EFFICIENT SEMIPARAMETRIC GARCH MODELING OF FINANCIAL VOLATILITY

Li Wang\textsuperscript{1}, Cong Feng\textsuperscript{1}, Qiongxia Song\textsuperscript{2} and Lijian Yang\textsuperscript{3,4}

\textsuperscript{1}University of Georgia, \textsuperscript{2}University of Texas at Dallas, \textsuperscript{3}Soochow University and \textsuperscript{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 real 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 many 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 $\{\xi_t\}_{t=1}^\infty$ are i.i.d with mean 0 and variance 1, and $\{\sigma_t^2\}_{t=1}^\infty$ denotes the conditional volatility series. Engle (1982) introduced the 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

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

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):

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

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 et al. (1993) proposed the following 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 the non/semi-parametric (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, et al. (1998), Bühlmann and McNeil (2002), Linton and Mammen (2005) and Yang (2006). These models generalize and outperform the parametric GARCH models when applied to real 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, et al. (1999) analyzed a multiplicative form of volatility using nonparametric smoothing. Carroll et al. (2002) and Yang (2002) proposed a truncated version of the nonparametric GARCH model with a finite number of lags \( J \)

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

However, for small \( J \), it 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 algorithm (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 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 polynomial spline estimates for the generalized additive model. In volatility studies, Engle and Ng (1993) employed the 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 real 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 \( Y_t = \sigma_t \xi_t \), \( t = 1, 2, \ldots, T \). We rewrite model (1.1) as the following additive autoregressive model,

\[
Y_t^2 = c + \sum_{j=1}^{J} m_j(Y_{t-j}) + \epsilon_t, \quad \epsilon_t = \sigma_t^2 (\xi_t^2 - 1),
\]  

(2.1)

where the component functions \( m_1(\cdot), \ldots, m_J(\cdot) \) are linked by a scalar parameter \( \beta_0 \) 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,

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

(2.2)

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

Our interest is to estimate the news impact function \( m_1 \) and dynamic coefficient parameter \( \beta_0 \). 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 \( \beta_0 \) by using the link restriction between the additive components \( \hat{m}_j(\cdot) \) \( (j = 1, \ldots, J) \). For simplicity of notation, let us call the above approach the spline additive GARCH (GARCH-ADD) approach.

We now 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 applying to real data, one can use fixed truncation to satisfy this condition. Let \( S_n \) be the space of polynomial
splines on $[a, b]$ of degree $p \geq 1$. We introduce a knot sequence with $N$ interior knots

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

where $N \equiv N_n$ increases when sample size $n$ increases, whose precise order is given in Assumption (A5). The spline of degree $p$ for the $j$th variable is denoted as $\{b_{j,k}\}_{k=-p}^{N}$ (de Boor (2001)). Then $S_n$ 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 of proof, while adaptively choosing the locations of the knots could have been done for real data analysis. Let $h = (b - a)/(N + 1)$ be the distance between neighboring knots. Define next the space $G$ as the linear space spanned by

$$G^{p, 2} = \{g(\cdot) : g(\cdot) \text{ is } p - \text{time continuously differentiable on } [a, b]\}.$$  

where

$$\lambda_0, \lambda_{1,-p}, \ldots, \lambda_{J,N}$$

be the solutions of the least squares problem

$$\left(\lambda_0, \lambda_{1,-p}, \ldots, \lambda_{J,N}\right)^T = \arg\min_{R^{J(N+p)}} \sum_{t=J+1}^T \left\{Y_t^2 - \lambda_0 - \sum_{j=1}^J \sum_{k=-p}^N \lambda_{j,k} b_{j,k}(Y_{t-j})\right\}^2.$$  

Denote $n = T - J$. 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

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

To estimate the parameter $\beta_0$, 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 all the components, so we define the sample least squares criterion,

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

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 \( \alpha \)-mixing with exponentially decaying mixing coefficients \( \alpha(k) \leq K_0 e^{-\lambda_0 k} \) for some positive constants \( K_0 \) and \( \lambda_0 \). The \( \alpha \)-mixing coefficients for \( \{Y_t\}_{t=1}^T \) is defined as

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

(A2) Function \( m_1 \) is a \( p \)th degree continuously differentiable function on interval \([a, b]\).

(A3) For any \( t, t' = 1, 2, ..., T, t \neq t' \), 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) The noise \( \xi_t \) satisfies \( E(\xi_t | \mathcal{F}_{t-1}) = 0, E(\xi_t^2 | \mathcal{F}_{t-1}) = 1 \), and \( E\left(|\xi_t|^{5+\delta} | \mathcal{F}_{t-1}\right) < M_\delta \) for some \( \delta > 0 \) and a finite positive \( M_\delta \).

(A5) The number of interior knots of the spline basis functions with degree \( p > 1 \) satisfies:

\[
c_{N}n^{1/(2p)} \log n \leq N \leq C_{N}n^{1/2}/\log^3 n,
\]

for some positive constants \( c_{N} \) and \( C_{N} \).

**Remark 1.** Assumption (A1) is a standard assumption in time series literature; see Linton and Mammen (2005), Wang and Yang (2007). Assumption (A2) is very relaxed in our paper compared with marginal integration method; see Linton and Nielsen (1995). Assumption (A3) only requires that the pairwise joint density is bounded away from 0 and \( \infty \). So it is a much weaker assumption compared with Assumption (iv) in Carroll et al. (2002) and Assumption (c) of Huang and Yang (2004) which require the boundedness of the joint density of the \( J \) variables. Assumption (A4) is comparable with Assumption (vi) in Carroll et al. (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, and the consistency result for the nonparametric news impact curve is given in the Appendix.
Theorem 1. Under Assumptions (A1)-(A5), as $n \to \infty$, $\hat{\beta} \to \beta_0$, 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 $D^{-2} \sum_t \text{Cov}(V_0, V_t)$, where $V_t = \varepsilon_t H(\beta_0, m_1(Y_t))$, and $H(\beta_0, m_1(Y_t))$ is given in (A.9) in Appendix, and $D = \sum_{j=2}^{J} (j-1)^2 \beta_j^2 - 4 \mu_1^2 \int [m_1^2(Y)]$.

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

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

(2.5)

As discussed in Carroll et al. (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 we can improve the efficiency of the estimators by taking the advantage of the structure of model (1.1). Define $\sigma_t^2(\beta, m) = \sum_{j=1}^{J} \beta^{j-1} m(Y_{t-j})$, and let $\beta_0$ and $m_0$ be defined as the minimizers of the population least squares (LS) criterion function $E \{Y_t^2 - \sigma_t^2(\beta, m)\}^2$, or be the minimizers of the negative likelihood (NL) criterion function $E \left[ \log \left( \sigma_t^2(\beta, m) + \frac{Y_t^2}{\sigma_t^2(\beta, m)} \right) \right]^2$. 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 $\sum_{t=J+1}^{T} \{Y_t^2 - \hat{\sigma}_t^2(\beta, \lambda)\}^2$ or $\sum_{t=J+1}^{T} \left[ \log \left( \hat{\sigma}_t^2(\beta, \lambda) + \frac{Y_t^2}{\hat{\sigma}_t^2(\beta, \lambda)} \right) \right]^2$, where $\lambda = \{\lambda_1, ..., \lambda_N\}$ and $\hat{\sigma}_t^2(\beta, \lambda) = \sum_{j=1}^{J} \sum_{k=1-p}^{N} \beta^j \lambda_k b_k(Y_{t-j})$.

The minimizer of $\beta$ based on the above LS or NL criterion is the estimator of $\beta$, 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 have been studied in a comprehensive Monte Carlo study; see Section 4.

2.4. Selection of knots and lags
An important aspect for regression splines is the choice of the knots. Splines with few knots are generally smoother than splines with many knots; however, increasing the knots usually can improve the fit of the spline function to the data. The number of knots used in our simulation is 

\[ N = \lfloor c_1 n^{1/(2p)} \log(n) \rfloor + c_2, \]

where \( \lfloor a \rfloor \) denotes the integer part of \( a \), and \( c_1 \) and \( c_2 \) are positive integers. As pointed out in Wang and Yang (2007), there is no optimal method to select \((c_1, c_2)\). In our simulation, the simple choice \( c_1 = c_2 = 1 \) works well, so these are set as default values.

For all the above 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) and the BIC is defined as

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

Numerical results of knots and lags selection in a simulation study 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 nonlinear additive autoregressive model, Song and Yang (2010) proposed a two-step spline smoothing method to estimate each additive component: the first step spline smoothing does a quick initial estimation of all additive components and removes all except the ones of interest; the second smoothing is then applied to the cleaned univariate data to refine the estimator of each component with the asymptotically oracle efficiency. They also established an asymptotic $100(1 - \alpha)\%$ conservative confidence band

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

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 \( Q_N(\alpha) \) 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), which may cause the results less accurate; see Song and Yang (2009). In this article we propose a bootstrap version of (3.1) similar to Song and Yang (2009). The following are the detailed procedures of 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 $\hat{W}_t = Y_t^2 - \hat{c} - \sum_{j=2}^J \hat{m}_j (Y_{t-j})$ 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 $\hat{m}_1 (\cdot) = \arg \min_{g(\cdot) \in S_n} \sum_{t=J+1}^T \{\hat{W}_t - g(Y_{t-1})\}^2$, and denote residual $\hat{\varepsilon}_t = \hat{W}_t - \hat{m}_1(Y_{t-1})$.

Step 3. Let $\{\delta_{t,b}\}_{J+1 \leq t \leq T}$ be i.i.d. mean 0 and variance 1 samples of the following discrete distribution $\delta_{t,b} = \frac{1+\sqrt{5}}{2}$ with probability $\frac{5+\sqrt{5}}{10}$.

Step 4. For any $1 \leq b \leq n_B$, define the $b$-th wild bootstrap sample $\hat{W}_{t,b}^* = \hat{m}_1(Y_{t-1}) + \delta_{t,b}\hat{\varepsilon}_t, J + 1 \leq t \leq T$. Then the bootstrap estimator of $m_1 (y)$ is $\hat{m}_1^{(b)} (y) = \sum_{k=-1}^{N_2} \hat{\varphi}_k^{(b)} B_k (y)$, where $\left(\hat{\varphi}_{-1}^{(b)}, \hat{\varphi}_2^{(b)}, \ldots, \hat{\varphi}_{N_2}^{(b)}\right)^T$ 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 $\left\{\hat{m}_1^{(b)} (y)\right\}_{b=1}^{n_B}$. The wild bootstrap 100$(1-\alpha)$% pointwise confidence interval for function value $m_1 (y)$ at one point $y$ is $\{L_{\alpha/2} (y), U_{\alpha/2} (y)\}$.

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 $F_\alpha = 2\varepsilon_{1-\alpha/2}^{-1} \{\log (N_2 + 1)\}^{1/2} Q_N (\alpha)$. We define the $(1-\alpha)$% bootstrap confidence band for $m_1 (y)$ as $[\hat{m}_1 (y) + \{L_{\alpha/2} (y) - \hat{m}_1 (y)\} F_\alpha, \hat{m}_1 (y) + \{U_{\alpha/2} (y) - \hat{m}_1 (y)\} F_\alpha]$. 

**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 $N_2 = \left[c_3 n^{1/5} \log(n)\right] + c_4$ and 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 defined in 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 $\{\sigma_t^2\}_{t=1}^T$ was from the following models:

$A : \sigma_t^2 = 0.10 + 0.20 Y_{t-1}^2 + 0.75 \sigma_{t-1}^2,$

$B : \sigma_t^2 = 0.05 + 0.20 Y_{t-1}^2 + 0.05 Y_{t-1}^2 I (Y_{t-1} < 0) + 0.75 \sigma_{t-1}^2,$

$C : \sigma_t^2 = 1 - 0.90 \exp(-2Y_{t-1}^2) + 0.70 \sigma_{t-1}^2,$

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 and a similar model has been studied by Carroll, et al. (2002) and Bühlmann and McNeil (2002).

We first considered time series from models $A$, $B$ and $C$ with $J_{\text{model}} = 5$. For $T = 500$, 1000, 2000 and 3000, we generated 200 replications for the above three processes of size $T + 1000$. Then the first 1000 observations were discarded to make sure the time series behave like 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 $\beta_0$ 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.

(Insert Table 1 about here)

The 3rd to the 5th columns in Table 1 provide the sample mean (MEAN), standard deviation (STD) and mean squared errors (MSE) of $\hat{\beta}$ based on the GARCH-ADD, GARCH-LS and GARCH-NL methods. As we expected, when the sample size increases, the parameter $\beta_0$ is more accurately estimated, with smaller MSE, confirmative to the conclusions of The-
orem 1. As one referee expected, the GARCH-LS and GARCH-NL estimators provide more accurate estimation in some cases, especially for Model C. While, we did not see obvious advantage using these model structures for Models A and B. The mean and median of selected number of lags $J_{\text{fit}}$ were reported in the last column of Table 1, and one sees that $J_{\text{fit}}$ is close to $J_{\text{model}} = 5$ for moderately large sample size. For the news impact curve estimation, in our simulation 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 $[-2.0, 2.0]^{J_{\text{fit}}}$ for $\hat{m}_1^*$.

Next, 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 three different models above. 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.

(Insert Table 2 about here)

We also carried out simulation considering model misspecification, and we generated time series from models A and B with $J_{\text{model}} = \infty$. Remember that for $J_{\text{model}} = \infty$, process A is a GARCH(1,1) process, so clearly GARCH(1,1) is the preferred estimator in this case. For process B, a 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 selected finite number of lags; see results in Table 3. For $T = 1000$, the nonparametric methods lost a small amount of efficiency relative to the parametric ones. But that effect decreases as the sample size increases for both processes A and B. Overall, we find that the GARCH-ADD works quite robust though $\beta_0$ is not the true parameter anymore. One explanation is that the selected number of lags based on our method is usually also large when $J_{\text{model}} = \infty$. 
In all our simulation experiments, our proposed GARCH-ADD method worked very fast, and we provide the time in seconds for all the methods in the last column 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 works much faster compared to its competitors which involve 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 1955 to January 30th 1994. There were a total of 2000 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 $\frac{1}{n} \sum_{t=J+1}^{T} (\hat{\sigma}_t^2 - Y_t^2)^2$ and the log-likelihood $-\sum_{t=J+1}^{T} \log \left\{ \hat{\sigma}_t^{-1} \varphi \left( \frac{Y_t}{\hat{\sigma}_t} \right) \right\}$ with $J_{fit} = 50$. 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 the 95% confidence band. From Figure 2, we find that the spline estimated news impact curve stands obvious contrast to the GARCH(1,1) fit, which shows strong evidence of the asymmetry of the news impact curve. But it seems that all three models can be fully covered by the
bootstrap band.

\( (\text{Insert Figure 1 about here}) \)

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

\( (\text{Insert Figure 2 about here}) \)

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 analysis. In this paper, we have advanced semi-parametric methods as flexible, computationally efficient and theoretically attractive tools for studying the 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 spline basis is chosen. Thus our method is particularly computationally efficient compared to its competitors which have to solve big 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 it as future research work. All the proposed estimators are easily implemented in commonly used software/package such as \texttt{lm()} in R.

There is more future work ahead. For example, it is interesting to consider the issue of model misspecification. In this paper, instead of estimating the true dynamic coefficient for \( J = \infty \), we estimate a parameter \( \beta_0 \) that approximates the true parameter by using some finite \( J \). If \( J = \infty \), \( \beta_0 \) would not be the true dynamic coefficient anymore. The asymptotic
results for the misspecified case has to be more fully explored.

**Acknowledgment**

The research work of the first author was partially supported by NSF grant DMS-0905730. The research work of the fourth author was supported in part by NSF grants DMS-0706518 and DMS-1007594.

**Appendix**

Throughout this section, we denote $c, C$ any positive constants, without distinction. Denote $\|\phi\|_2^n$ the theoretical $L^2$ norm of a function $\phi$ on $[a, b]$, $\|\phi\|_2^n = \int_a^b \phi^2(y) f(y) dy$, and define the empirical $L^2$ norm as $\|\phi\|_2^n = \frac{1}{n-1} \sum_{i=1}^n \phi^2(Y_i)$. The corresponding inner products are defined by $\langle \phi, \varphi \rangle_2 = \int_a^b \phi(y) \varphi(y) f(y) dy$ and $\langle \phi, \varphi \rangle_2^n = \frac{1}{n-1} \sum_{i=1}^n \phi(Y_i) \varphi(Y_i)$.

Define the centered version spline basis

$$b^*_j,k(y) = b_j,k(y) - E(b_j,k), j = 1, \ldots, J, \quad k = 1 - p, \ldots, N,$$

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

$$B_j,k(y) = \frac{b^*_j,k(y)}{\|b^*_j,k\|_2}. \quad (A.1)$$

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

$$B(x) = \{1, B_{1-1}(x_1), \ldots, B_{J,N}(x_J)\}^T, \quad B = \{B(X_{J+1}), \ldots, B(X_T)\}^T.$$

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

$$A_j = \text{diag}\{0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0\}$$

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

\[
\begin{align*}
\hat{m}_j(y) &= B(y)^T \Lambda_j (B^T B)^{-1} B^T \bar{m} - \frac{1}{n} 1_n^T B \Lambda_j (B^T B)^{-1} B^T m, \\
\hat{\epsilon}_j(y) &= B(y)^T \Lambda_j (B^T B)^{-1} B^T \epsilon - \frac{1}{n} 1_n^T B \Lambda_j (B^T B)^{-1} B^T \epsilon,
\end{align*}
\]

(A.2)

where \(1_n\) is a length \(n\) dimensional vector with all elements 1.

Defining \(Z = \{Y_{J+1}^2, ..., Y_T^2\}\), we can rewrite \(\hat{m}_j(y)\) in (2.3) using matrix as

\[
\hat{m}_j(y) = B(y)^T \Lambda_j (B^T B)^{-1} B^T Z - \frac{1}{n} 1_n^T B \Lambda_j (B^T B)^{-1} B^T Z.
\]

Then one has the following crucial decomposition for proving Theorem 1,

\[
\hat{m}_j(y) = \hat{m}_j(y) + \hat{\epsilon}_j(y), \quad j = 1, ..., J.
\]

(A.3)

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

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

\[
\|\hat{m}_1 - m_1\|_{2,n} + \|\hat{m}_1 - m_1\|_2 = O_{a.s.} \left( h^p + \log n / \sqrt{nh} \right).
\]

(A.4)

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

The following corollary states the asymptotic property of \(\hat{m}_1^*\) given in (2.5) to \(m_1\).

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

\[
\|\hat{m}_1^* - m_1\|_{2,n} + \|\hat{m}_1^* - m_1\|_2 = O_P \left( h^p + \log n / \sqrt{nh} \right).
\]
Proof. The proof is quite straightforward from the above lemma and Theorem 2.

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

For each \( j \), \( \| \hat{m}_j - m_j \|_2 \) has order \( OP \left( h^p + \log n/\sqrt{nh} \right) \). Combining with the result that \( \sum_{j=1}^J \beta^{j-1} (\beta_0^{j-1} - \beta^{j-1}) \) is with the order \( OP \left( h^p + \log n/\sqrt{nh} \right) \) from Theorem 2, we can easily obtain that \( \| \hat{m}_1^* - m_1 \|_2 \) is with the order \( OP \left( h^p + \log n/\sqrt{nh} \right) \), 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 on \( \beta \) and hence, consistency for \( \beta \) can be 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

\[
\hat{R}(\beta) = \sum_{j=1}^J \| \hat{m}_j - m_j + \beta^{j-1} m_1 - \beta^{j-1} \hat{m}_1 \|_{2,n}^2 + \sum_{j=1}^J \| m_j - \beta^{j-1} m_1 \|_{2,n}^2 \\
+ \sum_{j=1}^J 2 \langle m_j - \beta^{j-1} m_1, \hat{m}_j - m_j + \beta^{j-1} m_1 - \beta^{j-1} \hat{m}_1 \rangle_{2,n} = P_1(\beta) + P_2(\beta) + P_3(\beta).
\]

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

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

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

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

While

\[
\sup_{\beta \in [\beta_1, \beta_2]} | P_2(\beta) - R(\beta) | \\
\leq \left\| \frac{1}{n} \sum_{t=J+1}^T \left\{ \sum_{j=1}^J m_j^2(Y_t) \right\} - E \left\{ \sum_{j=1}^J m_j^2(Y_t) \right\} \right\| + \frac{1 - \beta_2^J}{1 - \beta_2} \left\| \frac{1}{n} \sum_{t=J+1}^T m_1^2(Y_t) - E m_1^2(Y_t) \right\| \\
+ 2 \frac{1 - \beta_2^J}{1 - \beta_2} \left\| \frac{1}{n} \sum_{t=J+1}^T \left\{ \sum_{j=1}^J m_j(Y_t) m_1(Y_t) \right\} - E \left\{ \sum_{j=1}^J m_j(Y_t) m_1(Y_t) \right\} \right\|. 
\]
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
\[
\sup_{\beta \in [\beta_1, \beta_2]} \left| \hat{R}(\beta) - R(\beta) \right| \leq \sup_{\beta \in [\beta_1, \beta_2]} \left| \hat{R}(\beta) - P_2(\beta) \right| + \sup_{\beta \in [\beta_1, \beta_2]} |P_2(\beta) - R(\beta)| = o_{a.s.}(1),
\]
and \(\hat{\beta}\) converges to \(\beta_0\) a.s. is followed.

**A.2 Proof of Theorem 2**

We make a Taylor expansion about \(\frac{d}{d\beta} \hat{R}(\beta)\) at \(\beta_0\),
\[
\sqrt{n} \frac{d}{d\beta} \hat{R}(\hat{\beta}) = \sqrt{n} \frac{d}{d\beta} \hat{R}(\beta) \bigg|_{\beta=\beta_0} + \frac{d^2}{d\beta^2} \hat{R}(\hat{\beta}) \bigg|_{\beta=\hat{\beta}} \sqrt{n} (\hat{\beta} - \beta_0),
\]
where \(\hat{\beta}\) is between \(\hat{\beta}\) and \(\beta_0\). Thus one has
\[
\sqrt{n} (\hat{\beta} - \beta_0) = \sqrt{n} \left\{ \frac{d^2}{d\beta^2} \hat{R}(\hat{\beta}) \bigg|_{\beta=\hat{\beta}} \right\}^{-1} \left\{ \frac{d}{d\beta} \hat{R}(\hat{\beta}) - \frac{d}{d\beta} \hat{R}(\beta) \bigg|_{\beta=\beta_0} \right\}.
\]
We need the following two lemmas to deal with \(-\sqrt{n} \frac{d}{d\beta} \hat{R}(\beta)\bigg|_{\beta=\beta_0}\) and \(\frac{d^2}{d\beta^2} \hat{R}(\beta)\) respectively.

**Lemma A.2.** Under Assumptions (A1)-(A5),
\[
-\frac{\sqrt{n}}{2} \frac{d}{d\beta} \hat{R}(\beta) \bigg|_{\beta=\beta_0} = n^{-1/2} \sum_{t=J+1}^{T} \epsilon_t H(\beta_0, m_1(Y_t)),
\]
where \(H(\beta_0, m_1(Y_t))\) is given in (A.9).

**Proof.** Note that
\[
-\frac{\sqrt{n}}{2} \frac{d}{d\beta} \hat{R}(\beta) \bigg|_{\beta=\beta_0} = n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^{j-2} \left\{ \hat{m}_j(Y_t) - \beta_0^{j-1} \hat{m}_1(Y_t) \right\} \hat{m}_1(Y_t)
\]
\[
= n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^{j-2} \left\{ \hat{m}_j(Y_t) - \beta_0^{j-1} \hat{m}_1(Y_t) \right\} [\hat{m}_1(Y_t) - m_1(Y_t)]
\]
\[
+ n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^{j-2} \left\{ \hat{m}_j(Y_t) - \beta_0^{j-1} \hat{m}_1(Y_t) \right\} m_1(Y_t)
\]
\[
= I + II,
\]

\[
I = n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^{j-2} \left\{ \hat{m}_j(Y_t) - \beta_0^{j-1} \hat{m}_1(Y_t) \right\} \hat{m}_1(Y_t)
\]
\[
II = n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^{j-2} \left\{ \hat{m}_j(Y_t) - \beta_0^{j-1} \hat{m}_1(Y_t) \right\} [\hat{m}_1(Y_t) - m_1(Y_t)]
\]
and the first term $I$ can be written as
\[
I = n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^j \{ \hat{m}_j(Y_t) - m_j(Y_t) \} \{ \hat{m}_1(Y_t) - m_1(Y_t) \}
\]
\[-n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^{j-2} \{ \hat{m}_j(Y_t) - m_j(Y_t) \}^2
\]
\[= I_1 + I_2.
\]
As in the proof of Theorem 1, we have both $I_1$ and $I_2$ with order $O_{a.s.}(h^{2p} + \frac{\log^2 n}{nh})$. With the order of $h$ in Assumption (A5), we have $I = O_{a.s.}\left \{ n^{1/2} \left( h^{2p} + \frac{\log^2 n}{nh} \right) \right \} = o_{a.s.}(1)$. For part $II$, noting that $\| \hat{m}_1 - m_1 \|_{2,n} = O(h^p)$, a.s., so by (A.3) and Assumption (A5) we have
\[
II = n^{-1/2} \sum_{t'=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^j \{ \hat{m}_j(Y_{t'}) + \hat{e}_j(Y_{t'}) - \beta_0^{j-2} \hat{m}_1(Y_{t'}) \}
\]
\[= n^{-1/2} \sum_{t'=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^j \{ \hat{e}_j(Y_{t'}) - \beta_0^{j-1} \hat{e}_1(Y_{t'}) \} \{ m_1(Y_{t'}) + o_{a.s.}(1) \}.
\]
Let
\[
\tilde{V} = \frac{1}{n} B^\top B = \begin{pmatrix} 1 & 0 \\ 0 & \langle B_{j,k}, B_{j',k'} \rangle_{2,n} \end{pmatrix}_{1 \leq j,j' \leq J, 1-p \leq k,k' \leq N}, \quad V = \begin{pmatrix} 1 & 0 \\ 0 & \langle B_{j,k}, B_{j',k'} \rangle_{2,n} \end{pmatrix}_{1 \leq j,j' \leq L - \frac{1}{p} \leq k,k' \leq N}.
\]
With $\hat{e}_j$ defined in (A.2), the main term in $II$ is
\[
n^{-1/2} \sum_{t'=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^j \{ \hat{m}_1(Y_{t'}) \}
\]
\[-\beta_0^{j-1} \left\{ \mathbf{B}(Y_{t'}) \mathbf{A}_1 - \frac{1}{n} 1_n^\top \mathbf{B} \mathbf{A}_1 \right\} \tilde{V}^{-1} \left\{ \frac{1}{n} \sum_{t=J+1}^{T} B_{j,k}(Y_t) \hat{e}_t \right\}_{j,k},
\]
\[= \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^j \{ \mathbf{B}(Y_t) \mathbf{A}_{j-1} - \frac{1}{n} 1_n^\top \mathbf{B} \mathbf{A}_{j-1} \}
\]
\[-\beta_0^{j-1} \left\{ \mathbf{B}(Y_t) \mathbf{A}_1 - \frac{1}{n} 1_n^\top \mathbf{B} \mathbf{A}_1 \right\} \tilde{V}^{-1} \left\{ \frac{1}{n} \sum_{t'=J+1}^{T} B_{j,k}(Y_{t'}) \hat{m}_1(Y_{t'}) \right\}_{j,k}.
\]
(7.7)

According to Lemma A.10 in Wang and Yang (2007), we can replace $\tilde{V}$ by $V$ in equation (A.7) with a negligible term $O_{a.s.} \left \{ n^{-1/2} N (\log n)^2 \right \}$. Next we interchange the indices $t$ and $t'$ of (A.7), thus the main term in $II$ can be approximated by
\[
n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1) \beta_0^j \{ \mathbf{B}(Y_t) \mathbf{A}_{j-1} - \frac{1}{n} 1_n^\top \mathbf{B} \mathbf{A}_{j-1} \}
\]
\[-\beta_0^{j-1} \left\{ \mathbf{B}(Y_t) \mathbf{A}_1 - \frac{1}{n} 1_n^\top \mathbf{B} \mathbf{A}_1 \right\} V^{-1} \left\{ \frac{1}{n} \sum_{t'=J+1}^{T} B_{j,k}(Y_{t'}) \hat{m}_1(Y_{t'}) \right\}_{j,k}.
\]
(8.8)
Denote
\[
H(\beta_0, m_1(Y_t)) = \sum_{j=1}^{J} \left( (j - 1) \beta_0^{j-2} \left[ \left\{ B(Y_t) A_j - \frac{1}{n} J_n^2 B A_j \right\} \right. \right. \\
\left. \left. - \beta_0^{j-1} \left\{ B(Y_t) A_1 - \frac{1}{n} J_n^2 B A_1 \right\} \right] V^{-1} \left\{ \frac{1}{n} \sum_{t'=J+1}^{T} B_{j,k}(Y_{t'}) m_1(Y_{t'}) \right\} \right].
\] (A.9)

and equation (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), \( \frac{d^2}{d\beta^2} \hat{R}(\beta) = Em_1^2(Y_t) \sum_{j=2}^{J} (j - 1) \beta^{j-2} \left\{ \hat{m}_j(Y_t) - \beta^j \hat{m}_1(Y_t) \right\} \) + o.a.s. (1).

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

where \( I_2 = o.a.s.(1) \) similarly as for \( I \) in Lemma (A.2), and for \( I_1 \),
\[
n^{-1} \sum_{t=J+1}^{T} \hat{m}_1^2(Y_t) - n^{-1} \sum_{t=J+1}^{T} m_1^2(Y_t) = n^{-1} \sum_{t=J+1}^{T} \{ \hat{m}_1(Y_t) - m_1(Y_t) \} \{ \hat{m}_1(Y_t) + m_1(Y_t) \} \\
\leq \left( n^{-1} \sum_{t=J+1}^{T} (\hat{m}_1(Y_t) - m_1(Y_t))^2 \right)^{1/2} \left( n^{-1} \sum_{t=J+1}^{T} (\hat{m}_1(Y_t) + m_1(Y_t))^2 \right)^{1/2} \\
\leq \| \hat{m}_1(x) - m_1(x) \|_{2,n} \sup_x |\sqrt{\theta} \hat{m}_1(x)| = O_a.s. \left( \frac{h^p + \log n}{\sqrt{n h}} \right) = o_a.s.(1).
\] (A.10)

By a law of large numbers, we have \( n^{-1} \sum_{t=J+1}^{T} m_1^2(Y_t) \rightarrow E \left[ m_1^2(Y_t) \right] \), as \( n \) goes to infinity. Combining with (A.10), we have \( \lim_{n \rightarrow \infty} n^{-1} \sum_{t=J+1}^{T} \hat{m}_1^2(Y_t) = E \left[ m_1^2(Y_t) \right] \).

Now, we continue the proof of Theorem 2. Combining (A.5), (A.6) and Lemma A.3, and noting that as \( n \rightarrow \infty, \sum_{j=2}^{J} (j - 1)^2 \beta_0^{2j-4} \rightarrow \sum_{j=2}^{J} (j - 1)^2 \beta_0^{2j-4}, \) a.s., we have,
\[
n^{1/2} \left( \hat{\beta} - \beta_0 \right) = n^{-1/2} \sum_{t} \epsilon_t H(\beta_0, m_1(Y_t)) \left\{ \sum_{j=2}^{J} (j - 1)^2 \beta_0^{2j-4} E \left[ m_1^2(Y_t) \right] \right\}^{-1} + o_a.s.(1).
\]
Asymptotic normality of \( n^{1/2} \left( \hat{\beta} - \beta_0 \right) \) follows from a Slutsky theorem and a central limit theorem for strongly mixing sequences (see, e.g., Bosq (1996), Theorem 1.7). We have to verify that for some \( \nu > 2 \), \( E |\epsilon_t H(\beta_0, m_1(Y_t))|^{\nu} < \infty \), which can be obtained with our Assumptions (A2) and (A4).

References


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, People’s Republic of China; and Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

E-mail: (yanglijian@suda.edu.cn)
Table 1: Monte 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 $J_{\text{fit}}$ and the oracle $J_{\text{oracle}} = 5$.

<table>
<thead>
<tr>
<th>Size</th>
<th>Estimator</th>
<th>Parametric component</th>
<th>Nonparametric component</th>
<th>$J_{\text{fit}}$</th>
<th>mean (median)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MEAN</td>
<td>STD</td>
<td>MSE</td>
<td>AMSE</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>GARCH-ADD</td>
<td>0.68 (0.63)</td>
<td>0.17 (0.14)</td>
<td>0.035 (0.033)</td>
<td>0.026 (0.025)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.81 (0.78)</td>
<td>0.15 (0.15)</td>
<td>0.027 (0.024)</td>
<td>0.034 (0.031)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.80 (0.79)</td>
<td>0.16 (0.15)</td>
<td>0.027 (0.024)</td>
<td>0.033 (0.032)</td>
</tr>
<tr>
<td>1000</td>
<td>GARCH-ADD</td>
<td>0.74 (0.70)</td>
<td>0.14 (0.11)</td>
<td>0.019 (0.017)</td>
<td>0.016 (0.017)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.80 (0.80)</td>
<td>0.11 (0.10)</td>
<td>0.014 (0.013)</td>
<td>0.025 (0.024)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.80 (0.82)</td>
<td>0.11 (0.10)</td>
<td>0.015 (0.016)</td>
<td>0.025 (0.026)</td>
</tr>
<tr>
<td>2000</td>
<td>GARCH-ADD</td>
<td>0.75 (0.73)</td>
<td>0.10 (0.08)</td>
<td>0.009 (0.008)</td>
<td>0.011 (0.012)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.78 (0.80)</td>
<td>0.08 (0.08)</td>
<td>0.008 (0.008)</td>
<td>0.017 (0.017)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.79 (0.81)</td>
<td>0.08 (0.08)</td>
<td>0.008 (0.010)</td>
<td>0.018 (0.019)</td>
</tr>
<tr>
<td>3000</td>
<td>GARCH-ADD</td>
<td>0.77 (0.75)</td>
<td>0.07 (0.07)</td>
<td>0.005 (0.004)</td>
<td>0.011 (0.011)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.78 (0.80)</td>
<td>0.08 (0.07)</td>
<td>0.007 (0.007)</td>
<td>0.015 (0.015)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.79 (0.81)</td>
<td>0.08 (0.06)</td>
<td>0.008 (0.008)</td>
<td>0.015 (0.016)</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>GARCH-ADD</td>
<td>0.73 (0.71)</td>
<td>0.15 (0.14)</td>
<td>0.022 (0.022)</td>
<td>0.413 (0.348)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.81 (0.80)</td>
<td>0.13 (0.13)</td>
<td>0.021 (0.019)</td>
<td>0.173 (0.170)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.81 (0.82)</td>
<td>0.12 (0.11)</td>
<td>0.019 (0.016)</td>
<td>0.175 (0.145)</td>
</tr>
<tr>
<td>1000</td>
<td>GARCH-ADD</td>
<td>0.77 (0.77)</td>
<td>0.10 (0.10)</td>
<td>0.011 (0.010)</td>
<td>0.177 (0.163)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.81 (0.81)</td>
<td>0.10 (0.09)</td>
<td>0.013 (0.012)</td>
<td>0.100 (0.108)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.80 (0.83)</td>
<td>0.09 (0.07)</td>
<td>0.011 (0.011)</td>
<td>0.095 (0.089)</td>
</tr>
<tr>
<td>2000</td>
<td>GARCH-ADD</td>
<td>0.77 (0.78)</td>
<td>0.08 (0.07)</td>
<td>0.006 (0.006)</td>
<td>0.118 (0.116)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.79 (0.81)</td>
<td>0.08 (0.07)</td>
<td>0.008 (0.008)</td>
<td>0.079 (0.088)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.79 (0.82)</td>
<td>0.07 (0.05)</td>
<td>0.007 (0.008)</td>
<td>0.074 (0.076)</td>
</tr>
<tr>
<td>3000</td>
<td>GARCH-ADD</td>
<td>0.78 (0.79)</td>
<td>0.06 (0.06)</td>
<td>0.005 (0.005)</td>
<td>0.093 (0.098)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.79 (0.81)</td>
<td>0.07 (0.06)</td>
<td>0.007 (0.007)</td>
<td>0.075 (0.082)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.79 (0.82)</td>
<td>0.06 (0.04)</td>
<td>0.006 (0.007)</td>
<td>0.069 (0.072)</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>GARCH-ADD</td>
<td>0.57 (0.55)</td>
<td>0.16 (0.12)</td>
<td>0.042 (0.036)</td>
<td>0.179 (0.079)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.79 (0.72)</td>
<td>0.18 (0.18)</td>
<td>0.040 (0.031)</td>
<td>0.106 (0.085)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.76 (0.71)</td>
<td>0.19 (0.19)</td>
<td>0.038 (0.034)</td>
<td>0.102 (0.083)</td>
</tr>
<tr>
<td>1000</td>
<td>GARCH-ADD</td>
<td>0.60 (0.58)</td>
<td>0.16 (0.11)</td>
<td>0.036 (0.029)</td>
<td>0.147 (0.055)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.75 (0.71)</td>
<td>0.15 (0.13)</td>
<td>0.026 (0.018)</td>
<td>0.075 (0.057)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.73 (0.71)</td>
<td>0.15 (0.14)</td>
<td>0.024 (0.020)</td>
<td>0.069 (0.057)</td>
</tr>
<tr>
<td>2000</td>
<td>GARCH-ADD</td>
<td>0.64 (0.61)</td>
<td>0.13 (0.09)</td>
<td>0.021 (0.016)</td>
<td>0.107 (0.033)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.73 (0.71)</td>
<td>0.10 (0.04)</td>
<td>0.011 (0.010)</td>
<td>0.044 (0.038)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.72 (0.71)</td>
<td>0.11 (0.05)</td>
<td>0.013 (0.012)</td>
<td>0.046 (0.040)</td>
</tr>
<tr>
<td>3000</td>
<td>GARCH-ADD</td>
<td>0.67 (0.64)</td>
<td>0.11 (0.07)</td>
<td>0.013 (0.009)</td>
<td>0.069 (0.020)</td>
</tr>
<tr>
<td></td>
<td>GARCH-LS</td>
<td>0.71 (0.70)</td>
<td>0.09 (0.08)</td>
<td>0.007 (0.006)</td>
<td>0.030 (0.027)</td>
</tr>
<tr>
<td></td>
<td>GARCH-NL</td>
<td>0.71 (0.71)</td>
<td>0.09 (0.08)</td>
<td>0.008 (0.007)</td>
<td>0.033 (0.030)</td>
</tr>
</tbody>
</table>
Table 2: Coverage probabilities from 500 replications.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95%</td>
<td>99%</td>
<td>95%</td>
</tr>
<tr>
<td>500</td>
<td>0.982</td>
<td>0.998</td>
<td>0.924</td>
</tr>
<tr>
<td>1000</td>
<td>0.970</td>
<td>0.996</td>
<td>0.962</td>
</tr>
<tr>
<td>2000</td>
<td>0.982</td>
<td>0.996</td>
<td>0.956</td>
</tr>
<tr>
<td>3000</td>
<td>0.976</td>
<td>0.996</td>
<td>0.932</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Size</th>
<th>Estimator</th>
<th>Parametric component</th>
<th>Nonparametric component</th>
<th>Time (secs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MEAN</td>
<td>STD</td>
<td>MSE</td>
</tr>
<tr>
<td>A</td>
<td>1000</td>
<td>GARCH(1,1)</td>
<td>0.77</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GJR(1,1)</td>
<td>0.77</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-ADD</td>
<td>0.71</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-LS</td>
<td>0.81</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-NL</td>
<td>0.82</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>GARCH(1,1)</td>
<td>0.77</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GJR(1,1)</td>
<td>0.77</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-ADD</td>
<td>0.75</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-LS</td>
<td>0.80</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-NL</td>
<td>0.81</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>GARCH(1,1)</td>
<td>0.77</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GJR</td>
<td>0.77</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-ADD</td>
<td>0.76</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-LS</td>
<td>0.80</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-NL</td>
<td>0.81</td>
<td>0.03</td>
</tr>
<tr>
<td>B</td>
<td>1000</td>
<td>GARCH(1,1)</td>
<td>0.76</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GJR(1,1)</td>
<td>0.76</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-ADD</td>
<td>0.76</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-LS</td>
<td>0.82</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-NL</td>
<td>0.84</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>GARCH(1,1)</td>
<td>0.76</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GJR(1,1)</td>
<td>0.76</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-ADD</td>
<td>0.78</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-LS</td>
<td>0.82</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-NL</td>
<td>0.83</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>GARCH(1,1)</td>
<td>0.76</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GJR(1,1)</td>
<td>0.76</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-ADD</td>
<td>0.79</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-LS</td>
<td>0.82</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH-NL</td>
<td>0.83</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Table 4: Fitting the BMW daily returns

<table>
<thead>
<tr>
<th>Model</th>
<th>−Log-likelihood</th>
<th>Volatility prediction error</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>3394.667</td>
<td>22.589</td>
</tr>
<tr>
<td>GJR(1,1)</td>
<td>3387.449</td>
<td>22.065</td>
</tr>
<tr>
<td>GARCH-ADD</td>
<td>3387.310</td>
<td>21.759</td>
</tr>
</tbody>
</table>
Figure 1: BMW daily returns: (a) original series; (b) the estimated volatility function.
Figure 2: Spline ARCH(∞) modelling of BMW daily returns: (a) estimated news impact curve; (b) the estimated ACF along with 95% Bartlett intervals for $\hat{\epsilon}^2$. 