

Bayesian estimation of the autocovariance of a model error in time series

Yoon Bae Jun

Iowa State University, United States

Chae Young Lim

Seoul National University, Korea

Kun Ho Kim*

Concordia University, Canada

Abstract

Autocovariance of the error term in a time series model plays a key role in the estimation and inference for the model that it belongs to. Typically, some arbitrary parametric structure is assumed upon the error to simplify the estimation, which inevitably introduces potential model-misspecification. We thus conduct nonparametric estimation of it. To avoid the difficult bandwidth selection issue under the traditional nonparametric truncation approach, this paper conducts the Bayesian estimation of its spectral density in a frequency domain. To this end, we consider two cases: fixed error variance and time-varying one. Each approach is taken to estimate the spectral density of the autocovariance and the model parameters. The methodology is applied to exchange rate forecasting and proves to compete favorably against some benchmark models, including the random walk without drift.

Keywords: Bayesian time series, Nonparametric estimation, Spectral density, Time-varying volatility, Exchange rate forecasting

*Corresponding author

Email addresses: junpeea@gmail.com (Yoon Bae Jun), twinwood@snu.ac.kr (Chae Young Lim), kunhokim8@gmail.com (Kun Ho Kim)

URL: <https://github.com/junpeea> (Yoon Bae Jun)

1. Introduction

Heteroskedasticity has played a key role in economic time series analysis. It is well-known that a number of important macroeconomic or financial return processes exhibit the pattern of heteroskedasticity (Engle, 1982) [1]. It has been one of the central issues in the literature how to model such processes to capture their underlying dynamics successfully. Classical time-varying volatility models, such as *ARCH* and *GARCH*, that accommodate conditional heteroskedasticity are typically stationary with some restrictions on the parameter values. The *ARCH* and *GARCH* processes are unconditionally uncorrelated while they are still dependent. For the uncorrelated error, the corresponding spectral density is constant and estimating unconditional variability of classical *ARCH*- or *GARCH*-type models may not be of interest. To that end, one can incorporate a heteroskedastic time-varying volatility within the model where the error terms are not only dependent but also unconditionally correlated. For example, see Kim and Kim (2016) [2]. In this study, we illustrate how such a framework can be useful in practice.

Specifically, we consider the following framework:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{y} = (y_1, \dots, y_T)'$ be a variable of interest, $\mathbf{X} = (X_1; \dots; X_p)$ be a covariate matrix where X_i is a $(T \times 1)$ vector of the i -th predictor, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown regression parameters and $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_T)'$ is an error vector. Here we assume $\epsilon_t = \sigma_{\epsilon,t} e_t$, where e_t is a weakly stationary and stochastically continuous Gaussian process with $E(e_t) = 0$ and $Var(e_t) = 1$ and $\sigma_{\epsilon,t}^2$ is a time-varying volatility of ϵ_t . Here $\sigma_{\epsilon,t}^2$ is also a marginal variance function of the error ϵ_t . We denote $\gamma(\cdot)$ by the autocorrelation function for e_t . Inference for the autocorrelation will be carried out in a frequency domain by utilizing the corresponding spectral density, $\lambda(\cdot)$. In such a setting, a log transformation of the time-varying volatility $\sigma_{\epsilon,t}^2$ can be represented by *B*-spline functions. Then, a hierarchical Bayesian framework can be adopted such that some appropriate

prior distributions can yield posterior distributions of the spectral density and of the time-varying volatility as well as those for the regression parameters.

As shown in Dey *et al.* (2018) [3], adoption of a Gaussian process prior for the log transformation of a spectral density not only helps capture the time dependence structure in the error process, but also enhances the precision of forecasting through a better forecast of the model error. The current study employs the prior choice for the spectral density suggested by Dey *et al.* (2018) [3], but also extends their work by allowing the volatility in (1) to change over time. It is well-known that a model with heteroskedasticity captures the underlying dynamics of an economic/financial time series variable more efficiently than a naive constant-volatility model. By adopting a flexible modeling for the volatility of such a process, one can expect potentially more precise estimation and forecasting for the variable of interest.

Estimation of the autocorrelation function for the error process via the spectral density and of the related model parameters is conducted under a Bayesian framework. In particular, the hierarchical Bayesian framework is known to be useful for macro-finance forecasting (Kim, 2011 [4]; Dey *et al.*, 2018 [3]). Moving into the frequency domain from the time domain, we use an approximate likelihood for the parameters including the spectral density.

Unlike the traditional time-domain-based one, the *frequency*-domain-based approach adopted in this study allows us to avoid the difficult issue of how to choose a nonparametric bandwidth for the truncation of long-run variance. In addition, the approach ensures the Bayesian estimate of the covariance matrix to be *positive-definite*, which is a non-trivial advantage in a data-rich environment. Furthermore, the proposed Bayesian estimation does not assume a specific parametric form for any time-dependent structure, by allowing one to incorporate relevant information into the covariance matrix estimation through the prior distribution of the spectral density, which helps reduce the likelihood of model mis-specification.

The methodology proposed in this study is applied to the issue of *exchange rate forecasting*. One of the crying needs in the modern macro-finance literature

is to develop a methodology or a model that can forecast foreign exchange rates with decent accuracy (Rossi, 2013 [5]). Over the years, many approaches with various degrees of success have been proposed in the literature. Meese and Rogoff (1983) [6] compare out-of-sample forecasting accuracy of various exchange rate models and find out that a random walk model outperforms the rest at one- to twelve-month horizons for the dollar/pound, dollar/mark and dollar/yen. Engel and Hamilton (1990) [7] develop a long-swing model that explains the uni-directional move in the dollar value for long periods of time. They claim that their model generates better forecasts than the random walk process.

Mark (1995) [8] shows that fundamentals, such as money stock and output, are useful in predicting exchange rates and that the regression with bias-adjusted coefficients can outperform a random walk process in forecasting the spot exchange rate. Kilian (1999) [9] reports no evidence of higher predictability for exchange rates at longer forecast horizons and concludes that the linear model framework underlying many empirical studies are likely to be mis-specified, suggestive of a non-linear data-generating process. To that end, Kilian and Taylor (2003) [10] propose a non-linear, exponential smooth transition autoregressive model to approximate the time series behavior of real exchange rate. Molodtsova and Papell (2009) [11] find strong evidence of exchange rate predictability with the Taylor rule fundamentals at the one-month horizon for 12 OECD countries over the post-Bretton Woods period. Ince *et al.* (2016) [12] further show that the Taylor-rule-fundamental model outperforms the Taylor-rule-differential model (Engel *et al.*, 2008) [13] and other traditional models in out-of-sample exchange rate forecasting.

On the Bayesian side, Wright (2008) [14] suggests pooling forecasts from several different exchange-rate models, called the Bayesian model averaging, for the problem of pseudo-out-of-sample exchange rate predictions. He shows that the forecasts generated by the model averaging methodology are very close to those based on the random walk process. Through the flexible nonparametric modeling approach proposed in this paper, the frequency-domain-based Bayesian

estimation of the error auto-covariance and the time-varying volatility would yield further progress in forecasting exchange rate variables.

The paper is organized as follows: Section 2 introduces some background relevant to our proposal. Specifically, we present particulars of the proposed non-parametric modeling of the error covariance and of the time-varying-volatility approach. The estimation procedure is described as well. A simulation study supporting our approach and a real data example involving foreign exchange rate data are provided in Section 3. The paper ends with its conclusion and discussion on potential extensions in Section 4. Details of the proposed methodology are relegated to an appendix.

2. Proposed Method

Recall that we consider the time series regression setting given in (1) with $\epsilon_t = \sigma_{\epsilon,t}e_t$:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\sigma}_t \odot \mathbf{e}, \quad (2)$$

where we let $\boldsymbol{\sigma}_\epsilon = (\sigma_{\epsilon,1}, \dots, \sigma_{\epsilon,T})'$ and $\mathbf{e} = (e_1, \dots, e_T)'$ to explicitly decompose the error into a *time-varying* volatility and a normalized stationary error process. Here \odot represents an element-wise multiplication. We further assume that e_t follows the stationary Gaussian process with zero mean, unit variance and autocorrelation function $\gamma(h)$.

Meese and Rogoff (1983) [6] and Rossi (2013) [5] mention *instabilities* as one of the potential reasons why traditional economic models with time-invariant parameters have not been successful in out-of-sample forecasting of exchange rates. Our framework in (2) partially addresses the issue by introducing the time-varying volatility, which could potentially yield more precise forecasts of exchange rates than the traditional ones.

Given the assumption of weakly stationarity for e_t in our model (2), Wold's theorem (Priestley, 1981 [15]) states that the function $\gamma(\cdot)$ is an autocorrelation

function, equivalent to the spectral measure $\Lambda(w)$ such that:

$$\gamma(h) = \int_{-\pi}^{\pi} \exp(iwh)\Lambda(dw) \quad (3)$$

where $i = \sqrt{-1}$ and $\Lambda(w)$ is a distribution function on $[-\pi, \pi)$. When $\Lambda(w)$ is differentiable everywhere, the spectral density function $\lambda(w) = \frac{d\Lambda(w)}{dw}$ exists and (3) becomes

$$\gamma(h) = \int_{-\pi}^{\pi} \exp(iwh)\Lambda(dw) = \int_{-\pi}^{\pi} \exp(iwh)\lambda(w)dw. \quad (4)$$

The spectral density function $\lambda(w)$ is recovered by the Fourier transformation of $\gamma(\cdot)$. That is, $\lambda(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-ihw)$ for $w \in [-\pi, \pi)$. Following the relation in (4), modeling the spectral density $\lambda(w)$ is equivalent to modeling the autocorrelation function $\gamma(h)$.

It is well-known that the truncated version of the Fourier transformation of a sample autocorrelation yields an estimate of the spectral density, which is called the *periodogram*. The periodogram is defined by:

$$\mathcal{I}_n(w) = \frac{1}{2n\pi} \left| \sum_{t=1}^n e_t \exp(-itw) \right|^2,$$

where $w \in [-\pi, \pi)$. From the Gaussianity assumption on e , $\mathcal{I}_n(w)$ is asymptotically exponentially distributed with mean $\lambda(w)$ and \mathcal{I}_n are asymptotically independent at Fourier frequencies, $w_j = 2\pi j/n$, for $j = -\lfloor \frac{n}{2} \rfloor, \dots, \lfloor \frac{n-1}{2} \rfloor$; $\lfloor x \rfloor$ being the greatest integer less than or equal to x . Since $\lambda(-w) = \lambda(w)$ as well as $\mathcal{I}_n(-w) = \mathcal{I}_n(w)$, we will only consider $j = 1, \dots, \lfloor \frac{n-1}{2} \rfloor (= m)$. Carter and Kohn(1997) [16] utilize that the logarithm of a standard exponential random variable can be approximated by a mixture of five Gaussian random variables to have the following *approximated* relationship:

$$\log \mathcal{I}_n(w) = \log \lambda(w) + \xi(w) \quad (5)$$

with ξ having a distribution $\pi(\xi)$ such that

$$\pi(\xi) \propto \sum_{l=1}^5 p_l \phi_{v_l}(\xi - k_l), \quad (6)$$

where $\phi_v(\xi - k)$ is normal density with mean k and variance v^2 . Here (p_l, k_l, v_l) for $l = 1, \dots, 5$ are known and are provided by Carter and Kohn (1997) [16].

We use (5) at Fourier frequencies to obtain an approximate likelihood for the spectral density in Bayesian inference. That is, the autocorrelation function $\gamma(h)$ for \mathbf{e} is then implicitly considered in $\boldsymbol{\varphi} = \boldsymbol{\theta} + \boldsymbol{\xi}$ in the frequency domain, where $\boldsymbol{\varphi} = (\log(\mathcal{I}_n(w_1)), \dots, \log(\mathcal{I}_n(w_m)))'$ is a vector of the logarithm of the periodogram and $\boldsymbol{\theta} = (\log(\lambda(w_1)), \dots, \log(\lambda(w_m)))'$ is a vector of the logarithm of the spectral density at Fourier frequencies. Here $\boldsymbol{\xi}$ is a vector of the five component mixture Gaussian random variables at Fourier frequencies from (5). Moreover, $\log(\sigma_{\epsilon,t}^2) \equiv \eta_{\epsilon,t}$ is represented by cubic B -spline basis functions such that $\eta_{\epsilon,t} = \sum_{b=1}^d \delta_b \phi_b(t)$.

In this model setting, the parameters to be estimated are $\boldsymbol{\beta}$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)'$, $\boldsymbol{\theta}$ as well as latent label information, denoted by $\boldsymbol{\psi}$, for $\boldsymbol{\xi}$. For the Bayesian inference, we consider prior specifications of the parameters as follows: For $\boldsymbol{\beta}$, we consider $\boldsymbol{\beta} \sim N(\mu_\beta \mathbf{1}, \sigma_\beta^2 \mathbf{I})$, which is a multivariate Gaussian distribution with mean vector $\mu_\beta \mathbf{1}$ and covariance matrix $\sigma_\beta^2 \mathbf{I}$, where $\mathbf{1}$ is a vector of one's and \mathbf{I} is an identity matrix. Similarly, for $\boldsymbol{\delta}$, we assume $\boldsymbol{\delta} \sim N(\mu_\delta \mathbf{1}, \sigma_\delta^2 \mathbf{I})$. For $\boldsymbol{\theta}$, the log spectral density at Fourier frequencies, we consider the prior from the Gaussian process, $GP(\nu(\cdot), \tau(\cdot, \cdot))$, where mean function $\nu(w)$ and covariance kernel $\tau(w_i, w_j) = (1/\tau_0^2) \exp(-\rho|w_i - w_j|)$. For the hyper-parameters in the prior distribution of $\boldsymbol{\theta}$, we further consider $\rho \propto 1$ and $\tau_0 \sim G(a, b)$, where $G(a, b)$ is the Gamma distribution with mean ab . With the prior distributions, we can obtain conditional posterior distributions of the parameter given the others, which will be used to obtain the Bayesian Markov Chain Monte Carlo (MCMC) samples. Detailed derivation of the conditional posterior distributions for each parameter is relegated to the appendix.

The k th-step-ahead forecast based on the above model for $1 \leq k \leq M$ is obtained by estimating the conditional expectation given the data. Specifically, we use the estimated conditional expectation given the data, i.e. $\hat{y}_{T+k} = \hat{E}(y_{T+k} | \mathbf{y})$ as the forecast of y_{T+k} . With the R MCMC samples from the posterior distri-

butions, we approximate:

$$\hat{y}_{T+k} = \frac{1}{R} \sum_{r=1}^R \left[X_{T+k} \hat{\boldsymbol{\beta}}^{(r)} + \hat{\sigma}_{\epsilon, T+k}^{(r)} \mathbf{h}^{(r)'} \tilde{\Gamma}_n^{(r)(-1)} (\mathbf{y} - X \hat{\boldsymbol{\beta}}^{(r)}) \right],$$

where super-script (r) refers to the r -th MCMC sample, X_{T+k} is the covariate vector at time $T+k$, $\mathbf{h} = \widehat{Cov}(\mathbf{e}, e_{T+k})$ and $\tilde{\Gamma}_n = \widehat{Cov}(\mathbf{e})$. Note that $\hat{\sigma}_{\epsilon, T+k}^{(r)}$ is obtained by $\hat{\sigma}_{\epsilon, T+k}^{(r)} = \left(e^{\hat{\boldsymbol{\eta}}_{\epsilon, T+k}^{(r)}} \right)^{\frac{1}{2}}$.

3. Empirical Results

3.1. Simulation Study

A simulated data set consists of three components: covariate (\mathbf{X}), error ($\boldsymbol{\epsilon}$), and the outcome (\mathbf{y}). For simplicity, we consider two covariates, X_1 and X_2 , so that $\mathbf{X} = (\mathbf{1}, X_1, X_2)$ has three attributes including intercept $\mathbf{1}$. Here we assume that X_1 follows a bivariate mixture Gaussian distribution such that $X_1 \sim 0.5N(0, 1) + 0.5N(0, 2)$ and that X_2 follows the standard exponential distribution so that $X_2 \sim Exp(1)$. For $\boldsymbol{\epsilon} = \boldsymbol{\sigma}_{\epsilon} \odot \mathbf{e}$, we construct a total of 6 different models based on two cases of volatility $\sigma_{\epsilon, t}$ and three cases of the normalized error e_t .

Specifically, we let $\sigma_{\epsilon, t}$ be (1) constant over time ($\sigma_{\epsilon, t}^2 \equiv 1$); (2) time-varying with a sinusoidal function, $\sigma_{\epsilon, t}^2 = 0.5 \sin\left(\frac{\pi t}{T}\right)$. For e_t , we consider (a) autoregressive model with order 2 [AR(2)]; (b) autoregressive-moving average model with order (1,1) [ARMA(1,1)]; and (c) autoregressive conditional heteroskedasticity model with order 1 [i.e. ARCH(1)]. When $e_t \sim AR(2)$, we let the coefficients be $(a_1, a_2) = (0.5, -0.3)$, so that it satisfies $e_t - 0.5e_{t-1} + 0.3e_{t-2} = \xi z_t$, where $z_t \sim N(0, 1)$, and ξ is a known scaling factor $\sqrt{\frac{1-a_2}{1+a_2} \frac{var(z_t)}{(1-a_2)^2 - a_1^2}} \approx 1.14$ to match its marginal variance one, i.e. $var(e_t) = 1$. The corresponding spectral density is $\lambda(w) = 1/|1 - 0.5e^{i2\pi w} + 0.3e^{i4\pi w}|^2$. When $e_t \sim ARMA(1,1)$, we set the coefficients to $(a_1, b_1) = (0.5, -0.6)$, so that $e_t - 0.5e_{t-1} = -0.6z_{t-1} + z_t$ with $\lambda(w) = 1|1 - 0.6e^{i2\pi w}|^2/|1 - 0.5e^{i2\pi w}|^2$. When $e_t \sim ARCH(1)$, we let $(\alpha_0, \alpha_1) = (0.1, 0.9)$, so that it satisfies $\sigma_{\epsilon, t}^2 = 0.1 + 0.9e_{t-1}^2$ and $z_t \sim N(0, 1)$

with $\lambda(w) \equiv 1$. Additionally, we let $\beta' = (\beta_0, \beta_1, \beta_2) = (1, 2, 3)$. Finally, \mathbf{y} is generated from the model in (2).

We generate each data set with the length of 205, so that the sample size for the *in-sample* estimation is set to $T = 200$. That is, the last five sample points are used for *out-of-sample* forecasting. We generate a total of 100 data sets for each case. For each data, the three MCMC chains with different initial values are run to generate 10,000 different sets of estimates. The results from the last 1,000 iterations are used for posterior analysis. We report the estimation and prediction results for each set of simulated data. Based on the parameter estimates, we compare our proposed model, the Bayesian Time-varying Volatility model (BTV), to some alternative approaches: (1) Bayesian Autorgressive with order 1 (BAR(1)) (2) Bayesian AutoRegressive Conditional Heteroskedasticity with order 1 (BARCH(1)) and (3) Bayesian Fixed Volatility model (BFV) proposed by Dey et al. (2018) [3].

Table 1 shows the bias and root mean squared error (RMSE) for the estimates of β_0, β_1 and β_2 under the six simulation settings. The first two columns specify the structures on the volatility $\sigma_{\epsilon,t}^2$ and on the normalized stationary error e_t . As discussed before, we let the volatility be either fixed ($\sigma_{\epsilon,t}^2 \equiv 1$) or time-varying ($\sigma_{\epsilon,t}^2 = 0.5 \sin(\frac{\pi t}{T})$). Secondly, the underlying structure for the stationary error e_t is either AR(2), ARMA(1,1) or ARCH(1). For all six combinations, we compare the performance of our proposed approach (i.e. BTV) to those of the other three specifications (i.e. BAR(1), BARCH(1) and BFV).

Under the fixed volatility, the BAR(1) produces a relatively large bias and a large RMSE, compared to the other specifications. In contrast, the BARCH(1) generates a small bias but a relatively large RMSE than the others. In general, the BFV provides both a small bias and a small RMSE and BTV are comparable to BFV although bias for β_0, β_1 are larger than those of BFV. In addition, it is interesting to see that the nonparametric modeling of the spectral density (i.e. BFV, BTV) provides better outcomes than its parametric counterpart

DGP		Fitted Model	$\hat{\beta}_0$		$\hat{\beta}_1$		$\hat{\beta}_2$	
$\sigma_{\epsilon,t}^2$	e_t		Bias	RMSE	Bias	RMSE	Bias	RMSE
Fixed	AR(2)	BAR(1)	-.0656	.1219	-.0882	.2235	-.1864	.2168
		BARCH(1)	.0306	.1170	.0078	.0385	-.0141	.0826
		BFV	-.0109	.0925	.0022	.0277	-.0066	.0550
		BTV	-.0248	.0962	.0193	.0340	-.0047	.0582
	ARMA (1,1)	BAR(1)	-.0225	.0998	-.0693	.0768	-.1555	.1697
		BARCH(1)	.0152	.1243	.0076	.0373	-.0055	.0851
		BFV	-.0138	.0831	.0019	.0267	-.0037	.0631
		BTV	-.0284	.0886	.0192	.0333	-.0031	.0656
	ARCH(1)	BAR(1)	-.0122	.1168	-.0558	.0656	-.1380	.1506
		BARCH(1)	.0152	.1243	.0076	.0373	-.0055	.0851
		BFV	.0009	.0868	.0052	.0261	-.0130	.0531
		BTV	-.0190	.0614	.0217	.0293	-.0095	.0415
Time- Varying	AR(2)	BAR(1)	-.0684	.1256	-.0661	.0796	-.1832	.1976
		BARCH(1)	.0032	.1283	.0032	.0394	.0042	.0905
		BFV	.0080	.1042	.0003	.0291	-.0092	.0607
		BTV	-.0277	.0942	.0172	.0312	-.0059	.0581
	ARMA (1,1)	BAR(1)	-.0146	.1004	-.0724	.0794	-.1616	.1752
		BARCH(1)	-.0019	.1261	-.0001	.0403	.0035	.0878
		BFV	-.0112	.0918	-.0009	.0281	-.0080	.0677
		BTV	-.0312	.0899	.0174	.0308	-.0029	.0641
	ARCH(1)	BAR(1)	-.0117	.1155	-.0628	.0735	-.1582	.1752
		BARCH(1)	-.0019	.1261	-.0001	.0403	.0035	.0878
		BFV	.0041	.0954	.0057	.0303	-.0151	.0615
		BTV	-.0161	.0644	.0208	.0303	-.0135	.0461

Table 1: Bias and Root Mean Squared Error (RMSE) based on the estimates from 100 repeated sets of data for each simulation setting.

(BARCH(1)) even when the parametric specification matches the true data generating process (ARCH(1)). We have similar findings in the case of time-varying volatility as well.

For those with time-varying volatility, BAR(1) and BARCH(1) provide relatively large RMSEs. However, BARCH(1) shows a small bias. BFV provides an overall small bias and RMSE. On the other hand, BTV shows its relative advantage in terms of RMSE for the majority of the simulation settings.

Figure 1 shows estimated log-scaled volatility $\sigma_{\epsilon,t}^2$ and log-scaled spectral

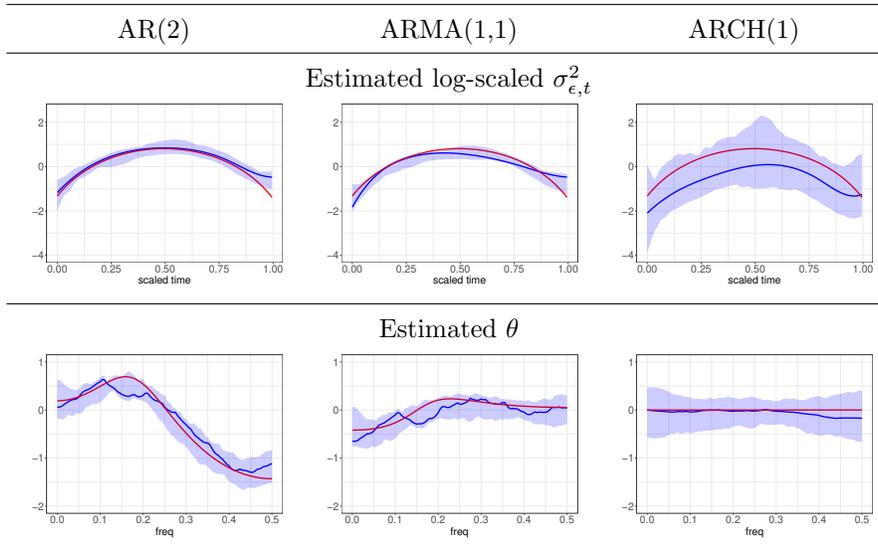


Figure 1: Estimated log-scaled volatility $\sigma_{\epsilon,t}^2$ (first row) and log-spectral density θ (second row) from BTV with 95% credible intervals. Each column corresponds to a different data generating process. The x-axis in the first row is the time scaled to $(0, 1)$ and the x-axis in the second row is the frequency scaled to $(0, 0.5)$. The red lines indicate the true curves, blue lines indicate the estimated curves. The shaded regions indicate the 95% credible intervals.

density $\theta(w)$ from the BTV when the data are simulated from the time-varying volatility ones. The estimated log-scaled $\sigma_{\epsilon,t}^2$ is somewhat biased, especially near the boundary points. However, it does capture the global pattern in the time-varying characteristics of $\sigma_{\epsilon,t}^2$ in all three cases. Regarding the spectral density estimation in the second row of Figure 1, the density curve estimates are reasonably close to the true ones in all cases. In addition, the credible interval of the log-scaled spectral density covers the true curve of the log-scaled spectral density in all three cases.

In Figure 2, we compare the mean squared prediction errors (MSPE) of BTV to those from the other specifications at time $T + k$ for $k = 1, \dots, 5$. In each figure, the black line refers to the MSPE of BTV, the red line refers to BAR(1), the purple line refers to BARCH(1) and the blue line refers to BFV. When the true models have a time-invariant σ_{ϵ} (first column), MSPE of BTV is similar or better to those of the BFV and smaller than the other specifications. When

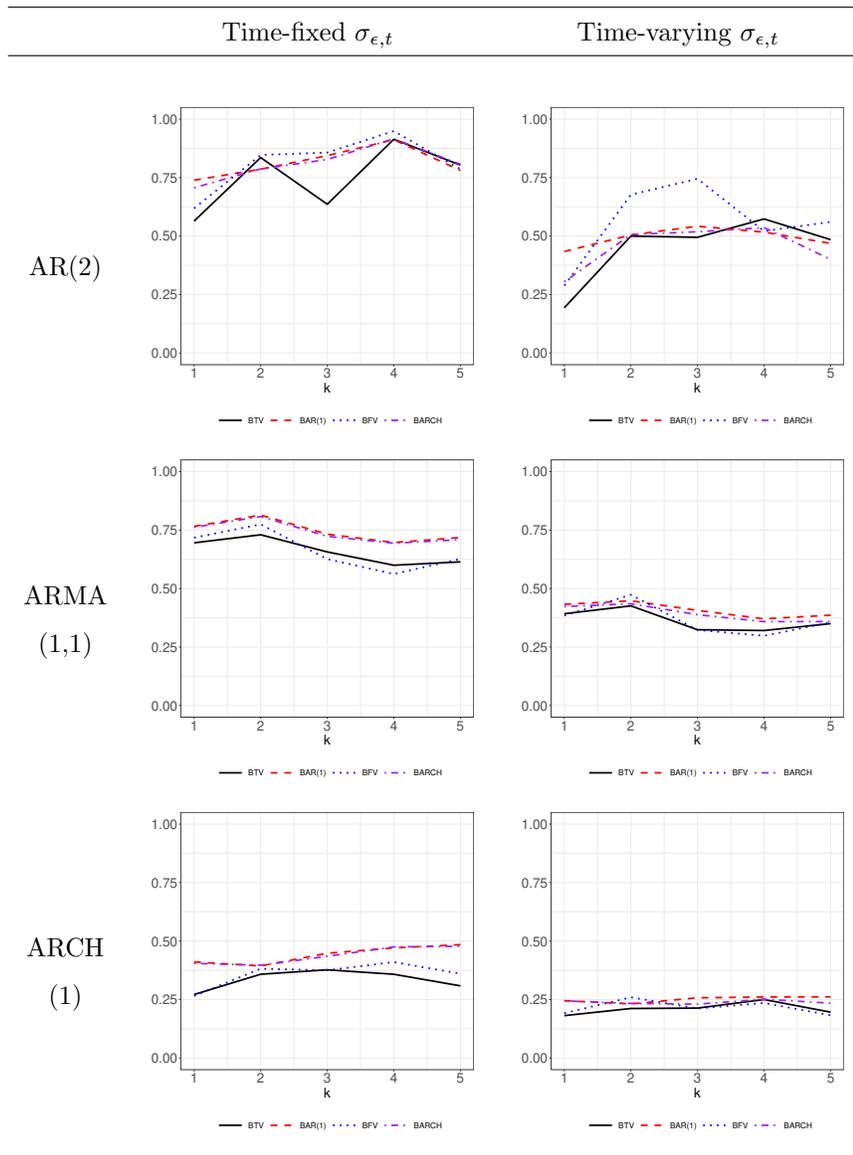


Figure 2: The estimated mean squared prediction errors for y_{n+k} , $k = 1, 2, 3, 4, 5$. The model is fitted up to time $T=200$ and then forecast y_{T+k} , $k = 1, 2, 3, 4, 5$. Each row corresponds to a different data generating process. We compare the proposed model, BTV (black), to the other models: BAR(1) (red); BARCH(1) (purple); and BFV (blue).

true models have a time-varying σ_ϵ , BTV generally shows a lower MSPE than the rest of specifications for most of cases. Interestingly, the BFV and the BTV behave in a rather similar fashion, although the BFV has a higher MSPE than the BTV in the majority of the cases.

3.2. Real Data Analysis

As an application of the proposed methodology, we consider the monetary model introduced by Mark (1995) [8]:

$$s_{t+k} - s_t = \alpha + \beta(f_t - s_t) + u_{t+k}, \quad (7)$$

where s_t is log of spot exchange rate at time t , s_{t+k} is log of spot exchange rate at time $t+k$. One can obtain the k -th-step-ahead forecast of s_t from the model. Here f_t is the *fundamental* defined by:

$$f_t := m_t - m_t^* - \lambda(y_t - y_t^*)$$

where m_t is log of the *U.S.* money supply and y_t is log of the *U.S.* output, while m_t^* and y_t^* are their foreign equivalents. For simplicity, we let $\lambda = 1$ here.

This study employs spot exchange rate data and their corresponding macroeconomic data, involving seven currencies of Canadian Dollar (*CAD*), Swiss Franc (*CHF*), Danish Krone (*DKK*), British Pound (*GBP*), Japanese Yen (*JPY*) and New Zealand Dollar (*NZD*) vis a vis U S Dollar (*USD*), which is the numeraire currency here. We use end-of-month observations from December 1988 through December 2016, including the 2008-2009 financial crisis period.

We have a total of $n = 336$ monthly observations for each currency. Note that *DKK* has missing values in m_t from January 2001 to March 2001. We thus impute these values by the sample average of m_t in their neighborhood. We start with the first $T = 320$ observations from December 1998 until August 2015 to fit the model. That is, the estimation window length is 320. The remaining observations are used to evaluate the model's *out-of-sample* forecasting performance. We try four different forecast horizons, i.e. $k = 1$ (monthly), $k = 3$ (quarterly), $k = 6$ (half yearly), and $k = 12$ (yearly). Given the starting point l , we conduct k -step-ahead forecasting as follows. That is, we forecast y_{l+T+k} based on the model fitted using the observations from $(l+1)$ -th to $(l+T)$ -th time points. We move l from 0 to 5 for the four forecast horizons in each currency.

We compare the forecasting performance of our proposed method to those of a pack of competing models, such as the random walk without drift (RW), Bayesian Autoregressive model (BAR(1)), Bayesian Autoregressive Conditional Heteroskedasticity model (BARCH(1)) and the Bayesian fixed volatility model using the spectral density (BFV). Throughout the Bayesian *MCMC* implementation, we run 3 chains with 10,000 iterations per chain. The last 1,000 iterations of the 3 chains are saved for estimation and inference.

Table 2 illustrates the results based on forecast error measures for the logarithm of spot return. The numbers represent the root mean squared prediction errors (RMSPE) for the forecasts, while those inside the parenthesis are the proportions that the corresponding model outperforms the random walk process in terms of the forecast error. We denote this number by RW_r. That is, the closer to one the RW_r is, the more frequently the corresponding model outperforms the random walk without drift.

Empirical results in Table 2 indicate that RMSPE tends to increase naturally as the forecast horizon k grows. Interestingly, BTV shows relatively low RMSPE values among the competing statistical methods for the majority of currencies, namely *CHF*, *DKK*, *GBP*, *JPY*, and *NZD*. Table 2 also shows that the RMSPE for the BTV is mostly either the lowest or the second lowest among all the competing ones. Except for the case of one-step-ahead forecasts, RW_r of the BTV is mostly more than or equal to 50.0% and remains relatively high in comparison to the other competing ones.

Table 3 shows the *ratios* of each model's RMSPE to the lowest RMSPE at the corresponding horizon. Hence the best performance at each horizon is marked by *one in bold*. Naturally, the further away from one the ratio is, the worse the forecasting performance of the corresponding model is. As we can see from the table, the ratios for the BTV are generally close to one and frequently take on ones in bold, except for *CAD*. This means that the BTV's forecasting performance is relatively better than the other approaches/models. In particular, the BTV's ratios are typically much closer to one than those for the random walk without drift, which illustrates the relative advantage of BTV

over the widely used benchmark model. By presenting the ratios instead of the original RMSPE's, one can more easily figure out the relative advantage of the BTV approach over the rest in forecasting the future spot rates.

Our results indicate that the nonparametric spectral-density-based method in the current study can generate more precise forecasts for foreign exchange rates because both BFV and BTV provide lower RMSPEs and higher RWrs than the alternative methods in the majority of cases. The relative advantage can be attributed to that, in contrast to the BAR(1) and BARCH(1) relying on their parametric assumptions, the BFV and BTV do not depend on any parametric covariance structure. Hence the latter could be relatively free from potential model mis-specification. Given that the BARCH(1) generally outperforms the BAR(1), modeling a time-varying volatility appears to further improve the corresponding model's forecasting performance. It should be noted, however, that it is rather hard to conclude which method is truly superior to the rest because the prediction gain for BARCH(1) over BAR(1) could be potentially due to the model's over-fitting.

4. Discussion

In this paper, we propose Bayesian estimation for the spectral density of an error auto-covariance in a frequency domain, without assuming any structure on the model error. The estimates are applied for the prediction of the model's dependent variable. The main motivation is to avoid any arbitrary parametric structure on the error so that estimation of the corresponding model and the forecasts of the dependent variable based on the model estimates become more precise. The approach enables us to avoid the notorious issue of nonparametric bandwidth selection by conducting the estimation in a frequency domain. The relevant Bayesian posterior distributions are derived and the estimation and prediction strategies are explained in detail. To check on its finite-sample performance, a simulation study is carefully designed and conducted. To illustrate how it works in practice, we apply the methodology to exchange rate forecast-

ing and show that it generally outperforms the well-known benchmark in the literature, the random walk without drift.

The current project leaves a couple of interesting issues that deserve future research on: Firstly, the current framework can be extended to accommodate *time-varying coefficients*. To focus on the role of time-varying volatility, our study assumes that the model coefficients are invariant over time. One can remove the restriction to make our findings more robust. Secondly, we can apply the proposed methodology to *density forecasting*. By forecasting the entire distribution of the dependent variable, one can forecast not only the mean but also any percentile of the underlying distribution. This would result in more comprehensive forecasts, such as density and/or interval forecasts, for the corresponding variable. Further insight can be gained by extending the current work in these and other directions.

Acknowledgement

This work is supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (No.2019R1A2C1002213, No.2020R1A4A1018207).

References

- [1] C. Engel, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica* 50 (1982) 987–1007.
- [2] K. Kim, T. Kim, Capital asset pricing model: a time-varying volatility approach, *Journal of Empirical Finance* 37 (2016) 268–281.
- [3] T. Dey, K. H. Kim, C. Y. Lim, Bayesian time series regression with non-parametric modeling of autocorrelation, *Computational Statistics* (2018) 1–17.
- [4] K. H. Kim, Density forecasting through disaggregation, *International Journal of Forecasting* 27 (2011) 394–412.
- [5] B. Rossi, Exchange rate predictability, *Journal of Economic Literature* 51 (2013) 1063–1119.
- [6] R. A. Meese, K. Rogoff, Empirical exchange rate models of the seventies: Do they fit out of sample?, *Journal of International Economics* 14 (1983) 3–24.
- [7] C. Engel, J. D. Hamilton, Long swings in the dollar: Are they in the data and do markets know it?, *American Economic Review* 80 (1990) 689–713.
- [8] N. C. Mark, Exchange rates and fundamentals: Evidence on long-horizon predictability, *American Economic Review* 85 (1995) 201–218.
- [9] L. Kilian, Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions?, *Journal of Applied Econometrics* 14 (1999) 491–510.
- [10] L. Kilian, M. P. Taylor, Why is it so difficult to beat the random walk forecast of exchange rates?, *Journal of International Economics* 60 (2003) 85–107.

- [11] T. Molodtsova, D. H. Papell, Out-of-sample exchange rate predictability with Taylor rule fundamentals, *Journal of International Economics* 77 (2009) 167–180.
- [12] O. Ince, T. Molodtsova, D. H. Papell, Taylor rule deviations and out-of-sample exchange rate predictability, *Journal of International Money and Finance* 69 (2016) 22–44.
- [13] C. Engel, N. C. Mark, K. D. West, Exchange rate models are not as bad as you think, *NBER Macroeconomics Annual 2007* (2008) 381–441.
- [14] J. H. Wright, Bayesian model averaging and exchange rate forecasts, *Journal of Econometrics* 146 (2008) 329–341.
- [15] M. B. Priestley, *Spectral analysis and time series*, Academic press, 1981.
- [16] C. K. Carter, R. Kohn, Semiparametric Bayesian inference for time series with mixed spectra, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59 (1) (1997) 255–268.

Appendix: Conditional posterior distributions

Conditional posterior distribution of parameters for Gibbs samplers are described below.

Posterior of β .

With a Gaussian prior $\beta \sim \mathbf{N}(\mu_\beta \mathbf{1}, \sigma_\beta^2 \mathbf{I})$, we have

$$\begin{aligned}\beta \mid \dots &\sim \mathbf{N}(\mu_*, \mathbf{T}_*), \\ \mathbf{T}_* &= \mathbf{X}^t \tilde{\mathbf{\Gamma}}^{-1} \mathbf{X} + \sigma_\beta^{-2} \mathbf{I}_p, \\ \mu_* &= \mathbf{T}_*^{-1} \left(\mathbf{X}^t \tilde{\mathbf{\Gamma}}^{-1} \mathbf{Y} + \sigma_\beta^{-2} \mu_\beta \mathbf{I} \right),\end{aligned}$$

where $\tilde{\mathbf{\Gamma}}$ is the covariance matrix of the data constructed by $\tilde{\mathbf{\Gamma}} = F_n \Lambda_n F_n^*$, F_n , Λ_n diagonal matrix with u -th entry $\lambda(w_{u-1})$, F_n $n \times n$ FFT-matrix with (u,v) -th entry $q_{u,v} = \frac{1}{\sqrt{(n)}} e^{i(u-1)(v-1)2\pi/n}$, and F_n^* its conjugate. This operation is performed effectively by the use of Fast Fourier Transform algorithm.

Posterior of τ_ϵ .

Let $\mathbf{r} = \mathbf{Y} - \mathbf{X}\beta$. With a Gamma prior for τ_ϵ we have

$$\tau_\epsilon \mid \dots \sim G(a_*, b_*),$$

where $a_* = a + \frac{n_1 n_2}{2}$, $b_* = b + \frac{1}{2} \mathbf{r}^t \tilde{\mathbf{\Gamma}}^{-1} \mathbf{r}$.

Posterior of θ .

Let the normalized residual $\mathbf{r}_* = \tau_\epsilon^{1/2} (\mathbf{Y} - \mathbf{X}\beta)$. Given \mathbf{r}_* , we compute φ by (5). Let $\mathbf{\Upsilon}$ be a covariance kernel matrix of the process θ . With a Gaussian process prior of θ , we obtain

$$\begin{aligned}\theta \mid \dots &\sim \mathbf{N}(\nu_*, \mathbf{\Upsilon}_*), \\ \mathbf{\Upsilon}_* &= (\mathbf{\Upsilon}^{-1} + \mathbf{V}_\psi^{-1})^{-1}, \\ \nu_* &= \mathbf{\Upsilon}_* \mathbf{V}_\psi^{-1} (\varphi - \kappa_\psi - \nu) + \nu\end{aligned}$$

where $\mathbf{V}_\psi = \text{diag}\{v_{\psi_0}^2, \dots, v_{\psi_{n^h}}^2\}$, and $\kappa_\psi = (\kappa_{\psi_0}, \dots, \kappa_{\psi_{n^h}})'$ for the assigned n^h Fourier frequencies.

Posterior of ψ .

Given the prior $P(\psi = l) = p_l$, for $l = 1, \dots, 5$,

$$P(\psi_s = l \mid \dots) = p_l \phi_{\nu_l}(\varphi_s - \theta_s - \kappa_l),$$

for $s = 1, 2, \dots, n^h$, the assigned n^h locations of Fourier frequencies.

Posterior of τ_θ .

$$\tau_\theta \mid \dots \sim G(c_*, d_*),$$

where $c_* = c + \frac{n^h}{2}$ and $d_* = d + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta} - \boldsymbol{\nu})$.

Posterior of ρ_θ .

We use a “grid-search” method to sample ρ_θ . That is, for $j = 1, 2$, we sample ρ_θ among its candidates $\{\rho_\theta^{(1)}, \rho_\theta^{(2)}, \dots, \rho_\theta^{(M)}\}$ with probability weights $\pi(\rho_\theta^{(l)}) / \sum_{m=1}^M \pi(\rho_\theta^{(m)})$, satisfying

$$\pi(\rho_\theta^{(l)}) \propto \exp\left(-\frac{1}{2} \log \left| \boldsymbol{\Upsilon}_{(\rho_\theta^{(l)})}^{-1} \right| - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}_{(\rho_\theta^{(l)})}^{-1} (\boldsymbol{\theta} - \boldsymbol{\nu})\right),$$

Posterior of δ .

Note that $\eta_{\epsilon,t} \equiv \log(\sigma_{\epsilon,t}^2)$ is represented by cubic B -spline functions such that $\eta_{\epsilon,t} = \sum_{b=1}^d \delta_b \phi_b(t)$. In this model setting, let Φ be a $n \times d$ matrix of which b -th column is $(\phi_b(1), \dots, \phi_b(n))$. We set the i.i.d. prior on each of the coefficients $\delta_b \sim N(0, \sigma_\delta^2 I)$, i.e. for $b = 1, \dots, d$. For $\boldsymbol{\delta} = (\delta_1, \dots, \delta_d)'$, we obtain

$$\boldsymbol{\delta} \mid \dots \propto \exp\left(-\frac{1}{2} \log \left| \Lambda_{\exp \eta}^{-1} \right| - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t (\Phi^t \Lambda_{\exp \eta} \Phi)^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2\sigma_\delta^2} \boldsymbol{\delta}^t \Phi^t \Phi \boldsymbol{\delta}\right)$$

where $\Lambda_{\exp \eta}$ is a diagonal matrix with u -th entry $\exp \eta(u)$. The posterior sampling is made from the Metropolis-Hastings algorithm with proposal multivariate normal of which mean is a zero vector, and standard deviation is a $d \times d$ identity matrix with all diagonal entry 1s.

Currency	k	RW	OLS	BAR(1)	BARCH(1)	BFV	BTV
CAD	1	0.025	<u>0.023</u>	0.025(0.40)	0.025(0.54)	0.091(0.8)	0.088(0.00)
	3	0.078	0.049	0.058(0.73)	<u>0.036</u> (0.69)	0.063(0.9)	0.061(1.00)
	6	0.106	0.060	0.059(0.20)	0.047(0.62)	0.038 (1.00)	0.042 (1.00)
	12	0.174	0.052	<u>0.036</u> (0.27)	0.058(0.62)	0.064(1.00)	0.131(1.00)
CHF	1	0.026	0.020	0.047(0.33)	<u>0.017</u> (0.38)	0.035(0.60)	0.019 (0.25)
	3	0.039	0.043	0.065(0.67)	0.025 (0.54)	0.037(1.00)	<u>0.017</u> (0.50)
	6	0.062	0.064	0.040(0.27)	0.025 (0.31)	0.044(0.40)	<u>0.019</u> (0.25)
	12	0.035	0.107	0.034(0.53)	<u>0.018</u> (0.77)	0.052(0.60)	0.023 (0.50)
DKK	1	0.024	<u>0.018</u>	0.035(0.27)	<u>0.018</u> (0.54)	0.034(0.70)	0.020 (1.00)
	3	0.047	0.031	0.043(0.60)	0.031 (0.77)	0.041(1.00)	0.020 (1.00)
	6	0.062	0.036	0.045(0.33)	0.035 (0.54)	0.043(0.50)	<u>0.018</u> (0.25)
	12	0.129	0.030	0.047(0.47)	<u>0.019</u> (0.69)	0.054(0.50)	0.026 (0.50)
GBP	1	0.038	0.030	0.051(0.33)	<u>0.029</u> (0.23)	0.053(0.10)	0.030 (0.00)
	3	0.058	0.067	0.078(0.60)	0.059(0.31)	0.062(0.20)	<u>0.032</u> (1.00)
	6	0.066	0.120	0.138(0.27)	0.102(0.31)	0.067(0.30)	<u>0.034</u> (1.00)
	12	0.145	0.209	0.207(0.53)	0.195(0.69)	0.075 (0.40)	<u>0.029</u> (1.00)
JPY	1	0.022	0.029	<u>0.021</u> (0.53)	0.03(0.46)	0.049(0.50)	0.025(1.00)
	3	0.073	0.066	0.068(0.47)	0.066(0.46)	0.064 (0.40)	<u>0.032</u> (1.00)
	6	0.114	0.100	0.103(0.47)	0.100(0.46)	0.071 (0.50)	<u>0.034</u> (0.75)
	12	0.209	0.112	0.130(0.47)	0.129(0.62)	0.070 (0.50)	<u>0.030</u> (0.75)
NZD	1	0.043	<u>0.023</u>	0.041(0.80)	0.024 (0.62)	0.036(0.50)	0.036(1.00)
	3	0.064	0.034	0.051(0.67)	0.034 (0.62)	0.035(0.60)	<u>0.028</u> (1.00)
	6	0.111	0.038	0.065(0.53)	0.052(0.69)	0.030 (1.00)	0.026 (0.50)
	12	0.278	0.081	0.099(0.53)	0.085(0.77)	<u>0.036</u> (0.50)	0.042 (1.00)

Table 2: Root mean square prediction error (RMSPE) of log spot exchange rate (s_{t+k}) for $k = 1, 3, 6, 12$ -step-ahead forecasts from the proposed model (BTV) and other models (RW, OLS, BAR(1), BARCH(1), BFV) assuming the exchange rate model by Mark (1995). The ones in bold indicate the lowest RMSE for each horizon. The rest are the ratios of the model's RMSE to the lowest RMSE at the corresponding horizon.

Currency	k	RW	OLS	BAR(1)	BARCH(1)	BFV	BTV
CAD	1	1.087	1.000	1.087	1.087	3.957	3.826
	3	2.167	1.361	1.611	1.000	1.750	1.694
	6	2.789	1.579	1.553	1.237	1.000	1.105
	12	4.833	1.444	1.000	1.611	1.778	3.639
CHF	1	1.529	1.176	2.765	1.000	2.059	1.118
	3	2.294	2.529	3.824	1.471	2.176	1.000
	6	3.263	3.368	2.105	1.316	2.316	1.000
	12	1.944	5.944	1.889	1.000	2.889	1.278
DKK	1	1.333	1.000	1.944	1.000	1.889	1.111
	3	2.350	1.550	2.150	1.550	2.050	1.000
	6	3.444	2.000	2.500	1.944	2.389	1.000
	12	6.789	1.579	2.474	1.000	2.842	1.368
GBP	1	1.310	1.034	1.759	1.000	1.828	1.034
	3	1.813	2.094	2.438	1.844	1.938	1.000
	6	1.941	3.529	4.059	3.000	1.971	1.000
	12	5.000	7.207	7.138	6.724	2.586	1.000
JPY	1	1.048	1.381	1.000	1.429	2.333	1.190
	3	2.281	2.063	2.125	2.063	2.000	1.000
	6	3.353	2.941	3.029	2.941	2.088	1.000
	12	6.967	3.733	4.333	4.300	2.333	1.000
NZD	1	1.870	1.000	1.783	1.043	1.565	1.565
	3	2.286	1.214	1.821	1.214	1.250	1.000
	6	4.269	1.462	2.500	2.000	1.154	1.000
	12	7.722	2.250	2.750	2.361	1.000	1.167

Table 3: Root mean square prediction error (RMSPE) of log spot exchange rate (s_{t+k}) for $k = 1, 3, 6, 12$ -step-ahead forecasts from the proposed model (BTV) and other models (RW, OLS, BAR(1), BARCH(1), BFV) assuming the exchange rate model by Mark (1995). The ones in bold indicate the lowest RMSPE for each horizon. The rest are the ratios of the model's RMSPE to the lowest RMSPE at the corresponding horizon.

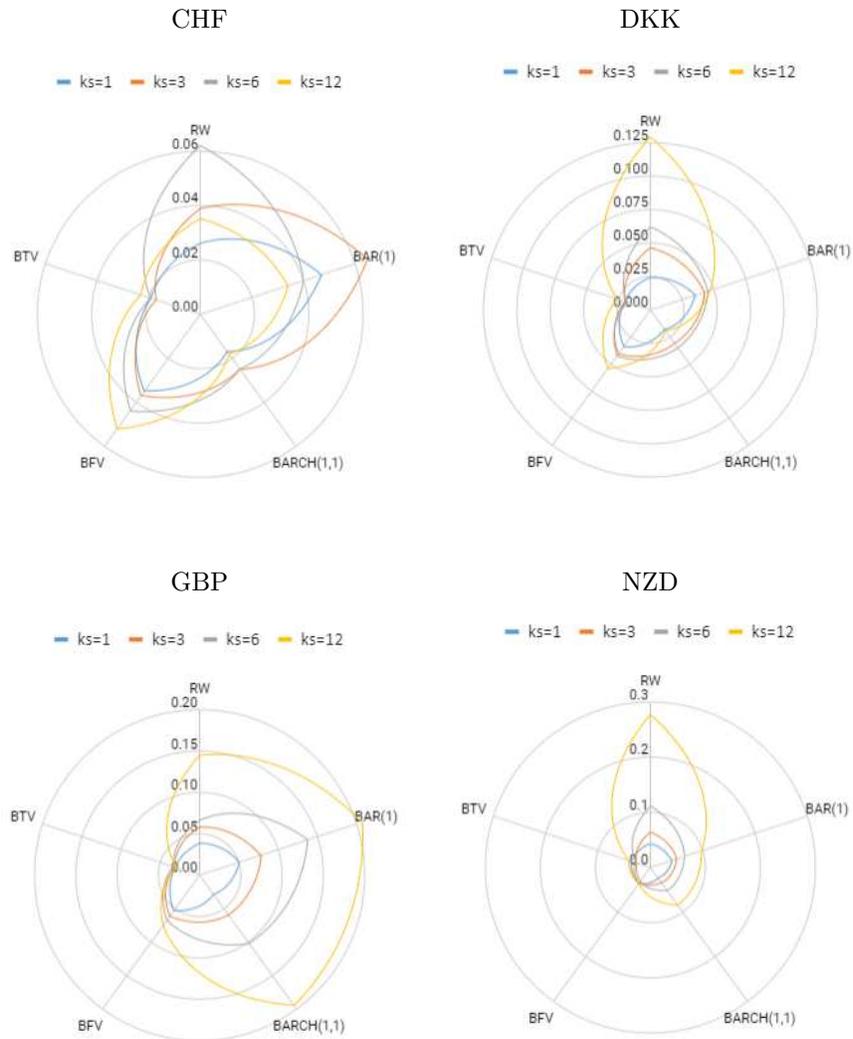


Figure 3: Radar charts of the root mean squared prediction errors from forecasting the log spot exchange rate for the five competing models based on four selected currencies: Switzerland Franc (CHF, top left); Danish Krone (DKK, top right); British Pound (GBP, bottom left); and New Zealand Dollar (NZD, bottom right). The different colors represent 1-month forecasting (blue), 3-month forecasting (orange), 6-month forecasting (gray), and 12-month forecasting (yellow), respectively.