Bayesian Analysis (2014) **TBA**, Number TBA, pp. 1–26

Bayesian Analysis of the Functional-Coefficient Autoregressive Heteroscedastic Model

Xin-Yuan Song, * Jing-Heng Cai, † Xiang-Nan Feng ‡ and Xue-Jun Jiang §

Abstract. In this paper, we propose a new model called the functional-coefficient autoregressive heteroscedastic (FARCH) model for nonlinear time series. The FARCH model extends the existing functional-coefficient autoregressive models and double-threshold autoregressive heteroscedastic models by providing a flexible framework for the detection of nonlinear features for both the conditional mean and conditional variance. We propose a Bayesian approach, along with the Bayesian P-splines technique and Markov chain Monte Carlo algorithm, to estimate the functional coefficients and unknown parameters of the model. We also conduct model comparison via the Bayes factor. The performance of the proposed methodology is evaluated via a simulation study. A real data set derived from the daily S&P 500 Composite Index is used to illustrate the methodology.

Keywords: Nonlinear time series, Autoregressive heteroscedastic models, Bayesian P-splines;, MCMC methods, Bayes factor

1 Introduction

Since the pioneering work of the autoregressive moving average (ARMA) models (Box et al. 2008) to analyze time series data, the original ARMA framework has been extended in numerous aspects. Examples of such extensions include the fractional ARMA (Granger and Joyeux 1980; Dahlhaus 1989) and (multivariate) vector ARMA models with exogenous variables (Hannan and Deistler 1988). However, most of these analyses were restricted to linear modeling. Given the existence of various nonlinear phenomena, such as asymmetric cycles and bimodality, nonlinear relationships among lagged variables have been observed in real time series data sets (see Tjøstheim 1994; Tong 1990, 1995). Moreover, these nonlinear features are beyond the capacity of linear models. Thus, nonlinear time series analysis has received a great deal of attention over the past thirty years. At the early stage of nonlinear time series development, analyses focused on known parametric forms, such as the threshold autoregressive (TAR) model (Tong 1990) and the exponential autoregressive (EXPAR) model (Haggan and Ozaki 1981). An evident limitation of these parametric (both linear and nonlinear) models is that they are too restrictive for many applications since the functional forms for the

*Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, xysong@sta.cuhk.edu.hk

†Department of Statistics, Sun Yat-sen University, Guangzhou, China, caijheng@mail.sysu.edu.cn

‡Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, fengxiangnan123@gmail.com

§Department of Financial Mathematics and Financial Engineering, South University of Science and Technology, Shenzhen, China, jiangxjyn05@gmail.com

 DOI:10.1214/13-BA865

relationships among variables have to be specified. However, nonlinear functions have many different types, and the functional form of a specific relationship is seldom known in advance. In addition, strict parametric functional forms are likely to miss subtle patterns. To relax the restrictions in parametric models, functional-coefficient autoregressive (FAR) models have been developed recently. The advantage of FAR models lies in the flexibility to accommodate most nonlinear features with only minimal prior information assumed. Further, such models can be employed as an exploratory tool for investigating functional forms. Given their flexibility in dealing with nonlinear relationships, FAR models have been widely used over the past decades (see, for example, Chen and Tsay 1993; Cai et al. 2000; Fan and Yao 2003, among others).

Volatility is important in asset pricing, monetary policymaking, proprietary trading, portfolio management, and risk analysis. Thus, modeling and predicting volatility is of great importance. A widely used approach for modeling volatility is the discrete time method, in which the volatility is considered to be a conditional variance of the return. Since the autoregressive conditional heteroscedasticity (ARCH) model was first introduced by Engle (1982) as an important tool for modeling volatility, numerous variants of the ARCH model have been proposed. Bollerslev (1986) extended the ARCH model into a generalized ARCH (GARCH) model. To capture the nonlinear features in real applications, Tong (1990) proposed a self-exciting threshold ARCH (SETAR-ARCH) model with changing conditional variance, which has a piecewise linear conditional mean and an ARCH innovation. Glosten et al. (1993) and Zakoian (1994) developed the threshold GARCH (TGARCH) model to capture different effects of the positive and negative parts of the past noises on the conditional variance. Li and Li (1996) subsequently extended the TAR model to the double-threshold ARCH (DTARCH) model, which can address conditions where both the conditional mean and the conditional variance specifications are piecewise linear given previous information. They studied model identification, estimation, and diagnostic verification using the maximum likelihood (ML) method with the normal assumption of innovations. The DTARCH model is useful for detecting nonlinear structures such as asymmetric behavior in the mean and the volatility of an asset return as well as heteroscedasticity with clustering in the volatility. Brooks (2001) extended the DTARCH model to the double-threshold GARCH (DTGARCH) model which uses a GARCH rather than ARCH specification for the conditional variance. Further, Hui and Jiang (2005) relaxed the normal assumption and investigated similar problems based on L_1 regression in the DTARCH setting. Although their method is resistant to outliers and robust against error distribution, it is inefficient when the error is normal. More recently, Jiang et al. (2013) proposed a weighted composite quantile regression method to analyze DTARCH models. They highlighted that the proposed method uniformly dominated the L_1 estimation and nearly reached the efficiency of the oracle ML method with known innovations.

Although the aforementioned models are useful for describing the conditional mean and/or conditional variance, they share the limitation of parametric models in that specific parametric forms may be too restrictive to reveal the true conditions. In this paper, we consider the functional-coefficient autoregressive heteroscedastic (FARCH) model. The proposed integrated model framework is more general. Moreover, the

FARCH model can accommodate most nonlinear relationships and does not require a priori specified parametric functions for both the conditional mean and conditional variance. Local polynomial regression (Fan and Gijbels 1996) coupled with the ML method is apparently a natural choice for analyzing the proposed model. However, bandwidth selection in the estimation procedure is difficult to implement for models with a nonparametric formulation of volatility. Moreover, the nonparametric specification of both the conditional mean and conditional variance make the problems of hypothesis testing and other statistical inferences highly non-trivial.

A sampling-based Bayesian approach is a potential alternative to the ML method and has been widely applied in the analysis of various statistical models (see, for example, Gelman et al. 2004; Congdon 2006, among others). The most attractive features of the Bayesian approach are as follows. First, this approach enables the use of genuine prior information to achieve better results. Second, as highlighted by Scheines et al. (1999) and Dunson (2000), sampling-based Bayesian methods rely less on large-sample asymptotic theory, thereby having the potential to produce reliable results even with small sample sizes. Finally, compared with ML-based methods, the Bayesian approach is conceptually simple and easy to implement. For instance, the ML-based hypothesis testing procedures require the derivation of the asymptotic distribution of the test statistic, whereas Bayesian hypothesis testing can be implemented more easily through model selection with Bayesian statistics, such as the Bayes factor (Kass and Raftery 1995) and the deviance information criterion (DIC; Spiegelhalter et al. 2002). In the estimation of nonparametric functions, the local polynomial or other conventional smoothing techniques require a tedious selection of bandwidth or other tuning parameters, whereas the Bayesian estimation can be obtained via posterior sampling with Markov chain Monte Carlo (MCMC) methods, thereby avoiding the difficulties that may be encountered in ML-based methods. Owing to the rapid development of powerful computational technologies, the Bayesian P-splines approach (Lang and Brezger 2004), a Bayesian analogue of penalized splines, has been widely used to model unknown smoothing functions in nonparametric or semiparametric models (see, for example, Berry et al. 2002; Fahrmeir and Raach 2007; Song and Lu 2012, among others). In this paper, we use the Bayesian P-splines technique to estimate the nonparametric functions in the proposed FARCH model. In addition to the estimation, we also consider model selection (hypothesis testing). Natural questions about the current nonparametric modeling are whether the conditional mean and variance (scale) are actually varying, and whether the nonparametric model is actually better than its parametric and semiparametric counterparts. These questions motivate us to develop an efficient procedure for Bayesian model selection.

The main contributions of this paper are summarized. First, we extend the FAR and DTARCH models to the FARCH model. This modeling framework is novel and more flexible to capture the nonlinear phenomena of financial time series data. To the best of our knowledge, no study has been conducted on the statistical analysis of the FARCH model. Second, we develop a Bayesian approach coupled with MCMC methods to analyze the proposed model. The Bayesian estimation of nonparametric functions is obtained using the Bayesian P-splines approach, whereas the model selection for the

conditional mean and conditional variance (scale) is conducted via a Bayesian model selection statistic, the Bayes factor (Kass and Raftery 1995). Finally, we apply the proposed methodology to a study of the daily S&P 500 Composite Index. The nonlinear features of return and volatility are investigated.

The remainder of this paper is organized as follows. Section 2 defines the proposed FARCH models. Section 3 introduces the Bayesian P-splines approach, as well as the Bayesian estimation procedure coupled with MCMC algorithms. Bayesian model selection based on the Bayes factor and Bayesian forecasting via parametric bootstrap method are also discussed in this section. In Section 4, we conduct a simulation study to demonstrate the empirical performance of the Bayesian approach. The result of a real life example concerning the daily S&P 500 Composite Index is also reported. Section 5 concludes the paper with a discussion.

2 FARCH Model

Let $\{Y_t,\ t = 1, \cdots, T\}$ be a stationary and ergodic time series with $E(Y_t^2) < \infty$, we assume that Y_t's are generated from the following model,

$$Y_t = \alpha_1(Y_{t-d})Y_{t-1} + \alpha_2(Y_{t-d})Y_{t-2} + \cdots + \alpha_p(Y_{t-d})Y_{t-p} + \varepsilon_t, \tag{1}$$

where the delay parameter d is a positive integer, p is the AR order, and $\alpha_j(\cdot)$'s are unknown smoothing functions with second order derivatives. In addition, let $\mathcal{F}_{t-1}$ be the σ-field generated by the random variables $\{\varepsilon_{t-j},\ j = 1, 2, \cdots\}$. For each t, when information $\mathcal{F}_{t-1}$ is given, we assume that the stochastic error ε_t satisfies $\varepsilon_t = h_t(Y_{t-d})u_t$ with the conditional scale $h_t(Y_{t-d})$ defined as follows,

$$h_t(Y_{t-d}) = \beta_0(Y_{t-d}) + \beta_1(Y_{t-d})|\varepsilon_{t-1}| + \cdots + \beta_q(Y_{t-d})|\varepsilon_{t-q}|, \tag{2}$$

where q denotes the ARCH order, $\beta_j(\cdot)$'s are unknown smoothing functions satisfying $\beta_0(\cdot) > 0$ and $\beta_j(\cdot) \geq 0$ for $j = 1, \cdots, q$. The innovations u_t's are independently and identically distributed as $N(0, 1)$. Inspired by Li and Li (1996) and Fan and Yao (2003), we denote the model defined in (1) and (2) as FARCH(p, d, q). Notably, $h_t(Y_{t-d})$ is defined as a conditional scale rather than a conditional variance. Bickel and Lehmann (1976) emphasized that such a scale provides a more natural dispersion concept than variance as well as offers substantial advantages in terms of robustness. A detailed discussion can be found in Koenker and Zhao (1996) and Jiang et al. (2001). Other alternative modeling approaches for the volatility include

$$\log h_t(Y_{t-d}) = \beta_0(Y_{t-d}) + \beta_1(Y_{t-d})|\varepsilon_{t-1}| + \cdots + \beta_q(Y_{t-d})|\varepsilon_{t-q}|, \tag{3}$$

and

$$h_t(Y_{t-d}) = \beta_0(Y_{t-d}) + \beta_1(Y_{t-d})\varepsilon_{t-1}^2 + \cdots + \beta_q(Y_{t-d})\varepsilon_{t-q}^2. \tag{4}$$

While there are certainly advantages to using these alternatives, for instance, (3) releases the assumption of $\beta_0(\cdot) > 0$, $\beta_j(\cdot) \geq 0, j = 1, \cdots, q$ and (4) leads to a simpler posterior computation for the volatility model, our extensive simulation study shows that the

multiplicative modeling of volatility in (3) and the quadratic forms of ε_{t-j} in (4) make the conditional scale $h_t(\cdot)$ much more sensitive to the change of lagged stochastic errors, which tends to result in non-stationary series or unstable estimation and prediction results.

Yau and Kohn (2003) developed a functional-coefficient heteroscedastic regression model in a Bayesian framework. However, substantial differences exist between our model and Yau and Kohn (2003). First, Yau and Kohn (2003) considered a cross-sectional rather than a time series model. Second, the conditional mean and variance/scale are regressed on independent explanatory variables in Yau and Kohn (2003) but on lagged responses Y_{t-j} and lagged stochastic errors ε_{t-j} in our paper. The latter is much more difficult to handle in a Bayesian framework. Finally, the conditional variance/scale is modeled in the log scale in Yau and Kohn (2003) but in the original scale in our paper.

The proposed FARCH model naturally extends the FAR model in Chen and Tsay (1993) and the DTARCH model in Li and Li (1996). First, the conditional scale changes over time in the FARCH model but remains constant in the FAR model. This feature makes the FARCH model more appealing in capturing the dynamic change of volatility. Second, the unknown functions are piecewise linear in the DTARCH model but nonlinear and unspecified in the FARCH model. The nonparametric modeling framework provides more flexibility for reflecting the true condition in reality. The proposed model is able to reveal how the dynamic effects of historical time series values on the future mean value vary according to the lagged variable Y_{t-d} while also investigating how historical volatilities influence the future volatility dynamically according to the lagged variable Y_{t-d}. Moreover, unlike the DTARCH model, it is unnecessary to carefully choose/estimate the thresholds in a FARCH model. Thus, the FARCH model is proposed to reveal nonlinear phenomena, such as asymmetric cycles, jump resonance, and amplitude-frequency dependence, especially in financial time series data analysis.

One important issue in time series modeling is to impose appropriate constraints to ensure stationarity. Unfortunately, it is difficult to derive such general conditions on the functional-coefficients $\alpha_j(\cdot)$ and $\beta_j(\cdot)$. Inspired by the stationary conditions on the FAR and ARCH models, we propose an empirical rule for the FARCH model as follows: (i) $\alpha_j(\cdot)$ and $\beta_j(\cdot)$ are bounded such that $|\alpha_j(\cdot)|\leq c_j$, $j=1,\cdots,p$, and $|\beta_j(\cdot)|\leq d_j$, $j=1,\cdots,q$; (ii) all the roots of the characteristic function $x^p - c_1 x^{p-1} - \cdots - c_{p-1}x - c_p = 0$ are inside the unit circle, and (iii) $\sum_{j=1}^{q} d_j < 1$. Our extensive simulation studies show that a FARCH model satisfying the above conditions produces stationary and ergodic time series. Instead, when one or more of the above conditions are violated, for instance, taking $\alpha_j(u) = 1.5\cos(u-0.5)$ or $\beta_j(u) = 0.8 + (u-1)^2/30$ in a FARCH(2,2,1) model, the time series produced by the FARCH model are not stationary and most likely to diverge to infinity.

3 Bayesian Analysis of the FARCH Model

3.1 Bayesian P-splines

In the analysis of the FARCH model defined in the previous section, one important issue is the modeling of the unknown smooth functions in (1) and (2). Inspired by Eilers and Marx (1996) and Lang and Brezger (2004), we consider a Bayesian P-splines approach to estimate the unknown smooth functions. The basic idea of B-splines smoothing is to approximate the smooth functions $\alpha_j(\cdot)$ and $\beta_j(\cdot)$ in (1) and (2) by using a sum of B-splines (De Boor 2001) with a large number of knots in the domains of the Y_{t-d}'s. Specifically, $\alpha_j(Y_{t-d})$ in (1) can be approximated by

$$\alpha_j(Y_{t-d}) = \sum_{k=1}^{K_\lambda} \lambda_{jk} B_k^\lambda(Y_{t-d}) = \boldsymbol{\lambda}_j^T \mathbf{B}^\lambda(Y_{t-d}), \tag{5}$$

where K_λ is the number of splines determined by the number of knots, $\boldsymbol{\lambda}_j = (\lambda_{j1}, \cdots, \lambda_{jK_\lambda})^T$ is a vector of unknown parameters, $\mathbf{B}^\lambda(Y_{t-d}) = \left(B_1^\lambda(Y_{t-d}), \cdots, B_{K_\lambda}^\lambda(Y_{t-d})\right)^T$, and the functions $B_k^\lambda(\cdot)$ are B-splines basis functions with appropriate order. A natural choice of $B_k^\lambda(\cdot)$ is the cubic B-splines. Consequently, $\alpha_j(Y_{t-d})$ is a nonlinear function of Y_{t-d}. In practice, K_λ ranging from 10 to 30 provides sufficient flexibility for modeling $\alpha(\cdot)$. Similarly, $\beta_j(Y_{t-d})$ in (2) is approximated by

$$\beta_j(Y_{t-d}) = \sum_{k=1}^{K_\gamma} \gamma_{jk} B_k^\gamma(Y_{t-d}) = \boldsymbol{\gamma}_j^T \mathbf{B}^\gamma(Y_{t-d}), \tag{6}$$

where K_γ, $\boldsymbol{\gamma}_j = (\gamma_{j1}, \cdots, \gamma_{jK_\gamma})^T$, and $\mathbf{B}^\gamma(Y_{t-d}) = \left(B_1^\gamma(Y_{t-d}), \cdots, B_{K_\gamma}^\gamma(Y_{t-d})\right)^T$ are defined in a similar manner as those in (5). To satisfy the model assumption of $\beta_0(\cdot) > 0$, $\beta_j(\cdot) \geq 0$, $j = 1, \cdots, q$, we impose the constraints: $\boldsymbol{\gamma}_0^T \mathbf{B}^\gamma(Y_{t-d}) > 0$, $\boldsymbol{\gamma}_j^T \mathbf{B}^\gamma(Y_{t-d}) \geq 0, j = 1, \cdots, q$ for $t = d+1, \cdots, T$.

Let $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1^T, \cdots, \boldsymbol{\lambda}_p^T)^T$, $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_0^T, \cdots, \boldsymbol{\gamma}_q^T)^T$, $\mathbf{B}_{Yj}^\lambda(Y_{t-d}) = \mathbf{B}^\lambda(Y_{t-d})Y_{t-j}$ for $j = 1, \cdots, p$, $\mathbf{B}_{\varepsilon j}^\gamma(Y_{t-d}) = \mathbf{B}^\gamma(Y_{t-d})|\varepsilon_{t-j}|$ for $j = 1, \cdots, q$, and $\mathbf{B}_{\varepsilon 0}^\gamma(Y_{t-d}) = \mathbf{B}^\gamma(Y_{t-d})$. With the constant terms being disregarded, the conditional log-likelihood function can be written as

$$\begin{aligned} L(\boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = & -\frac{1}{2} \sum_{t=s+1}^{T} \left[\log(h_t^2(Y_{t-d})) + \varepsilon_t^2 h_t(Y_{t-d})^{-2} \right] \\ = & -\frac{1}{2} \sum_{t=s+1}^{T} \left[2 \log\left(\sum_{j=0}^{q} \boldsymbol{\gamma}_j^T \mathbf{B}_{\varepsilon j}^\gamma(Y_{t-d}) \right) + \frac{\left(Y_t - \sum_{j=1}^{p} \boldsymbol{\lambda}_j^T \mathbf{B}_{Yj}^\lambda(Y_{t-d}) \right)^2}{\left[\sum_{j=0}^{q} \boldsymbol{\gamma}_j^T \mathbf{B}_{\varepsilon j}^\gamma(Y_{t-d}) \right]^2} \right], \end{aligned} \tag{7}$$

where $s = \max\{p, d, q\}$.

The problem of over-fitting may occur if an excessive number of knots are used in (5) and (6). Eilers and Marx (1996) proposed the P-spline by penalizing the coefficients of adjacent B-splines to prevent over-fitting of the B-splines approximation and to guarantee sufficient smoothness of the fitted curves. This process facilitates the penalized likelihood estimation, of which the penalized likelihood

$$L_p = L(\mathbf{\Lambda}, \mathbf{\Gamma}) - \sum_{j=1}^{p} \rho_{\lambda j} \sum_{l=k+1}^{K} (\Delta^k \lambda_{jl})^2 - \sum_{j=0}^{q} \rho_{\gamma j} \sum_{l=k+1}^{K} (\Delta^k \gamma_{jl})^2 \tag{8}$$

is maximized with respect to the unknown parameters λ and γ. In (8), $\rho_{\lambda j}$ and $\rho_{\gamma j}$ are smoothness tuning parameters for controlling the amount of penalty, and Δ^k is the difference operator of order k and defined in a recursive manner. For example, $\Delta\lambda_{jl} = \lambda_{jl} - \lambda_{j,l-1}$, $\Delta^2\lambda_{jl} = \Delta\lambda_{jl} - \Delta\lambda_{j,l-1} = \lambda_{jl} - 2\lambda_{j,l-1} + \lambda_{j,l-2}$, $\cdots$, $\Delta^k\lambda_{jl} = \Delta^{k-1}\lambda_{jl} - \Delta^{k-1}\lambda_{j,l-1}$. Using the matrix notation, (8) can be rewritten as:

$$L_p = L(\mathbf{\Lambda}, \mathbf{\Gamma}) - \sum_{j=1}^{p} \rho_{\lambda j} \boldsymbol{\lambda}_j^T \mathbf{M}_\lambda \boldsymbol{\lambda}_j - \sum_{j=0}^{q} \rho_{\gamma j} \boldsymbol{\gamma}_j^T \mathbf{M}_\gamma \boldsymbol{\gamma}_j, \tag{9}$$

where $\mathbf{M}_\lambda$ and $\mathbf{M}_\gamma$ are the penalty matrices derived from the specified difference penalty. For example, with the penalty order k, $\mathbf{M}_\lambda = (\mathbf{D}_{k-1} \times \cdots \times \mathbf{D}_0)^T \times (\mathbf{D}_{k-1} \times \cdots \times \mathbf{D}_0)$, where $\mathbf{D}_l$ is a $(K-l-1) \times (K-l)$ matrix:

$$\mathbf{D}_l = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}, \quad l = 0, \cdots, k-1.$$

The trade-off between flexibility and smoothness is controlled by the smoothing parameters $\rho_{\lambda j}$ and $\rho_{\gamma j}$. In the context of ML estimation, these smoothing parameters are chosen via a cross-validation procedure. However, the computational burden for determining the optimal values of $\rho_{\lambda j}$ and $\rho_{\gamma j}$ is heavy when the number of smooth functions in the model is large. More importantly, identifying the explicit form of cross-validation for the conditional variance model is difficult. Therefore, for the proposed FARCH model, the optimal values of $\rho_{\lambda j}$ and $\rho_{\gamma j}$ are difficult to obtain using the ML-based methods. In the Bayesian framework, the coefficients λ and γ are regarded as random, and the difference penalties in (9) are replaced by their stochastic analogues, $\Delta^k\lambda_{jl} = \Delta^k\lambda_{j,l-1} + e_{\lambda,jl}$ and $\Delta^k\gamma_{jl} = \Delta^k\gamma_{j,l-1} + e_{\gamma,jl}$, where $e_{\lambda,jl}$ and $e_{\gamma,jl}$ are independently distributed as $N[0, \tau_{\lambda j}]$ and $N[0, \tau_{\gamma j}]$, respectively. The amount of smoothness is then controlled by the additional variance parameters $\tau_{\lambda j}$ and $\tau_{\gamma j}$, which correspond to the inverse of the smoothing parameters in (9). In this work, $\tau_{\lambda j}$ and $\tau_{\gamma j}$ can be considered as new smoothing parameters. In the Bayesian model framework, these smoothing parameters, along with the regression coefficients in $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$, can be obtained via the MCMC algorithm without much difficulty.

The stochastic analogues to the difference penalties are equivalent to assigning the

following Gaussian prior distributions to the coefficients of the B-splines basis functions:

$$p(\boldsymbol{\lambda}_j|\tau_{\lambda j}) = \left(\frac{1}{2\pi\tau_{\lambda j}}\right)^{(K_\lambda^*/2)} \exp\left\{-\frac{1}{2\tau_{\lambda j}}\boldsymbol{\lambda}_j^T\mathbf{M}_\lambda\boldsymbol{\lambda}_j\right\}, \quad j = 1, \cdots, p, \tag{10}$$

$$p(\boldsymbol{\gamma}_0|\tau_{\gamma 0}) = \left(\frac{1}{2\pi\tau_{\gamma 0}}\right)^{(K_\gamma^*/2)} \exp\left\{-\frac{1}{2\tau_{\gamma 0}}\boldsymbol{\gamma}_0^T\mathbf{M}_\gamma\boldsymbol{\gamma}_0\right\} \mathrm{I}(\boldsymbol{\gamma}_0^T\mathbf{B}^\gamma > \mathbf{0}), \tag{11}$$

$$p(\boldsymbol{\gamma}_j|\tau_{\gamma j}) = \left(\frac{1}{2\pi\tau_{\gamma j}}\right)^{(K_\gamma^*/2)} \exp\left\{-\frac{1}{2\tau_{\gamma j}}\boldsymbol{\gamma}_j^T\mathbf{M}_\gamma\boldsymbol{\gamma}_j\right\} \mathrm{I}(\boldsymbol{\gamma}_j^T\mathbf{B}^\gamma \geq \mathbf{0}), \quad j = 1, \cdots, q, \tag{12}$$

where $K_\lambda^* = \mathrm{rank}(\mathbf{M}_\lambda)$, $K_\gamma^* = \mathrm{rank}(\mathbf{M}_\gamma)$, $\mathbf{B}^\gamma = (\mathbf{B}^\gamma(Y_1), \cdots, \mathbf{B}^\gamma(Y_{T-d}))$, and $\mathrm{I}(\cdot)$ is an indicator function. For the smoothing parameters $\tau_{\lambda j}$ and $\tau_{\gamma j}$, we follow the existing literature (Lang and Brezger 2004; Song and Lu 2012) in assigning highly dispersed (but proper) inverse gamma priors

$$\begin{aligned} p(\tau_{\lambda j}^{-1}) &\overset{D}{=} \mathrm{Gamma}[\alpha_{\lambda 0}, \beta_{\lambda 0}], \quad \text{for } j = 1, \cdots, p, \\ p(\tau_{\gamma j}^{-1}) &\overset{D}{=} \mathrm{Gamma}[\alpha_{\gamma 0}, \beta_{\gamma 0}], \quad \text{for } j = 0, \cdots, q, \end{aligned} \tag{13}$$

where $\alpha_{\lambda 0}$, $\beta_{\lambda 0}$, $\alpha_{\gamma 0}$, and $\beta_{\gamma 0}$ are hyperparameters with preassigned values. To obtain highly dispersed priors, common choices for these hyperparameters are $\alpha_{\lambda 0} = \alpha_{\gamma 0} = 1$, whereas $\beta_{\lambda 0}$ and $\beta_{\gamma 0}$ are small. In this article, we set $\alpha_{\lambda 0} = \alpha_{\gamma 0} = 1$, and $\beta_{\lambda 0} = \beta_{\gamma 0} = 0.005$.

3.2 Posterior inference

Let $\mathbf{Y} = \{Y_1, \cdots, Y_T\}$ be the set of observed time series. Let $\boldsymbol{\tau}_\lambda = \{\tau_{\lambda 1}, \cdots, \tau_{\lambda p}\}$, $\boldsymbol{\tau}_\gamma = \{\tau_{\gamma 0}, \cdots, \tau_{\gamma_q}\}$, and $\boldsymbol{\theta} = \{\boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\tau}_\lambda, \boldsymbol{\tau}_\gamma\}$ include all unknown parameters in the model. The Bayesian estimate of $\boldsymbol{\theta}$ can be obtained via the sample mean of a sufficiently large number of observations drawn from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y})$. However, this posterior distribution is intractable because of model complexity. To solve this problem, we use the Gibbs sampler (Geman and Geman 1984) algorithm to simulate each component of $\{\boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\tau}_\lambda, \boldsymbol{\tau}_\gamma\}$ given others from its full conditional distribution iteratively. The full conditional distributions required in the Gibbs sampler are derived in the Appendix. Considering the nonlinearity and complexity of the model, several full conditional distributions are nonstandard, and sampling observations from such distributions is not straightforward. Hence, the Metropolis-Hastings (MH) algorithm (Metropolis et al. 1953; Hastings 1970) is adopted to simulate observations from nonstandard distributions. The implementation of the MH algorithm is also provided in the Appendix.

3.3 Model comparison

A specific question regarding model selection is whether an advanced (more complex) model is actually "better" than an elementary (simpler) one. This issue is particularly

relevant here: If a parametric model with simple constant coefficients can provide a better fit to the observed data, then the nonparametric model with functional coefficients is unnecessary; or if a model with constant variance is adequate, then modeling both the conditional mean and conditional variance is unnecessary. One simple Bayesian model selection statistic is the deviance information criterion (DIC; Spiegelhalter et al. 2002). While DIC is advantageous due to its ease of computation, it may encounter the problem of always selecting the most complex model as the sample size becomes large. As a remedy, Ando (2007) proposed the Bayesian predictive information criterion (BPIC), which evaluates the model complexity more accurately and thus lessens the overfitting problem. However, the computation of BPIC in our model framework is difficult because it requires the maximization of an extremely complex joint posterior distribution. In this article, we use the Bayes factor (Kass and Raftery 1995), which is defined by the ratio of marginal likelihoods, as the model comparison statistic. Given that the marginal likelihood in the current model setting involves high-dimensional integration, the computation of the Bayes factor is challenging. Although the path sampling procedure proposed by Gelman and Meng (1998) has been demonstrated to be useful for computing the Bayes factor in the comparison of many statistical models (see, for example, Song and Lee 2012), it cannot be directly applied to the current model because defining a linked model between the nonparametric FARCH model and its parametric or semiparametric counterparts is difficult. To solve this difficulty, instead of directly working on the competing models, we compare each of the candidate models with the following simple intermediate model M_0:

$$M_0: \quad Y_t = u_t, \quad u_t \overset{iid}{\sim} N(0,1). \tag{14}$$

The Bayes factors between the competing models and M_0 is obtained via the path sampling procedure. For example, the competing models are given below:

M_1: FARCH model defined by (1) and (2).

M_2: Parametric ARCH model

$$\begin{aligned} Y_t &= \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t, \\ h_t &= \beta_0 + \beta_1 |\varepsilon_{t-1}| + \cdots + \beta_q |\varepsilon_{t-q}|, \end{aligned}$$

where α's and β's are unknown parameters.

The linked model between M_1 and M_0 is defined by

$$\begin{aligned} M_{1l}: \quad Y_t = l &\left[\alpha_1(Y_{t-d}) Y_{t-1} + \alpha_2(Y_{t-d}) Y_{t-2} + \cdots + \alpha_p(Y_{t-d}) Y_{t-p} \right] \\ &+ \left[l h_t(Y_{t-d}) + (1-l) \right] u_t, \end{aligned}$$

$$h_t(Y_{t-d}) = \beta_0(Y_{t-d}) + \beta_1(Y_{t-d}) |\varepsilon_{t-1}| + \cdots + \beta_q(Y_{t-d}) |\varepsilon_{t-q}|.$$

It can be seen that $M_{1l} = M_1$ when $l = 1$ and $M_{1l} = M_0$ when $l = 0$. Hence, the log Bayes factor, $\log B_{10}$, can be computed via the path sampling procedure (see Song and Lee 2012). Similarly, the linked model between M_2 and M_0 can be defined as

$$\begin{aligned} M_{2l}: \quad & Y_t = l\left[\alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p}\right] + \left[lh_t + (1-l)\right]u_t, \\ & h_t = \beta_0 + \beta_1|\varepsilon_{t-1}| + \cdots + \beta_q|\varepsilon_{t-q}|. \end{aligned}$$

And $\log B_{20}$ can be obtained accordingly. Thus, the log Bayes factor between M_1 and M_2 can be calculated as

$$\log B_{12} = \log B_{10} - \log B_{20}. \tag{15}$$

The interpretation of the Bayes factor between candidate models can be found in Kass and Raftery (1995).

3.4 Bayesian Forecasting

Following Tsay (2010), we apply parametric bootstraps to compute nonlinear forecasts. Let T be the forecast origin and $l > 0$ be the forecast horizon. That is, we are at time index T and interested in forecasting Y_{T+l}. Given the observed data $Y_1, \cdots, Y_T$, the parametric bootstrap computes realizations $Y_{T+1}, \cdots, Y_{T+l}$ as follows. (i) Draw a new stochastic error ε_{T+1} from $N(0, h_{T+1}(Y_{T+1-d}))$, where $h_{T+1}(Y_{T+1-d})$ is calculated based on (2). (ii) Compute Y_{T+1} based on (1). (iii) Repeat steps (i) and (ii) to obtain $Y_{T+2}, \cdots, Y_{T+l}$. The above procedure produces a realization for Y_{T+l}. The procedure is repeated M times to obtain M realizations for Y_{T+l}, denoted by $Y_{T+l}^{(m)}$, $m = 1, \cdots, M$. The point forecast of Y_{T+l} and the estimated variance of the forecast are calculated as follows,

$$\hat{Y}_{T+l} = \frac{1}{M}\sum_{m=1}^{M} Y_{T+l}^{(m)}, \quad \widehat{\mathrm{Var}}(\hat{Y}_{T+l}) = \frac{1}{M-1}\sum_{m=1}^{M}(Y_{T+l}^{(m)} - \hat{Y}_{T+l})^2.$$

Tsay (2010) pointed out that $M = 3000$ could provide satisfactory results. In this paper, we employ mean square error (MSE) to measure the performance of point forecasts. For an l-step-forecast, the MSE is defined as follows,

$$\mathrm{MSE}(l) = \frac{1}{n}\sum_{j=0}^{n-1}(\hat{Y}_{T+j+l} - Y_{T+j+l})^2, \tag{16}$$

where n is the number of l-step-ahead forecasts available in the forecasting subsample. In application, the model with the smallest magnitude on that measure is regarded as the best l-step-ahead forecasting model. However, it is possible that different l may result in the selection of different models. In addition, we employ mean variance values (MV) to assess the robustness of the forecasting. For l-step-forecast, the MV is defined as follows,

$$\mathrm{MV}(l) = \frac{1}{n}\sum_{j=0}^{n-1}\widehat{\mathrm{Var}}(\hat{Y}_{T+j+l}). \tag{17}$$

The model with the smallest MV value provides the most reliable forecasts.

4 Numerical Results

4.1 Simulation study

The main objective of this subsection is to demonstrate the empirical performance of the proposed methodology. We considered the following FARCH(2,2,1) model:

$$\begin{aligned} Y_t &= \alpha_1(Y_{t-2})Y_{t-1} + \alpha_2(Y_{t-2})Y_{t-2} + \varepsilon_t, \\ h_t &= \beta_0(Y_{t-2}) + \beta_1(Y_{t-2})|\varepsilon_{t-1}|, \end{aligned} \tag{18}$$

where $\alpha_1(u) = u/20 - 0.15$, $\alpha_2(u) = 0.5\cos(u - 0.5)$, $\beta_0(u) = 0.5 + (u - 0.5)^2/20$, and $\beta_1(u) = 0.2 + (u - 1)^2/30$. A total of 25 equidistant knots were employed to construct cubic P-splines in the domains of Y_{t-2}. The second-order random-walk penalty was used in the Bayesian P-splines for estimating the unknown smooth functions. We conducted 100 replications with two different sample sizes $T = 1,000$ and 2,000. The convergence was assessed by the estimated potential scale reduction (EPSR) values (Gelman et al. 1996). Figure 1 shows the EPSR values against the iteration numbers in an arbitrarily selected replication. Based on the EPSR plot, we found that the algorithm converged within 10,000 iterations. We therefore took a burn-in phase of 10,000 iterations and used an additional 10,000 observations to obtain the Bayesian results. Figures 2 and 3 present the averages of the pointwise posterior means of functional coefficients, along with the 5%- and 95%- pointwise quantiles. Compared with their true functions (represented by solid curves), the estimated curves capture the peaks, valleys, and change patterns of the true functions correctly. The accuracy of the estimated curves is further assessed through the deviance of the estimated curve $\hat{f}(u)$ from its true curve $f(u)$, which is measured by the square root of average squared errors (RASE)

$$\text{RASE} = \left[n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} \left(\hat{f}(u_k) - f(u_k) \right)^2 \right]^{1/2}, \tag{19}$$

where $\{u_k,\ k = 1, \cdots, n_{\text{grid}}\}$ are grid points taken in the domain of u. The boxplots for 100 RASE values are presented in Figure 4. The RASE values are reasonably small, which indicates that the performance of our proposed methodology is satisfactory.

To evaluate the performance of the Bayes factor, we conducted model selection based on the above simulated data sets. The following competing models are considered:

M_1: FARCH model defined in (18).

M_2: Semiparametric model with functional-coefficient mean model but constant-coefficient

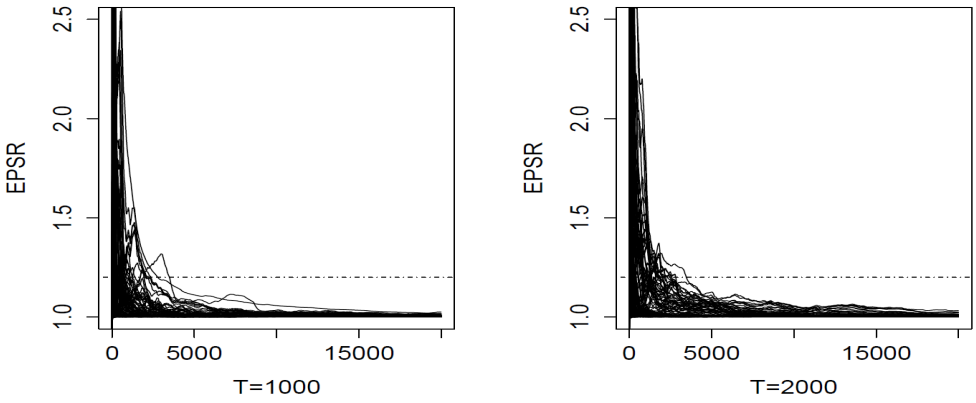

Figure 1: EPSR values against the number of iterations in the simulation study.

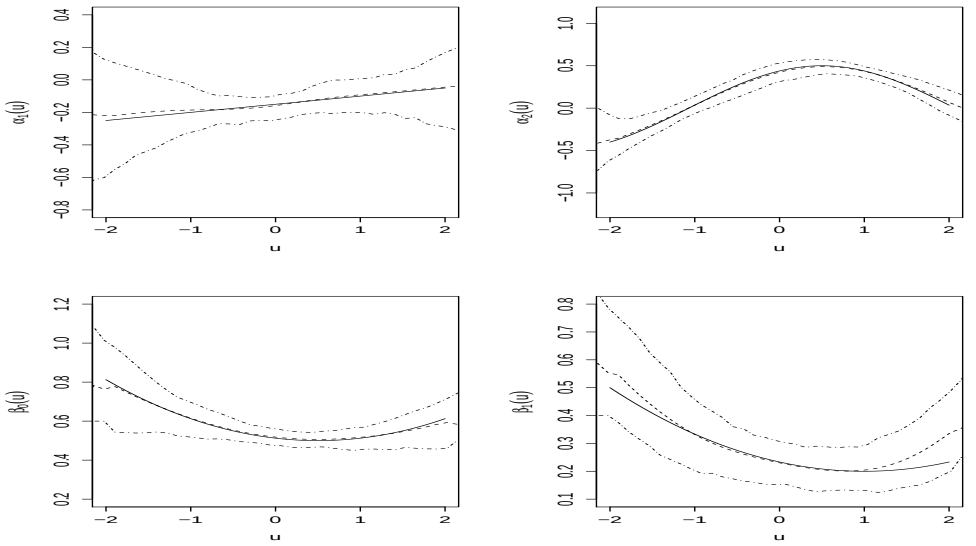

Figure 2: Estimates of the unknown smooth functions in the simulation study with sample size $T = 1000$. The solid curves represent the true curves, the dashed and dot-dashed curves respectively represent the estimated posterior means and the 5%- and 95%- pointwise quantiles on the basis of 100 replications.

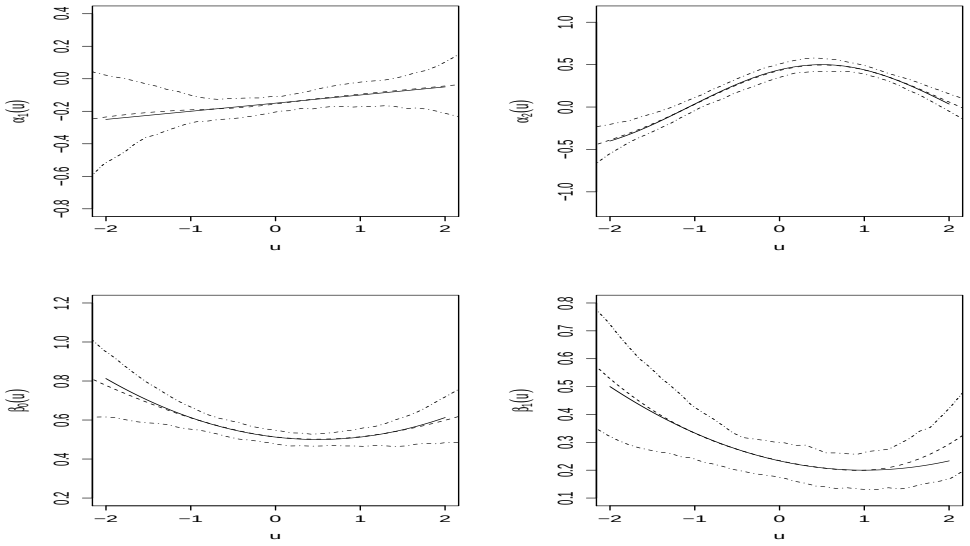

Figure 3: Estimates of the unknown smooth functions in the simulation study with sample size $T = 2000$. The solid curves represent the true curves, the dashed and dot-dashed curves respectively represent the estimated posterior means and the 5%- and 95%- pointwise quantiles on the basis of 100 replications.

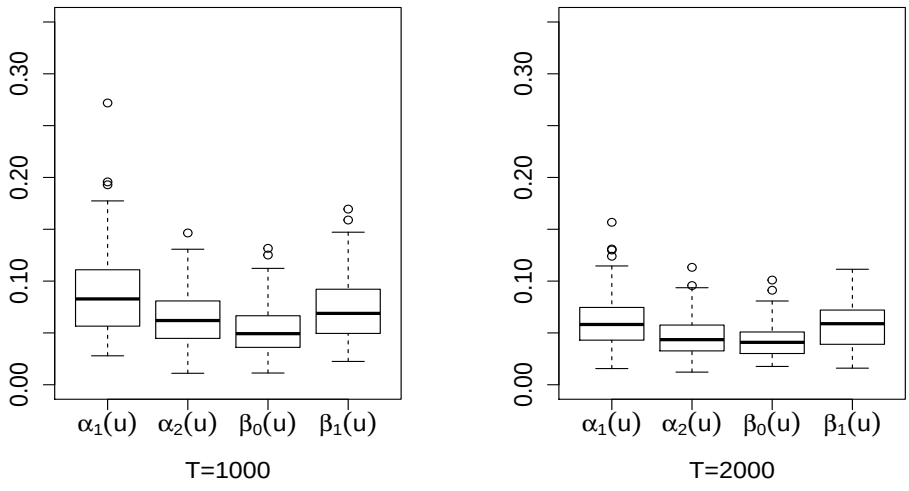

Figure 4: Boxplots of the RASE values in the simulation study. The data sets are generated from the FARCH(2,2,1) model with sample sizes $T = 1000$ and $T = 2000$.

variance model:

$$Y_t = \alpha_1(Y_{t-2})Y_{t-1} + \alpha_2(Y_{t-2})Y_{t-2} + \varepsilon_t,$$
$$h_t = \beta_0 + \beta_1|\varepsilon_{t-1}|,$$

where $\alpha(\cdot)$'s are unknown functions, and β's are unknown parameters.

M_3: DTARCH(2,2,1) model

$$Y_t = [\alpha_{11}Y_{t-1} + \alpha_{21}Y_{t-2}] \cdot \mathrm{I}(Y_{t-2} \le 0) + [\alpha_{12}Y_{t-1} + \alpha_{22}Y_{t-2}] \cdot \mathrm{I}(Y_{t-2} > 0) + \varepsilon_t,$$
$$h_t = \Big[\beta_{01} + \beta_{11}|\varepsilon_{t-1}|\Big] \cdot \mathrm{I}(Y_{t-2} \le 0) + \Big[\beta_{02} + \beta_{12}|\varepsilon_{t-1}|\Big] \cdot \mathrm{I}(Y_{t-2} > 0),$$

where α's and β's are unknown parameters.

M_4: Parametric ARCH(2,1) model

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t,$$
$$h_t = \beta_0 + \beta_1|\varepsilon_{t-1}|,$$

where α's and β's are unknown parameters.

We compared each of the competing models with the intermediate model M_0 (see (14)) using the Bayes factor. The boxplots of the log Bayes factors between the above candidate models and M_0 for 100 replicated data sets are presented in Figure 5. It can be seen that the means of the log Bayes factors follow the order: $M_1 > M_2 > M_3 > M_4$. When checking each of the 100 replications further, we found that the Bayes factor selected M_1 in 95 replications when $T = 1,000$ and in 97 replications when $T = 2,000$. Moreover, we used 100 out-of-sample observations to assess the performance of the Bayesian forecasting. Table 1 presents the means of MSE and MV based on an l-step-forecast ($l = 1, 2$) for 100 replications. The means of MSE and MV corresponding to M_1 are the smallest, indicating that the FARCH model provides more accurate and robust forecasts. Based on the results of the Bayes factor and the performances of forecasting, we conclude that M_1 is the best one among the candidate models. In addition, we compared the FARCH models with different p, d, and q values. We found that the Bayes factor chose the FARCH(2,2,1) model consistently in each replication. The details are not presented here to save space.

In order to check whether the FARCH model can provide a good approximation when the true model is relatively simple, we generated time series from a parametric DTARCH model and analyzed the data sets via the FARCH model. Then, the relevant questions might be (i) whether the FARCH model produces reasonable estimation results, and (ii) whether the Bayes factor selects the true (simpler) model. To address these issues, we generated 100 data sets with the sample size of $T = 2,000$ from the following DTARCH(1,2,1) model:

$$\begin{aligned} Y_t &= -0.3Y_{t-1} \cdot \mathrm{I}(Y_{t-2} \le 0) + 0.35Y_{t-1} \cdot \mathrm{I}(Y_{t-2} > 0) + \varepsilon_t, \\ h_t &= \Big[0.2 + 0.42|\varepsilon_{t-1}|\Big] \cdot \mathrm{I}(Y_{t-2} \le 0) + \Big[0.4 + 0.24|\varepsilon_{t-1}|\Big] \cdot \mathrm{I}(Y_{t-2} > 0). \end{aligned} \tag{20}$$

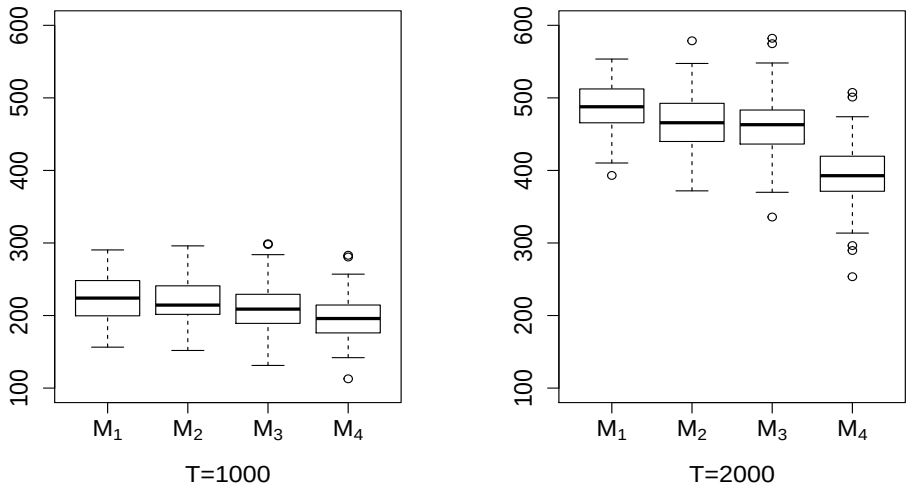

Figure 5: Boxplots of the log Bayes factor values in the simulation study with sample sizes $T = 1000$ and $T = 2000$.

		1-step		2-step	
Sample Sizes	Model	MSE	MV	MSE	MV
$T = 1000$	M_1	0.463	0.475	0.472	0.488
	M_2	0.533	0.488	0.490	0.499
	M_3	0.549	0.493	0.494	0.508
	M_4	0.563	0.507	0.509	0.519
$T = 2000$	M_1	0.464	0.472	0.469	0.484
	M_2	0.484	0.480	0.486	0.492
	M_3	0.492	0.493	0.499	0.508
	M_4	0.500	0.506	0.505	0.518

Table 1: Out-of-sample forecast comparison among M_1 to M_4 in the simulation study.

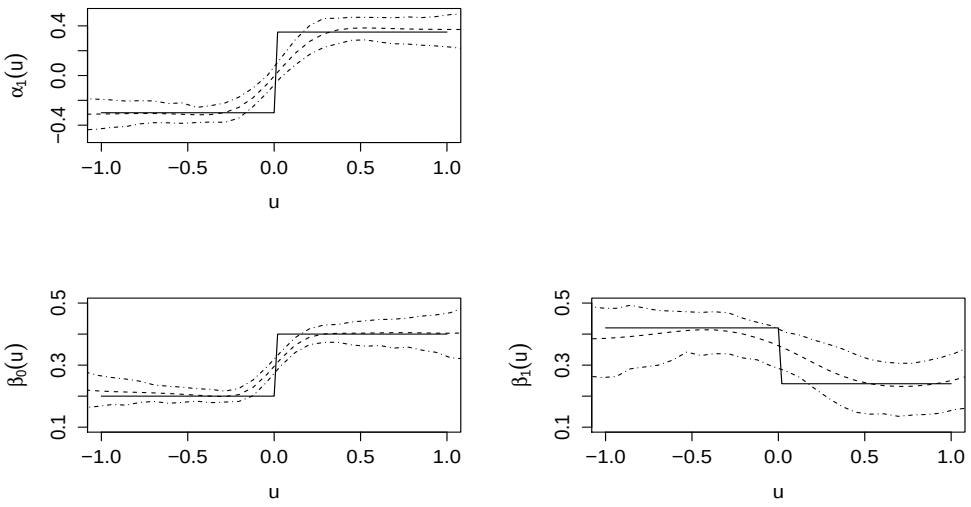

Figure 6: Estimates of the unknown smooth functions in the simulation study. The data sets are generated from the DTARCH(1,2,1) model with sample size $T = 2000$. The solid curves represent the true curves, the dashed and dot-dashed curves respectively represent the estimated posterior means and the 5%- and 95%- pointwise quantiles on the basis of 100 replications.

The generated data sets were analyzed using the FARCH(1,2,1) model. Figure 6 shows the true curves, the averages of the pointwise posterior means of functional coefficients, and the 5%- and 95%- pointwise quantiles. Figure 7 (left panel) presents the boxplots for 100 RASE values, most of which are not far from zero, indicating that FARCH(1,2,1) provides reasonable estimation for piecewise constant coefficients. Figure 7 (right panel) presents the boxplots of the log Bayes factor values between the candidate models and the intermediate model M_0. The mean value of the log Bayes factors for DTARCH(1,2,1) is larger than that for FARCH(1,2,1). When checking each of the 100 replications further, we found that the true model, DTARCH(1,2,1), was consistently selected in all replications. Therefore, the Bayes factor picks the true model correctly regardless of model size and complexity.

4.2 Real data analysis

As an illustration, we apply the proposed methodology to the study of the daily S&P 500 Composite Index from Jan 3, 2000 to July 27, 2011. The S&P 500 Composite index is a stock market index based on the common stock prices of 500 top publicly traded American companies, which is one of the most commonly followed indices and is a good representation of the market of the U.S. economy. Our main goal is to

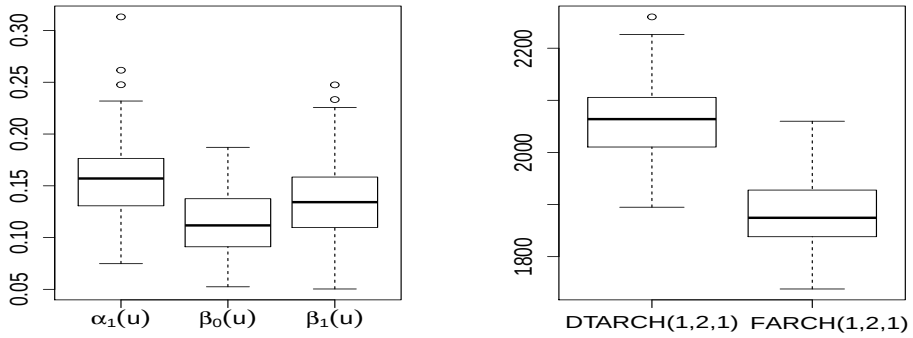

Figure 7: Boxplots of the RASE values (left panel) and the log Bayes factor values (right panel) in the simulation study. The data sets are generated from the DTARCH(1,2,1) model with sample size $T = 2000$.

investigate the nonlinear features of return and volatility in terms of the S&P 500 Composite Index. Let X_t be the closing price at time t, the return series Y_t was defined as $Y_t = 100 \log(X_t/X_{t-1})$. The total sample size was $T = 2,909$. We used the first 2,700 observations to develop the model and used the remaining 209 observations as the out-of-sample set to assess the performance of the Bayesian forecasting. We proposed a FARCH(p, d, q) model to investigate the asymmetry of the conditional mean and conditional variance as well as to determine how this asymmetric feature changes dynamically according to lagged returns.

First, we determined the delay parameter d, the AR order p, and the ARCH order q by comparing FARCH models with different values of p, d, and q. By rotating their values from 1 to 3, a total of 27 candidate models were considered. We first compared each of the candidate models with the intermediate model M_0 (see (14)) using the Bayes factor. The results are presented in Table 2, from which we can calculate the log Bayes factors between different FARCH models via (15) and conclude that FARCH(1,2,1) is the best model. We also compared FARCH models with larger values of p, d, and q (the results are not presented here). The Bayes factors again chose FARCH(1,2,1), which is defined as

$$\begin{aligned} Y_t &= \alpha_1(Y_{t-2})Y_{t-1} + \varepsilon_t, \\ h_t &= \beta_0(Y_{t-2}) + \beta_1(Y_{t-2})|\varepsilon_{t-1}|. \end{aligned} \tag{21}$$

Next, we considered several candidate models, including M_1: FARCH(1,2,1) and its semiparametric and parametric counterparts:

M_2: Semiparametric model with functional-coefficient mean model but constant-coefficient

	$d=1$			$d=2$			$d=3$		
q \ p	1	2	3	1	2	3	1	2	3
1	293	275	290	**478**	275	290	427	363	350
2	381	294	246	320	294	246	426	393	377
3	311	307	230	259	239	225	298	289	205

Table 2: The log Bayes factors between FARCH(p,d,q) models and the intermediate model M_0 in the analysis of the S&P 500 data set.

variance model:

$$Y_t = \alpha_1(Y_{t-2})Y_{t-1} + \varepsilon_t,$$
$$h_t = \beta_0 + \beta_1|\varepsilon_{t-1}|,$$

where $\alpha_1(\cdot)$ is an unknown function, β's are unknown parameters.

M_3: Semiparametric model with AR mean model but functional-coefficient variance model:

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t,$$
$$h_t = \beta_0(Y_{t-2}) + \beta_1(Y_{t-2})|\varepsilon_{t-1}|,$$

where α_1 is an unknown parameter, $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are unknown functions.

M_4: DTARCH(1,2,1) model

$$Y_t = \alpha_{11}Y_{t-1}\cdot \mathrm{I}(Y_{t-2}\le 0) + \alpha_{12}Y_{t-1}\cdot \mathrm{I}(Y_{t-2}>0) + \varepsilon_t,$$
$$h_t = \Big[\beta_{01} + \beta_{11}|\varepsilon_{t-1}|\Big]\cdot \mathrm{I}(Y_{t-2}\le 0) + \Big[\beta_{02} + \beta_{12}|\varepsilon_{t-1}|\Big]\cdot \mathrm{I}(Y_{t-2}>0),$$

where α's and β's are unknown parameters.

M_5: Parametric ARCH$(1,1)$ model

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t,$$
$$h_t = \beta_0 + \beta_1|\varepsilon_{t-1}|,$$

where α_1 and β's are unknown parameters.

The log Bayes factor values between M_1 and M_0 to M_5 and M_0 are equal to 478, 63, 358, 167, and 105, respectively. Again, M_1 was selected. Further, we used the out-of-sample observations to assess the performance of the Bayesian forecasting of M_1 to M_5. Table 3 presents the MSE and MV values of l-step-forecast ($l=1,2$) for M_1 to M_5. Both the MSE and MV values of FARCH(1,2,1) are the smallest, reconfirming the best fit of the selected model.

Model	1-step		2-step	
	MSE	MV	MSE	MV
M_1	0.628	1.232	0.633	1.368
M_2	0.657	1.566	0.635	1.956
M_3	0.628	1.237	0.633	1.375
M_4	0.630	9.251	0.634	7.744
M_5	0.642	9.511	0.643	7.982

Table 3: Out-of-sample forecast comparison among M_1 to M_5 for the S&P 500 data set.

In this study, a total of 25 equidistant knots were used to construct cubic P-splines in the domains of Y_{t-2}. The second-order random-walk penalty was used for the Bayesian P-splines in the estimation of the unknown smooth functions. Figure 8 shows the EPSR values against the iteration numbers. We found that the Gibbs sampler converged within 20,000 iterations. Hence, we discarded the first 20,000 burn-in iterations and took 20,000 additional iterations to obtain the estimation. The estimated curves, along with the 5%- and 95%- pointwise quantiles, are depicted in Figure 9. Several interesting findings can be derived from this analysis. First, the estimated values of $\hat{\alpha}_1(\cdot)$ are slightly below zero, implying mild negative associations between the future mean return and the historical returns. Thus, the future mean return tends to be positive (negative) if the historical returns are negative (positive). This patten may explain why the index values tend to be stable as well as why the overall market is robust to minor financial events. Second, the estimated curve $\hat{\beta}_1(\cdot)$ exhibits a U-shaped pattern, which means the persistence in volatility tends to be high when the historical returns are high or low, reaching the minimum when Y_{t-d} is approximately zero. In addition, the estimated curve presents obvious asymmetry; the values of $\hat{\beta}_1(\cdot)$ at the negative range of Y_{t-d} are clearly larger than those at the positive range of Y_{t-d}. That is, the volatility persistence tends to be higher in bear markets than in bull markets. Third, the Bayes factors selected the model with $d = 2$. Fan and Yao (2003) highlighted that Y_{t-d} should be regarded as the "model-dependent variable," which provides useful information on the modeling and the dependence structure of the observed data. In this study, the delay parameter $d = 2$ indicates that the historical returns will be internally reflected in the model with one day lag. Finally, the dynamic coefficients, $\hat{\beta}_0(\cdot)$ and $\hat{\beta}_1(\cdot)$ in the volatility model are clearly neither linear nor piecewise linear. These nonlinear patterns cannot be well captured by the existing parametric or semiparametric time series models.

To assess the sensitivity of the Bayesian results to inputs of hyperparameters in the prior distributions, the above analysis was repeated with disturbances of the current prior input. Two different choices of $\alpha_{\lambda 0} = \alpha_{\gamma 0} = 1$, $\beta_{\lambda 0} = \beta_{\gamma 0} = 0.05$ and $\alpha_{\lambda 0} = \alpha_{\gamma 0} = \beta_{\lambda 0} = \beta_{\gamma 0} = 0.001$ were used. We obtained the same model selection results, close Bayesian estimates of the unknown parameters, and similar estimated curves of the unknown smooth functions.

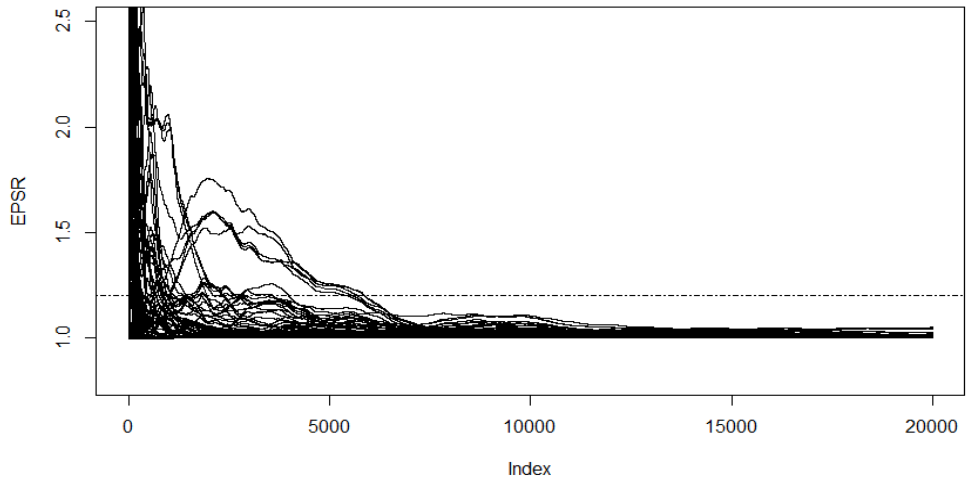

Figure 8: EPSR values against the number of iterations in the analysis of the S&P 500 data set under FARCH$(1, 2, 1)$.

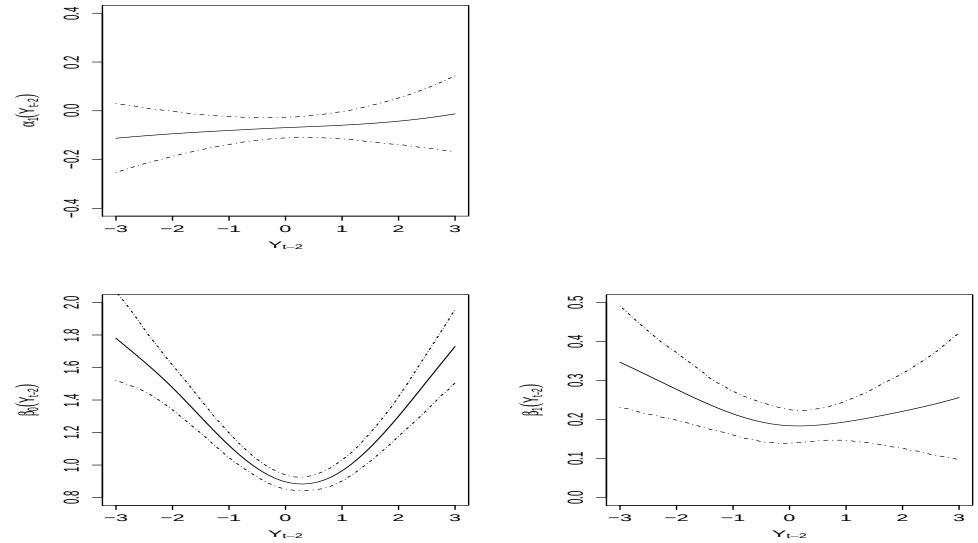

Figure 9: Estimated curves in the analysis of the S&P 500 data set under FARCH$(1, 2, 1)$. The solid curves represent the pointwise mean curves, and the dot-dashed curves represent the 5%- and 95%- pointwise quantiles.

5 Discussion

In this paper, we proposed a novel FARCH model to analyze time series data. The FARCH model generalizes both FAR and DTARCH models, thus providing a flexible model framework to capture various nonlinear phenomena for both the conditional mean and conditional variance. We developed a Bayesian approach coupled with Bayesian P-splines and MCMC algorithms to obtain the estimation of the functional coefficients. We employed a Bayesian model selection statistic, the Bayes factor, to address the hypothesis testing problem. Simulation results showed that our proposed methodology had satisfactory performance. Given that volatility is an important factor for understanding and predicting reality and that the true relationships among time series observations are seldom known a priori especially in the financial market, the proposed FARCH model and the Bayesian approach are useful and important. The work in this paper expands the scope of Bayesian nonparametric time series modeling in the statistics and economics literature.

This research has limitations. First, we did not provide a sufficient and/or a necessary condition to theoretically ensure the stationarity of the FARCH model because of the model complexity. This important issue is worthy of further investigation. Second, our proposed method cannot satisfactorily estimate a very wiggly function like $\beta(u) = 0.2 + 0.1\sin(2\pi u)$, in the conditional variance/scale model. The unsatisfactory performance may be due to the nature of volatility. Exploring an efficient way to address this problem is also important. Third, the innovations were assumed to follow the standard normal distribution in our model. This assumption may not be valid in practice. Therefore, more robust methods such as L_1 regression and quantile regression may be considered for the FARCH model. Fourth, as suggested by an anonymous referee, the current model framework can be extended to functional-coefficient GARCH (FGARCH) models to achieve more flexibility. However, substantial effort will be required to solve the technical challenges in the posterior computation of such models. Finally, the current research only handles univariate time series. A generalization to accommodate multivariate time series data might be interesting for further research.

Acknowledgments

The authors are very thankful to the Editor, Associate Editor, and the two reviewers for their valuable comments and suggestions, which have helped greatly in improving our paper. This paper is supported by the grants GRF 403109 and GRF 404711 from the Research Grant Council of the Hong Kong Special Administration Region; NSFC 11101432, NSFC 11101443, NSFC 11271383, and NSFC 11301555 from the National Natural Science Foundation of China. This study is also supported by the high-performance grid computing platform of Sun Yat-sen University.

References

Ando, T. (2007). Bayesian predictive information criterion for evaluation of hierachical Bayesian and empirical Bayes models. *Biometrika*, 94: 443–458. 9

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97: 160–169. 3

Bickel, P. J. and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. dispersion. *The Annals of Stastistics*, 4: 1139–1158. 4

Bollerslev, T. (1986). A generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31: 307–327. 2

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time Series Analysis : Forecasting and Control*. John Wiley, New Jersey, 4th edition. 1

Brooks, C. (2001). A double-threshold GARCH model for the French Franc/Deutschmark exchange rate. *Journal of Forecasting*, 20: 135–145. 2

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95: 941–956. 2

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88: 298–308. 2, 5

Congdon, P. (2006). *Bayesian Statistical Modelling*. John Wiley & Sons, New York. 3

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Stastistics*, 17: 1749–1766. 1

De Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, New York, revised edition. 6

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B*, 62: 355–366. 3

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B–splines and penalties. *Statistical Science*, 11: 89–121. 6, 7

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50: 987–1008. 2

Fahrmeir, L. and Raach, A. (2007). A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika*, 72: 327–346. 3

Fan, J. and Gijbels, I. (2003). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London. 3

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. 2, 4, 19

Gelman, A. (1996). Inference and monitoring convergence. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 131–144. Chapman & Hall, London. 11, 25

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall, Boca Raton, 2nd edition. 3

Gelman, A. and Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13: 163–185. 9

Gelman, A., Robert, G. O., and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Simth, A. F. M., editors, *Bayesian Statistics 5*, pages 599–607. Oxford University Press, New York. 11, 25

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: 721–741. 8

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48: 1779–1801. 2

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1: 15–29. 1

Haggan, V. and Ozaki, T. (1981). Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 68: 189–196. 1

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York. 1

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57: 97–109. 8, 25

Hui, Y. V. and Jiang, J. C. (2005). Robust modeling of DTARCH models. *Econometrics Journal*, 8: 143–158. 2

Jiang, J. C., Zhao, Q., and Hui, Y. V. (2001). Robust modelling of ARCH models. *Journal of Forecasting*, 20: 111–133. 4

Jiang, X. J., Jiang, J. C., and Song, X. Y. (2013). Weighted composite quantile regression estimation of DTARCH models. *Econometrics Journal*, accepted. 2

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90: 773–795. 3, 4, 9, 10

Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12: 793–813. 4

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13: 183–212. 3, 6, 8

Li, C. W. and Li, W. K. (1996). On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics*, 11: 253–274. 2, 4, 5

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *The Journal of Chemical Physics*, 21: 1087–1091. 8, 25

Scheines, R., Hoijtink, H., and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64: 37–52. 3

Song, X. Y. and Lee, S. Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. Wiley, London. 9, 10

Song, X. Y. and Lu, Z. H. (2012). Semiparametric transformation models with Bayesian P-splines. *Statistics and Computing*, 22: 1085–1098. 3, 8

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64: 583–639. 3, 9

Tjøstheim, D. (1994). Non-linear time series: A selective review. *Scandinavian Journal of Statistics*, 21: 97–130. 1

Tong, H. (1990). *Nonlinear Time Series Analysis: A Dynamical System Approach*. Oxford University Press, London. 1, 2

Tong, H. (1995). A personal overview of non-linear time series analysis from a chaos perspective (with discussion). *Scandinavian Journal of Statistics*, 22: 399–445. 1

Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons, New Jersey, 3rd edition. 10

Yau, P. and Kohn, R. (2003). Estimation and variable selection in nonparametric heteroscedastic regression. *Statistics and Computing*, 13: 191–208. 5

Zakoian, J. M. (1994). Threshold heteroscedastic models. *Journal of Economic Dynamics and Control*, 18: 931–955. 2

Appendix: Full Conditional Distributions and Implementation of the MH Algorithm.

(1) For $j = 1, \cdots, p$, the full conditional distribution of $\boldsymbol{\lambda}_j$ is:

$$p(\boldsymbol{\lambda}_j|\cdot) \propto \exp\left\{ -\frac{1}{2}\sum_{t=s+1}^{T}\left[2\log\Big(\sum_{j=0}^{q}\boldsymbol{\gamma}_j^T\mathbf{B}_{\varepsilon j}^{\gamma}(Y_{t-d})\Big) + \frac{\Big(Y_t - \sum_{j=1}^{p}\boldsymbol{\lambda}_j^T\mathbf{B}_{Yj}^{\lambda}(Y_{t-d})\Big)^2}{\Big[\sum_{j=0}^{q}\boldsymbol{\gamma}_j^T\mathbf{B}_{\varepsilon j}^{\gamma}(Y_{t-d})\Big]^2}\right] - \frac{1}{2\tau_{\lambda j}}\boldsymbol{\lambda}_j^T\mathbf{M}_{\lambda}\boldsymbol{\lambda}_j\right\}. \tag{22}$$

The MH algorithm (Metropolis et al. 1953; Hastings 1970) is employed to sample observations from this complicated distribution. The details of the implementation are as follows. At the rth iteration with the current values of $\{\boldsymbol{\Lambda}^{(r)}, \boldsymbol{\Gamma}^{(r)}, \boldsymbol{\tau}_{\lambda}^{(r)}, \boldsymbol{\tau}_{\gamma}^{(r)}\}$, we first generate the candidate $\boldsymbol{\lambda}_j^*$ from the proposed distribution

$$N[\boldsymbol{\lambda}_j^{(r)}, \sigma_{\lambda j}^2\boldsymbol{\Sigma}_{\lambda j}^*],$$

where

$$\boldsymbol{\Sigma}_{\lambda_j}^{*-1} = \sum_{t=s+1}^{T}\mathbf{B}_{Yj}^{\gamma}(Y_{t-d})\mathbf{B}_{Yj}^{\gamma T}(Y_{t-d}) + \mathbf{M}_{\lambda}/\tau_{\lambda j}.$$

Then the candidate $\boldsymbol{\lambda}_j^*$ is accepted as the new value of $\boldsymbol{\lambda}_j^{(r+1)}$ with probability

$$\min\left\{1, \frac{p(\boldsymbol{\lambda}_j^*|\cdot)}{p(\boldsymbol{\lambda}_j^{(r)}|\cdot)}\right\}.$$

If the candidate is rejected, then $\boldsymbol{\lambda}_j^{(r+1)} = \boldsymbol{\lambda}_j^{(r)}$, and the chain does not move. Finally, $\sigma_{\lambda j}^2$ is selected such that the acceptance rate is 0.25 or more (Gelman et al. 1996).

(2) For $j = 0, \cdots, q$, the full conditional distribution of $\boldsymbol{\gamma}_j$ is:

$$p(\boldsymbol{\gamma}_j|\cdot) \propto \exp\left\{ -\frac{1}{2}\sum_{t=s+1}^{T}\left[2\log\Big(\sum_{j=0}^{q}\boldsymbol{\gamma}_j^T\mathbf{B}_{\varepsilon j}^{\gamma}(Y_{t-d})\Big) + \frac{\Big(Y_t - \sum_{j=1}^{p}\boldsymbol{\lambda}_j^T\mathbf{B}_{Yj}^{\lambda}(Y_{t-d})\Big)^2}{\Big[\sum_{j=0}^{q}\boldsymbol{\gamma}_j^T\mathbf{B}_{\varepsilon j}^{\gamma}(Y_{t-d})\Big]^2}\right] - \frac{1}{2\tau_{\gamma j}}\boldsymbol{\gamma}_j^T\mathbf{M}_{\gamma}\boldsymbol{\gamma}_j\right\}\mathrm{I}(\boldsymbol{\gamma}_j^T\mathbf{B}^{\gamma} \geq \mathbf{0}). \tag{23}$$

Note that the constraint is $\mathrm{I}(\boldsymbol{\gamma}_j^T\mathbf{B}^\gamma > \mathbf{0})$ for $j = 0$. The given distributions are non-standard. Similarly, the MH algorithm is employed. We first generate the candidate $\boldsymbol{\gamma}_j^*$ from the proposed distribution

$$N[\boldsymbol{\gamma}_j^{(r)}, \sigma_{\gamma j}^2\boldsymbol{\Sigma}_{\gamma j}^*]\mathrm{I}(\boldsymbol{\gamma}_j\mathbf{B}^\gamma > \mathbf{0}), \quad j = 0,$$
$$N[\boldsymbol{\gamma}_j^{(r)}, \sigma_{\gamma j}^2\boldsymbol{\Sigma}_{\gamma j}^*]\mathrm{I}(\boldsymbol{\gamma}_j\mathbf{B}^\gamma \geq \mathbf{0}), \quad j = 1, \cdots, q,$$

where

$$\boldsymbol{\Sigma}_{\gamma_j}^{*-1} = \frac{1}{2}\sum_{t=s+1}^{T}\mathbf{B}_{\varepsilon j}^{\gamma}(Y_{t-d})\mathbf{B}_{\varepsilon j}^{\gamma T}(Y_{t-d})\Big(Y_t - \sum_{j=1}^{p}\boldsymbol{\lambda}_j^T\mathbf{B}_{Yj}^{\lambda}(Y_{t-d})\Big)^2 + \mathbf{M}_\gamma/\tau_{\gamma j}.$$

This can be done by the rejection method. That is, we keep generating $\boldsymbol{\gamma}_j^*$ from $N[\boldsymbol{\gamma}_j^{(r)}, \sigma_{\gamma j}^2\boldsymbol{\Sigma}_{\gamma j}^*]$ until $\boldsymbol{\gamma}_j^*\mathbf{B}^\gamma > \mathbf{0}$ when $j = 0$ or $\boldsymbol{\gamma}_j^*\mathbf{B}^\gamma \geq \mathbf{0}$ when $j = 1, \cdots, q$. Then the candidate $\boldsymbol{\gamma}_j^*$ is accepted as the new value of $\boldsymbol{\gamma}_j^{(r+1)}$ with probability

$$\min\left\{1, \frac{p(\boldsymbol{\gamma}_j^*|\cdot)}{p(\boldsymbol{\gamma}_j^{(r)}|\cdot)}\right\}.$$

Finally, $\sigma_{\gamma j}^2$ is chosen such that the acceptance rate is 0.25 or more.

(3) The full conditional distributions of $\boldsymbol{\tau}_\lambda$ and $\boldsymbol{\tau}_\gamma$ are as follows:

$$p(\tau_{\lambda j}^{-1}|\cdot) \stackrel{D}{=} \mathrm{Gamma}\Big[\alpha_{\lambda 0} + K_\lambda^*/2,\ \beta_{\lambda 0} + \boldsymbol{\lambda}_j^T\mathbf{M}_{\lambda j}\boldsymbol{\lambda}_j/2\Big], \quad j = 1, \cdots, p, \tag{24}$$

$$p(\tau_{\gamma j}^{-1}|\cdot) \stackrel{D}{=} \mathrm{Gamma}\Big[\alpha_{\gamma 0} + K_\gamma^*/2,\ \beta_{\gamma 0} + \boldsymbol{\gamma}_j^T\mathbf{M}_{\gamma j}\boldsymbol{\gamma}_j/2\Big], \quad j = 0, \cdots, q. \tag{25}$$