Bayesian Analysis (2021) **16**, Number 4, pp. 1157–1185

Bayesian Modelling of Time-Varying Conditional Heteroscedasticity

Sayar Karmakar* and Arkaprava Roy†

Abstract. Conditional heteroscedastic (CH) models are routinely used to analyze financial datasets. The classical models such as ARCH-GARCH with time-invariant coefficients are often inadequate to describe frequent changes over time due to market variability. However, we can achieve significantly better insight by considering the time-varying analogs of these models. In this paper, we propose a Bayesian approach to the estimation of such models and develop a computationally efficient MCMC algorithm based on Hamiltonian Monte Carlo (HMC) sampling. We also established posterior contraction rates with increasing sample size in terms of the average Hellinger metric. The performance of our method is compared with frequentist estimates and estimates from the time constant analogs. To conclude the paper we obtain time-varying parameter estimates for some popular Forex (currency conversion rate) and stock market datasets.

Keywords: autoregressive model, B-splines, Hamiltonian Monte Carlo (HMC), non-stationary, posterior contraction, volatility.

1 Introduction

For datasets observed over a long period, stationarity turns out to be an oversimplified assumption that ignores systematic deviations of parameters from constancy. Thus time-varying parameter models have been studied extensively in the literature of statistics, economics, and related fields. For example, the financial datasets, due to external factors such as war, terrorist attacks, economic crisis, political events, etc. exhibit deviation from time-constant stationary models. Accounting for such changes is crucial as otherwise time-constant models can lead to incorrect policy implications as pointed out by Bai (1997). Thus functional estimation of unknown parameter curves using time-varying models has become an important research topic today. In this paper, we analyze popular conditional heteroscedastic models such as AutoRegressive Conditional Heteroscedasticity (ARCH) and Generalized ARCH (GARCH) from a Bayesian perspective in a time-varying setup. Before discussing our new contributions in this paper, we provide a brief overview of some previous works in this area.

In the regression regime, time-varying models have garnered a lot of recent attention; see, for instance, Fan and Zhang (1999), Fan and Zhang (2000), Hoover et al. (1998), Huang et al. (2004), Lin and Ying (2001), Ramsay and Silverman (2005), Zhang et al. (2002) among others. The models show time-heterogeneous relationship between response and predictors. Consider the following two regression models

$$\text{Model I: } y_i = x_i^\intercal \theta_i + e_i, \quad \text{Model II: } y_i = x_i^\intercal \theta_0 + e_i, \qquad i = 1, \ldots, n,$$

*Department of Statistics, University of Florida, sayarkarmakar@ufl.edu
†Department of Biostatistics, University of Florida, ark007@ufl.edu

 https://doi.org/10.1214/21-BA1267

where $x_i \in \mathbb{R}^d$ $(i = 1, \ldots, n)$ are the covariates, $^\intercal$ is the transpose, θ_0 and $\theta_i = \theta(i/n)$ are the regression coefficients. Here, θ_0 is a constant parameter and $\theta : [0, 1] \to \mathbb{R}^d$ is a smooth function. Estimation of $\theta(\cdot)$ has been considered by Hoover et al. (1998), Cai (2007) and Zhou and Wu (2010) among others. One popular way to decide if there is an evidence to favor time-varying models over the time-constant analogue is to perform hypothesis testing. See, for instance, Zhang and Wu (2012), Zhang and Wu (2015), Chow (1960), Brown et al. (1975), Nabeya and Tanaka (1988), Leybourne and McCabe (1989), Nyblom (1989), Ploberger et al. (1989), Andrews (1993) and Lin and Teräsvirta (1999). Zhou and Wu (2010) discussed obtaining simultaneous confidence bands (SCB) in model I, i.e. with additive errors. However their treatment is heavily based on the closed-form solution and it does not extend to processes defined by a more general recursion.

For time-varying AR, MA, or ARMA processes, the results from time-varying linear regression can be naturally extended. However, such an extension is not obvious for conditional heteroscedastic (CH hereafter) models. These are, by the simple definition of evolution is difficult to estimate even in the time-constant case. However, one cannot possibly ignore its usefulness in analyzing and predicting financial datasets. These models (even the simple time-constant ones) have remained primary tools for analyzing and forecasting certain trends for stock market datasets since Engle (1982) introduced the classical ARCH model and Bollerslev (1986) extended it to a more general GARCH model. However, with the rapid dynamics of market vulnerability, the simple classical time-constant models fail in terms of estimation or prediction due to over-compensating the past data. Several references point out the necessity of extending these classical models to a set-up where the parameters can change across time, for example Stărică and Granger (2005), Engle and Rangel (2005) and Fryzlewicz et al. (2008a). Consider the simple tvARCH(1) model

$$X_i = \sigma_i \zeta_i, \zeta_i \sim N(0, 1), \sigma_i^2 = \mu_0(i/n) + a_1(i/n) X_{i-1}^2.$$

Similar models can be defined for tvGARCH(1,1) as well where σ_i^2 has an additional recursive term involving σ_{i-1}^2

$$X_i = \sigma_i \zeta_i, \zeta_i \sim N(0, 1), \sigma_i^2 = \mu_0(i/n) + a_1(i/n) X_{i-1}^2 + b_1(i/n) \sigma_{i-1}^2.$$

When the two recursive parameters in a GARCH model sum up to 1, i.e. $a_1 + b_1 = 1$ it is usually called an integrated GARCH (iGARCH; or bubble garch/explosive garch by some authors) process which employing the above display can also be extended towards a time-varying analog i.e. $b_1(\cdot) = 1 - a_1(\cdot)$. A wide range of financial datasets exhibits iGARCH phenomena.

In the parlance of time-varying parameter models in the CH setting, numerous works discussed the CUSUM-type procedure, for instance, Kim et al. (2000) for testing change in parameters of GARCH(1,1). Kulperger et al. (2005) studied the high moment partial sum process based on residuals and applied it to residual CUSUM tests in GARCH models. Interested readers can find some more change-point detection results in the context of CH models in James Chu (1995), Chen and Gupta (1997), Lin et al. (1999), Kokoszka et al. (2000) or Andreou and Ghysels (2006).

A time-varying framework and a pointwise curve estimation using M-estimators for locally stationary ARCH models were provided by Dahlhaus and Subba Rao (2006). Since then, while several pointwise approaches were discussed in the tvARMA and tvARCH case (cf. Dahlhaus and Polonik (2009), Dahlhaus and Subba Rao (2006), Fryzlewicz et al. (2008a)), pointwise theoretical results for estimation in tvGARCH processes were discussed in Rohan and Ramanathan (2013) and Rohan (2013) for GARCH(1,1) and GARCH(p, q) models respectively. In a series of recent works Karmakar et al. (2021); Karmakar (2018) such models were discussed in wide generality. However, the focus remained frequentist, and the main goal accomplished there was to build simultaneous inference. One strong criticism for the CH type models remained that one needs a relatively large sample size ($n \sim 2000$) to achieve nominal coverage levels. The recursive definition of the models and a subsequent kernel-based method of estimating make it difficult to achieve satisfying results for relatively smaller sample sizes. This motivated us to explore a Bayesian way of building and estimating these models and use the posteriors to construct posterior estimates of the coefficient curves $\theta(\cdot)$.

In this paper, we develop a Bayesian estimation method for time-varying analogs of ARCH, GARCH, and iGARCH models. We model the time-varying functional parameters using cubic B-splines. In the context of general varying-coefficient modeling, spline bases are a popular choice for its convenience and flexibility (Hastie and Tibshirani, 1993; Gu and Wahba, 1993; Cai et al., 2000; Biller and Fahrmeir, 2001; Huang et al., 2002; Huang and Shen, 2004; Amorim et al., 2008; Fan and Zhang, 2008; Yue et al., 2014; Franco-Villoria et al., 2019). Specific to the literature of time-varying volatility modeling, B-spline-based models have also been explored (Engle and Rangel, 2008; Audrino and Bühlmann, 2009; Liu and Yang, 2016).

Our contributions in this paper are two-fold. Towards the methodological development, note that the tvARCH, tvGARCH, and tviGARCH models require complex shape constraints on the coefficient functions. We achieve those by imposing different hierarchical structures on B-spline coefficients. The constraints are designed to be able to develop an efficient sampling algorithm based on gradient-based Hamiltonian Monte Carlo (HMC) (Neal et al., 2011; Betancourt and Girolami, 2015; Betancourt, 2017; Livingstone et al., 2019). Strong motivation towards implementing such a Bayesian methodology was to circumvent the requirement of a huge sample size which is almost essential for effective estimation using the frequentist and kernel-based methods. This requirement on sample size has been frequently pointed out in the literature of ARCH/GARCH models and thus this was one of our main motivations to see if a reasonable estimation scheme can be designed in a Bayesian way.

Secondly, the existing literature on obtaining posterior concentration rates for dependent data is thin, even for an extremely simple model. To the best of our knowledge, ours is the first such attempt towards a theoretical development for these models under Gaussian-link. Posterior contraction rates for these models with respect to the average Hellinger metric are established. The main challenge therein is to construct exponentially consistent tests for these classes of models. Using some recently developed tools from Jeong (2019); Ning et al. (2020) we have developed such tests. We first establish posterior contraction rates with respect to average log-affinity and then the same rate is transferred to the average Hellinger metric. The frequentist literature on inference about time-varying needs very stringent moment assumption and local stationarity assumptions which are often difficult to verify. Moreover, for econometric datasets, the existence of even the fourth moment is often questionable. Thus this paper offers some alternative way to estimate coefficients under lesser assumptions.

The rest of the paper is organized as follows. Section 2 describes the proposed Bayesian model in detail. Section 3 discusses an efficient computational scheme for the proposed method. We calculate posterior contraction rate in Section 4. In Section 5 we study the performance of our proposed method in the light of. Section 6 deals with real data application of the proposed methods for the three separate models and concludes with a brief interpretation of the results. We wrap the paper up with discussions, some concluding remarks, and possible future directions in Section 7. The supplementary materials Karmakar and Roy (2021) contain theoretical proofs and some additional results.

2 Modeling

We elaborate on the models and our Bayesian framework for time-varying analogs of three specific cases that are popularly used to analyze econometric datasets.

2.1 tvARCH Model

Let $\{X_i\}$ satisfy the following time-varying ARCH(p) model for X_i given $\mathcal{F}_{i-1} = \{X_j : j \leq (i-1)\}$,

$$X_i | \mathcal{F}_{i-1} \sim \mathrm{N}(0, \sigma_i^2), \tag{2.1}$$

$$\sigma_i^2 = \mu(i/n) + \sum_{k=1}^{p} a_k(i/n) X_{i-k}^2, \tag{2.2}$$

where the parameter functions $\mu(\cdot), a_i(\cdot)$ satisfy

$$\mathcal{P} = \{\mu, a_k : \mu(x) \geq 0, 0 \leq a_k(x) \leq 1, \sup_x \sum_k a_k(x) < 1\}. \tag{2.3}$$

In a Bayesian regime we put priors on $\mu(\cdot)$ and $a_i(\cdot)$. To respect the shape-constraints as imposed by $\mathcal{P}$ we reformulate the problem. With B_j as the B-spline basis functions, let

$$\mu(x) = \sum_{j=1}^{K_1} \exp(\beta_j) B_j(x),$$

$$a_k(x) = \sum_{j=1}^{K_2} \theta_{kj} M_k B_j(x), \quad 0 \leq \theta_{kj} \leq 1,$$

$$\begin{aligned}
M_i =& \frac{\exp(\delta_i)}{\sum_{k=0}^{p}\exp(\delta_k)}, \quad i = 1, \ldots, p, \\
\delta_l \sim& N(0, c_1), \text{ for } 0 \le l \le p, \\
\beta_j \sim& N(0, c_2) \text{ for } 1 \le j \le K_1, \\
\theta_{kj} \sim& U(0,1) \text{ for } 1 \le k \le p, 1 \le j \le K_2.
\end{aligned}$$

The prior induced by above construction is $\mathcal{P}$-supported. The verification is very straightforward. In above construction, $\sum_{j=0}^{P} M_j = 1$. Thus $\sum_{j=1}^{P} M_j \le 1$. Since $0 \le \theta_{kj} \le 1, \sup_x a_i(x) \le M_i$. Thus $\sup_x \sum_{i=1}^{P} a_i(x) \le \sum_{i=1}^{P} M_i \le 1$. We have $\sum_{j=1}^{P} M_j \le 1$ if and only if $\delta_0 = -\infty$, which has probability zero. On the other hand, we also have $\mu(\cdot) \ge 0$ as we have $\exp(\beta_j) \ge 0$. Thus, the induced priors, described above are well supported in $\mathcal{P}$.

2.2 tvGARCH Model

Let $\{X_i\}$ satisfy the following time-varying GARCH(p,q) model for X_i given $\mathcal{F}_{i-1} = \{X_j : j \le (i-1)\}$,

$$\begin{aligned}
&X_i|\mathcal{F}_{i-1} \sim \mathrm{N}(0, \sigma_i^2), \\
&\sigma_i^2 = \mu(i/n) + \sum_{k=1}^{p} a_k(i/n) X_{i-k}^2 + \sum_{j=1}^{q} b_j(i/n)\sigma_{i-j}^2. \qquad (2.4)
\end{aligned}$$

Additionally we impose the following constraints on parameter space for the time-varying parameters,

$$\mathcal{P}_1 = \{\mu, a_i : \mu(x) \ge 0, 0 \le a_i(x), 0 \le b_j(x), \sup_x \sum_k a_k(x) + \sum_j b_j(x) < 1\}. \qquad (2.5)$$

The condition on the AR parameters imposed by (2.5) is somewhat popular in time-varying AR literature. See Dahlhaus and Subba Rao (2006); Fryzlewicz et al. (2008b); Karmakar et al. (2021) for example. Different from these references, we additionally do not assume existence of any unobserved local-stationary process that are close to the observed process.

To proceed with Bayesian computation, we again put priors on the unknown functions $\mu(\cdot), a_i(\cdot)$ and $b_j(\cdot)$'s such that they are supported in $\mathcal{P}_1$. Again the restrictions imposed by (2.5) are respected. The complete description of prior is

$$\begin{aligned}
\mu(x) =& \sum_{j=1}^{K_1} \exp(\beta_j) B_j(x), \\
a_k(x) =& \sum_{j=1}^{K_2} \theta_{kj} M_k B_j(x), \quad 0 \le \theta_{kj} \le 1, 1 \le k \le p,
\end{aligned}$$

$$
\begin{aligned}
b_k(x) =& \sum_{j=1}^{K_3} \eta_{kj} M_{k+p} B_j(x), \quad 0 \le \eta_{ij} \le 1, 1 \le k \le q, \\
M_i =& \frac{\exp(\delta_i)}{\sum_{k=0}^{p} \exp(\delta_k)}, \quad i = 1, \dots, p+q, \\
\delta_l \sim& N(0, c_1), \text{ for } 0 \le l \le p+q, \\
\beta_j \sim& N(0, c_2) \text{ for } 1 \le j \le K_1, \\
\theta_{kj} \sim& U(0,1) \text{ for } 1 \le k \le p, 1 \le j \le K_2, \\
\eta_{kj} \sim& U(0,1) \text{ for } 1 \le k \le q, 1 \le j \le K_3.
\end{aligned}
$$

Here B_j's are the B-spline basis functions. The parameters δ_j's are unbounded. The verification of support condition 2.5 for the proposed prior is similar.

2.3 tviGARCH Model

Although the GARCH(1,1) remains one of the most popular models to analyze econometric datasets, empirical evidence shows that these datasets regularly raise suspicion to the parameter space restriction $\sum_i a_i + \sum_j b_j < 1$. Note that we used a time-varying analog of this restriction for the tvGARCH modeling in Section 2.2. This often creates a very stringent condition as the validity of $\sum_i a_i(t) + \sum_j b_j(t) < 1$ is questionable. The special case for a time-constant GARCH model where this restriction fails is called an iGARCH model in the literature. We consider the following time-varying analog of iGARCH.

$$
\begin{aligned}
&X_i | \mathcal{F}_{i-1} \sim \mathrm{N}(0, \sigma_i^2), \\
&\sigma_i^2 = \mu(i/n) + \sum_{k=1}^{p} a_k(i/n) X_{i-k}^2 + \sum_{j=1}^{q} b_j(i/n) \sigma_{i-j}^2.
\end{aligned}
\tag{2.6}
$$

We impose the following constraints on parameter space for the time-varying parameters,

$$
\mathcal{P} = \{\mu, a_i : \mu(x) \ge 0, 0 \le a_k(x) \le 1, \sum_k a_k(x) + \sum_j b_j(x) = 1\}. \tag{2.7}
$$

The prior functions that allow us to reformulate the problem keeping it consistent with (2.7) is described below:

$$
\begin{aligned}
\mu(x) =& \sum_{j=1}^{K_1} \exp(\beta_j) B_j(x), \\
a_k(x) =& \sum_{j=1}^{K_2} \theta_{kj} M_k B_j(x), \quad 0 \le \theta_{kj} \le 1, 1 \le k \le p, \\
b_i(x) =& \sum_{j=1}^{K_3} \eta_{kj} M_{k+p} B_j(x), \quad 0 \le \eta_{ij} \le 1, 1 \le i \le (q-1),
\end{aligned}
$$

$$
\begin{aligned}
b_q(x) =& 1 - \left\{ \sum_{k=1}^{p} a_k(x) + \sum_{j=1}^{q-1} b_j(x) \right\}, \\
M_i =& \frac{\exp(\delta_i)}{\sum_{k=0}^{p+q-1} \exp(\delta_k)}, \quad i = 1, \ldots, p+q-1, \\
\delta_l \sim& N(0, c_1), \text{ for } 0 \leq l \leq p+q-1, \\
\beta_j \sim& N(0, c_2) \text{ for } 1 \leq j \leq K_1, \\
\theta_{kj} \sim& U(0,1) \text{ for } 1 \leq k \leq p, 1 \leq j \leq K_2, \\
\eta_{kj} \sim& U(0,1) \text{ for } 1 \leq k \leq (q-1), 1 \leq j \leq K_3.
\end{aligned}
$$

3 Posterior Computation and Implementation

3.1 tvARCH Structure

The complete likelihood L of the proposed Bayesian method is given by

$$
\begin{aligned}
L \propto \exp\Bigg(\sum_{i=p}^{n} \Big[-\{\mu(i/n) + \sum_{k=1}^{p} a_k(i/n) X_{i-k}^2\} + X_i^2 \log\big\{\mu(i/n) \\
+ \sum_{i=1}^{p} a_i(i/n) X_{i-i}^2\big\}\Big] - \sum_{j=1}^{K_1} \beta_j^2/(2c_2) - \sum_{l=0}^{p} \delta_l^2/(2c_1) \Bigg) \mathbf{1}_{0 \leq \theta_{kj} \leq 1},
\end{aligned}
$$

where $\mu(x) = \sum_{j=1}^{K_1} \exp(\beta_j) B_j(x), a_k(x) = \sum_{j=1}^{K_2} \theta_{kj} M_k B_j(x)$ and $M_j = \frac{\exp(\delta_j)}{\sum_{k=0}^{p} \exp(\delta_k)}$. We develop efficient Markov Chain Monte Carlo (MCMC) algorithm to sample the parameter β, θ and δ from the above likelihood. The computation of derivatives allows us to develop an efficient gradient-based MCMC algorithm to sample these parameters. We calculate the gradients of negative log-likelihood ($-\log L$) with respect to the parameters β, θ and δ. The gradients are given below,

$$
-\frac{d \log L}{\beta_j} = \exp(\beta_j) \left(1 - \sum_i \frac{B_j(i/n) X_i^2}{(\mu(i/n) + \sum_j a_j(i/n) X_{i-j}^2)} \right) + \beta_j / c_2, \tag{3.1}
$$

$$
-\frac{d \log L}{\theta_{kj}} = M_k \left(1 - \sum_i \frac{B_j(i/n) X_i^2}{(\mu(i/n) + \sum_j a_j(i/n) X_{i-j}^2)} \right), \tag{3.2}
$$

$$
\begin{aligned}
-\frac{d \log L}{\delta_j} =& \delta_j / c_1 + \sum_k (M_j \mathbf{1}_{\{j=k\}} - M_j M_k) \sum_i \theta_{kj} B_j(x) \\
& \times \left(1 - \sum_i \frac{B_j(i/n) X_i^2}{(\mu(i/n) + \sum_j a_j(i/n) X_{i-j}^2)} \right),
\end{aligned} \tag{3.3}
$$

where $\mathbf{1}_{\{j=k\}}$ stands for the indicator function which takes the value 1 when $j = k$.

3.2 tvGARCH / tviGARCH Structure

The complete likelihood L_2 of the proposed Bayesian method of (2.4) is given by

$$L_2 \propto \exp\Bigg(\sum_{t=p}^{n} \big[-\{\mu(i/n) + \sum_{i=1}^{p} a_i(i/n) X_{t-i} + \sum_{i=1}^{q} b_i(i/n)\lambda_{t-i}\big\} + X_t \log\big\{\mu(i/n)$$
$$+ \sum_{i=1}^{p} a_i(i/n) X_{t-i} + \sum_{i=1}^{q} b_i(i/n)\lambda_{t-i}\}\big] - \sum_{j=1}^{K_1} \beta_j^2/(2c_2) - \sum_{l=0}^{p} \delta_l^2/(2c_1)$$
$$- (d_1+1)\log\lambda_0 - d_1/\lambda_0 \Bigg) \mathbf{1}_{0\le\theta_{ij},\eta_{ij}\le 1}.$$

We calculate the gradients of negative log-likelihood $(-\log L_2)$ with respect to the parameters β, θ, η and δ. The gradients are given below,

$$-\frac{d\log L_2}{\beta_j} = \exp(\beta_j)\Bigg(1 - \sum_t \frac{B_j(i/n)X_{i-j}^2}{(\mu(i/n) + \sum_j a_j(i/n)X_{i-j}^2) + \sum_k b_k(i/n)\sigma_{i-k}^2)}\Bigg) + \beta_j/c_2,$$
$$-\frac{d\log L_2}{\theta_{lj}} = M_l\Bigg(1 - \sum_t \frac{B_j(i/n)X_{i-j}^2}{(\mu(i/n) + \sum_j a_j(i/n)X_{i-j}^2) + \sum_k b_k(i/n)\sigma_{i-k}^2)}\Bigg),$$
$$-\frac{d\log L_2}{\eta_{kj}} = M_{p+k}\Bigg(1 - \sum_t \frac{B_j(i/n)\sigma_{i-j}^2}{(\mu(i/n) + \sum_j a_j(i/n)X_{i-j}^2) + \sum_k b_k(i/n)\sigma_{i-k}^2)}\Bigg),$$
$$-\frac{d\log L_2}{\delta_j} = \delta_j/c_1 + \sum_k (M_j \mathbf{1}_{\{j=k\}} - M_j M_k)\times$$
$$\Bigg[\sum_{i\le p} \theta_{ij} B_j(x)\Bigg(1 - \sum_t \frac{B_j(i/n)X_{i-j}^2}{(\mu(i/n) + \sum_j a_j(i/n)X_{i-j}^2) + \sum_k b_k(i/n)\sigma_{i-k}^2)}\Bigg)\mathbf{1}_{\{j\le p\}} +$$
$$\sum_{1\le k\le q} \eta_{kj} B_j(x)\Bigg(1 - \sum_t \frac{B_j(i/n)\sigma_t^2}{(\mu(i/n) + \sum_j a_j(i/n)X_{i-j}^2) + \sum_k b_k(i/n)\sigma_{i-k}^2)}\Bigg)\mathbf{1}_{\{j>p\}}\Bigg].$$

While fitting tvGARCH(p,q), we assume for any $t<0$, $X_t^2 = 0, \sigma_t^2 = 0$. Thus, we need to additionally estimate the parameter σ_0^2. The derivative of the likelihood concerning σ_0^2 is calculated numerically using the `jacobian` function from R package `pracma`. For the tviGARCH, the derivatives are similar so we avoid computing them for the sake of brevity.

Based on these gradient functions, we develop gradient-based Hamiltonian Monte Carlo (HMC) sampling. Note that, parameter spaces of θ_{kj}'s have bounded support. We circumvent this by mapping any Metropolis candidate falling outside the parameter space back to the nearest boundary. HMC has two parameters, required to be specified. These are the leap-frog step and the step-size parameter. It is difficult to tune both of them simultaneously. We choose to tune the step size parameter to maintain an acceptance range between 0.6 to 0.8. After every 100 iterations, the step-length is adjusted (increased or reduced) accordingly if it falls outside. Neal et al. (2011) showed that a higher leapfrog step is better for estimation accuracy at the expense of greater computation. To maintain a balance between accuracy and computational complexity, we keep it fixed at 30 and obtain good results.

4 Large-Sample Properties

We now focus on obtaining posterior contraction rates for our proposed Bayesian models. The posterior consistency is studied in the asymptotic regime of increasing number of time points n. We study the posterior consistency with respect to the average Hellinger distance on the coefficient functions which is

$$d_{1,n}^2 = \frac{1}{n} d_H^2(\kappa_1, \kappa_2) = \frac{1}{n}\int (\sqrt{f_1} - \sqrt{f_2})^2,$$

where $f_1 = \prod_{i=1}^n P_{\kappa_1}(X_i|X_{i-1})$ and f_2 denotes the corresponding likelihoods.

Definition: For a sequence ϵ_n if $\Pi_n(d(f,f_0)|X^{(n)} \geq M_n\epsilon_n|X^{(n)}) \to 0$ in $F_{\kappa_0}^{(n)}$-probability for every sequence $M_n \to \infty$, then the sequence ϵ_n is called the *posterior contraction rate.*

All the proofs are postponed to the supplementary materials. The proof is based on the general contraction rate result for independent non-i.i.d. observations (Ghosal and Van der Vaart, 2017) and some results on B-splines based finite random series. The exponentially consistent tests are constructed leveraging on the famous Neyman-Pearson Lemma as in Ning et al. (2020). Thus the first step is to calculate posterior contraction rate with respect to average log-affinity $r_n^2(f_1, f_2) = -\frac{1}{n}\log\int f_1^{1/2} f_2^{1/2}$. Then we show that $r_n^2(f_1, f_2) \lesssim \epsilon_n^2$ implies $\frac{1}{n} d_H^2(f_1, f_2) \lesssim \epsilon_n^2$. We also consider following simplified priors for α_j and τ_i to get better control over tail probabilities,

$$\alpha_j \sim \text{Gamma}(g_1, g_1), \quad \tau_i \sim U(0,1). \tag{4.1}$$

4.1 tvARCH Model

Let $\kappa = (\mu, a_1)$ stand for the complete set of parameters. For sake of generality of the method, we put a prior on K_1 and K_2 with probability mass function given by,

$$\Pi(K_i = k) = b_{i1}\exp[-b_{i2}k(\log k)^{b_{i3}}], \tag{4.2}$$

for $i = 1, 2$. These priors have not been considered while fitting the model as it would require computationally expensive reversible jump MCMC strategy. The contraction rate will depend on the smoothness of true coefficient functions μ and a and the parameters b_{13} and b_{23} from the prior distributions of K_1 and K_2. Let $\kappa_0 = (\mu_0, a_{01})$ be the truth of κ.

Assumptions (A): There exists constants $M_X > 1, 0 < M_\mu < M_X$ such that,

(A.1) The coefficient functions satisfy $\sup_x \mu_0(x) < M_\mu$ and $\sup_x a_{01}(x) < 1 - M_\mu/M_X$.

(A.2) $\inf_x \min(\mu_0(x), a_{01}(x)) > \rho$ for some small $\rho > 0$.

(A.3) $E(X_0^2) < M_X$.

Assumptions (A.1)–(A.3) ensure

$$\mathbb{E}_{\kappa_0}(X_i^2) = \mathbb{E}_{\kappa_0}(\mathbb{E}_{\kappa_0}(X_i^2|X_{i-1})) < M_\mu + \left(1 - \frac{M_\mu}{M_X}\right) M_X < M_X$$

by recursion.

Theorem 1. *Under assumptions (A.1)–(A.3), let the true functions $\mu_0(\cdot)$ and $a_{10}(\cdot)$ be Hölder smooth functions with regularity level ι_1 and ι_2 respectively, then the posterior contraction rate with respect to the distance $d_{1,n}^2$ is*

$$\max\left\{n^{-\iota_1/(2\iota_1+1)}(\log n)^{\iota_1/(2\iota_1+1)+(1-b_{13})/2}, n^{-\iota_2/(2\iota_2+1)}(\log n)^{\iota_2/(2\iota_2+1)+(1-b_{23})/2}\right\},$$

where b_{ij} are specified in (4.2).

4.2 tvGARCH Model

Let $\kappa = (\mu, a_1, b_1)$ stand for the complete set of parameters. For sake of generality of the method, we put a prior on K_1, K_2 and K_3 with probability mass function given by,

$$\Pi(K_i = k) = b_{i1} \exp[-b_{i2} k (\log k)^{b_{i3}}], \tag{4.3}$$

for $i = 1, 2$. These priors have not been considered while fitting the model as it would require computationally expensive reversible jump MCMC strategy. The contraction rate will depend on the smoothness of true coefficient functions μ and a and the parameters b_{13} and b_{23} from the prior distributions of K_1 and K_2. Let $\kappa_0 = (\mu_0, a_{01})$ be the truth of κ.

Assumptions (B): There exists constants $M_X > 1, 0 < M_\mu < M_X$ such that,

(B.1) The coefficient functions satisfy $\sup_x \mu_0(x) < M_\mu$ and $\sup_x(a_{01}(x) + b_{01}(x)) < 1 - M_\mu/M_X$.

(B.2) $\inf_x \min(\mu_0(x), a_{01}(x), b_{01}(x)) > \rho$ for some small $\rho > 0$.

(B.3) $E(X_0^2) < M_X, \sigma_{00}^2 < M_X$.

Assumptions (B.1) and (B.3) ensure

$$\mathbb{E}_{\kappa_0}(X_i^2) = \mathbb{E}_{\kappa_0}(\mathbb{E}_{\kappa_0}(X_i^2|X_{i-1})) < M_\mu + \left(1 - \frac{M_\mu}{M_X}\right) M_X < M_X$$

by recursion. Similarly we have $\mathbb{E}(\sigma_i^2) = \mathbb{E}_{\kappa_0}(\mathbb{E}_{\kappa_0}(X_i^2|\mathcal{F}_i)) = \mathbb{E}_{\kappa_0}(X_i^2) < M_X$.

Theorem 2. *Under assumptions (B.1)–(B.3), let the true functions $\mu_0(\cdot)$, $a_{10}(\cdot)$ and $b_{10}(\cdot)$ be Hölder smooth functions with regularity level ι_1. ι_2 and ι_3 respectively, then the posterior contraction rate with respect to the distance $d^2_{1,n}$ is*

$$\max\Big\{n^{-\iota_1/(2\iota_1+1)}(\log n)^{\iota_1/(2\iota_1+1)+(1-b_{13})/2}, n^{-\iota_2/(2\iota_2+1)}(\log n)^{\iota_2/(2\iota_2+1)+(1-b_{23})/2},$$
$$n^{-\iota_3/(2\iota_3+1)}(\log n)^{\iota_2/(2\iota_3+1)+(1-b_{33})/2}\Big\},$$

where b_{ij} are specified in (4.3).

4.3 tviGARCH Model

Let $\kappa = (\mu, a_1)$ stand for the complete set of parameters. For sake of generality of the method, we put a prior on K_1 and K_2 with probability mass function given by,

$$\Pi(K_i = k) = b_{i1}\exp[-b_{i2}k(\log k)^{b_{i3}}], \tag{4.4}$$

for $i = 1, 2$. These priors have not been considered while fitting the model as it would require computationally expensive reversible jump MCMC strategy. The contraction rate will depend on the smoothness of true coefficient functions μ and a and the parameters b_{13} and b_{23} from the prior distributions of K_1 and K_2. Let $\kappa_0 = (\mu_0, a_{01})$ be the truth of κ.

(C.1) The coefficient functions satisfy $\sup_x \mu_0(x) < M_\mu < \infty$ for some M_μ.

(C.2) $\inf_x(\mu_0(x)) > \rho, \inf_x a_{01}(x) > \rho, \sup_x a_{0,1}(x) < 1-\rho$ for some $\rho > 0$.

Theorem 3. *Under assumptions (C.1)–(C.2), let the true functions $\mu_0(\cdot)$ and $a_{10}(\cdot)$ be Hölder smooth functions with regularity level ι_1 and ι_2 respectively, then the posterior contraction rate with respect to the distance $d^2_{1,n}$ is*

$$\max\Big\{n^{-\iota_1/(2\iota_1+1)}(\log n)^{\iota_1/(2\iota_1+1)+(1-b_{13})/2}, n^{-\iota_2/(2\iota_2+1)}(\log n)^{\iota_2/(2\iota_2+1)+(1-b_{23})/2}\Big\},$$

where b_{ij} are specified in (4.4).

5 Simulation

We run simulations to study the performance of our proposed Bayesian method in capturing the true coefficient functions under different true models. The hyperparameters c_1 and c_2 of the normal prior are all set 100, which makes the prior weakly informative. We consider 4, 5 and 6 equidistant knots for the B-splines when $n = 200, 500$ and 1000 respectively. We collect 10000 MCMC samples and consider the last 5000 as post burn-in samples for inferences. We shall compare the estimated functions with the true functions in terms of the posterior estimates of functions along with its 95% pointwise credible bands. The credible bands are calculated from the MCMC samples at each point $t = 1/T, 2/T, \ldots, 1$. We take the posterior mean as the posterior estimate of the unknown functions.

Since, to the best of our knowledge, there is no other Bayesian model for these time-varying conditional heteroscedastic models, we compare our Bayesian estimates with corresponding frequentist time-varying estimates. For computing the time-varying estimates of these models, we use the kernel-based method from Karmakar et al. (2021). The M-estimator of the parameter vector $\theta(t)$ are obtained using the conditional quasi log-likelihood. For instance, in the tvARCH(1) case, say $\theta(t) = (\mu(t), a_1(t))$

$$\hat{\theta}_{b_n}(t) = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=2}^{n} K\left(\frac{t - i/n}{b_n}\right) \ell(X_i|\mathcal{F}_{i-1}, \theta) \qquad t \in [0, 1],$$

where $\ell(\cdot)$ denotes the Gaussian log-likelihood. Note that these methods are fast but usually need a cross-validated choice of bandwidth b_n. We use $K(x) = 3/4(1-x^2)\mathbf{I}(|x| \leq 1)$ and an appropriately chosen bandwidth as suggested by the authors therein. Since our discussion also involves iGARCH formulation, we wrote a separate kernel-based frequentist estimation for iGARCH models analogously. Apart from these two time-varying estimates, we also obtain a time-constant fit on the same data to help initiate a discussion on whether there was a necessity of introducing coefficients varying with time. For this `tseries` and `rugarch` R packages are used respectively for ARCH/GARCH and iGARCH fits.

To compare these estimates, we evaluate the average mean square errors (AMSE) for the three estimates. Note that in an usual linear regression of response y on predictor X scenario, the fitted MSE is often defined as $\frac{1}{n}\sum(y_i - \hat{y}_i)^2$. Since here, $X_i|\mathcal{F}_{i-1} \sim N(0, \sigma_i^2)$, we use the following definition of AMSE

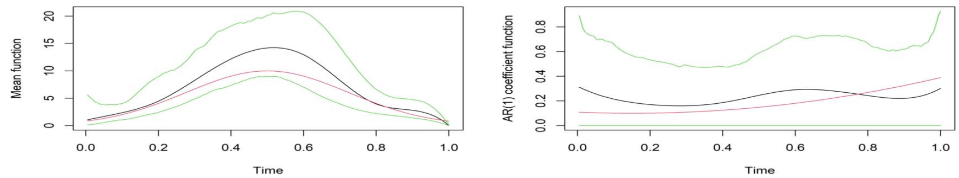
$$\text{AMSE} = \frac{1}{n}\sum_i (X_i^2 - \hat{\sigma}_i^2)^2,$$

where the $\hat{\sigma}_i^2$ is computed with the fitted parameter values as per the model under consideration. For example, for a tvGARCH(1,1) model we have

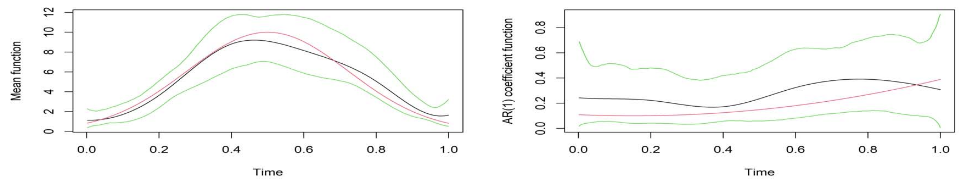
$$\hat{\sigma}_i^2 = \hat{\mu}(i/n) + \hat{a}(i/n)X_{i-1}^2 + \hat{b}(i/n)\hat{\sigma}_{i-1}^2,$$

where $\hat{\mu}(\cdot), \hat{a}(\cdot)$ and $\hat{b}(\cdot)$ are the estimated curves from the posterior. Replacing the response y_i by X_i^2 is natural as often autocorrelations of X_i^2 are checked to gauge presence of CH effect. Moreover, one of the early methods to deal with CH models was to view X_i^2 approximated by an TVAR(1) process. See Bose and Mukherjee (2003) and references therein. Similar estimators as our proposed AMSE to evaluate the fitting accuracy has been used in the literature previously. See Starica (2003); Fryzlewicz et al. (2008a); Rohan and Ramanathan (2013); Karmakar et al. (2021) for example.

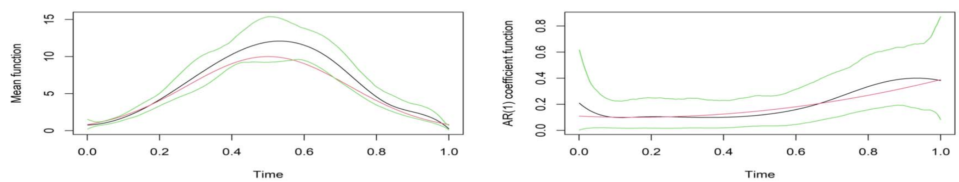
In the next three subsections, we provide the results for the three models, namely, tvARCH, tvGARCH, and tviGARCH. Our conclusions from these results are illustrated at the end of the section.



(a) $n = 200$



(b) $n = 500$



(c) $n = 1000$

Figure 1: tvARCH(1): True coefficient functions (red), estimated curve (black) along with the 95% pointwise credible bands (green) are shown for $T = 200, 500, 1000$ from top to bottom.

5.1 tvARCH Case

We start by considering the following tvARCH(1) model from 2.2. Three different choices for n are considered, $n = 200, 500$ and 1000. The true functions are,

$$\begin{aligned}\mu_0(x) =& 10 \exp\big(-(x-0.5)^2/0.1\big),\\ a_{10}(x) =& 0.4(x-0.15)^2 + 0.1.\end{aligned}$$

We compare the estimated functions with the truth for sample size 1000 in Figures 1. Table 1 illustrates the performance of our method with respect to other competing methods.

	ARCH(1)	Frequentist tvARCH(1)	Bayesian tvARCH(1)
$n = 200$	96.42	90.34	**85.22**
$n = 500$	128.07	122.53	**118.45**
$n = 1000$	138.06	130.33	**127.06**

Table 1: AMSE comparison for different sample sizes across different methods when the true model is tvARCH with $p = 1$.

5.2 tvGARCH Case

Next we explore the following GARCH(1,1) model (cf. 2.4)for different choices of n. The true functions are,

$$\begin{aligned}\mu_0(x) =& 1 - 0.8\sin(\pi x/2),\\ a_{10}(x) =& 0.5 - (x - 0.3)^2,\\ b_{10}(x) =& 0.4 - 0.5(x - 0.4)^2.\end{aligned}$$

Note that, estimation of GARCH, due to the additional $b_i(\cdot)$ parameter curves is a significantly more challenging problem and often requires a much higher sample size to have a reasonable estimation. We show by the means of the following pictures in Figure 2 that the estimation looks reasonable even for smaller sample sizes. The AMSE score comparisons are shown in Tables 2. The performance of our method is also contrasted with other competing methods.

	GARCH(1,1)	Frequentist tvGARCH(1,1)	Bayesian tvGARCH(1,1)
$n = 200$	33.99	31.84	**29.43**
$n = 500$	45.46	34.77	**33.33**
$n = 1000$	42.60	37.09	**36.55**

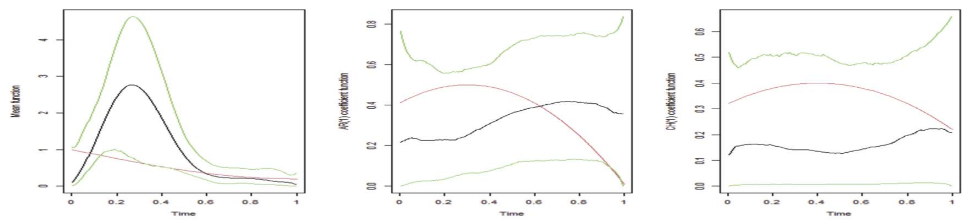
Table 2: AMSE comparison for different sample sizes across different methods when the true model is tvGARCH(1,1).

5.3 tviGARCH Case

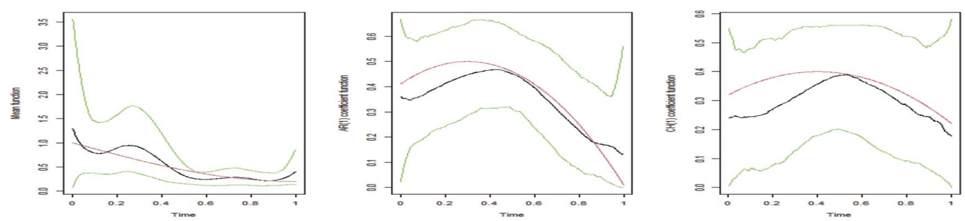
Finally we consider the tviGARCH(1,1) model (cf. 2.6) a special case of GARCH. Note that due to the constraint $a_1(\cdot) + b_1(\cdot) = 1$ we only consider the mean function and AR(1) function for plotting. For this case, our true functions are as follows

$$\begin{aligned}\mu_0(x) =& \exp\big(-(x - 0.5)^2/0.1\big),\\ a_{10}(x) =& 0.4(x - 1)^2 + 0.1.\end{aligned}$$

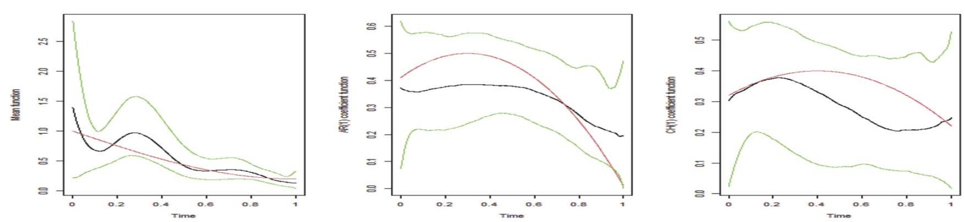
The frequentist computation for tviGARCH method is carried out based on a kernel-based estimation scheme along the same line as Karmakar et al. (2021). The estimated plots along with the 95% credible intervals are shown in Figure 3 for three sample sizes $n = 200, 500, 1000$ and the AMSE scores in Table 3.



(a) $n = 200$



(b) $n = 500$



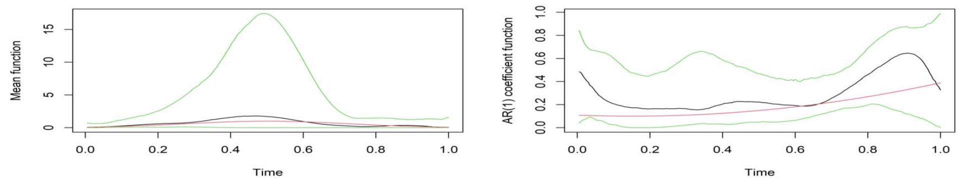
(c) $n = 1000$

Figure 2: tvGARCH(1,1): True coefficient functions (red), estimated curve (black) along with the 95% pointwise credible bands (green) are shown for $T = 200, 500, 1000$ from top to bottom.

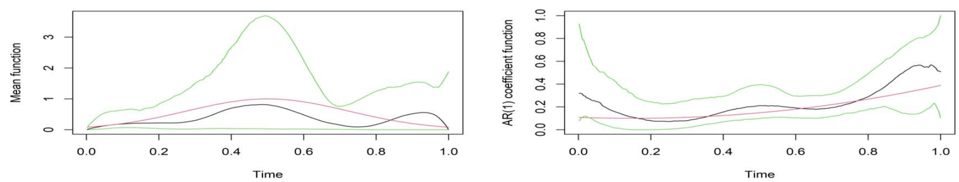
To summarize, our estimated functions are close to true functions for all the cases. We also find that the credible bands are getting tighter with increasing sample size. Thus estimates are improving in precision as sample size increases as shown in Figures 1 to 3. AMSEs of our Bayesian estimates are at least better for all the cases as in Tables 1 to 3. For tviGARCH, AMSE* is considered due to the huge and somewhat incomparable values of AMSE due to non-existent variance.

6 Real Data Application

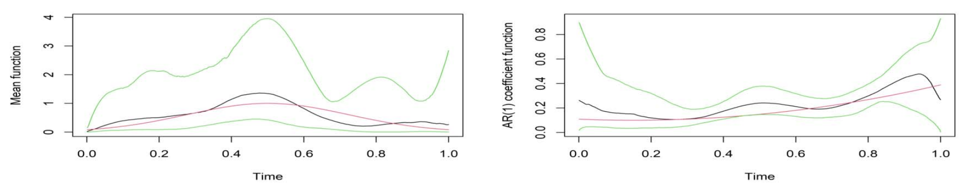
Towards applying our methods on real-life datasets we stick to econometric data for varying time horizons. These datasets show considerable time-variation justifying our models to be suitable for understanding how the parameter functions have evolved.



(a) $n = 200$



(b) $n = 500$



(c) $n = 1000$

Figure 3: tviGARCH(1,1): True coefficient functions (red), estimated curve (black) along with the 95% pointwise credible bands (green) are shown for $T = 200, 500, 1000$ from top to bottom.

Typically we model the log-return data of the daily closing price of these data to avoid the unit-root scenario. The log-return is defined as follows and is close to the relative return

$$Y_i = \log P_i - \log P_{i-1} = \log\left(1 + \frac{P_i - P_{i-1}}{P_{i-1}}\right) \approx \frac{P_i - P_{i-1}}{P_{i-1}},$$

where P_i is the closing price on the i^{th} day. Conditional heteroscedastic models are popularly used for model building, analysis and forecasting. Here we extend such models to a more sophisticated and general scenario by allowing the coefficient functions to vary.

In this section, we analyze two datasets: USD to JPY conversion and NASDAQ, a popular US stock market data. We analyze the NASDAQ data through tvGARCH(1,1) and tviGARCH(1,1) models and USDJPY conversion rate data through tvARCH(1) models. We just fit one lag for these models as multiple lag fits are similar and larger lags seem to be insignificant. This result is consistent with the findings in Karmakar

	iGARCH(1,1)	Frequentist tviGARCH(1,1)	Bayesian tviGARCH(1,1)
200	8.20	23.86	**8.14**
500	9.06	18.72	**9.06**
1000	10.59	25.92	**10.59**

Table 3: AMSE* comparison for different sample sizes across different methods when the true model is tviGARCH with $p = 1, q = 1$. AMSE* stands for mean of the log(AMSE).

et al. (2021), Fryzlewicz et al. (2008b) etc. Moreover, as Fryzlewicz et al. (2008b) claims, stock indices and Forex rates are more suited to GARCH and ARCH type of models respectively for their superior predictive performance. Each of these datasets was collected up to 31 July 2020. We exhibit our results for the last 200, 500 and 1000 days which capture the last 6 months, around 1.5 years, and around 3 years of data respectively. All these datasets were collected from `www.investing.com`. Note that these datasets are usually available for weekdays barring holidays and typically there are about 260 data points every single year.

6.1 USDJPY Data: tvARCH(1) Model

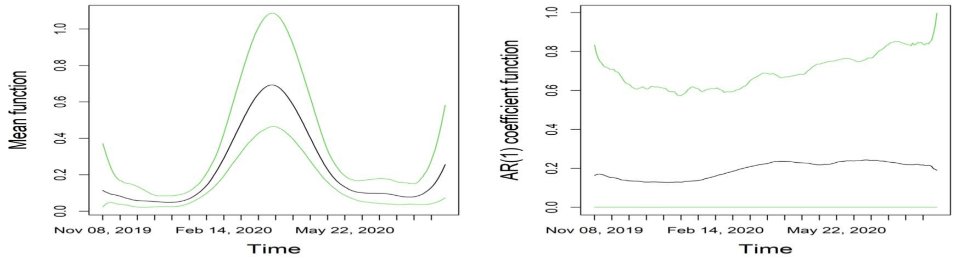
We obtain the following Figure 4 that shows our estimation for fitting a tvARCH(1) model on the USD to JPY conversion data for the last 200, 500 and 1000 days ending on 31 July 2020. The AMSE is also computed and contrasted with other competing methods in Table 4. Figure 4 depicts the estimated functions with 95% credible bands for different sample sizes. One can see the bands become much shorter for larger sample sizes. The mean coefficient function $\mu(\cdot)$ is generally time-varying for all three cases as one cannot fit a horizontal line through the 95% posterior bands. There seems to be a rise in the mean value around 100 days ago from July 31, 2020, which is right around the time the COVID-19 pandemic hit the world. With the analysis of $n = 1000$ days, we see that the volatility is quite high around October 2016 which coincides with the presidential election time of 2016. The AR(1) coefficient does not show the huge time-varying property. We also tabulate the AMSE for the three sample sizes in Table 4 and one can see for smaller sample sizes such as $n = 200$, the proposed Bayesian tvARCH achieves a significantly lower score but when the sample size grows then the performance becomes similar.

	ARCH(1)	Frequentist tvARCH(1)	Bayesian tvARCH(1)
$n = 200$	1.4572	1.2259	**1.1712**
$n = 500$	0.6281	0.5313	**0.5218**
$n = 1000$	0.4265	**0.3773**	0.3785

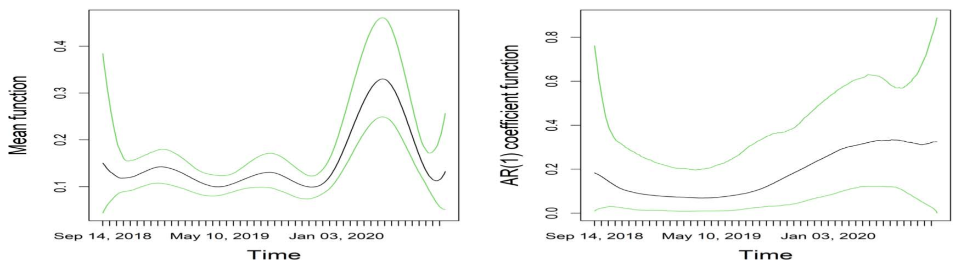
Table 4: AMSE comparison: tvARCH(1) model – USDJPY data.

6.2 NASDAQ Data: tvGARCH(1,1) Model

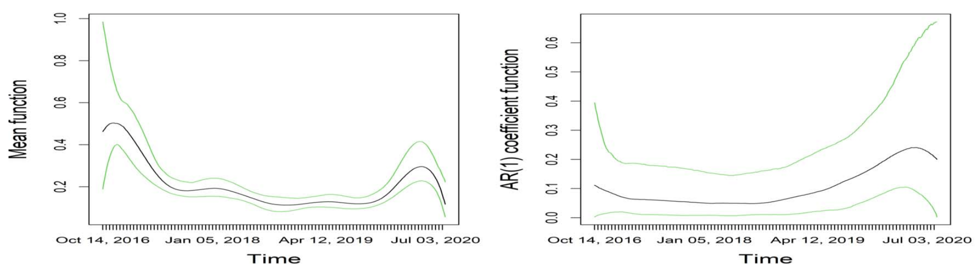
As has become standard in analyzing stock market datasets using GARCH models, we use time-varying GARCH for small orders. We obtain the following Figure 5 for fitting



(a) $n = 200$



(b) $n = 500$



(c) $n = 1000$

Figure 4: USDJPY data (tvARCH(1) model) Estimated curve (black) along with the 95% pointwise credible bands (green) are shown for $T = 200, 500, 1000$ from top to bottom.

a tvGARCH(1,1) model on the NASDAQ data for the last 200, 500, 1000 days ending on 31 July 2020. One can see the $a_1(\cdot)$ values are generally low and the $b_1(\cdot)$ values are higher which is consistent with how these outcomes turn out for time-constant estimates for econometric datasets. One can also see the role sample size plays in curating these time-varying estimates. For $n = 200$, the $b_1(\cdot)$ achieves high value of 0.6 around mid-March 2020 but for higher sample sizes it shows values as high as 0.8. One can also note the striking similarity for the analysis of the last 500 and 1000 days which is fairly consistent with the idea that estimation is more stable for such CH type models with a higher sample size. Nonetheless, the estimates for $n = 200$ seem quite smooth as well which can be seen as a benefit of our methodology. Table 5 provides a comparison of AMSE scores across the three methods for three sample sizes. The Bayesian tv-GARCH(1,1) performs relatively better than other methods and estimated curves have smaller credible bands with a growing sample size. The behavior of the mean function also shows higher volatility around the pandemic.

	GARCH(1,1)	Frequentist tvGARCH(1,1)	Bayesian tvGARCH(1,1)
$n = 200$	203.5917	203.5917	**202.6192**
$n = 500$	104.7443	90.5395	**90.3126**
$n = 1000$	46.16759	46.9225	**45.5618**

Table 5: AMSE comparison: tvGARCH(1,1) model – NASDAQ data.

6.3 NASDAQ Data: tviGARCH(1,1) Model

In Figure 5 the sum of estimated coefficient functions $a(\cdot) + b(\cdot)$ is close to 1 for a significant time-horizon. This motivates us to also fit tviGARCH(1,1) to analyze the same NASDAQ data. The estimated functions are presented in Figure 6 for the last $n = 200, 500$ and 1000 days. Table 6 compares the AMSE scores for the same three methods as before with varying sample sizes. The estimated mean and AR(1) functions of Figure 6 change a little from the estimated functions of tvGARCH(1,1) fit in Figure 5. Moreover, the effect of the three sample sizes is clear here with $n = 1000$ showing very precise bands and can reveal an interesting time-varying pattern.

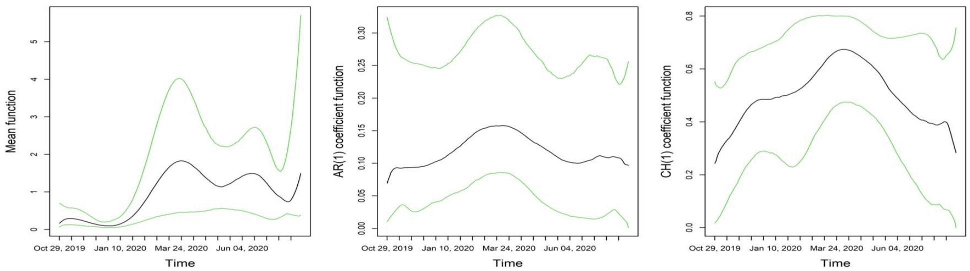
In terms of AMSE, one can see in Table 6 that the frequentist methods did worse than even the time-constant versions. The time-constant estimates were computed using the `rugarch` package in R. The Bayesian tviGARCH method provides significantly better AMSE uniform overall sample sizes. Here the mean function also shows higher volatility around the time when the pandemic struck us. Volatility due to the presidential election in 2016 can also be observed here.

	iGARCH(1,1)	Frequentist tviGARCH(1,1)	Bayesian tviGARCH(1,1)
$n = 200$	217.4988	278.4635	**206.6886**
$n = 500$	96.5001	132.544	**90.4456**
$n = 1000$	54.1171	260.4696	**46.3704**

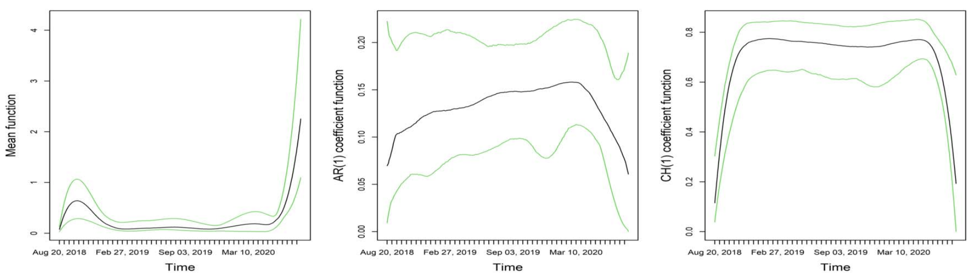
Table 6: AMSE comparison: tviGARCH(1,1) model – Nasdaq data.

6.4 Model Comparison

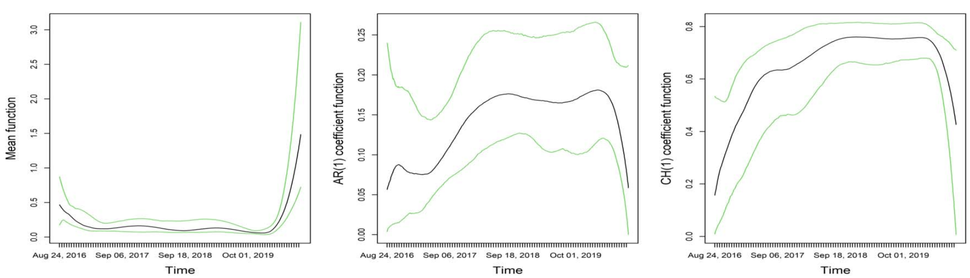
For the analysis of NASDAQ data, we have used two different models and thus it is pertinent to answer how should one choose between a competing class of models. We provide some measures in this subsection to decide between these two competing models. We start by comparing the performance of tvGARCH and tviGARCH models



(a) $n = 200$



(b) $n = 500$

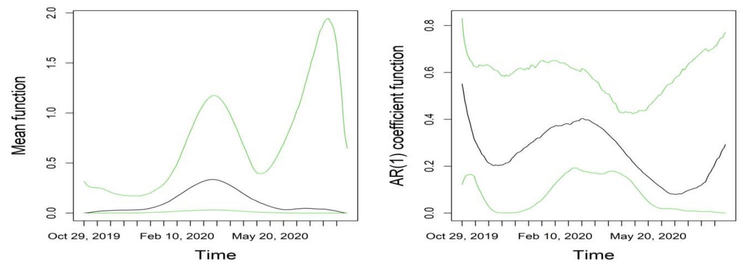


(c) $n = 1000$

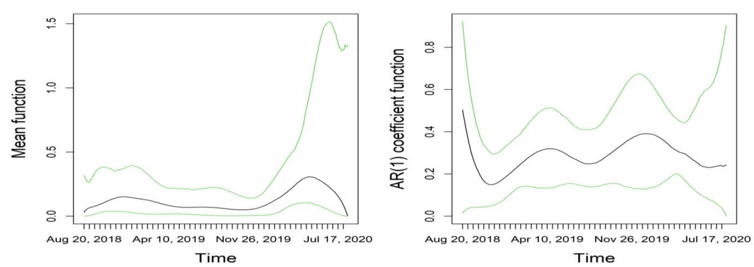
Figure 5: NASDAQ data (tvGARCH(1,1) model) Estimated curve (black) along with the 95% pointwise credible bands (green) are shown for $T = 200, 500, 1000$ from top to bottom.

in terms of Bayes factor (Kass and Raftery, 1995). Our calculation of the Bayes factor is based on the posterior samples using the harmonic mean identity of Neton and Raftery (1994). Let us denote B_{200}, B_{500} and B_{1000} as the Bayes factors for three sample sizes where

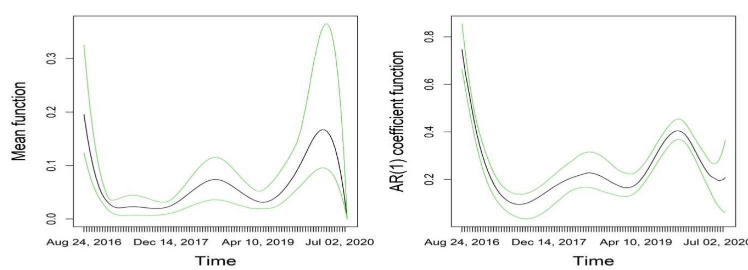
$$B_i = \frac{P(D^{(i)}|\text{tvGARCH})}{P(D^{(i)})|\text{tviGARCH})},$$



(a) $n = 200$



(b) $n = 500$



(c) $n = 1000$

Figure 6: NASDAQ data (tviGARCH(1,1) model) Estimated curve (black) along with the 95% pointwise credible bands (green) are shown for $T = 200, 500, 1000$ from top to bottom.

for sample size i and the corresponding dataset $D^{(i)}$. The values we obtain are $2\log(B_{200}) = 8.16$, $2\log(B_{500}) = 19.08$ and $2\log(B_{1000}) = 24.14$. According to guidelines from section 3.2 of Kass and Raftery (1995), there is 'positive' evidence in favor of tvGARCH for sample sizes 200 and 500. However, the same evidence becomes 'strong' for sample size 1000.

We also try to address out-of-sample predictive performance comparison here. Note that for a time-varying GARCH or time-varying iGARCH model this is generally a difficult task due to the assumed non-stationarity of the model. Thus, we take following approach to calculate out of sample joint predictive log-likelihoods for model comparison. Let us assume we have the data $D^{(n)}$ with n data points. To evaluate the joint predictive log-likelihood for the last $m(< n)$ most recent data points, we fit the models in (2.4) and (2.6) with the first $n - m$ data. Note that the assumed time horizons for these two models are n. Based on the estimated B-spline coefficients, and other parameters from each model, we can compute the joint predictive log-likelihood of the last m data points as

$$L_m^{(n)} = \frac{1}{m} \sum_{i=n-m+1}^{n} \frac{1}{2} \left\{ -X_i^2/\hat{\sigma}_i^2 - \log(\hat{\sigma}_i) - \log(2\pi) \right\},$$

where $\hat{\sigma}_i^2 = \hat{\mu}(i/n) + \hat{a}_1(i/n)X_{i-1}^2 + \hat{b}_1(i/n)\hat{\sigma}_{i-1}^2$. For the tviGARCH model, we have $\hat{b}_1(\cdot) = 1 - \hat{a}_1(\cdot)$.

Using this predictive log-likelihood we decide to evaluate the two fits from tvGARCH and tviGARCH in the following manner. For each of the sample sizes, we run it on three separate regimes of the data, the full data, and two halves of the data. In all these 8 settings, (three sample sizes, three possible regions of the data, but the latest half of the 1000-sized data is the same as the full data for sample size 500) we compute 10, 20, and 50 steps ahead forecast. We tabulate these results in Table 7. One can see that generally speaking, there is somewhat conclusive evidence towards the iGARCH model for a smaller sample size. This supports our motivation why we additionally provide a tviGARCH(1,1) modeling on the same dataset.

Based on our model comparison exercises, we have an interesting phenomenon where for in-sample model fit, tvGARCH is better but in terms of out-of-sample prediction, tviGARCH outperforms tvGARCH in almost all the cases. Note that, tvGARCH has one additional free parameter and thus is expected to fit the data better but since the estimated $a_1(\cdot)$ and $b_1(\cdot)$ coefficients are close to one satisfying the iGARCH formulation, the out-of-sample performance for tviGARCH may have exceeded that for tvGARCH.

As per the suggestion from a reviewer, we also add a one-step-ahead point forecasting exercise between these models. Here the computation method remains the same as outlined in the predictive log-likelihood computation however we only restrict ourselves to $m = 1$ to make the discussion concise. For this part of the exercise we choose to compute posterior mean of $(X_n^2 - \hat{\sigma}^2)^2$ where to ensure out-of-sample prediction $\hat{\sigma}^2$ is estimated solely based on $X_1, \ldots, X_{n-1}$. As one-step-ahead forecasts can be prohibitively misleading given it depends so much on one single location, we decide to take an average of over 15 random starting points over the entire time spectrum of 10 years resulting in 2518 points. For each of the sample sizes, we tabulate the performance in the following Table 8. Note that, here we are only comparing the two Bayesian time-varying models to see which one fits our data better. The advantage of predicting the future coefficients using B-spline is not available in the kernel-based frequentist method and thus is not included here in the discussion.

n	Steps (m)	Full GARCH	Full iGARCH	First Half GARCH	First Half iGARCH	Second (Latest) Half GARCH	Second (Latest) Half iGARCH
200	10	-2.1×10^8	$\mathbf{-1.982}$	-3.532	$\mathbf{-3.154}$	-8.9×10^6	$\mathbf{-2.251}$
	20	-2.689	$\mathbf{-1.881}$	-14.889	$\mathbf{-2.475}$	-90281	$\mathbf{-2.278}$
	50	-3.640	$\mathbf{-2.487}$	-86.81	$\mathbf{-2.877}$	-5.839	$\mathbf{-4.221}$
500	10	-2.842	$\mathbf{-2.068}$	-10.260	$\mathbf{-1.898}$	-1161	$\mathbf{-2.079}$
	20	-2.341	$\mathbf{-1.848}$	-3897	$\mathbf{-1.407}$	-343.49	$\mathbf{-1.893}$
	50	-2.147	$\mathbf{-2.112}$	-44.381	$\mathbf{-3.061}$	-932.7	$\mathbf{-2.371}$
1000	10	$\mathbf{-1.856}$	-1.936	$\mathbf{-2.499}$	-2.790	-2.842	$\mathbf{-2.068}$
	20	-1.911	$\mathbf{-1.789}$	-1.999	$\mathbf{-1.903}$	-2.341	$\mathbf{-1.848}$
	50	-2.266	$\mathbf{-2.214}$	-1.893	$\mathbf{-1.536}$	-2.147	$\mathbf{-2.112}$

Table 7: Joint log-likelihood for 10, 20, 50 steps ahead: Comparing tv-GARCH(1,1)/tviGARCH(1,1) model – NASDAQ data. Better model is in bold.

	Bayesian tvGARCH(1,1)	Bayesian tviGARCH(1,1)
$n = 200$	0.9914	**0.3477**
$n = 500$	1.5208	**1.4102**
$n = 1000$	**1.6378**	1.7033

Table 8: One-step-ahead out-of-sample forecast for NASDAQ data: Comparing tv-GARCH(1,1)/tviGARCH(1,1) model.

One can see, we again observe the same advantage of tviGARCH modeling over the tvGARCH one for smaller sample sizes. This is an interesting find of this paper in the context of the Bayesian model fitting to these datasets.

7 Discussion and Conclusion

In this paper, we consider a Bayesian estimation framework for time-varying conditional heteroscedastic models. Our prior specifications are amenable to Hamiltonian Monte Carlo for efficient computation. One of the key motivations towards going to a Bayesian regime was to achieve reasonable estimation even for a small sample size. Our simulation coverage shows good performance for all three models tvARCH, tvGARCH, tviGARCH for both small and large sample sizes. Importantly, in all three of the cases, we were able to establish posterior contraction rates. These calculations are, to the best of our knowledge, the first such work in even simple dependent models let alone the complicated recursions that these conditional heteroscedastic models demand. Moreover, the assumptions on the true functions and the number of moments needed were very minimal. An interesting future theoretical work would be to calculate posterior contraction rate with respect to empirical ℓ_2-distance which is a more desirable metric for function estimation. While analyzing real data, we see that the parameter curves vary significantly for the intercept terms, but not that much for AR or CH parameters. The associated codes to fit the three models are available at https://github.com/royarkaprava/tvARCH-tvGARCH-tviGARCH.

As future work, it will be interesting to explore multivariate time-varying volatility modeling (Tse and Tsui, 2002; Kwan et al., 2005) through a Bayesian framework similar to ours. Another interesting time-heterogeneity that we plan to explore through the glass of a Bayesian framework is regime-switching CH models where instead of the smooth time-varying functions the parameters change abruptly. We have a brief discussion in Section 6.4 about how to choose between competing models. Those discussions can easily be extended to choose a proper number of lags or to choose between different types of ARCH/GARCH models. We believe this would provide an interesting parallel to the usual penalized likelihood-based methods for model selection in time-series. Finally note that, even though we do provide some insights onto future prediction for these datasets for real data applications, that was not the main focus in this paper. Forecasting for the time-varying model is extremely tricky since it requires 'estimation' of the future parameter values. Although in-filled asymptotics can help in this regard, still the literature so far is very sparse in this direction for both Bayesian and frequentist regimes. We plan to explore this extensively in near future.

Supplementary Material

Proof of Theorems (DOI: 10.1214/21-BA1267SUPP; .pdf). The supplementary material includes the proof of Theorems 1, 2 and 3 and a general discussion of the main strategy behind them. We also include the traceplots for the MCMC chain from our simulations.

References

Amorim, L. D., Cai, J., Zeng, D., and Barreto, M. L. (2008). Regression splines in the time-dependent coefficient rates model for recurrent event data. *Statistics in Medicine*, 27(28):5890–5906. MR2597750. doi: `https://doi.org/10.1002/sim.3400`. 1159

Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in financial markets. *Journal of Econometrics*, 135(1-2):77–124. MR2328397. doi: `https://doi.org/10.1016/j.jeconom.2005.07.023`. 1158

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856. MR1231678. doi: `https://doi.org/10.2307/2951764`. 1158

Audrino, F. and Bühlmann, P. (2009). Splines for financial volatility. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):655–670. MR2749912. doi: `https://doi.org/10.1111/j.1467-9868.2009.00696.x`. 1159

Bai, J. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics*, 79(4):551–563. 1157

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*. MR1699395. doi: `https://doi.org/10.1017/CBO9780511470813.003`. 1159

Betancourt, M. and Girolami, M. (2015). Hamiltonian Monte Carlo for hierarchical models. *Current Trends in Bayesian Methodology with Applications*, 79(30):2–4. MR3644666. 1159

Biller, C. and Fahrmeir, L. (2001). Bayesian varying-coefficient models using adaptive regression splines. *Statistical Modelling*, 1(3):195–211. 1159

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327. MR0853051. doi: https://doi.org/10.1016/0304-4076(86)90063-1. 1158

Bose, A. and Mukherjee, K. (2003). Estimating the arch parameters by solving linear equations. *Journal of Time Series Analysis*, 24(2):127–136. MR1965808. doi: https://doi.org/10.1111/1467-9892.00296. 1168

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 37:149–192. With discussion by D. R. Cox, P. R. Fisk, Maurice Kendall, M. B. Priestley, Peter C. Young, G. Phillips, T. W. Anderson, A. F. M. Smith, M. R. B. Clarke, A. C. Harvey, Agnes M. Herzberg, M. C. Hutchison, Mohsin S. Khan, J. A. Nelder, Richard E. Quant, T. Subba Rao, H. Tong and W. G. Gilchrist and with reply by J. Durbin and J. M. Evans. MR0378310. 1158

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188. MR2328589. doi: https://doi.org/10.1016/j.jeconom.2005.08.004. 1158

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956. MR1804449. doi: https://doi.org/10.2307/2669476. 1159

Chen, J. and Gupta, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association*, 92(438):739–747. MR1467863. doi: https://doi.org/10.2307/2965722. 1158

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28:591–605. MR0141193. doi: https://doi.org/10.2307/1910133. 1158

Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39. MR2546797. doi: https://doi.org/10.3150/08-BEJ137. 1159

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Annals of Statistics*, 34(3):1075–1114. MR2278352. doi: https://doi.org/10.1214/009053606000000227. 1159, 1161

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007. MR0666121. doi: https://doi.org/10.2307/1912773. 1158

Engle, R. F. and Rangel, J. G. (2005). The spline garch model for unconditional volatility and its global macroeconomic causes. 1158

Engle, R. F. and Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, 21(3):1187–1222. 1159

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*, 27(5):1491–1518. MR1742497. doi: https://doi.org/10.1214/aos/1017939139. 1157

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27(4):715–731. MR1804172. doi: https://doi.org/10.1111/1467-9469.00218. 1157

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface*, 1(1):179. MR2425354. doi: https://doi.org/10.4310/SII.2008.v1.n1.a15. 1159

Franco-Villoria, M., Ventrucci, M., Rue, H., et al. (2019). A unified view on Bayesian varying coefficient models. *Electronic Journal of Statistics*, 13(2):5334–5359. MR4047589. doi: https://doi.org/10.1214/19-EJS1653. 1159

Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2008a). Normalized least-squares estimation in time-varying ARCH models. *Annals of Statistics*, 36(2):742–786. MR2396814. doi: https://doi.org/10.1214/07-AOS510. 1158, 1159, 1168

Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2008b). Normalized least-squares estimation in time-varying ARCH models. *Annals of Statistics*, 36(2):742–786. MR2396814. doi: https://doi.org/10.1214/07-AOS510. 1161, 1173

Ghosal, S. and Van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press. MR3587782. doi: https://doi.org/10.1017/9781139029834. 1165

Gu, C. and Wahba, G. (1993). Smoothing spline anova with component-wise Bayesian "confidence intervals". *Journal of Computational and Graphical Statistics*, 2(1):97–117. MR1272389. doi: https://doi.org/10.2307/1390957. 1159

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779. MR1229881. 1159

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822. MR1666699. doi: https://doi.org/10.1093/biomet/85.4.809. 1157, 1158

Huang, J. Z. and Shen, H. (2004). Functional coefficient regression models for nonlinear time series: a polynomial spline approach. *Scandinavian Journal of Statistics*, 31(4):515–534. MR2101537. doi: https://doi.org/10.1111/j.1467-9469.2004.00404.x. 1159

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128. MR1888349. doi: https://doi.org/10.1093/biomet/89.1.111. 1159

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14(3):763–788. MR2087972. 1157

James Chu, C.-S. (1995). Detecting parameter shift in garch models. *Econometric Reviews*, 14(2):241–266. 1158

Jeong, S. (2019). Frequentist properties of Bayesian procedures for high-dimensional sparse regression. MR4094193. 1159

Karmakar, S. (2018). Asymptotic theory for simultaneous inference under dependence. Technical report, University of Chicago. 1159

Karmakar, S. and Roy, A. (2021). "Supplementary Material of "Bayesian Modelling of Time-Varying Conditional Heteroscedasticity"." *Bayesian Analysis*. doi: https://doi.org/10.1214/21-BA1267SUPP. 1160

Karmakar, S., Richter, S., and Wu, W. B. (2021). Simultaneous inference for time-varying models. *Journal of Econometrics*, doi: https://doi.org/10.1016/j.jeconom.2021.03.002. 1159, 1161, 1168, 1170, 1172

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795. MR3363402. doi: https://doi.org/10.1080/01621459.1995.10476572. 1176, 1177

Kim, S., Cho, S., and Lee, S. (2000). On the cusum test for parameter changes in garch (1, 1) models. *Communications in Statistics – Theory and Methods*, 29(2):445–462. MR1749743. doi: https://doi.org/10.1080/03610920008832494. 1158

Kokoszka, P., Leipus, R., et al. (2000). Change-point estimation in ARCH models. *Bernoulli*, 6(3):513–539. MR1762558. doi: https://doi.org/10.2307/3318673. 1158

Kulperger, R., Yu, H., et al. (2005). High moment partial sum processes of residuals in garch models and their applications. *Annals of Statistics*, 33(5):2395–2422. MR2211090. doi: https://doi.org/10.1214/009053605000000534. 1158

Kwan, C., Li, W., and Ng, K. (2005). A multivariate threshold garch model with time-varying correlations. *Econometric Reviews*. 1180

Leybourne, S. J. and McCabe, B. P. M. (1989). On the distribution of some test statistics for coefficient constancy. *Biometrika*, 76(1):169–177. MR0991435. doi: https://doi.org/10.1093/biomet/76.1.169. 1158

Lin, C.-F. J. and Teräsvirta, T. (1999). Testing parameter constancy in linear models against stochastic stationary parameters. *Journal of Econometrics*, 90(2):193–213. MR1703341. doi: https://doi.org/10.1016/S0304-4076(98)00041-4. 1158

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–126. With comments and a rejoinder by the authors. MR1952726. doi: https://doi.org/10.1198/016214501750333018. 1157

Lin, S.-J., Yang, J., et al. (1999). *Testing shifts in financial models with conditional heteroskedasticity: an empirical distribution function approach*. School of Finance and Economics, University of Technology, Sydney. 1158

Liu, R. and Yang, L. (2016). Spline estimation of a semiparametric garch model. *Econometric Theory*, 32(4):1023. MR3530459. doi: https://doi.org/10.1017/S0266466615000055. 1159

Livingstone, S., Betancourt, M., Byrne, S., and Girolami, M. (2019). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli*, 25(4A):3109–3138. MR4003576. doi: https://doi.org/10.3150/18-BEJ1083. 1159

Nabeya, S. and Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Annals of Statistics*, 16(1):218–235. MR0924867. doi: https://doi.org/10.1214/aos/1176350701. 1158

Neal, R. M. et al. (2011). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2. MR2858447. 1159, 1164

Neton, M. and Raftery, A. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 1–48. MR1257793. 1176

Ning, B., Jeong, S., Ghosal, S., et al. (2020). Bayesian linear regression for multivariate responses under group sparsity. *Bernoulli*, 26(3):2353–2382. MR4091112. doi: https://doi.org/10.3150/20-BEJ1198. 1159, 1165

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230. MR0999682. 1158

Ploberger, W., Krämer, W., and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40(2):307–318. MR0994952. doi: https://doi.org/10.1016/0304-4076(89)90087-0. 1158

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition. MR2168993. 1157

Rohan, N. (2013). A time varying garch (p, q) model and related statistical inference. *Statistics & Probability Letters*, 83(9):1983–1990. MR3079033. doi: https://doi.org/10.1016/j.spl.2013.04.030. 1159

Rohan, N. and Ramanathan, T. V. (2013). Nonparametric estimation of a time-varying GARCH model. *Journal of Nonparametric Statistics*, 25(1):33–52. MR3039969. doi: https://doi.org/10.1080/10485252.2012.728600. 1159, 1168

Starica, C. (2003). Is GARCH (1, 1) as good a model as the accolades of the Nobel prize would imply? *Available at SSRN 637322*. 1168

Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87(3):503–522. 1158

Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3):351–362. MR1939906. doi: https://doi.org/10.1198/073500102288618496. 1180

Yue, Y. R., Simpson, D., Lindgren, F., Rue, H., et al. (2014). Bayesian adaptive smoothing splines using stochastic differential equations. *Bayesian Analysis*, 9(2):397–424. MR3217001. doi: https://doi.org/10.1214/13-BA866. 1159

Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *Annals of Statistics*, 40(3):1376–1402. MR3015029. doi: https://doi.org/10.1214/12-AOS1010. 1158

Zhang, T. and Wu, W. B. (2015). Time-varying nonlinear regression models: nonparametric estimation and model selection. *Annals of Statistics*, 43(2):741–768. MR3319142. doi: https://doi.org/10.1214/14-AOS1299. 1158

Zhang, W., Lee, S.-Y., and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, 82(1):166–188. MR1918619. doi: https://doi.org/10.1006/jmva.2001.2012. 1157

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society. Series B. (Statistical Methodology)*, 72(4):513–531. MR2758526. doi: https://doi.org/10.1111/j.1467-9868.2010.00743.x. 1158

Acknowledgments

We would like to thank the editor, the associate editor, and two anonymous referees for their constructive suggestions that improved the quality of the manuscript.