_t^2-1\right) = 0$, so

$$\left|Z_{t,3}^{D_n}\right| = \left|E\left[g(U_t)\left(\xi_t^2-1\right) I\left\{\left|g(U_t)\left(\xi_t^2-1\right)\right| > D_n\right\}\right]\right| \le E\left|g(U_t)\left(\xi_t^2-1\right)\right|^3 D_n^{-2}.$$

Then

$$\sup_{\gamma\in\Gamma}\left|(n'')^{-1}\sum\nolimits_{t=n'}^{n} B_{j,4}(U_{\gamma,t})\,Z_{t,3}^{D_n}\right| = O\left(D_n^{-2}\right) = o\left(n^{-2/3}\right).$$

One has $\sum_{n=n'+1}^{\infty} P\left\{\left|g(U_{n-1})\left(\xi_n^2-1\right)\right| > D_n\right\} \le \sum_{n=n'+1}^{\infty} D_n^{-3} < \infty$ according to the assumption that $E\left(\xi_t^6\right) = m_6 < +\infty$, and Borel–Cantelli lemma implies that the tail part

$$\sup_{\gamma\in\Gamma}\left|(n'')^{-1}\sum\nolimits_{t=n'+1}^{n} B_{j,4}(U_{\gamma,t})\,Z_{t,1}^{D_n}\right| = O\left(n^{-k}\right), \quad \text{for any } k > 0.$$

For the truncated part, similar to the proof of Lemma A.5, using Lemma A.1 and the discretization technique in Fan and Yao (2003, p. 266), one has

$$\sup_{\gamma\in\Gamma}\left|(n'')^{-1}\sum\nolimits_{t=n'+1}^{n} B_{j,4}(U_{\gamma,t})\,Z_{t,2}^{D_n}\right| = O\left(\log n/\sqrt{Nn}\right).$$

Therefore the first equation in (A.27) is established with probability 1. To prove the second equation of (A.28), notice that

$$\frac{\mathbf{B}_\gamma^T \mathbf{E}_\gamma}{n''} = \left[(n'')^{-1} \sum_{t=n'+1}^{n} B_{j,4}\left(U_{\gamma,t}\right) \left\{ g\left(U_t\right) - g_\gamma \left(U_{\gamma,t}\right) \right\} \right]_{j=-3}^{N},$$

$$\frac{\partial}{\partial\theta}\left(\frac{\mathbf{B}_\gamma^T \mathbf{E}_\gamma}{n''} \right) = \left[(n'')^{-1} \sum_{t=n'+1}^{n} \frac{\partial}{\partial\theta} \left[B_{j,4}\left(U_{\gamma,t}\right) \left\{ g\left(U_t\right) - g_\gamma \left(U_{\gamma,t}\right) \right\} \right] \right]_{j=-3}^{N}.$$

While $E\left[B_{j,4}\left(U_{\gamma,t}\right) \left\{ g\left(U_t\right) - g_\gamma \left(U_{\gamma,t}\right) \right\} \right] = 0, \quad -3 \le j \le N$ implies that

$$E\left\{ \frac{\partial}{\partial\theta} \left[B_{j,4}\left(U_{\gamma,t}\right) \left\{ g\left(U_t\right) - g_\gamma \left(U_{\gamma,t}\right) \right\} \right] \right\} = 0, \quad -3 \le j \le N, \ \gamma \in \Gamma,$$

which allows one to apply Lemma A.1 to obtain that with probability one

$$\sup_{\gamma \in \Gamma} \left\| \frac{\partial}{\partial\theta}\left(\frac{\mathbf{B}_\gamma^T \mathbf{E}_\gamma}{n''} \right) \right\|_\infty = O\left(\log n / \sqrt{nh} \right).$$

■

LEMMA A.14. *Under Assumptions (A2)–(A4) and (A6), as $n \to \infty$*

$$\sup_{\gamma \in \Gamma} \sup_{u \in [0,1]} \left| \hat{\varepsilon}_\gamma (u) \right| = O\left(\log n / \sqrt{nh} \right), a.s., \tag{A.30}$$

$$\sup_{\gamma \in \Gamma} \sup_{u \in [0,1]} \left| \tilde{\varepsilon}_\gamma (u) \right| = O\left(\log n / \sqrt{nh} \right), a.s.. \tag{A.31}$$

Proof. We only prove (A.30), the proof of (A.31) is similar. Denote $\hat{\mathbf{a}} = \left(\hat{a}_{-3}, \ldots, \hat{a}_N \right)^T = \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma \right)^{-1} \mathbf{B}_\gamma^T \mathbf{E} = \mathbf{V}_{n,\gamma}^{-1} \left\{ (n'')^{-1} \mathbf{B}_\gamma^T \mathbf{E} \right\}$, then $\hat{\varepsilon}_\gamma (u) = \sum_{j=-3}^{N} \hat{a}_j B_{j,4}(u)$.

$$\begin{aligned} \sup_{\gamma \in \Gamma} \sup_{u \in [0,1]} \left| \hat{\varepsilon}_\gamma (u) \right| \le \sup_{\gamma \in \Gamma} \left\| \hat{\mathbf{a}} \right\|_\infty &= \sup_{\gamma \in \Gamma} \left\| \mathbf{V}_{n,\gamma}^{-1} \left(n^{-1} \mathbf{B}_\gamma^T \mathbf{E} \right) \right\|_\infty \\ &\le CN \sup_{\gamma \in \Gamma} \left\| (n'')^{-1} \mathbf{B}_\gamma^T \mathbf{E} \right\|_\infty , a.s., \end{aligned}$$

where the last inequality follows from Lemmas A.8 and A.13. ■

LEMMA A.15. *Under Assumptions (A2)–(A4) and (A6), as $n \to \infty$*

$$\sup_{\gamma \in \Gamma} \max_{n'+1 \le t \le n} \left| \frac{\partial}{\partial\theta} \hat{\varepsilon}_\gamma \left(U_{\gamma,t}\right) \right| = O\left(n^{-1/2} N^{3/2} \log n \right), a.s., \tag{A.32}$$

$$\sup_{\gamma \in \Gamma} \max_{n'+1 \le t \le n} \left| \frac{\partial}{\partial\theta} \tilde{\varepsilon}_\gamma \left(U_{\gamma,t}\right) \right| = O\left(n^{-1/2} N^{3/2} \log n \right), a.s., \tag{A.33}$$

$$\sup_{\gamma \in \Gamma} \max_{n'+1 \le t \le n} \left| \frac{\partial^2}{\partial\theta^2} \hat{\varepsilon}_\gamma \left(U_{\gamma,t}\right) \right| = O\left(n^{-1/2} N^{5/2} \log n \right), a.s., \tag{A.34}$$

$$\sup_{\gamma \in \Gamma} \max_{n'+1 \le t \le n} \left| \frac{\partial^2}{\partial\theta^2} \tilde{\varepsilon}_\gamma \left(U_{\gamma,t}\right) \right| = O\left(n^{-1/2} N^{5/2} \log n \right), a.s.. \tag{A.35}$$

Proof. We only prove (A.32) and (A.33), the proofs of (A.34) and (A.35) are similar. One has

$$\left\{\frac{\partial}{\partial\theta}\hat{\varepsilon}_\gamma(U_{\gamma,t})\right\}_{t=n'+1}^{n} = (\mathbf{I}-\mathbf{P}_\gamma)\dot{\mathbf{B}}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T\mathbf{E}+\mathbf{B}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\dot{\mathbf{B}}_\gamma^T(\mathbf{I}-\mathbf{P}_\gamma)\mathbf{E}$$
$$= (\mathbf{I}-\mathbf{P}_\gamma)\dot{\mathbf{B}}_\gamma\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n}\right)^{-1}\frac{\mathbf{B}_\gamma^T\mathbf{E}}{n}+\mathbf{B}_\gamma\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n}\right)^{-1}\frac{\dot{\mathbf{B}}_\gamma^T(\mathbf{I}-\mathbf{P}_\gamma)\mathbf{E}}{n}.$$

According to (A.13), (A.24), (A.25), and (A.27), one has (A.32). To prove (A.33), note that

$$\left\{\frac{\partial}{\partial\theta}\tilde{\varepsilon}_\gamma(U_{\gamma,t})\right\}_{t=n'+1}^{n} = (\mathbf{I}-\mathbf{P}_\gamma)\dot{\mathbf{B}}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T\mathbf{E}_\gamma+\mathbf{B}_\gamma^T\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\dot{\mathbf{B}}_\gamma(\mathbf{I}-\mathbf{P}_\gamma)\mathbf{E}_\gamma$$
$$+\mathbf{B}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T\frac{\partial}{\partial\theta}\mathbf{E}_\gamma$$
$$= (\mathbf{I}-\mathbf{P}_\gamma)\dot{\mathbf{B}}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T\mathbf{E}_\gamma-\mathbf{B}_\gamma^T\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\dot{\mathbf{B}}_\gamma\mathbf{P}_\gamma\mathbf{E}_\gamma$$
$$+\mathbf{B}_\gamma^T\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\dot{\mathbf{B}}_\gamma\mathbf{E}_\gamma+\mathbf{B}_\gamma\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T\frac{\partial}{\partial\theta}\mathbf{E}_\gamma$$
$$= T_1+T_2$$

where

$$T_1 = \left\{(\mathbf{I}-\mathbf{P}_\gamma)\dot{\mathbf{B}}_\gamma-\mathbf{B}_\gamma^T\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\dot{\mathbf{B}}_\gamma\mathbf{B}_\gamma^T\right\}\left(\mathbf{B}_\gamma^T\mathbf{B}_\gamma\right)^{-1}\mathbf{B}_\gamma^T\mathbf{E}_\gamma$$
$$= \left\{(\mathbf{I}-\mathbf{P}_\gamma)\dot{\mathbf{B}}_\gamma-\mathbf{B}_\gamma^T\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n}\right)^{-1}\frac{\dot{\mathbf{B}}_\gamma\mathbf{B}_\gamma^T}{n}\right\}\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n}\right)^{-1}\frac{\mathbf{B}_\gamma^T\mathbf{E}_\gamma}{n}$$
$$T_2 = \mathbf{B}_\gamma\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n}\right)^{-1}\frac{\partial}{\partial\theta}\left(\frac{\mathbf{B}_\gamma^T\mathbf{E}_\gamma}{n}\right).$$

By (A.13), (A.24), (A.25), (A.27), and (A.28), one has $\sup_{\gamma\in\Gamma}\|T_1\|_\infty = O\left(n^{-1/2}N^{3/2}\log n\right)$ and $\sup_{\gamma\in\Gamma}\|T_2\|_\infty = O\left(n^{-1/2}N^{3/2}\log n\right), a.s.$ which leads to (A.33). ■

Proof of Proposition 2. According to (A.6), one has (A.8) by (A.16), (A.30), and (A.31). Similarly, one has

$$\frac{\partial}{\partial\theta}\left\{\hat{g}_\gamma(U_{\gamma,t})-g_\gamma(U_{\gamma,t})\right\} = \frac{\partial}{\partial\theta}\left\{\tilde{g}_\gamma(U_{\gamma,t})-g_\gamma(U_{\gamma,t})\right\}+\frac{\partial}{\partial\theta}\tilde{\varepsilon}_\gamma(U_{\gamma,t})+\frac{\partial}{\partial\theta}\hat{\varepsilon}_\gamma(U_{\gamma,t}).$$

Thus one has (A.9) by (A.16), (A.32), and (A.33). The proof of (A.10) is similar. ■

A.3. Proof of Proposition 1

LEMMA A.16. *Under Assumptions (A1)–(A6), as $n\to\infty$,* $\sup_{\gamma\in\Gamma}\left|\hat{R}(\gamma)-R(\gamma)\right| = o(1), a.s.$.

Proof.

$$\hat{R}(\gamma) = \frac{1}{n''}\sum_{t=n'+1}^{n} \left\{Y_t^2 - \hat{g}_\gamma(U_{\gamma,t})\right\}^2$$
$$= \frac{1}{n''}\sum_{t=n'+1}^{n} \left\{g(U_t) + g(U_t)\left(\xi_t^2 - 1\right) - g_\gamma(U_{\gamma,t}) + g_\gamma(U_{\gamma,t}) - \hat{g}_\gamma(U_{\gamma,t})\right\}^2$$
$$= \frac{1}{n''}\sum_{t=n'+1}^{n} \left\{g_\gamma(U_{\gamma,t}) - \hat{g}_\gamma(U_{\gamma,t})\right\}^2 + \frac{1}{n''}\sum_{t=n'+1}^{n} \left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}^2$$
$$+ \frac{2}{n''}\sum_{t=n'+1}^{n} \left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}\left\{g(U_t)\left(\xi_t^2 - 1\right)\right\}$$
$$+ \frac{1}{n''}\sum_{t=n'+1}^{n} \left\{g(U_t)\left(\xi_t^2 - 1\right)\right\}^2$$
$$+ \frac{2}{n''}\sum_{t=n'+1}^{n} \left\{g_\gamma(U_{\gamma,t}) - \hat{g}_\gamma(U_{\gamma,t})\right\}\left\{g(U_t) - g_\gamma(U_{\gamma,t}) + g(U_t)\left(\xi_t^2 - 1\right)\right\},$$

$$R(\gamma) = E\left\{Y_t^2 - g_\gamma(U_{\gamma,t})\right\}^2$$
$$= E\left\{g(U_t) + g(U_t)\left(\xi_t^2 - 1\right) - g_\gamma(U_{\gamma,t})\right\}^2$$
$$= E\left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}^2 + E\left\{g(U_t)\left(\xi_t^2 - 1\right)\right\}^2.$$

Hence

$$\sup_{\gamma\in\Gamma}\left|\hat{R}(\gamma) - R(\gamma)\right| \le I_1 + I_2 + I_3 + I_4,$$

where

$$I_1 = \sup_{\gamma\in\Gamma}\left|\frac{1}{n-n'}\sum_{t=n'+1}^{n}\left\{g_\gamma(U_{\gamma,t}) - \hat{g}_\gamma(U_{\gamma,t})\right\}^2\right|,$$

$$I_2 = \sup_{\gamma\in\Gamma}\left|\frac{2}{n''}\sum_{t=n'+1}^{n}\left\{g_\gamma(U_{\gamma,t}) - \hat{g}_\gamma(U_{\gamma,t})\right\}\left\{g(U_t) - g_\gamma(U_{\gamma,t}) + g(U_t)\left(\xi_t^2 - 1\right)\right\}\right|,$$

$$I_3 = \sup_{\gamma\in\Gamma}\left|\frac{1}{n''}\sum_{t=n'+1}^{n}\left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}^2 - E\left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}^2\right|,$$

$$I_4 = \sup_{\gamma\in\Gamma}\left\{\left|\frac{1}{n''}\sum_{t=n'+1}^{n}\left\{g(U_t)\left(\xi_t^2 - 1\right)\right\}^2 - (m_4 - 1)Eg^2(U_t)\right| + \left|\frac{2}{n''}\sum_{t=n'+1}^{n}\left\{g(U_t) - g_\gamma(U_{\gamma,t})\right\}\left\{g(U_t)\left(\xi_t^2 - 1\right)\right\}\right|\right\}.$$

According to Lemma A.1, one has $I_3+I_4=o(1), a.s$, and (A.8) entails that $I_1=O\left\{\left(n^{-1/2}\log n\right)^2+\left(h^4\right)^2\right\}, a.s..$ One also has

$$I_2 \le O\left(n^{-1/2}\log n+h^4\right)\sup_{\gamma\in\Gamma}\frac{2}{n''}\sum_{t=n'+1}^{n}\left|\left\{g(U_t)-g_\gamma\left(U_{\gamma,t}\right)+g(U_t)\left(\xi_t^2-1\right)\right\}\right|,$$

which is $O\left(n^{-1/2}\log n+h^4\right), a.s..$ The lemma is proved by combining I_1, I_2, I_3, I_4. ■

LEMMA A.17. *Under Assumptions (A1)–(A6), as $n\to\infty$, one has for $k=1,2$*

$$\sup_{\gamma\in\Gamma}\left|\nabla^{(k)}\hat{R}(\gamma)-\nabla^{(k)}R(\gamma)\right|=O\left(n^{-1/2}h^{-1/2-k}\log n+h^{4-k}\right), a.s.. \qquad \textbf{(A.36)}$$

Proof. We only show the proof for the case $\frac{\partial}{\partial\theta}\hat{R}(\gamma)-\frac{\partial}{\partial\theta}R(\gamma)$ and the proof for other elements is similar.

$$\frac{1}{2}\frac{\partial}{\partial\theta}\hat{R}(\gamma)=\frac{1}{n''}\sum_{t=n'+1}^{n}\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right)-Y_t^2\right\}\frac{\partial}{\partial\theta}\hat{g}_\gamma\left(U_{\gamma,t}\right),$$

$$\frac{1}{2}\frac{\partial}{\partial\theta}R(\gamma)=E\left[\left\{g_\gamma\left(U_{\gamma,t}\right)-Y_t^2\right\}\frac{\partial}{\partial\theta}g_\gamma\left(U_{\gamma,t}\right)\right],$$

then

$$\frac{1}{2}\frac{\partial}{\partial\theta}\left(\hat{R}(\gamma)-R(\gamma)\right)=\frac{1}{n''}\sum_{t=n'+1}^{n}\xi_{\gamma,t}+J_{\gamma,1}+J_{\gamma,2}+J_{\gamma,3},$$

where $\xi_{\gamma,t}$ is defined in (A.38) and $E\xi_{\gamma,t}=\mathbf{0}$, and where

$$J_{\gamma,1}=\frac{1}{n''}\sum_{t=n'+1}^{n}\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right)-g_\gamma\left(U_{\gamma,t}\right)\right\}\frac{\partial}{\partial\theta}\left(\hat{g}_\gamma-g_\gamma\right)\left(U_{\gamma,t}\right),$$

$$J_{\gamma,2}=\frac{1}{n''}\sum_{t=n'+1}^{n}\left\{g_\gamma\left(U_{\gamma,t}\right)-Y_t^2\right\}\frac{\partial}{\partial\theta}\left(\hat{g}_\gamma-g_\gamma\right)\left(U_{\gamma,t}\right),$$

$$J_{\gamma,3}=\frac{1}{n''}\sum_{t=n'+1}^{n}\left\{\hat{g}_\gamma\left(U_{\gamma,t}\right)-g_\gamma\left(U_{\gamma,t}\right)\right\}\frac{\partial}{\partial\theta}g_\gamma\left(U_{\gamma,t}\right).$$

By Lemma A.1, $\sup_{\gamma\in\Gamma}\left|\left(n''\right)^{-1}\sum_{t=n'+1}^{n}\xi_{\gamma,t}\right|=O\left(n^{-1/2}\log n\right)$ $a.s..$ Meanwhile, (A.8) and (A.9) imply that $\sup_{\gamma\in\Gamma}\left|J_{\gamma,1}\right|=O\left(n^{-1}h^{-2}\log^2 n+h^7\right)$ $a.s..$ Note that

$$J_{\gamma,2}=\frac{1}{n''}\sum\nolimits_{t=n'+1}^{n}\left\{g_\gamma\left(U_{\gamma,t}\right)-Y_t^2\right\}\frac{\partial}{\partial\theta}\left(\tilde{g}_\gamma-g_\gamma\right)\left(U_{\gamma,t}\right)$$
$$-\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\frac{\partial}{\partial\theta}\left\{\mathbf{P}_\gamma\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right\}.$$

One has

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{n''}\sum\nolimits_{t=n'+1}^{n}\left\{g_\gamma\left(U_{\gamma,t}\right)-Y_t^2\right\}\frac{\partial}{\partial\theta}\left(\tilde{g}_\gamma-g_\gamma\right)\left(U_{\gamma,t}\right)\right|=O\left(h^3\right)\ a.s.$$

according to (A.16). Next

$$\begin{aligned}
&\left|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\frac{\partial}{\partial\theta}\left\{\mathbf{P}_\gamma\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right\}\right|\\
&\quad=\left|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\frac{\partial}{\partial\theta}\left\{\mathbf{B}_\gamma\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n''}\right)^{-1}\frac{\mathbf{B}_\gamma^T}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right\}\right|\\
&\quad\le\left|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\dot{\mathbf{B}}_\gamma\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n''}\right)^{-1}\frac{\mathbf{B}_\gamma^T}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right|\\
&\qquad+\left|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\mathbf{B}_\gamma\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n''}\right)^{-1}\frac{\partial}{\partial\theta}\frac{\mathbf{B}_\gamma^T}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right|\\
&\qquad+\left|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\mathbf{B}_\gamma\frac{\partial}{\partial\theta}\left\{\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n''}\right)^{-1}\right\}\frac{\mathbf{B}_\gamma^T}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right|.
\end{aligned}$$

Thus

$$\begin{aligned}
&\sup_{\gamma\in\Gamma}\left|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\frac{\partial}{\partial\theta}\left\{\mathbf{P}_\gamma\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right\}\right|\\
&\quad\le O(N)\times\sup_{\gamma\in\Gamma}\left\{\left\|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\dot{\mathbf{B}}_\gamma\right\|_\infty\left\|\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n''}\right)^{-1}\right\|_\infty\left\|\frac{\mathbf{B}_\gamma^T}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right\|_\infty\right\}\\
&\qquad+O(N)\times\sup_{\gamma\in\Gamma}\left\{\left\|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\mathbf{B}_\gamma\right\|_\infty\left\|\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n''}\right)^{-1}\right\|_\infty\left\|\frac{\partial}{\partial\theta}\frac{\mathbf{B}_\gamma^T}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right\|_\infty\right\}\\
&\qquad+O(N)\times\sup_{\gamma\in\Gamma}\left\{\left\|\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\mathbf{B}_\gamma\right\|_\infty\left\|\frac{\partial}{\partial\theta}\left\{\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n''}\right)^{-1}\right\}\right\|_\infty\left\|\frac{\mathbf{B}_\gamma^T}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)\right\|_\infty\right\}\\
&\quad=O(N)\times O\left(\log n/\sqrt{nh}\right)\times O(N)\times O\left(\log n/\sqrt{nN}\right)=O\left(n^{-1}N^2\log^2 n\right)\ a.s.
\end{aligned}$$

according to (A.27), (A.28), (A.29), (A.13), and (A.25). So $\sup_{\gamma\in\Gamma}\left|J_{\gamma,2}\right|=O\left(n^{-1}N^2\log^2 n+h^3\right), a.s.$. Similarly, one can write

$$\begin{aligned}
J_{\gamma,3}=&\frac{1}{n''}\sum_{t=n'+1}^{n}\left\{\tilde{g}_\gamma\left(U_{\gamma,t}\right)-g_\gamma\left(U_{\gamma,t}\right)\right\}\frac{\partial}{\partial\theta}g_\gamma\left(U_{\gamma,t}\right)\\
&+\frac{1}{n''}\left(\mathbf{E}_\gamma+\mathbf{E}\right)^T\mathbf{B}_\gamma\left(\frac{\mathbf{B}_\gamma^T\mathbf{B}_\gamma}{n}\right)^{-1}\frac{\mathbf{B}_\gamma^T}{n}\frac{\partial}{\partial\theta}g_\gamma.
\end{aligned}$$

and has

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{n''} \sum_{t=n'+1}^{n} \left\{ \tilde{g}_\gamma (U_{\gamma,t}) - g_\gamma (U_{\gamma,t}) \right\} \frac{\partial}{\partial \theta} g_\gamma (U_{\gamma,t}) \right| = O\left(h^4\right) \ a.s.,$$

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{n''} \left(\mathbf{E}_\gamma + \mathbf{E}\right)^T \mathbf{B}_\gamma \left(\frac{\mathbf{B}_\gamma^T \mathbf{B}_\gamma}{n} \right)^{-1} \frac{\mathbf{B}_\gamma^T}{n} \frac{\partial}{\partial \theta} g_\gamma \right|$$

$$= O\left\{ \frac{\log n}{\sqrt{nN}} \times N \times h \right\} = O\left\{ \frac{\log n}{\sqrt{nN}} \right\} \ a.s..$$

Thus (A.36) is proved for $k = 1$. One can prove that for the term $\xi_{\gamma,t}$ defined in (A.38), with probability 1

$$\sup_{\gamma \in \Gamma} \left| \frac{1}{2} \frac{\partial}{\partial \theta} \left\{ \hat{R}(\gamma) - R(\gamma) \right\} - \frac{1}{n''} \sum_{t=n'+1}^{n} \xi_{\gamma,t} \right| = o\left(n^{-1/2}\right). \qquad \textbf{(A.37)}$$

The proof of (A.36) for $k = 2$ follows from (A.8), (A.9), and (A.10), since

$$\frac{1}{2} \nabla^2 \hat{R}(\gamma) = \frac{1}{n''} \sum_{t=n'+1}^{n} \left[\left\{ \hat{g}_\gamma (U_{\gamma,t}) - Y_t^2 \right\} \nabla^2 \hat{g}_\gamma (U_{\gamma,t}) + \nabla \hat{g}_\gamma (U_{\gamma,t}) \nabla \hat{g}_\gamma (U_{\gamma,t}) \right],$$

$$\frac{1}{2} \nabla^2 R(\gamma) = E\left[\left\{ g_\gamma (U_{\gamma,t}) - Y_t^2 \right\} \nabla^2 g_\gamma (U_{\gamma,t}) + \nabla g_\gamma (U_{\gamma,t}) \nabla g_\gamma (U_{\gamma,t}) \right]. \qquad \blacksquare$$

Proof of Proposition 1. It follows from Lemmas A.16 and A.17. ■

Proof of Theorem 1. According to Proposition 1, one has $\sup_{\gamma \in \Gamma} \left| \hat{R}(\gamma) - R(\gamma) \right| \to 0, a.s.$ Thus there exists an integer n_0, such that $\hat{R}(\gamma_0) - R(\gamma_0) < \delta/2$ when $n > n_0$. Notice that $\hat{\gamma}$ is the minimizer of $\hat{R}(\gamma_0)$, so $\hat{R}(\hat{\gamma}) - R(\gamma_0) < \delta/2$. There also exists an integer n_1, such that $R(\hat{\gamma}) - \hat{R}(\hat{\gamma}) < \delta/2$ when $n > n_1$. Thus, when $n > \max(n_0, n_1)$,

$$R(\hat{\gamma}) - R(\gamma_0) = R(\hat{\gamma}) - \hat{R}(\hat{\gamma}) + \hat{R}(\hat{\gamma}) - R(\gamma_0) < \delta.$$

According to Assumption (A5), $R(\gamma)$ is locally convex at γ_0, so for any $\varepsilon > 0$, if $R(\hat{\gamma}) - R(\gamma_0) < \delta$, then $\left|\hat{\gamma} - \gamma_0\right| < \varepsilon$ for n large enough, which has proved the theorem. ■

Proof of Theorem 2. Denote $\hat{S}(\gamma) = \nabla \hat{R}(\gamma)$ and

$$\xi_{\gamma,t} = \left\{ g_\gamma (U_{\gamma,t}) - Y_t^2 \right\} \nabla g_\gamma (U_{\gamma,t}) - E\left[\left\{ g_\gamma (U_{\gamma,t}) - Y_t^2 \right\} \nabla g_\gamma (U_{\gamma,t}) \right], \qquad \textbf{(A.38)}$$

then because $\nabla R(\gamma_0) = \mathbf{0}$, one has

$$\left| \hat{S}(\gamma_0) - \frac{2}{n''} \sum\nolimits_{t=n'+1}^{n} \xi_{\gamma_0,t} \right| = o\left(n^{-1/2}\right), a.s.. \qquad \textbf{(A.39)}$$

according to (A.37). For some γ_1, γ_2 between $\hat{\gamma}$ and γ_0,

$$\hat{S}(\hat{\gamma}) - \hat{S}(\gamma_0) = \begin{pmatrix} \left(\partial^2/\partial\theta^2\right) \hat{R}(\gamma_1) & \left(\partial^2/\partial\theta\partial\eta\right) \hat{R}(\gamma_1) \\ \left(\partial^2/\partial\theta\partial\eta\right) \hat{R}(\gamma_2) & \left(\partial^2/\partial\eta^2\right) \hat{R}(\gamma_2) \end{pmatrix} (\hat{\gamma} - \gamma_0)$$

$$= A\left(\hat{\gamma} - \gamma_0\right),$$

and $\hat{S}(\hat{\gamma}) = 0$ because $\hat{R}(\gamma)$ attains its minimum at $\hat{\gamma}$. Thus, we have

$$\hat{\gamma} - \gamma_0 = -A^{-1}\hat{S}(\gamma_0).$$

According to Theorem 1 and Proposition 1, we have $A^{-1} \to \left\{\nabla^2 R(\gamma_0)\right\}^{-1} a.s.$, where $\nabla^2 R(\gamma_0)$ is given in (7). According to (A.39), one has $\sqrt{n''}\hat{S}(\gamma_0) \to_d N\{\mathbf{0}, \Psi(\gamma_0)\}$ by the Central Limit Theorem for strongly mixing processes (Bosq, 1998, Thm. 1.7), where $\Psi(\gamma_0)$ is given in (6). Then Theorem 2 is proved by formula (5) and Slutsky's theorem. ∎