Statistica Sinica **22** (2012), 249-270
doi:http://dx.doi.org/10.5705/ss.2009.285

EFFICIENT SEMIPARAMETRIC GARCH MODELING OF FINANCIAL VOLATILITY

Li Wang[1], Cong Feng[1], Qiongxia Song[2] and Lijian Yang[3,4]

[1] *University of Georgia,* [2] *University of Texas at Dallas,*
[3] *Soochow University and* [4] *Michigan State University*

Abstract: We consider a class of semiparametric GARCH models with additive autoregressive components linked together by a dynamic coefficient. We propose estimators for the additive components and the dynamic coefficient based on spline smoothing. The estimation procedure involves only a small number of least squares operations, thus it is computationally efficient. Under regularity conditions, the proposed estimator of the parameter is root-n consistent and asymptotically normal. A simultaneous confidence band for the nonparametric component is proposed by an efficient one-step spline backfitting. The performance of our method is evaluated by various simulated processes and a financial return series. For the empirical financial return series, we find further statistical evidence of the asymmetric news impact function.

Key words and phrases: B-spline, confidence band, knots, news impact curve, volatility.

1. Introduction

Forecasting financial market volatility is important in applications such as portfolio selection, asset management, pricing of primary and derivative assets. Consider a time series $\{Y_t\}_{t=1}^{\infty}$ of the form $Y_t = \sigma_t \xi_t$, where the $\{\xi_t\}_{t=1}^{\infty}$'s are i.i.d with mean 0 and variance 1, and $\left\{\sigma_t^2\right\}_{t=1}^{\infty}$ denotes the conditional volatility series. Engle (1982) introduced autoregressive heteroskedastic (ARCH) models for conditional volatility as a quadratic function of past observations. For example, an ARCH model of order q is defined as

$$\sigma_t^2 = \gamma + \alpha_1 Y_{t-1}^2 + \cdots + \alpha_q Y_{t-q}^2, \quad \gamma > 0,\ \alpha_i \geq 0,\ i = 1, \ldots, q.$$

Research on financial volatility models has grown tremendously since then, for example, the generalized autoregressive conditional heteroscedasticity (GARCH) models. The most popular version of the GARCH models is the GARCH$(1,1)$ model of Bollerslev (1986):

$$\sigma_t^2 = \gamma_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \quad \gamma_0 > 0,\ \alpha_0,\ \beta_0 \geq 0,$$

or equivalently $\sigma_t^2 = \beta_0 \sigma_{t-1}^2 + m_0(Y_{t-1})$, where $m_0(y) \equiv \alpha_0 y^2 + \gamma$ is the "news impact curve".

The quadratic form of the function $m_0(\cdot)$ had been questioned by many. For example, Glosten, Jaganathan, and Runkle (1993) proposed the GJR model

$$\sigma_t^2 = \gamma_0 + \alpha_0 Y_{t-1}^2 + \delta_0 Y_{t-1}^2 I(Y_{t-1} < 0) + \beta_0 \sigma_{t-1}^2$$

with $m_0(y) \equiv \gamma + \alpha y^2 + \delta y^2 I\,(y < 0)$, allowing different "leverages" of good and bad news on m_0. For this reason, recent studies have introduced non/semiparametric (G)ARCH models to increase the flexibility of the class of models; see for example, Pagan and Schwert (1990), Engle and Ng (1993), Masry and Tjøstheim (1995), Härdle and Tsybakov (1997), Hafner (1998), Härdle, Tsybakov, and Yang (1998), Bühlmann and McNeil (2002), Linton and Mammen (2005), and Yang (2006). These models generalize and outperform the parametric GARCH models when applied to data with many lagged variables. However, smoothing high dimensional and strongly correlated time series data still presents great challenges in both computation and theory.

As an alternative, additive models (Stone (1985)) overcome these difficulties while keeping the flexibility of the models. Yang, Härdle, and Nielsen (1999) analyzed a multiplicative form of volatility using nonparametric smoothing. Carroll, Härdle, and Mammen (2002) and Yang (2002) proposed a truncated version of the nonparametric GARCH model with a finite number of lags J:

$$\sigma_t^2 = \sum_{j=1}^{J} \beta_0^{j-1} m_0(Y_{t-j}), \quad \beta_0 \in [\beta_1, \beta_2]\,. \tag{1.1}$$

However, for small J, this may not capture the persistence of volatility for many time series; see Linton and Mammen (2005) and Yang (2006).

In this paper, we re-examine model (1.1) based on a data-driven lag selection procedure. Most of the existing methods rely on marginal integration kernel smoothing (Linton and Nielsen (1995)) or iterative approaches such as backfitting (Hastie and Tibshirani (1990)). The marginal integration can be computationally expensive if the selected number of lags J or sample size n is large, and it requires $O(n^3)$ operations (Hengartner and Sperlich (2005)). Moreover, n is required to be larger than $10,000$ for convergence when smoothing 10-dimensional data, so it is not routinely used in practice despite good theoretical properties. Widely used R/Splus packages `gam` and `mgcv`, based on backfitting with splines, provide convenient implementation in practice but lack theoretical justifications except for some special cases in Opsomer and Ruppert (1997).

Our goal is to develop a simple but flexible semiparametric method with a well-justified theory and a fast algorithm to implement the method in practice.

This is done by approximating the nonparametric components with polynomial splines. The use of spline smoothing goes back to Stone (1985), who first obtained the rate of convergence of the spline estimates for generalized additive models. In volatility studies, Engle and Ng (1993) employed linear spline smoothing to estimate the news impact function, without pursuing asymptotic results.

Our approach allows for formal derivation of the asymptotic properties of the proposed estimators. We establish the $\sqrt{n}$-consistency and the asymptotic normality for the parameter estimator, and L^2 convergence rate for the functional component. To examine the validity of certain forms of the volatility models, we provide a simultaneous confidence band for the news impact curve using the one-step spline-backfitted spline estimator in Song and Yang (2010).

The rest of the paper is organized as follows. Section 2 gives details of the model specification, proposed methods of estimation, and presents the asymptotic results. In addition, we discuss some alternative methods and the practical issue of lag selection. In Section 3 we describe a spline confidence band for the news impact curve. In Section 4 we report our findings in an extensive simulation study. An application to a financial return data is given in Section 5. Most of the technical proofs are contained in the Appendix.

2. The Method

2.1. Semiparametric GARCH models with additive autoregressive structure

Consider a stationary time series $\{Y_t\}_{t=1}^{T}$, with $Y_t = \sigma_t \xi_t$, $t = 1, \ldots, T$. We rewrite model (1.1) as the additive autoregressive model

$$Y_t^2 = c + \sum_{j=1}^{J} m_j(Y_{t-j}) + \epsilon_t, \quad \epsilon_t = \sigma_t^2 \left(\xi_t^2 - 1\right), \tag{2.1}$$

where the component functions $m_1(\cdot), \ldots, m_J(\cdot)$ are linked by a scalar parameter β_0 such that $m_j(y) = \beta_0^{j-1} m_1(y)$ for $j \geq 2$. Define the least squares risk function $R(\beta)$ over $[\beta_1, \beta_2]$ as

$$R(\beta) = E\Bigg[\sum_{j=1}^{J} \left\{m_j(Y_t) - \beta^{j-1} m_1(Y_t)\right\}^2\Bigg]. \tag{2.2}$$

Since $R(\beta) = \sum_{j=1}^{J}\{(\beta_0^{j-1} - \beta^{j-1})^2\}E\{m_1(Y_t)^2\}$ is a convex function with respect to β, β_0 is the unique minimizer of $R(\beta)$ over $[\beta_1, \beta_2]$. For identifiability, the component functions in (2.1) satisfy $E\left\{m_j\left(Y_t\right)\right\} = 0$, $j = 1, \ldots, J$.

Our intent is to estimate the news impact function m_1 and the dynamic coefficient parameter β_0. To reach this goal, first we employ the polynomial spline smoothing to obtain the estimates $\hat{m}_j(\cdot)$ of the additive components $m_j(\cdot)$ without taking into account the parametric link of the components, then we estimate the dynamic coefficient β_0 by using the link restriction between the additive components $\hat{m}_j(\cdot)$ $(j = 1, \ldots, J)$. For simplicity of notation, call the above approach the spline additive GARCH (GARCH-ADD) approach.

We only consider the estimation of $m_j(\cdot)$ based on all bounded measurable function on compact interval $[a, b]$, where a, b are some fixed constants. When facing data, one can use fixed truncation to satisfy this condition. Let $\mathcal{S}_n$ be the space of polynomial splines on $[a, b]$ of degree $p \geq 1$. We introduce a knot sequence with N interior knots,

$$u_{-p} = \cdots = u_{-1} = u_0 = a < u_1 < \cdots < u_N < b = u_{N+1} = \cdots = u_{N+p+1},$$

where $N \equiv N_n$ increases when sample size n increases, with precise order as given in Assumption (A5). The spline of degree p for the jth variable is denoted as $\{b_{j,k}\}_{k=-p}^{N}$ (de Boor (2001)). Then $\mathcal{S}_n$ consists of functions $g(\cdot)$ satisfying (i) $g(\cdot)$ is a polynomial of degree p on each of the subintervals $I_k = [u_k, u_{k+1})$, $k = 0, \ldots, N-1$, $I_N = [u_N, b]$; (ii) for $p \geq 2$, $g(\cdot)$ is $p-1$ time continuously differentiable on $[a, b]$.

Equally-spaced knots are used here for simplicity, while adaptively choosing the locations of the knots could have been done for data analysis. Let $h = (b-a)/(N+1)$ be the distance between neighboring knots. Take the space $G = G[a, b]$ of additive splines as the linear space spanned by the basis $\{1, b_{j,k}(y_j), j = 1, \ldots, J, k = -p, \ldots, N\}$. Let $(\hat{\lambda}'_0, \hat{\lambda}'_{1,-p}, \ldots, \hat{\lambda}'_{J,N})^{\mathrm{T}}$ be the solutions of the least squares problem

$$\left(\hat{\lambda}'_0, \hat{\lambda}'_{1,-p}, \ldots, \hat{\lambda}'_{J,N}\right)^{\mathrm{T}} = \underset{R^{1+J(N+p)}}{\operatorname{argmin}} \sum_{t=J+1}^{T} \left\{ Y_t^2 - \lambda_0 - \sum_{j=1}^{J} \sum_{k=-p}^{N} \lambda_{j,k} b_{j,k}\left(Y_{t-j}\right) \right\}^2.$$

Let $n = T - J$. Let $\hat{c} = n^{-1} \sum_{t=J+1}^{T} Y_t^2$, which is a $\sqrt{n}$-consistent estimator of c by the Central Limit Theorem. The centered spline estimator of each component function is

$$\hat{m}_j\left(y\right) = \sum_{k=-p}^{N} \hat{\lambda}_{j,k} b_{j,k}\left(y\right) - \frac{1}{n} \sum_{t=J+1}^{T} \sum_{k=-p}^{N} \hat{\lambda}_{j,k} b_{j,k}\left(Y_{t-j}\right), \quad 1 \leq j \leq J. \tag{2.3}$$

To estimate the parameter β_0, we regress $\{\hat{m}_2(Y_t)\}_{t=J+1}^{T}$ on $\{\hat{m}_1(Y_t)\}_{t=J+1}^{T}$ and solve the least squares $\sum_{t=J+1}^{T} \{\hat{m}_2(Y_t) - \beta \hat{m}_1(Y_t)\}^2$. The performance is improved by averaging over the deviation squares from all the components, so we define the sample least squares criterion

$$\hat{R}\left(\beta\right) = \frac{1}{n} \sum_{t=J+1}^{T} \sum_{j=1}^{J} \left\{\hat{m}_j(Y_t) - \beta^{j-1} \hat{m}_1(Y_t)\right\}^2, \tag{2.4}$$

and the minimizer of (2.4) $\hat{\beta}$ is the GARCH-ADD estimator of the dynamic coefficient.

2.2. Asymptotic properties of the GARCH-ADD estimators

For our theoretical results, we enforce the following technical assumptions.

(A1) *The data-generating process* $\{Y_t, t > 0\}$ *is strictly stationary and* α*-mixing with exponentially decaying mixing coefficients* $\alpha(k) \leq K_0 e^{-\lambda_0 k}$ *for some positive constants* K_0 *and* λ_0. *The* α*-mixing coefficients for* $\{Y_t\}_{t=1}^T$ *is*

$$\alpha(k) = \sup_{B \in \sigma\{Y_s, s \leq t\}, C \in \sigma\{Y_s, s \geq t+k\}} |P(B \cap C) - P(B) P(C)|, \ k \geq 1.$$

(A2) *Function* m_1 *is a pth degree continuously differentiable function on* $[a, b]$.

(A3) *For any* $t, t' = 1, \ldots, T, t \neq t'$, *the joint density* $f(y_t, y_{t'})$ *of* $(Y_t, Y_{t'})$ *is continuous, and* $0 < c_f \leq \inf_{(y_t, y_{t'}) \in [a,b]^2} f(y_t, y_{t'}) \leq \sup_{(y_t, y_{t'}) \in [a,b]^2} f(y_t, y_{t'}) \leq C_f < \infty$.

(A4) ξ_t *satisfies* $E(\xi_t|\mathcal{F}_{t-1}) = 0, E(\xi_t^2|\mathcal{F}_{t-1}) = 1$, and $E(|\xi_t|^{5+\delta}|\mathcal{F}_{t-1}) < M_\delta$ *for some* $\delta > 0$ *and a finite positive* M_δ.

(A5) *The number of interior knots of the spline basis functions with degree* $p > 1$ *is such that*: $c_N n^{1/(2p)} \log n \leq N \leq C_N n^{1/2} / \log^3 n$, *for some positive constants* c_N *and* C_N.

Remark 1. Assumption (A1) is standard in time series literature; see Linton and Mammen (2005), Wang and Yang (2007). Assumption (A2) is very relaxed here when compared with the marginal integration method; see Linton and Nielsen (1995). Assumption (A3) only requires that the pairwise joint density be bounded away from 0 and ∞; thus it is a much weaker assumption than Assumption (iv) in Carroll, Härdle, and Mammen (2002) and Assumption (c) of Huang and Yang (2004) that require the boundedness of the joint density of the J variables. Assumption (A4) is comparable with Assumption (vi) in Carroll, Härdle, and Mammen (2002). Assumption (A5) gives the order of the number of interior knots.

We now describe our asymptotic results for the parameter in Theorems 1 and 2; the consistency result for the nonparametric news impact curve is given in the Appendix.

Theorem 1. *Under Assumptions* (A1)$-$(A5), *as* $n \to \infty$, $\hat{\beta} \longrightarrow \beta_0$ *a.s.*.

Theorem 2. *Under Assumptions* (A1)$-$(A5), *as* $n \to \infty$, $\sqrt{n}(\hat{\beta} - \beta_0)$ *has an asymptotic normal distribution with mean* 0 *and variance* $D^{-2} \sum_t \mathrm{Cov}(V_0, V_t)$, *where* $V_t = \varepsilon_t H(\beta_0, m_1(Y_t))$, $H(\beta_0, m_1(Y_t))$ *is given in* (A.9) *in the Appendix, and* $D = \sum_{j=2}^J (j-1)^2 \beta_0^{2j-4} E\left[m_1^2(Y_t)\right]$.

As an added refinement, considering that the additive components are linked, we define

$$\hat{m}_1^*(y) = \frac{\sum_{j=1}^{J} \hat{\beta}^{(j-1)} \hat{m}_j(y)}{\sum_{j=1}^{J} \hat{\beta}^{2(j-1)}}. \tag{2.5}$$

As discussed in Carroll, Härdle, and Mammen (2002), the asymptotic variance of $\{\hat{m}_1^*(y) - m_1(y)\}$ is smaller than that of $\{\hat{\beta}^{-(j-1)}\hat{m}_j(y) - m_1(y)\}$ for all j. We show, in the Appendix, that $\hat{m}_1^*(y)$ has the same convergence rate as $\hat{m}_1(y)$.

2.3. The alternatives

There is a host of possible alternative methods for estimating the GARCH models nonparametrically, for example, a referee has suggested that one can improve the efficiency of the estimators by taking the advantage of the structure of model (1.1). Let $\sigma_t^2(\beta, m) = \sum_{j=1}^{J} \beta^{j-1} m(Y_{t-j})$, and let β_0 and m_0 be defined as the minimizers of the population least squares (LS) criterion function $E\left\{Y_t^2 - \sigma_t^2(\beta, m)\right\}^2$, or be the minimizers of the negative likelihood (NL) criterion function $E\left[\log\left(\sigma_t^2(\beta, m)\right) + Y_t^2/\sigma_t^2(\beta, m)\right]^2$. Similar to the method in Section 2.1, we approximate $m(\cdot)$ by polynomial splines. Thus the empirical version of the LS or NL problem is $\sum_{t=J+1}^{T}\left\{Y_t^2 - \hat{\sigma}_t^2(\beta, \lambda)\right\}^2$ or $\sum_{t=J+1}^{T}[\log\{\hat{\sigma}_t^2(\beta, \lambda)\} + Y_t^2/[\hat{\sigma}_t^2(\beta, \lambda)]]$, where $\lambda = \{\lambda_{1-p}, \ldots, \lambda_N\}$ and $\hat{\sigma}_t^2(\beta, \lambda) = \sum_{j=1}^{J}\sum_{k=1-p}^{N} \beta^{j-1}\lambda_k b_k(Y_{t-j})$.

The minimizer of β based on the LS or NL criterion is the estimator of β denoted by GARCH-LS and GARCH-NL, respectively. We have not investigated their asymptotic properties due to some technical challenges, but the numerical performance of these two estimators has been studied in a comprehensive Monte Carlo study; see Section 4.

2.4. Selection of knots and lags

An important aspect of regression splines is the choice of the knots: splines with few knots are generally smoother than splines with many knots; increasing the knots usually improves the fit of the spline function to the data. The number of knots used in our simulation is $N = [c_1 n^{1/(2p)} \log(n)] + c_2$, where $[a]$ denotes the integer part of a, and c_1 and c_2 are positive constants. As pointed out in Wang and Yang (2007), there is no optimal method for selecting (c_1, c_2). In our simulation, the simple choice $c_1 = c_2 = 1$ works well, so these are set as default values.

For all modeling approaches, we need to determine the number of lags J. For the GARCH-ADD approach we adopt the consistent BIC lag selection method for non-linear additive autoregressive models (Huang and Yang (2004)), where

$$BIC(J) = \log\left[\frac{1}{n}\sum_{t=J+1}^{T}\{Y_t^2 - \hat{c} - \sum_{j=1}^{J}\hat{m}_j\left(Y_{t-j}\right)\}^2\right] + \frac{\log\log(n)}{n}\left\{1 + J(N+p+1)\right\}.$$

Numerical results on knots and lags selection are reported in Section 4.

3. Confidence Band for the News Impact Curve

In this section, we introduce a simultaneous confidence band for the news impact curve. For the nonlinear additive autoregressive models, Song and Yang (2010) proposed a two-step spline smoothing method to estimate each additive component: the first spline smoothing does a quick initial estimation of the additive components and removes all except the one of interest; the second smoothing is then applied to the cleaned univariate data to refine the estimator of each component with asymptotically oracle efficiency. They established an asymptotic $100(1-\alpha)\%$ conservative confidence band

$$\hat{m}_j\left(y\right) \pm 2\hat{\sigma}_j\left(y\right)\left\{\log\left(N+1\right)\right\}^{1/2} Q_N\left(\alpha\right), \tag{3.1}$$

where $\hat{m}_j$ is the spline-backfitted spline estimator, $\hat{\sigma}_j$ is the estimator of the standard deviation function of $\hat{m}_j$, and $Q_N\left(\alpha\right)$ is an inflation factor; see Song and Yang (2010).

When constructing the confidence band in (3.1), one needs additional smoothing steps to estimate the functions $\hat{\sigma}_j$ in (3.1); this may make the results less accurate, see Song and Yang (2009). Here we propose a bootstrap version of (3.1) similar to Song and Yang (2009). The following is a detailed procedure for constructing the simultaneous confidence band. Denote a predetermined large integer by n_B. By default n_B is 500.

Step 1. Pre-estimate m_j by its centered pilot estimator $\hat{m}_j$, $j = 1, \ldots, J$, through an under-smoothed spline smoothing procedure with N_1 knots.

Step 2. Construct the pseudo-response $\hat{W}_t = Y_t^2 - \hat{c} - \sum_{j=2}^{J}\hat{m}_j\left(Y_{t-j}\right)$, and approximate m_1 by linear spline smoothing with N_2 knots based on $\{\hat{W}_t, Y_{t-1}\}_{t=J+1}^{T}$. Define the estimator $\breve{m}_1\left(\cdot\right) = \arg\min_{g(\cdot)\in\mathcal{S}_n}\sum_{t=J+1}^{T}\{\hat{W}_t - g\left(Y_{t-1}\right)\}^2$, and denote residual $\hat{\varepsilon}_t = \hat{W}_t - \breve{m}_1(Y_{t-1})$.

Step 3. Let $\{\delta_{t,b}\}_{J+1\le t\le T}^{1\le b\le n_B}$ be i.i.d. mean 0 and variance 1 samples of the discrete distribution $\delta_{t,b} = (1\pm\sqrt{5})/2$ with probability $(5\pm\sqrt{5})/10$.

Step 4. For any $1 \leq b \leq n_B$, define the b-th wild bootstrap sample $\hat{W}^*_{t,b} = \breve{m}_1(Y_{t-1}) + \delta_{t,b}\hat{\varepsilon}_t$, $J+1 \leq t \leq T$. Then the bootstrap estimator of $m_1(y)$ is $\breve{m}_1{}^{(b)}(y) = \sum_{k=-1}^{N_2} \hat{\varphi}_k^{(b)} B_k(y)$, where $(\hat{\varphi}_{-1}^{(b)}, \hat{\varphi}_2^{(b)}, \ldots, \hat{\varphi}_{N_2}^{(b)})^{\mathrm{T}}$ are the estimated spline coefficients.

Step 5. Denote by $L_{\alpha/2}(y)$ and $U_{\alpha/2}(y)$, respectively, the lower and upper $100(1-\alpha/2)\%$ quantiles of the set $\{\breve{m}_1^{(b)}(y)\}_{b=1}^{n_B}$. The wild bootstrap $100(1-\alpha)\%$ pointwise confidence interval for function value $m_1(y)$ at the point y is $\left\{L_{\alpha/2}(y), U_{\alpha/2}(y)\right\}$.

Step 6. According to Song and Yang (2010), when localized at any point y, the uniform confidence band in (3.1) is wider than the pointwise confidence interval in Huang (2003) by a common factor $F_\alpha = 2z_{1-\alpha/2}^{-1}\{\log(N_2+1)\}^{1/2} Q_N(\alpha)$. We take the $100(1-\alpha)\%$ wild bootstrap confidence band for $m_1(y)$ to be $\left[\breve{m}_1(y) + \left\{L_{\alpha/2}(y) - \breve{m}_1(y)\right\} F_\alpha, \breve{m}_1(y) + \left\{U_{\alpha/2}(y) - \breve{m}_1(y)\right\} F_\alpha\right]$.

Remark 3. Song and Yang (2010) proposed to use $N_1 \sim n^{2/5}\log n$ knots for the initial spline estimation in Step 1, and $N_2 \sim n^{1/5}$ knots for the backfitting spline in estimation Step 2. In our simulation, N_1 and N_2 for the spline estimation are calculated as $N_1 = \min\left\{[c_1 n^{2/5}\log(n)] + c_2, [n/4-1]/J\right\}$ and $N_2 = \left[c_3 n^{1/5}\log(n)\right] + c_4$, with tuning constants $c_1 = 1$, $c_2 = 1$, $c_3 = 0.5$, $c_4 = 1$ by default.

4. Simulation

We carried out some simulations to illustrate the finite-sample behavior of the proposed estimators of Section 2. We compared the performance of the GARCH-ADD, GARCH-LS and GARCH-NL estimators with the GARCH(1,1) and GJR(1,1) estimators.

We generated time series $Y_t = \sigma_t \xi_t$ with the noise sequence $\{\xi_t\}_{t=1}^T$ i.i.d standard normal random variables. The volatility $\left\{\sigma_t^2\right\}_{t=1}^T$ was from the models

$$
\begin{aligned}
&A: \sigma_t^2 = 0.10 + 0.20Y_{t-1}^2 + 0.75\sigma_{t-1}^2,\\
&B: \sigma_t^2 = 0.05 + 0.20Y_{t-1}^2 + 0.05Y_{t-1}^2 I(Y_{t-1} < 0) + 0.75\sigma_{t-1}^2,\\
&C: \sigma_t^2 = 1 - 0.90\exp(-2Y_{t-1}^2) + 0.70\sigma_{t-1}^2,
\end{aligned}
$$

where the news impact curve in model A is symmetric, and a switching asymmetry has been built into model B. Model C involves exponential curves; a similar model has been studied by Carroll, Härdle, and Mammen (2002) and Bühlmann and McNeil (2002).

We first considered time series from models A, B and C with $J_{\text{model}} = 5$. For $T = 500$, 1,000, 2,000 and 3,000, we generated 200 replications for the three processes of size $T+1{,}000$; the first 1,000 observations were discarded to make sure the time series close to strictly stationary. We truncated each time series according to its 2.5th and 97.5th percentile. For these truncated time series, we estimated the parameter β_0 and the news impact curve m_1 by cubic splines; the number of lags, J, was selected according to the BIC described in Section 2.4; the minimization of $\hat{R}(\beta)$ was based on a grid search of 100 points around the true value.

The 3rd to the 5th columns in Table 1 provide the sample mean (MEAN), standard deviation (STD), and mean squared error (MSE) of $\hat{\beta}$ based on the GARCH-ADD, GARCH-LS, and GARCH-NL methods. As we expected, when the sample size increased, the parameter β_0 was more accurately estimated, with smaller MSE, confirmative to the conclusions of Theorem 1. As one referee expected, the GARCH-LS and GARCH-NL estimators provided more accurate estimation in some cases, especially for Model C. We did not see any obvious advantage to using these model structures for Models A and B. The mean and median of the selected number of lags J_{fit} are reported in the last column of Table 1, and one sees that J_{fit} is close to $J_{\mathrm{model}} = 5$ for moderately large sample size. For the news impact curve estimation we tried both $\hat{m}_1$ in (2.3) and $\hat{m}_1^*$ in (2.5), and the refined $\hat{m}_1^*$ performed slightly better, as we expected. The 6th column in Table 1 shows the average MSEs (AMSE) in $[-2.0, 2.0]^{J_{\mathrm{fit}}}$ for $\hat{m}_1^*$.

To illustrate the finite-sample behavior of our confidence bands, we calculated the percentage of coverage of the true news impact function by the confidence bands for the three models. Two nominal confidence levels 0.99 and 0.95 were considered. We carried out 500 replications and, for each replication, 500 bootstrap samples were generated for the bootstrap band. Table 2 contains the Monte Carlo coverage probabilities of the proposed bands. One can see the coverage rate gets close to the nominal level for all three models as sample size increases.

We also carried out simulations allowing model misspecification, and we generated time series from models A and B with $J_{\mathrm{model}} = \infty$. Recall that for $J_{\mathrm{model}} = \infty$, process A is a GARCH(1,1) process, so clearly GARCH(1,1) is the preferred estimator in this case. For process B, GJR(1,1) is the desired model. It is thus interesting to see how much efficiency, if any, is lost by using the proposed nonparametric methods with some selected number of lags; see the results in Table 3. For T =1,000, the nonparametric methods lost a small amount of efficiency relative to the parametric ones, but that effect decreased as the sample size increased for both processes A and B. Overall, we see that the GARCH-ADD worked quite well though β_0 was no longer the true parameter. One explanation is that the selected number of lags based on our method is usually also large when $J_{\mathrm{model}} = \infty$.

Table 1. Monto Carlo performance results based on 200 replications ($J_{\text{model}} = 5$). The values outside and inside the parentheses are the results based on the fitted J_{fit} and the oracle $J_{\text{oracle}} = 5$.

Size	Estimator	Parametric component			Nonparametric component	J_{fit} mean(median)
		MEAN	STD	MSE	AMSE	
A						
500	GARCH-ADD	0.68(0.63)	0.17(0.14)	0.035(0.033)	0.026(0.025)	3.7(3)
	GARCH-LS	0.81(0.78)	0.15(0.15)	0.027(0.024)	0.034(0.031)	5.8(5)
	GARCH-NL	0.80(0.79)	0.16(0.15)	0.027(0.024)	0.033(0.032)	5.8(5)
1000	GARCH-ADD	0.74(0.70)	0.14(0.11)	0.019(0.017)	0.016(0.017)	3.9(4)
	GARCH-LS	0.80(0.80)	0.11(0.10)	0.014(0.013)	0.025(0.024)	6.2(6)
	GARCH-NL	0.80(0.82)	0.11(0.10)	0.015(0.016)	0.025(0.026)	6.1(5)
2000	GARCH-ADD	0.75(0.73)	0.10(0.08)	0.009(0.008)	0.011(0.012)	4.1(4)
	GARCH-LS	0.78(0.80)	0.08(0.08)	0.008(0.008)	0.017(0.017)	6.2(5)
	GARCH-NL	0.79(0.81)	0.08(0.08)	0.008(0.010)	0.018(0.019)	6.2(5)
3000	GARCH-ADD	0.77(0.75)	0.07(0.07)	0.005(0.004)	0.011(0.011)	4.6(5)
	GARCH-LS	0.78(0.80)	0.08(0.07)	0.007(0.007)	0.015(0.015)	6.3(5)
	GARCH-NL	0.79(0.81)	0.08(0.06)	0.008(0.008)	0.015(0.016)	6.2(5)
B						
500	GARCH-ADD	0.73(0.71)	0.15(0.14)	0.022(0.022)	0.413(0.348)	5.1(5)
	GARCH-LS	0.81(0.80)	0.13(0.13)	0.021(0.019)	0.173(0.170)	6.0(5)
	GARCH-NL	0.81(0.82)	0.12(0.11)	0.019(0.016)	0.175(0.145)	5.9(5)
1000	GARCH-ADD	0.77(0.77)	0.10(0.10)	0.011(0.010)	0.177(0.163)	5.5(5)
	GARCH-LS	0.81(0.81)	0.10(0.09)	0.013(0.012)	0.100(0.108)	6.1(5)
	GARCH-NL	0.80(0.83)	0.09(0.07)	0.011(0.011)	0.095(0.089)	6.1(5)
2000	GARCH-ADD	0.77(0.78)	0.08(0.07)	0.006(0.006)	0.118(0.116)	5.4(5)
	GARCH-LS	0.79(0.81)	0.08(0.07)	0.008(0.008)	0.079(0.088)	6.4(5)
	GARCH-NL	0.79(0.82)	0.07(0.05)	0.007(0.008)	0.074(0.076)	6.3(5)
3000	GARCH-ADD	0.78(0.79)	0.06(0.06)	0.005(0.005)	0.093(0.098)	5.6(5)
	GARCH-LS	0.79(0.81)	0.07(0.06)	0.007(0.007)	0.075(0.082)	6.2(5)
	GARCH-NL	0.79(0.82)	0.06(0.04)	0.006(0.007)	0.069(0.072)	6.3(5)
C						
500	GARCH-ADD	0.57(0.55)	0.16(0.12)	0.042(0.036)	0.179(0.079)	2.7(2)
	GARCH-LS	0.79(0.72)	0.18(0.18)	0.040(0.031)	0.106(0.085)	5.3(5)
	GARCH-NL	0.76(0.71)	0.19(0.19)	0.038(0.034)	0.102(0.083)	5.3(5)
1000	GARCH-ADD	0.60(0.58)	0.16(0.11)	0.036(0.029)	0.147(0.055)	2.6(2)
	GARCH-LS	0.75(0.71)	0.15(0.13)	0.026(0.018)	0.075(0.057)	5.7(5)
	GARCH-NL	0.73(0.71)	0.15(0.14)	0.024(0.020)	0.069(0.057)	5.6(5)
2000	GARCH-ADD	0.64(0.61)	0.13(0.09)	0.021(0.016)	0.107(0.033)	2.7(2)
	GARCH-LS	0.73(0.71)	0.10(0.04)	0.011(0.010)	0.044(0.038)	5.8(5)
	GARCH-NL	0.72(0.71)	0.11(0.05)	0.013(0.012)	0.046(0.040)	5.8(5)
3000	GARCH-ADD	0.67(0.64)	0.11(0.07)	0.013(0.009)	0.069(0.020)	2.9(3)
	GARCH-LS	0.71(0.70)	0.09(0.08)	0.007(0.006)	0.030(0.027)	6.2(5)
	GARCH-NL	0.71(0.71)	0.09(0.08)	0.008(0.007)	0.033(0.030)	6.2(5)

Table 2. Coverage probabilities from 500 replications.

Sample Size	Model A		Model B		Model C	
	95%	99%	95%	99%	95%	99%
500	0.982	0.998	0.924	0.982	0.904	0.982
1000	0.970	0.996	0.962	0.990	0.848	0.950
2000	0.982	0.996	0.956	0.984	0.868	0.934
3000	0.976	0.996	0.932	0.986	0.892	0.954

Table 3. Monto Carlo performance results based on 200 replications ($J_{\text{model}} = \infty$).

	Size	Estimator	Parametric component			Nonparametric component	Time (secs)
			MEAN	STD	MSE	AMSE	
A	1000	GARCH(1,1)	0.77	0.04	0.002	0.008	1.3
		GJR(1,1)	0.77	0.04	0.002	0.008	7.4
		GARCH-ADD	0.71	0.10	0.012	0.016	2.6
		GARCH-LS	0.81	0.06	0.008	0.030	11.1
		GARCH-NL	0.82	0.06	0.008	0.031	11.2
	2000	GARCH(1,1)	0.77	0.03	0.001	0.006	6.2
		GJR(1,1)	0.77	0.03	0.001	0.006	12.9
		GARCH-ADD	0.75	0.06	0.004	0.009	5.6
		GARCH-LS	0.80	0.04	0.004	0.023	23.2
		GARCH-NL	0.81	0.04	0.005	0.025	22.6
	3000	GARCH(1,1)	0.77	0.02	0.001	0.005	11.8
		GJR	0.77	0.02	0.001	0.006	19.0
		GARCH-ADD	0.76	0.04	0.002	0.008	8.0
		GARCH-LS	0.80	0.03	0.004	0.017	32.0
		GARCH-NL	0.81	0.03	0.005	0.019	32.8
B	1000	GARCH(1,1)	0.76	0.06	0.004	0.015	3.6
		GJR(1,1)	0.76	0.03	0.001	0.009	6.2
		GARCH-ADD	0.76	0.09	0.009	0.141	2.6
		GARCH-LS	0.82	0.06	0.010	0.066	11.1
		GARCH-NL	0.84	0.04	0.010	0.058	11.2
	2000	GARCH(1,1)	0.76	0.02	0.001	0.013	6.1
		GJR(1,1)	0.76	0.02	0.001	0.006	12.3
		GARCH-ADD	0.78	0.06	0.005	0.060	5.2
		GARCH-LS	0.82	0.04	0.007	0.035	21.0
		GARCH-NL	0.83	0.04	0.008	0.039	20.7
	3000	GARCH(1,1)	0.76	0.02	0.001	0.012	11.4
		GJR(1,1)	0.76	0.02	0.001	0.005	15.9
		GARCH-ADD	0.79	0.04	0.004	0.033	7.8
		GARCH-LS	0.82	0.03	0.006	0.028	31.0
		GARCH-NL	0.83	0.02	0.007	0.030	30.4

In all our simulation experiments, the proposed GARCH-ADD method worked quickly, and we provide the time in seconds for all the methods in the last column

Table 4. Fitting the BMW daily returns

Model	−Log-likelihood	Volatility prediction error
GARCH(1,1)	3394.667	22.589
GJR(1,1)	3387.449	22.065
GARCH-ADD	3387.310	21.759

in Table 3. The proposed GARCH-ADD method only needs to solve a moderate number of linear least squares and a simple univariate nonlinear optimization; so in most cases one can see that the GARCH-ADD worked much faster compared to its competitors that require high-dimensional nonlinear optimization.

5. Application

In this section, we investigate the news impact curve on BMW daily stock return series to discover the relationship between past return shocks and conditional volatility. We collected the samples of daily percentage returns on the BMW share price from June 1st 1986 to January 30th 1994, a total of 2,000 observations. We truncated Y_t by its 0.01 and 0.99 quantiles.

For comparison, we also fitted the classical GARCH(1,1) and GJR(1,1) models. We compared the goodness-of-fit of our model with these two models in terms of volatility prediction error $(1/n)\sum_{t=J+1}^{T}\left(\hat{\sigma}_t^2 - Y_t^2\right)^2$ and the log-likelihood $-\sum_{t=J+1}^{T}\log\left\{\hat{\sigma}_t^{-1}\varphi\left(Y_t/\hat{\sigma}_t\right)\right\}$ with $J_{\text{fit}} = 50$. Clearly, the semiparametric method had an edge over the two parametric models in terms of prediction error and log-likelihood. One can see from Table 4 that the leverage effects of the GJR model can be further enhanced by a nonlinear link to yield a much better volatility fit.

To examine the validity of the GARCH and GJR models, we constructed the spline bootstrap confidence band. Figure 2 plots the GARCH, GJR, GARCH-ADD fits with 95% confidence bands. From Figure 2, we see that the spline estimated news impact curve stands in obvious contrast to the GARCH(1,1) fit, that shows strong evidence of the asymmetry of the news impact curve. It does seem that all three models can be fully covered by the bootstrap band.

For diagnostic purposes, we show the estimated autocorrelation function (ACF) of the daily standardized residuals $\hat{\varepsilon}^2$ with 95% Bartlett intervals, and one sees that the autocorrelation in the daily returns series is very small.

6. Discussion

Non/semi-parametric methods enhance the flexibility of the volatility models that practitioners use. However, due to the limitations in either interpretability, computational complexity, or theoretical reliability, most of the nonparametric stochastic volatility models have not been widely used as general tools in volatility

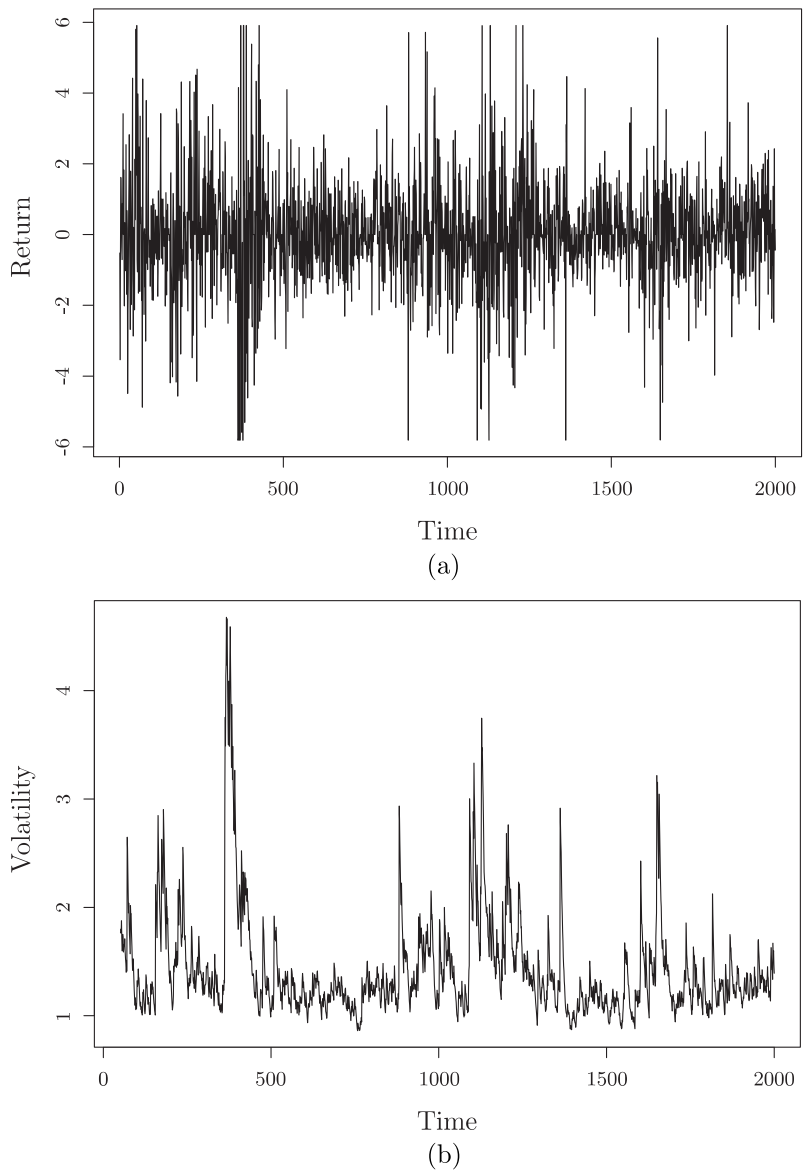

Figure 1. BMW daily returns: (a) original series; (b) the estimated volatility function.

analysis. In this paper, we have advanced semiparametric methods as flexible, computationally efficient and theoretically attractive tools for studying financial volatility.

We propose approximating the functional component in an additive volatility model by B-splines, which can be done by running OLS operations once the

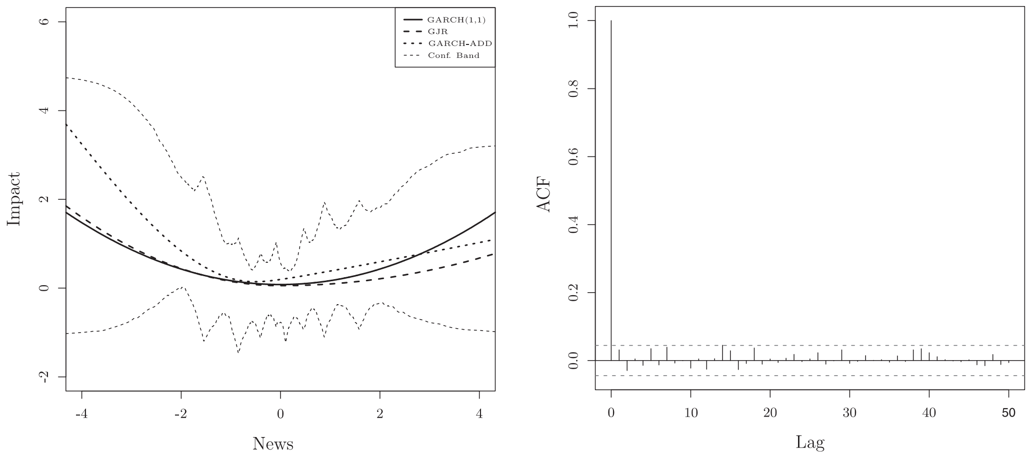

Figure 2. Analysis of BMW daily returns: (a) estimated news impact curve; (b) the estimated ACF using GARCH-ADD along with 95% Bartlett intervals for $\hat{\epsilon}^2$.

spline basis is chosen. Thus our method is particularly computationally efficient compared to its competitors that have to solve large system equations or optimize high-dimensional nonlinear functions. In addition, we introduced two alternative methods taking into account the model structure. These alternative methods are supposed to be more efficient in principle, but obtaining the asymptotics is likely to be difficult; we leave this as future research work. All the proposed estimators are easily implemented in commonly used software/package such as `lm()` in R.

There is other future work ahead. For example, it is interesting to consider the issue of model misspecification. Here, instead of estimating the true dynamic coefficient for $J = \infty$, we estimate a parameter β_0 that approximates the true parameter by using some finite J. If $J = \infty$, β_0 would not be the true dynamic coefficient anymore. The asymptotic results for the misspecified case has yet to be fully explored.

Acknowledgement

The research of the first author was partially supported by NSF grant DMS-0905730; research of the fourth author was supported in part by NSF grants DMS-0706518 and DMS-1007594. The authors would like to thank two referees, an associate editor and the editor for their insightful comments.

Appendix

Throughout this section, c and C are any positive constants, without distinction. Let $\|\phi\|_2$ be the L^2 norm of a function ϕ on $[a,b]$, $\|\phi\|_2^2 = \int_a^b \phi^2(y) f(y)\,dy$, and take the empirical L^2 norm as $\|\phi\|_{2,n}^2 = n^{-1}\sum_{i=1}^n \phi^2(Y_i)$. The corresponding inner products are $\langle \phi, \varphi\rangle_2 = \int_a^b \phi(y)\varphi(y) f(y)\,dy$ and $\langle \phi, \varphi\rangle_{2,n} = n^{-1}\sum_{i=1}^n \phi(Y_i)\varphi(Y_i)$.

Define the centered version spline basis

$$b_{j,k}^*(y) = b_{j,k}(y) - \frac{E(b_{j,k})}{E(b_{j,k-1})} b_{j,k-1}(y), \quad j = 1,\ldots,J, \quad k = 1-p,\ldots,N,$$

with the standardized version, given for any $j = 1,\ldots,J$, $k = 1-p,\ldots,N$,

$$B_{j,k}(y) = \frac{b_{j,k}^*(y)}{\|b_{j,k}^*\|_2}. \tag{A.1}$$

In practice, basis $\{b_{j,k},\ j = 1,\ldots,J,\ k = -p,\ldots,N\}^{\mathrm{T}}$ is used for data analysis, and the mathematically equivalent expression (A.1) is convenient for asymptotic analysis. Let $\mathbf{x} = (x_1,\ldots,x_J)^{\mathrm{T}}$. For a J-dimensional vector $\mathbf{X}_t = (Y_{t-1},\ldots, Y_{t-J})^{\mathrm{T}}$, define

$$\mathbf{B}(\mathbf{x}) = \{1, B_{1-p,1}(x_1), \ldots, B_{J,N}(x_J)\}^{\mathrm{T}}, \ \boldsymbol{B} = \{\mathbf{B}(\mathbf{X}_{J+1}), \ldots, \mathbf{B}(\mathbf{X}_T)\}^{\mathrm{T}}.$$

Let $m_t = c + \sum_{j=1}^J \beta_0^{j-1} m_1(Y_{t-j})$ and consider the signal vector $\mathbf{m} = \{m_{J+1},\ldots,m_T\}^{\mathrm{T}}$ and the noise vector $\epsilon = \{\epsilon_{J+1},\ldots,\epsilon_T\}^{\mathrm{T}}$. Let

$$\boldsymbol{\Lambda}_j = \mathrm{diag}\{0,\ldots,0, \underbrace{1,\ldots,1}_{\text{from } (N+p)(j-1)+2 \text{ to } (N+p)j+1}, 0,\ldots,0\}$$

be a diagonal matrix. Based on the relation $Y_t^2 = m_t + \epsilon_t$, one defines the signal spline smoothers and the noise spline components by

$$\begin{aligned}
\tilde{m}_j(y) &= \mathbf{B}(y)^{\mathrm{T}}\ \boldsymbol{\Lambda}_j (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\mathbf{m} - \frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Lambda}_j(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\mathbf{m},\\
\tilde{\epsilon}_j(y) &= \mathbf{B}(y)^{\mathrm{T}}\boldsymbol{\Lambda}_j(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\epsilon - \frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Lambda}_j(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\epsilon,
\end{aligned} \tag{A.2}$$

where $\mathbf{1}_n$ is a length n dimensional vector with all elements 1.

With $\mathbf{Z} = \{Y_{J+1}^2,\ldots,Y_T^2\}$, we can rewrite $\hat{m}_j(y)$ in (2.3) as

$$\hat{m}_j(y) = \mathbf{B}(y)^{\mathrm{T}}\ \boldsymbol{\Lambda}_j(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\mathbf{Z} - \frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Lambda}_j(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\mathbf{Z}.$$

Then one has the crucial decomposition for proving Theorem 1,

$$\hat{m}_j(y) = \tilde{m}_j(y) + \tilde{\epsilon}_j(y), \quad j = 1,\ldots,J. \tag{A.3}$$

To prove Theorems 1 and 2, we need a lemma on the L^2 convergence rate of the one-step spline estimator $\hat{m}_1$ to m_1.

Lemma A.1. *Under* (A1)−(A5)*, as* $n \to \infty$*,*

$$\|\hat{m}_1 - m_1\|_{2,n} + \|\hat{m}_1 - m_1\|_2 = O_{a.s.}\Big(h^p + \frac{\log n}{\sqrt{nh}}\Big). \tag{A.4}$$

Proof. Using page 149 of de Boor (2001), we have $\|\tilde{m}_1 - m_1\|_2 = O\left(h^p\right)$, and $\|\tilde{m}_1 - m_1\|_{2,n} = O\left(h^p\right)$. According to Lemma A.6 of Wang and Yang (2007), $\|\tilde{\epsilon}\|_2 = O_{a.s.}(\log n/\sqrt{nh})$ and $\|\tilde{\epsilon}\|_{2,n} = O_{a.s.}(\log n/\sqrt{nh})$. The result in Lemma A.1 follows from the decomposition in (A.3).

A corollary relates the asymptotic property of $\hat{m}_1^*$ given in (2.5) to m_1.

Corollary A.1. *Under* (A1)−(A5)*, as* $n \to \infty$*,*

$$\|\hat{m}_1^* - m_1\|_{2,n} + \|\hat{m}_1^* - m_1\|_2 = O_P\Big(h^p + \frac{\log n}{\sqrt{nh}}\Big).$$

Proof. The proof is quite straightforward from the lemma and Theorem 2.

$$\|\hat{m}_1^* - m_1\|_2 = \left\| \frac{1}{\sum_{j=1}^{J} \hat{\beta}^{2(j-1)}} \left\{ \sum_{j=1}^{J} \hat{\beta}^{(j-1)}(\hat{m}_j - m_j) + \sum_{j=1}^{J} \hat{\beta}^{j-1}(\beta_0^{j-1} - \hat{\beta}^{j-1}) m_1 \right\} \right\|_2.$$

For any $1 \le j \le J$, $\|\hat{m}_j - m_j\|_2$ has order $O_P(h^p + \log n/\sqrt{nh})$. Combined with the result that $\sum_{j=1}^{J} \hat{\beta}^{j-1}(\beta_0^{j-1} - \hat{\beta}^{j-1}) = O_P(n^{-1/2}) = o_P(h^p + \log n/\sqrt{nh})$ from Theorem 2, one has that $\|\hat{m}_1^* - m_1\|_2$ is $O_P(h^p + \log n/\sqrt{nh})$, and so is $\|\hat{m}_1^* - m_1\|_{2,n}$.

A.1. Proof of Theorem 1

Note that the risk function $R(\beta)$ given in (2.2) is locally convex in β and, hence, consistency for β is implied by $\sup_{\beta \in [\beta_1, \beta_2]} |\hat{R}(\beta) - R(\beta)| \to 0$ *a.s.*, where $\hat{R}(\beta)$ is given in (2.4). Note that

$$\begin{aligned}
\hat{R}(\beta) &= \sum_{j=1}^{J} \left\|\hat{m}_j - m_j + \beta^{j-1} m_1 - \beta^{j-1}\hat{m}_1\right\|_{2,n}^2 + \sum_{j=1}^{J} \left\|m_j - \beta^{j-1} m_1\right\|_{2,n}^2 \\
&\quad + \sum_{j=1}^{J} 2\left\langle m_j - \beta^{j-1} m_1, \hat{m}_j - m_j + \beta^{j-1} m_1 - \beta^{j-1}\hat{m}_1 \right\rangle_{2,n} \\
&= P_1(\beta) + P_2(\beta) + P_3(\beta).
\end{aligned}$$

By (A.4), we have $\sup_{\beta \in [\beta_1, \beta_2]} P_1(\beta) = O_{a.s.}(h^{2p} + \log^2 n/nh)$, and

$$\begin{aligned}
&\sup_{\beta \in [\beta_1, \beta_2]} P_3(\beta) \\
&\le 2J \max_{1 \le j \le J} \left\{ \left\|\hat{m}_j - m_j + \beta^{j-1} m_1 - \beta^{j-1}\hat{m}_1\right\|_{2,n} \sup_{x \in [a,b]} \left|m_j\left(x\right) - \beta^{j-1} m_1\left(x\right)\right| \right\},
\end{aligned}$$

which is of the order $O_{a.s.}\left(h^p + \log n/\sqrt{nh}\right)$. Thus,

$$\sup_{\beta\in[\beta_1,\beta_2]} \left|\hat{R}(\beta) - P_2(\beta)\right| = O_{a.s.}\left(h^p + \frac{\log n}{\sqrt{nh}}\right).$$

And

$$\begin{aligned}
&\sup_{\beta\in[\beta_1,\beta_2]} |P_2(\beta) - R(\beta)| \\
&\quad \le \left|\frac{1}{n}\sum_{t=J+1}^{T}\Big\{\sum_{j=1}^{J} m_j^2(Y_t)\Big\} - E\Big\{\sum_{j=1}^{J} m_j^2(Y_t)\Big\}\right| \\
&\qquad + \frac{2(1-\beta_2^{2J})}{1-\beta_2^2}\left|\frac{1}{n}\sum_{t=J+1}^{T}\Big\{\sum_{j=1}^{J} m_j(Y_t)m_1(Y_t)\Big\} - E\Big\{\sum_{j=1}^{J} m_j(Y_t)m_1(Y_t)\Big\}\right| \\
&\qquad + \frac{1-\beta_2^{2J}}{1-\beta_2^2}\left|\frac{1}{n}\sum_{t=J+1}^{T} m_1^2(Y_t) - Em_1^2(Y_t)\right|.
\end{aligned}$$

By a strong law of large numbers for mixing processes, $\sup_{\beta\in[\beta_1,\beta_2]} |P_2(\beta) - R(\beta)| = o_{a.s.}(1)$. Thus

$$\sup_{\beta\in[\beta_1,\beta_2]} \left|\hat{R}(\beta) - R(\beta)\right| \le \sup_{\beta\in[\beta_1,\beta_2]} \left|\hat{R}(\beta) - P_2(\beta)\right| + \sup_{\beta\in[\beta_1,\beta_2]} |P_2(\beta) - R(\beta)| = o_{a.s.}(1),$$

and $\hat{\beta}$ converges to β_0 almost surely.

A.2. Proof of Theorem 2

Write

$$\sqrt{n}\frac{d}{d\beta}\hat{R}\left(\hat{\beta}\right) = \sqrt{n}\,\frac{d}{d\beta}\hat{R}(\beta)\Big|_{\beta=\beta_0} + \frac{d^2}{d\beta^2}\hat{R}(\beta)\Big|_{\beta=\tilde{\beta}}\sqrt{n}\left(\hat{\beta}-\beta_0\right),$$

where $\tilde{\beta}$ is between $\hat{\beta}$ and β_0. Then one has

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}-\beta_0\right) &= \sqrt{n}\Big\{\frac{d^2}{d\beta^2}\hat{R}(\beta)\Big|_{\beta=\tilde{\beta}}\Big\}^{-1}\Big\{\frac{d}{d\beta}\hat{R}\left(\hat{\beta}\right) - \frac{d}{d\beta}\hat{R}(\beta)\Big|_{\beta=\beta_0}\Big\} \\
&= -\sqrt{n}\Big\{\frac{d^2}{d\beta^2}\hat{R}(\beta)\Big|_{\beta=\tilde{\beta}}\Big\}^{-1}\frac{d}{d\beta}\hat{R}(\beta)\Big|_{\beta=\beta_0}. \qquad \text{(A.5)}
\end{aligned}$$

We need the two lemmas to deal with $-\sqrt{n}\,\frac{d}{d\beta}\hat{R}(\beta)\Big|_{\beta=\beta_0}$ and $\frac{d^2}{d\beta^2}\hat{R}(\beta)$.

Lemma A.2. *Under* (A1)−(A5),

$$-\frac{\sqrt{n}}{2}\frac{d}{d\beta}\hat{R}(\beta)\bigg|_{\beta=\beta_0} = n^{-1/2}\sum_{t=J+1}^{T}\epsilon_t H(\beta_0, m_1(Y_t)), \tag{A.6}$$

where $H(\beta_0, m_1(Y_t))$ *is given in* (A.9).

Proof. Note that

$$\begin{aligned}
&-\frac{\sqrt{n}}{2}\frac{d}{d\beta}\hat{R}(\beta)\bigg|_{\beta=\beta_0}\\
&=\frac{1}{\sqrt{n}}\sum_{t=J+1}^{T}\sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{\hat{m}_j(Y_t)-\beta_0^{j-1}\hat{m}_1(Y_t)\right\}\hat{m}_1(Y_t)\\
&=\frac{1}{\sqrt{n}}\sum_{t=J+1}^{T}\sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{\hat{m}_j(Y_t)-\beta_0^{j-1}\hat{m}_1(Y_t)\right\}\left[\hat{m}_1(Y_t)-m_1(Y_t)\right]\\
&\quad+\frac{1}{\sqrt{n}}\sum_{t=J+1}^{T}\sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{\hat{m}_j(Y_t)-\beta_0^{j-1}\hat{m}_1(Y_t)\right\}m_1(Y_t)\\
&=I+II.
\end{aligned}$$

Here the first term I can be written as

$$\begin{aligned}
I&=\frac{1}{\sqrt{n}}\sum_{t=J+1}^{T}\sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{\hat{m}_j(Y_t)-m_j(Y_t)\right\}\left\{\hat{m}_1(Y_t)-m_1(Y_t)\right\}\\
&\quad-\frac{1}{\sqrt{n}}\sum_{t=J+1}^{T}\sum_{j=1}^{J}(j-1)\beta_0^{2j-3}\left\{\hat{m}_1(Y_t)-m_1(Y_t)\right\}^2\\
&=I_1+I_2.
\end{aligned}$$

As in the proof of Theorem 1, both I_1 and I_2 are of order $O_{a.s.}(h^{2p}+\log^2 n/nh)$. With the order of h in (A5), $I=O_{a.s.}\left\{n^{1/2}\left(h^{2p}+\log^2 n/(nh)\right)\right\}=o_{a.s.}(1)$. For II, noting that $\|\tilde{m}_1-m_1\|_{2,n}=O(h^p)$ *a.s.*, by (A.3) and Assumption (A5), one can write II as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t'=J+1}^{T}\sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{\tilde{m}_j(Y_{t'})+\tilde{\epsilon}_j(Y_{t'})-\beta_0^{j-1}\tilde{m}_1(Y_{t'})-\beta_0^{j-1}\tilde{\epsilon}_1(Y_{t'})\right\}m_1(Y_{t'})\\
&=\frac{1}{\sqrt{n}}\sum_{t'=J+1}^{T}\sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{\tilde{\epsilon}_j(Y_{t'})-\beta_0^{j-1}\tilde{\epsilon}_1(Y_{t'})\right\}m_1(Y_{t'})+o_{a.s.}(1).
\end{aligned}$$

Let

$$\widehat{\boldsymbol{V}} = \begin{pmatrix} 1 & 0 \\ 0 & \left\langle B_{j,k}, B_{j',k'} \right\rangle_{2,n} \end{pmatrix}_{\substack{1 \le j,j' \le J, \\ 1-p \le k,k' \le N}}, \quad \boldsymbol{V} = \begin{pmatrix} 1 & 0 \\ 0 & \left\langle B_{j,k}, B_{j',k'} \right\rangle_{2} \end{pmatrix}_{\substack{1 \le j,j' \le J, \\ 1-p \le k,k' \le N}}.$$

With $\tilde{\epsilon}_j$ defined in (A.2), the main term in II is

$$n^{-1/2} \sum_{t'=J+1}^{T} \sum_{j=1}^{J} \Bigg[(j-1)\, \beta_0^{j-2} m_1(Y_{t'}) \Big\{ \mathbf{B}(Y_{t'})\, \boldsymbol{\Lambda}_j - \frac{1}{n} \mathbf{1}_n^{\mathrm{T}} \boldsymbol{B} \boldsymbol{\Lambda}_j - \beta_0^{j-1} \times \Big(\mathbf{B}(Y_{t'})\, \boldsymbol{\Lambda}_1 - \frac{1}{n} \mathbf{1}_n^{\mathrm{T}} \boldsymbol{B} \boldsymbol{\Lambda}_1 \Big) \Big\} \widehat{\boldsymbol{V}}^{-1} \Big\{ \frac{1}{n} \sum_{t=J+1}^{T} B_{j,k}(Y_t) \epsilon_t \Big\}_{j,k} \Bigg]. \tag{A.7}$$

According to Lemma A.10 in Wang and Yang (2007), we can replace $\widehat{\boldsymbol{V}}$ by $\boldsymbol{V}$ in (A.7) with a negligible error term $O_{a.s.}\left\{ n^{-1/2} N (\log n)^2 \right\}$. Next we interchange the indices t and t' of (A.7) to find the main term in II approximated by

$$n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} \Bigg[(j-1)\, \beta_0^{j-2} \epsilon_t \Big\{ \mathbf{B}(Y_t)\, \boldsymbol{\Lambda}_j - \frac{1}{n} \mathbf{1}_n^{\mathrm{T}} \boldsymbol{B} \boldsymbol{\Lambda}_j - \beta_0^{j-1} \times \Big(\mathbf{B}(Y_t)\, \boldsymbol{\Lambda}_1 - \frac{1}{n} \mathbf{1}_n^{\mathrm{T}} \boldsymbol{B} \boldsymbol{\Lambda}_1 \Big) \Big\} \boldsymbol{V}^{-1} \Big\{ \frac{1}{n} \sum_{t'=J+1}^{T} B_{j,k}(Y_{t'}) m_1(Y_{t'}) \Big\}_{j,k} \Bigg]. \tag{A.8}$$

Let

$$\begin{aligned} H(\beta_0, m_1(Y_t)) &= \sum_{j=1}^{J} \Bigg[(j-1)\, \beta_0^{j-2} \Big\{ \mathbf{B}(Y_t)\, \boldsymbol{\Lambda}_j - \frac{1}{n} \mathbf{1}_n^{\mathrm{T}} \boldsymbol{B} \boldsymbol{\Lambda}_j - \beta_0^{j-1} \\ &\quad \times \Big(\mathbf{B}(Y_t)\, \boldsymbol{\Lambda}_1 - \frac{1}{n} \mathbf{1}_n^{\mathrm{T}} \boldsymbol{B} \boldsymbol{\Lambda}_1 \Big) \Big\} \boldsymbol{V}^{-1} \Big\{ \frac{1}{n} \sum_{t'=J+1}^{T} B_{j,k}(Y_{t'}) m_1(Y_{t'}) \Big\}_{j,k} \Bigg], \end{aligned} \tag{A.9}$$

so that (A.8) can be written as $n^{-1/2} \sum_{t=J+1}^{T} \epsilon_t H(\beta_0, m_1(Y_t))$, which leads to (A.6).

Lemma A.3. *Under* (A1)$-$(A5), $\frac{d^2}{d\beta^2} \hat{R}(\beta) = E m_1^2(Y_t) \sum_{j=2}^{J} \left\{ (j-1)\, \beta^{(j-2)} \right\}^2 + o_{a.s.}(1)$.

Proof. Note that

$$\frac{d^2}{d\beta^2}\hat{R}(\beta) = n^{-1}\sum_{t=J+1}^{T}\sum_{j=2}^{J}\hat{m}_1^2(Y_t)\left\{(j-1)\beta^{(j-2)}\right\}^2$$
$$+n^{-1}\sum_{t=J+1}^{T}\sum_{j=2}^{J}(j-1)(j-2)\beta^{j-3}\hat{m}_1(Y_t)\left\{\hat{m}_j(Y_t)-\beta^{j-1}\hat{m}_1(Y_t)\right\}$$
$$= I_1 + I_2,$$

where $I_2 = o_{a.s.}(1)$, as for I in Lemma (A.2), and for I_1,

$$\frac{1}{n}\sum_{t=J+1}^{T}\hat{m}_1^2(Y_t) - \frac{1}{n}\sum_{t=J+1}^{T}m_1^2(Y_t) \tag{A.10}$$
$$= \frac{1}{n}\sum_{t=J+1}^{T}\{\hat{m}_1(Y_t)-m_1(Y_t)\}\{\hat{m}_1(Y_t)+m_1(Y_t)\}$$
$$\leq \left\{\frac{1}{n}\sum_{t=J+1}^{T}(\hat{m}_1(Y_t)-m_1(Y_t))^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{t=J+1}^{T}(\hat{m}_1(Y_t)+m_1(Y_t))^2\right\}^{1/2}$$
$$\leq \|\hat{m}_1 - m_1\|_{2,n}\sup_x\left|\sqrt{6}m_1(x)\right| = O_{a.s.}\left(h^p + \frac{\log n}{\sqrt{nh}}\right) = o_{a.s.}(1). \tag{A.11}$$

By a law of large numbers, $n^{-1}\sum_{t=J+1}^{T} m_1^2(Y_t) \to E\left[m_1^2(Y_t)\right]$ as n goes to infinity. Combined with (A.11), we have $\lim_{n\to\infty} n^{-1}\sum_{t=J+1}^{T}\hat{m}_1^2(Y_t) = E\left[m_1^2(Y_t)\right]$.

For the proof of Theorem 2, combine (A.5), (A.6), and Lemma A.3, and note that as $n \to \infty$, $\sum_{j=2}^{J}(j-1)^2\tilde{\beta}^{2j-4} \to \sum_{j=2}^{J}(j-1)^2\beta_0^{2j-4}$ $a.s.$. Then

$$n^{1/2}\left(\hat{\beta}-\beta_0\right) = n^{-1/2}\sum_t \epsilon_t H(\beta_0, m_1(Y_t))\left\{\sum_{j=2}^{J}(j-1)^2\beta_0^{2j-4}E\left[m_1^2(Y_t)\right]\right\}^{-1}$$
$$+o_{a.s.}(1).$$

Asymptotic normality of $n^{1/2}(\hat{\beta}-\beta_0)$ follows from a Slutsky theorem and a central limit theorem for strongly mixing sequences (see, e.g., Bosq (1996, Thm. 1.7)). We have to verify that for some $\nu > 2$, $E\left|\epsilon_t H(\beta_0, m_1(Y_t))\right|^{\upsilon} < \infty$, which can be obtained from (A2) and (A4).

References

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag, New York.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307-327.

Bühlmann, P. and McNeil, A. J. (2002). An algorithm for nonparametric GARCH modelling. *Comput. Statist. Data Anal.* **40**, 665-683.

Carroll, R., Härdle, W. and Mammen, E. (2002). Estimation in an additive model when the components are linked parametrically. *Econ. Theory* **18**, 886-912.

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

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

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

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

Härdle, W. and Tsybakov, A. B. (1997). Locally polynomial estimators of the volatility Function. *J. Econometrics* **81**, 223-242.

Härdle, W., Tsybakov, A. B. and Yang, L. (1998). Nonparametric vector autoregression. *J. Statist. Plann. Inference* **68**, 221-245.

Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *J. Multivariate Anal.* **95**, 246-272.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31**, 1600-1635.

Huang, J. Z. and Yang, L. (2004). Identification of nonlinear additive autoregressive models. *J. Roy. Statist. Soc. Ser. B* **66**, 463-477.

Hafner, C. (1998). *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility*. Physica-Verlag, Heidelberg.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-101.

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

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. *Econ. Theory* **11**, 258-289.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25**, 186-211.

Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatility. *J. Econometrics* **45**, 267–290.

Song, Q. and Yang, L. (2009). Spline confidence bands for variance function. *J. Nonparametr. Stat.* **21**, 589-609.

Song, Q. and Yang, L. (2010). Oracally efficient spline smoothing of nonlinear additive autoregression model with simultaneous confidence band. *J. Multivariate Anal.*, **101**, 2008-2025.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 689-705.

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35**, 2474-2503.

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

Yang, L. (2006). Semiparametric GARCH model and foreign exchange volatility. *J. Econometrics* **130**, 365-384.

Yang, L., Härdle, W. and Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *J. Time Ser. Anal.* **20**, 597-604.

Department of Statistics, University of Georgia, Athens, GA 30602, USA.

E-mail: lilywang@uga.edu

Department of Statistics, University of Georgia, Athens, GA 30602, USA.

E-mail: congfeng@uga.edu

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080, USA.

E-mail: song@utdallas.edu

Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou 215006, P. R. China.

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA.

E-mail: yanglijian@suda.edu.cn

(Received October 2009; accepted September 2010)