Nonstationary Bayesian Time Series Models with Time-Varying Parameters and Regime-Switching

---

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Yuelei Sui

Advisor: Scott H. Holan

APRIL 2021

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

Nonstationary Bayesian Time Series Models with Time-Varying Parameters and Regime-Switching

presented by Yuelei Sui,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Scott H. Holan

Dr. Christopher K. Wikle

Dr. Athanasios C. Micheas

Dr. Lori Thombs

Dr. E. John Sadler

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and thanks to my advisor, Dr. Scott H. Holan, for introducing me to this challenging yet interesting topic. Without his continuous encouragement and inspiration, throughout my dissertation this work will never be possible. I also thank him for his kind help and wise guidance throughout my PhD study, which make my research life here much easier and more enjoyable.

I am grateful to Dr. Wen-Hsi Yang for his help in conquering hard background materials and his inspiring suggestions in breakthrough obstacle throughout in the development of this work.

I extend my thanks to my committee members: Dr. Christopher K. Wikle, Dr. Athanasios C. Micheas, Dr. Lori Thombs and Dr. E. John Sadler for their insightful comments and suggestions on this work

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Nonstationary Bayesian Time Series Models with Time-Varying Parameters and Regime-Switching

Yuelei Sui

Scott H. Holan Dissertation Supervisor

ABSTRACT

Nonstationary time series data exist in various scientific disciplines, including environmental science, biology, signal processing, econometrics, among others. Many Bayesian models have been developed to handle nonstationary time series. Some of the models characterize nonstationarity through the mean, e.g., regime-switching time series models and time-varying coefficients models, whereas other models characterize nonstationarity through modeling the variance/covariance. Models of nonstationary behavior can be viewed both in the time domain (e.g., stochastic volatility) or in the time-frequency domain.

This dissertation proposes a multi-regime smooth transition stochastic volatility model based on ordered categorical variables and is illustrated using additional information in the form of covariates (e.g., trading volume). This model can handle the non-Gaussian behavior often found in return data, address the asymmetric effects of financial returns, and provide regime-specific inference. The time-varying autoregressive (TV-VAR) model is a well-established model for multivariate nonstationary time series. Nevertheless, in most cases, the large number of parameters presented by the model results in a high computational burden, ultimately limiting its usage. This dissertation proposes a computationally efficient multivariate Bayesian Circular

Lattice Filter to extend the usage of the TV-VAR model to a broader class of high-dimensional problems. Finally, modeling and forecasting nonstationary count time series presents many challenges. To address these challenges, this dissertation introduces a computationally efficient Poisson model with a latent Gaussian time-varying autoregressive process. The model is shown to provide effective multi-step ahead forecasts for nonstationary count time series, e.g., COVID-19 count data. The utility of each proposed method is illustrated through simulation and through case-studies.

Chapter 1

Introduction

Nonstationary time series data exist in various scientific disciplines, including environmental science, biology, signal processing, econometrics, among others. Many Bayesian models have been developed to handle nonstationary time series. Some of the models characterize nonstationarity through the mean, e.g., regime-switching time series models, and some through modeling the variance/covariance. Modeling nonstationary variances can occur both in the time domain (e.g., stochastic volatility) or in the time-frequency domain (Jacquier et al. (1994); Kim et al. (1998); Primiceri (2005); Zhao and Prado (2020) among others).

1.1 Stochastic Volatility Models with Regime Switching

Stochastic volatility models can be used to characterize the time-varying variance typically found in many financial time series data. For example, Bayesian inference

for stochastic volatility models have been proposed by Jacquier et al. (1994) and Kim et al. (1998), with many extensions having been proposed to date. One class of SV model extensions are regime-switching SV models, including threshold stochastic volatility models (So et al., 2002; So and Choi, 2009), Markov-switching stochastic volatility models (So et al., 1998; Kalimipalli and Susmel, 2004; Carvalho and Lopes, 2007; Lopes and Carvalho, 2007; Abanto-Valle et al., 2010; Virbickaitė et al., 2019b), and smooth transition stochastic volatility models (Lopes and Salazar, 2006; Livingston Jr and Nur, 2017). Since parametric SV models may not fully capture the non-Gaussian behaviors of typical of return data (Gallant et al., 1997; Mahieu and Schotman, 1998; Liesenfeld and Jung, 2000; Durham, 2006), another class of SV models have been proposed; that is, semiparametric SV models (Jensen and Maheu, 2010, 2014; Yu, 2012; Virbickaitė et al., 2019a).

In Chapter 2, we propose an SV model with volatility following a regime-switching autoregressive process associated with some auxiliary covariates. The proposed approach can 1) account for the asymmetry of financial returns and 2) utilize the trading volume information, which is not typically used in SV models. We use a semiparametric prior, proposed by Jensen and Maheu (2010, 2014), in our model to increase the flexibility and to account for the potential non-Gaussian behavior of the return innovations.

1.2 Bayesian Circular Lattice Filters for Multiple Time Series

Parametric models of nonstationary multivariate time series play an increasingly important role as high-dimensional data are increasingly available for analysis. Some of the parametric models that apply to economic and financial data are cast in the time domain (Primiceri, 2005; Del Negro and Primiceri, 2015; Nakajima et al., 2011; Nakajima and West, 2013). In contrast, some work in the frequency domain reveals features of the spectrum of each time series as well as the coherence between them (e.g., Zhao and Prado (2020)). In this dissertation, we are most interested in the time-varying vector autoregressive (TV-VAR) model. However, when it comes to high-dimensional data, the computation cost becomes an obstacle in terms of estimation.

In Chapter 3, we develop a computationally efficient Bayesian methodology for analyzing multivariate nonstationary time series. The approach offers a fast way to fit time-varying vector autoregressive (TV-VAR) models based on the "one channel at-a-time" paradigm (Pagano, 1978). Our approach utilizes a *Bayesian Circular Lattice Filter* and, thus, accelerates parameter estimation. Compared with existing frequentist and Bayesian approaches, our proposed method can simultaneously estimate the multivariate time-varying coefficients and the time-varying innovation covariance. We apply the proposed methodology in both the time and frequency domain. We demonstrate the effectiveness of our approach through simulation and applications to U.S. economic data and Northern California wind data.

1.3 Time-varying Poisson Autoregressive Models

Different types of count time series models have been proposed and applied to pandemic incidences, insurance claims, integer financial data, and many others. These models include the classes of observation-driven and parameter-driven models. The observation-driven models include the INGARCH model (Ferland et al., 2006; Fokianos et al., 2009), integer-valued autoregressive model, also called Poisson autoregressive model (Al-Osh and Alzaid, 1987), generalized linear ARMA (GLARMA) model (Zeger, 1988; Dunsmuir, 2016) and Poisson AR model (Brandt and Williams, 2001), among others (see Davis et al. (2016) for a comprehensive overview). The parameter-driven (process-driven) models include the Poisson state space model (Smith and Miller, 1986), Poisson exponentially weighted moving average (PEWMA) model (Brandt et al., 2000), and dynamic count mixture model (Berry and West, 2019), among others. Some of this research proceeds under a Bayesian framework (Berry and West, 2019; Bradley et al., 2020).

For nonstationary count time series with changing trends, e.g., daily new COVID cases data, many traditional methods (Ferland et al., 2006; Fokianos et al., 2009; Brandt et al., 2000; Brandt and Williams, 2001) may not capture local trends and also may not provide a good multi-step ahead forecast. The motivation of our study is to propose an efficient method to capture the time-varying pattern of the mean of such nonstationary count time series and, therefore, make better forecasts than traditional methods. To capture the evolutionary properties, a parameter-driven model with a time-varying coefficient latent process provides one good choice. Moreover, a latent process with appropriate innovations can also address issues surrounding over-dispersion, which is common in count time series modeling.

Parameter estimation is an important part of Poisson time series modeling because the intensity parameter of the Poisson distribution often does not yield a closed-form estimator. One exception is Bradley et al. (2018), which proposed a conjugate multivariate log-gamma prior for the intensity and provided efficient multivariate spatio-temporal models for high-dimensional count-valued data. For estimation of the intensity following an autoregressive process, the particle filter (PF) is often used, e.g., Aktekin et al. (2018). However, the sampling difficulty of the PF increases exponentially when the dimension of the data increases and the PF cannot handle time-varying coefficients. Random-walk Metropolis also presents estimation challenges for high-dimensional data. For more efficient estimation, we use the No-U-Turn Sampler (NUTS) instead (Hoffman and Gelman, 2014), which speeds up the mixing of the sample chains.

We propose a time-varying Poisson autoregressive (TV-Pois-AR) model for nonstationary count time series. We use a time-varying autoregressive (TV-AR) latent process to model the nonstationary intensity of the Poisson process. This flexible model can capture the latent dynamics of the time series and, therefore, make superior forecasts. The estimation of such a TV-AR process is greatly sped up by using the Bayesian Lattice Filter (BLF, Yang et al. (2016)). Moreover, in our model and many other models, the intensity parameter has no closed-form full conditional distribution. To sample the intensity parameter, the No-U-Turn Sampler (NUTS) is used, instead of a random walk Metropolis algorithm, for faster mixing of the sample chains. Benefiting from the joint use of the Bayesian lattice filter and No-U-Turn Sampler, estimation of the TV-Pois-AR model is efficient and fast, especially for higher model orders and/or longer length time series.

In Chapter 4, we propose a Poisson autoregressive (TV-Pois-AR) model for non-stationary count time series. This model has time-varying parameters and enables us to forecast changing trends within the count time series. Moreover, this model benefits from the BLF and provides efficient Bayesian inference.

Chapter 2

Multi-regime Smooth Transition Stochastic Volatility Models

2.1 Introduction

Regime-switching has been observed in the volatility of financial returns. For example, Hamilton and Susmel (1994) proposed a Markov switching autoregressive conditional heteroskedasticity (SWARCH) model with regime-switching for the volatility level, whereas So et al. (2002) and Carvalho and Lopes (2007) developed a methodology for a threshold stochastic volatility model and Markov switching stochastic volatility (SV) models, respectively. The asymmetric effect of financial returns (i.e., the variance responds asymmetrically to the past returns) has been previously studied in many financial time series analyses. In particular, Black (1976), Christie (1982), and Engle and Ng (1993) believe this phenomenon happens because the "bad" news tends to drive the variance higher than "good" news does. Lam et al. (1990) and Li and

Lam (1995) also studied such asymmetric effects in stock returns. They successfully used a threshold model with conditional heteroscedasticity to capture the asymmetric behavior of stock returns during bear and bull markets. Specifically, Li and Lam (1995) showed that the stock returns on one day could significantly affect the variance of the stock returns on the next day. Many other multi-regime stochastic models were proposed to address the above multi-regime features, including threshold stochastic volatility models (So and Choi, 2009), Markov switching stochastic volatility models (So et al., 1998; Kalimipalli and Susmel, 2004; Lopes and Carvalho, 2007; Abanto-Valle et al., 2010; Virbickaitė et al., 2019b), and smooth transition stochastic volatility models (Lopes and Salazar, 2006; Livingston Jr and Nur, 2017).

There is a long history of studying the relationship between stock return volatility and trading volume in the finance literature. Clark (1973) introduced the mixture of distributions hypothesis (MDH), which assumes the return and the trading volume depends on the same underlying latent flow variable, i.e., the changes in the stock price and the trading volume are driven primarily by the release of new information. Karpoff (1987) also discussed the relationship between the return and the trading volume in various financial markets based on a literature review. We consider a smooth transition process with categorical time series (Davis and Ensor, 2007; Wang and Holan, 2012) as a good candidate to model such a relationship.

We propose a method to use the trading volume information to model the regime-switching in the volatility, where the regime-switching is observed through an ordered categorical variable that denotes the regime the series is in at each time. Considering that the trading volume might be quite variable across assets, a more stable relationship should hold for this standardization of volume. Specifically, we propose an

SV model with the volatility following an ordered regime-switching autoregressive process based on auxiliary covariates. The flexibility of our approach can account for the asymmetry of financial returns and utilize the trading volume information. The latter is not typically used in SV models.

Additionally, the non-Gaussian behavior of return data has not been fully captured by parametric SV models (Gallant et al., 1997; Mahieu and Schotman, 1998; Liesenfeld and Jung, 2000; Durham, 2006; Jensen and Maheu, 2010). Some flexible semiparametric SV models were proposed by Jensen and Maheu (2010, 2014) to capture these important characteristics. Building on this literature, we use the Dirichlet process mixture prior (DPM) to increase the flexibility and robustness to potential non-Gaussian behavior of the return innovations.

The remainder of this chapter is organized as follows. In Section 2.2, we formulate several Bayesian multi-regime smooth transition regression models with ordered categorical variables. In Section 2.3, we present simulation results that illustrate the frequentist properties of the proposed Bayesian estimators. In Section 2.4, we demonstrate an application to British Petroleum stock returns (1/1/2009-9/30/2016) and compare our smooth transition stochastic volatility model with a stochastic volatility threshold autoregressive model (SV-TAR). Section 2.5 contains concluding remarks. Finally, the details of our Markov chain Monte Carlo (MCMC) estimation algorithm are provided in Section 2.6.

2.2 Models

2.2.1 Smooth Transition Stochastic Volatility Models

Financial returns $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$ have been modeled by Jacquier et al. (1994) and Kim et al. (1998) using the following SV model

$$y_t = \exp(h_t/2)\epsilon_t \tag{2.1}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_h \eta_t, \tag{2.2}$$

where it is assumed that $\{\epsilon_t\}$ and $\{\eta_t\}$ are standard Gaussian white-noise innovations for $t \in \{1, \ldots, T\}$. We call (2.1) the observation equation and (2.2) the state equation. The latent variable $\mathbf{h} = (h_0, h_1, \ldots, h_T)'$ in the state equation follows an autoregressive process of order 1 (AR(1)), which is interpreted as the volatility of returns. To ensure stationarity of the latent process, the common AR(1) assumption is that $|\phi| < 1$ and that the initial state is distributed as $h_0|\mu, \phi, \sigma_h^2 \sim \mathrm{N}(\mu, \sigma_h^2/(1-\phi^2))$. To accommodate the volatility processes with regime switching, we replace $\{h_t\}$ in (2.2) with the multi-regime smooth transition autoregressive process with ordered categorical variables presented in Wang and Holan (2012). Before introducing the proposed model, we define the multi-regime autoregressive process of $\{h_t\}$ as a logistic smooth transition autoregressive model with ordered categorical variables (Wang and Holan, 2012, LSTAR-Cv model;). First, we define a latent transition variable ν_t as

$$\nu_t = \mathbf{z}_t'\boldsymbol{\gamma} + e_t, \quad e_t \overset{iid}{\sim} \mathrm{N}(0,1), \quad \text{for } t \in \{1, 2, \ldots, T\}, \tag{2.3}$$

where $\mathbf{z}_t' = (z_{1t}, z_{2t}, \ldots, z_{Kt})'$ is a vector of covariates associated with ν_t. The latent transition variable ν_t controls the transition in latent process

$$h_t = \phi_{1,1} + \phi_{1,2}h_{t-1} + (\phi_{2,1} + \phi_{2,2}h_{t-1})\frac{1}{1+\exp\{-\nu_{t-1}\}} + \sigma_h\eta_t, \quad \eta_t \overset{iid}{\sim} \mathrm{N}(0,1), \tag{2.4}$$

where for $j = 1, 2$, $\phi_{j,1}$ and $\phi_{j,2}$ denote the additive effect of Regime j on the intercept and the slope of the autoregressive process $\{h_t\}$, respectively; i.e., $\phi_{1,1}$ and $\phi_{1,2}$ is the baseline effect. Neither ν_t or h_t are observed. Instead, u_t is observed as the regime that restricts the range of the latent transition variable ν_t, such that,

$$\theta_{j-1} \leq \nu_t < \theta_j \quad \text{if} \quad u_t = j, \tag{2.5}$$

where $-\infty = \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_q = +\infty$ denote ordered threshold values. The unknown thresholds $\{\theta_0, \ldots, \theta_q\}$ divide the autoregressive process into $q+1$ regimes. The function $M(\nu_t; 1, 0) = 1/\{1+\exp(-\nu_{t-1})\}$ works as the transition function in this equation. To better understand the smooth transition between regimes, we consider the two-regime model. We reparametrize and let $a_1 = \phi_{1,1}$, $a_2 = \phi_{1,1} + \phi_{2,1}$, $b_1 = \phi_{1,2}$ and $b_2 = \phi_{1,2} + \phi_{2,2}$. As the regime smoothly transitions between Regime 1 and Regime 2, the intercept of the process $\{h_t\}$ varies between a_1 and a_2 and the slope of the process $\{h_t\}$ varies between b_1 and b_2 according to the value of ν_t.

Conditional on $\{h_t\}$, the returns $\{y_t\}$ are normally distributed as follows

$$y_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathrm{N}(0,1). \tag{2.6}$$

In the model, $\{y_t\}$ and $\{u_t\}$ are the only two observed variables, $\{\mathbf{z}_t\}$ are known

covariates. All of the other variables and parameters are unknown. The $\{\nu_t\}$ and $\{\boldsymbol{\theta}_t\}$ jointly control the regime switching.

In the context of stochastic volatility, to better facilitate the Bayesian sampling, we model a new variable y_t^* instead of y_t. Suggested by Shephard (1994) and Carter and Kohn (1994), (2.1) is approximated as

$$y_t^* = \log(y_t^2 + c) = h_t + \log(\epsilon_t^2) = h_t + \epsilon_t^*, \quad \epsilon_t \overset{iid}{\sim} \mathrm{N}(0,1), \tag{2.7}$$

where y_t^* denotes $\log(y_t^2 + c)$. A small positive offset c is here to deal with the very small value of y_t^2. We use $c = 10^{-4}$ throughout the remainder, while it is possible to let c depend on the actual value y_t^2 (Kim et al., 1998). The distribution of ϵ_t^* is approximated by a mixture distribution with 10 components (Omori et al., 2007).

Consequently, (2.1) and (2.2) can be rewritten in the form of a linear conditionally Gaussian state space model. As such, this transformation greatly increases the sampling efficiency of the MCMC simulation. The distribution of $\log(\epsilon_t^2)$ can be approximated by a mixture of normal distributions. Specifically, let $r_t \in \{1, \ldots, 10\}$ be the mixture component indicator at time t, while m_{r_t} and $w_{r_t}^2$ denote the mean and the variance of the r_tth mixture component as shown in Omori et al. (2007). Noting that $\epsilon_t^* = \log(\epsilon_t^2)$, the logarithm of the squared innovation is distributed as $\epsilon_t^*|r_t \sim \mathrm{N}(m_{r_t}, w_{r_t}^2)$. The distribution of $\epsilon_t^*|r_t$ is defined to be a mixture of 10 normal distributions and the values of p_k, m_k, w_k^2 are given in Table 2.1 (as in Omori et al. (2007) and further described in Section 2.6). Since the distribution of ϵ_t^* is approximated, Kim et al. (1998); Omori et al. (2007) proposed an additional step of reweighting the MCMC samples to correct for the approximation error. But according to their experiments, the reweighting has little impact on the parameters in the

Table 2.1: Selection of the mixing distribution to approximate $\log\chi_1^2$ by Omori et al. (2007).

k	$K = 10$		
	p_k	m_k	w_k^2
1	0.00609	1.92677	0.11265
2	0.04775	1.34744	0.17788
3	0.13057	0.73504	0.26768
4	0.20674	0.02266	0.40611
5	0.22715	-0.85173	0.62699
6	0.18842	-1.97278	0.98583
7	0.12047	-3.46788	1.57469
8	0.05591	-5.55246	2.54498
9	0.01575	-8.68384	4.16591
10	0.00115	-14.65000	7.33342

latent process, Moreover, the 10 component approximation of Omori et al. (2007) improved the accuracy and made the effect of reweighting even smaller. Therefore, we do not use reweighting, though it is straightforward to add a reweighting step.

Combining the observation process and the volatility process, we get the bivariate SV-LSTAR-Cv model for $(y_t, u_t), t = 1, \ldots, T$:

$$
\begin{aligned}
&y_t^* = \log(y_t^2 + c) = h_t + \epsilon_t^*, \quad \epsilon_t^* | r_t \sim N(m_{r_t}, w_{r_t}^2), \\
&Pr(r_t = k) = p_k, \quad k \in \{1, \ldots, 10\}, \\
&h_t = \phi_{1,1} + \phi_{1,2} h_{t-1} + (\phi_{2,1} + \phi_{2,2} h_{t-1}) \frac{1}{1 + \exp\{-\nu_{t-1}\}} + \sigma_h \eta_t, \quad \eta_t \overset{iid}{\sim} \mathrm{N}(0,1), \\
&\theta_{j-1} \leq \nu_t < \theta_j \quad \text{if} \quad u_t = j, \\
&\nu_t = \mathbf{z}_t' \boldsymbol{\gamma} + e_t, \quad e_t \overset{iid}{\sim} N(0,1), \quad e_t \perp \eta_t, \quad \text{for } t \in \{1, 2, \ldots, T\},
\end{aligned}
$$

where the transition function $M(\nu_t) = 1/\{1 + \exp(-\nu_t)\}$ is used.

We extend the above model by (1) specifying regime-wise intercept and slope ($\phi's$) to each regime; (2) introducing $M(s_t;\lambda,\theta_j) = 1/\left[1+\exp\{-\lambda(s_t-\theta_j)\}\right]$ to replace $1/\left[1+\exp\{-\theta_j\}\right]$. We get the bivariate SV-MRSTAR-C model for $(y_t, u_t), t = 1,\ldots,T$:

$$
\begin{aligned}
& y_t^* = \log(y_t^2 + c) = h_t + \epsilon_t^*, \quad \epsilon_t^* | r_t \sim \mathrm{N}(m_{r_t}, w_{r_t}^2), \\
& Pr(r_t = k) = p_k, \quad k \in \{1,\ldots,10\}, \\
& h_t = \phi_{1,1} + \phi_{1,2} h_{t-1} + (\phi_{2,1} + \phi_{2,2} h_{t-1}) M(\mathbf{z}_{t-1}'\boldsymbol{\gamma};\lambda,\theta_1) + \cdots \\
& \quad + (\phi_{q+1,1} + \phi_{q+1,2} h_{t-1}) M(\mathbf{z}_{t-1}'\boldsymbol{\gamma};\lambda,\theta_q) + \sigma_h \eta_t, \quad \eta_t \overset{iid}{\sim} \mathrm{N}(0,1), \\
& h_0 \sim N\left(0, \frac{\sigma^2}{1-\phi^{*2}}\right), \quad \phi^* = \sum_{j=1}^{q+1} \phi_{j,2}, \\
& \theta_{j-1} \le \nu_t < \theta_j \quad \text{if} \quad u_t = j, \\
& \nu_t = \mathbf{z}_t'\boldsymbol{\gamma} + e_t, \quad e_t \overset{iid}{\sim} \mathrm{N}(0,1), \quad \text{for } t \in \{1,2,\ldots,T\},
\end{aligned}
$$

where $\epsilon_t^* | r_t$ is from a mixture of 10 normal distributions and p_k, m_k, w_k^2 are specified in Table 2.1 (as in Omori et al. (2007) and further described in the Section 2.6), $M(\mathbf{z}_t'\boldsymbol{\gamma};\lambda,\theta_j) = 1/\left[1+\exp\{-\lambda(\mathbf{z}_t'\boldsymbol{\gamma}-\theta_j)\}\right]$ and λ is the parameter that controls the smoothness of the regime transition. Note that, $\mathbf{z}_t'\boldsymbol{\gamma}$ is the mean of ν_t. If instead we still use ν_t in the transition function rather than $\mathbf{z}_t'\boldsymbol{\gamma}$, the MCMC sample chains of such a model converges slowly and shows large autocorrelations, which is similar to fitting of MRSTAR-Cv in Wang and Holan (2012). Therefore, throughout the sequel, we only consider the SV-LSTAR-Cv model and the SV-MRSTAR-C model.

Conditional on the transition variable ν_t, smoothness parameter λ, and the threshold parameters θ_j, $j = 1,\ldots,q$, the model becomes a dynamic linear model with time-varying evolution parameter. Many well-established techniques have been de-

veloped to sample from the posterior distribution of the volatility $\{h_t\}$. Here we use the Forward Filtering Backward Sampling (FFBS) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) to sample $\{h_t\}$ from its full conditional distribution.

The volatility variable $\{h_t\}$ and the latent variable $\{\nu_t\}$ follow a typical multi-regime smooth transition autoregressive model. Conditional on $\{h_t\}$, θ_q, and $\{\nu_t\}$ can be sampled as in Wang and Holan (2012). As mentioned previously, even with a known transition variable, there is still some difficulty in estimating the smoothness parameter λ. Moreover, our model has an additional level in the hierarchy for the smooth transition autoregressive models, which makes λ difficult to estimate (i.e., λ is farther away from the data). Also, as noted in Lubrano (2000) and Lopes and Salazar (2006), for the logistic smooth transition autoregressive (LSTAR) model, the prior distribution for λ should assign small weight to the region near 0 and its density should decrease at the rate of $O(\lambda^{-1-d})$ for $d > 0$. Therefore, similar to Wang and Holan (2012), we assume that λ has an informative inverse gamma (IG) prior. See Section 2.6 for computational details.

2.2.2 Semiparametric SV-LRSTAR-Cv

As discussed previously, the asymmetry and leptokurtotic behavior of return data has not been fully captured by parametric SV models. Therefore, to model the skewness and kurtosis of returns better, we introduce an SV model with a flexible nonparametric innovation distribution (Jensen and Maheu, 2010, 2014; Virbickaitė et al., 2019a), which specifies both parametric and nonparametric features for the return process. The nonparametric part of this model consists of an infinitely ordered mixture of normal distributions with each component probability and associated pa-

rameters modeled through Dirichlet process mixture prior (DPM). In this context, we use the simplified version, DPM-P which is also defined in Jensen and Maheu (2010), which assumes that the mixture of normal distributions is centered at zero and that only the mixture probabilities and the mixture precision parameter follow a DP prior. The model comparison in Jensen and Maheu (2010) indicates the fitting of stochastic volatility DPM-P (SV-DPM-P) model is no worse than the stochastic volatility DPM (SV-DPM) model in terms of DIC. Considering that the latent smooth transition adds significant flexibility, we want the DPM in the model to be more restrictive to make the estimation more stable, and to make the MCMC sample chains converge faster. Therefore, SV-DPM-P is preferred and used throughout this context. The SV-DPM-P is defined in Jensen and Maheu (2010) as

$$y_t|h_t, \mu, \omega_t \overset{\perp}{\sim} \mathrm{N}\left(\mu, \omega_t^{-2}\exp\{h_t\}\right), \tag{2.8}$$

$$h_t|h_{t-1}, \phi, \sigma_h^2 \sim \mathrm{N}\left(\phi h_{t-1}, \sigma_h^2\right), \quad h_t \perp y_t \tag{2.9}$$

$$\omega_t|G \overset{iid}{\sim} G, \tag{2.10}$$

$$G|G_0, \alpha \sim \mathrm{DP}(G_0, \alpha), \tag{2.11}$$

$$G_0(\omega_t^2) \equiv \Gamma(\nu_0, s_0), \tag{2.12}$$

where $\overset{\perp}{\sim}$ denotes independently distributed. The latent log volatility $\{h_t\}$ follows the autoregressive (AR) process defined by (2.9) with AR parameter ϕ. The identifiability of this model requires the unconditional mean of $\{h_t\}$ to be zero. Thus, the intercept of the return's innovation variance is controlled by ω_t^2. The autoregressive coefficient ϕ is restricted to the interval $(-1, 1)$ to ensure stationary returns. This guarantees the volatility process $\{h_t\}$ has finite mean and variance. The unconditional return

distribution is given a nonparametric prior defined by (2.10)-(2.12), which consists of an overall mean and a potentially infinite ordered mixture of normals centered at zero. It is assumed that the mixture probabilities and parameters ω_t^2 follow the Dirichlet process prior (DP) of Ferguson (1973) as (2.10) and (2.11). The DP prior consists of the base distribution G_0 with nonnegative concentration parameter α, where G_0 is defined in (2.12) as a conjugate gamma distribution.

We embed this semiparametric prior into the proposed smooth transition SV models and specify the semiparametric bivariate SV-LSTAR-Cv model for $(y_t, u_t), t = 1, \ldots, T$ as

$$
\begin{aligned}
& y_t | h_t, \mu, \omega_t \sim \mathrm{N}\left(\mu, \omega_t^{-2} \exp\{h_t\}\right), \\
& h_t | h_{t-1}, \boldsymbol{\phi}, \sigma_h^2 \sim \mathrm{N}\left(\phi_1 h_{t-1} + \phi_2 h_{t-1} \frac{1}{1+\exp\{-\nu_{t-1}\}}, \sigma_h^2\right), \qquad (2.13) \\
& h_0 \sim N\left(0, \frac{\sigma_h^2}{1-(\phi_1+\phi_2)^2}\right), \\
& \theta_{j-1} \leq \nu_t < \theta_j \quad \text{if} \quad u_t = j, \qquad (2.14) \\
& \nu_t = \mathbf{z}_t' \boldsymbol{\gamma} + e_t, \quad e_t \overset{iid}{\sim} \mathrm{N}(0,1), \\
& \omega_t | G \sim G, \qquad (2.15) \\
& G | G_0, \alpha \sim \mathrm{DP}(G_0, \alpha), \qquad (2.16) \\
& G_0(\omega_t^2) \equiv \Gamma(\nu_0, s_0). \qquad (2.17)
\end{aligned}
$$

The latent log volatility $\{h_t\}$ follows the LSTAR-Cv process defined by (2.13) with AR parameter $\boldsymbol{\phi} = (\phi_1, \phi_2)'$, which has regimes and thresholds defined by (2.14). The identifiability of this model requires the unconditional mean of $\{h_t\}$ to be zero. Thus, the intercept of the return's innovation variance is controlled by ω_t^2. We assume

a restricted normal piror on π to ensure the latent autoregressive process is nonexplosive within each regime, such that, $\boldsymbol{\phi} \sim \mathrm{N}_{|\phi_1|<1,|\phi_1+\phi_2|<1}\left(\mathbf{0}, \mathrm{diag}\{\sigma^2_{\phi_1}, \sigma^2_{\phi_2}\}\right)$, where $\mathrm{N}_{\mathcal{A}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a restricted normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ and restriction conditions $\mathcal{A}$. We also assign priors $\sigma^2_h \sim \mathrm{IG}(a_\sigma, b_\sigma)$, $\boldsymbol{\gamma} \sim \mathrm{N}(\mathbf{0}, \delta^2_\gamma \mathbf{I})$, and $\mu \sim \mathrm{N}(m_\mu, \sigma^2_\mu)$, with fixed hyperparameters. Thus, the full conditional distributions of both $\boldsymbol{\phi}$ and σ^2_h have analytical forms. The volatility $\{h_t\}$ can not be sampled using FFBS because the current model is no longer a dynamic linear model. Instead, we use the random length block sampler proposed by Jensen and Maheu (2010). Conditional on $\{h_t\}$, θ_q, λ, and ν can be sampled in a similar way to Section 2.2.1. The DPM-P uses conjugate priors and the corresponding parameters are sampled using a two-step procedure. According to Escobar and West (1995), we assign $\alpha \sim \Gamma(a, b)$, the hyperparameters of which are selected to make these priors vague. Computational details are provided in Section 2.6.

As a mixture of distribution, DPM suffers from "label switching," which results in the drawback that the mixture parameters are unidentified (Jensen and Maheu, 2010). Nevertheless, our purpose is to use DPM to model the asymmetry and leptokurtotic distribution of $y_t \exp(-h_t/2)$. The "label switching" does not result in any problems for the inference of the other parameters of a stochastic volatility model. Geweke (2007); Frühwirth-Schnatter (2006) provide additional discussion of this issue in mixture models.

We embed the DPM-P into the MRSTAR-C model. However, the result of the simulation studies indicate that the smoothness parameter of the transition, λ, is not easily identified together within the flexible semiparametric sampler. Therefore, instead, we focus on the semiparametric SV-LSTAR-Cv model.

2.3 Simulation Studies

2.3.1 Simulation 1: SV-LSTAR-Cv

In the simulation study, we generate $\{y_t\}$, $\{h_t\}$, $\{\nu_t\}$ and $\{u_t\}$ series of length $T = 1000, 2000$, where the ordinal variable u_t has $q = 3$ regimes. The transition process $\{\nu_t\}$ and the categorical process $\{u_t\}$ are generated similar to the examples of Davis and Ensor (2007) and Wang and Holan (2012). That is, for $t = 1, \ldots, T$,

$$\nu_t = \mathbf{z}_t'\boldsymbol{\gamma} + \epsilon_t = \gamma_1 l_t + \gamma_2 u_{t-1,1} + \gamma_3 u_{t-1,2} + e_t, \tag{2.18}$$

$$u_t = j \iff \theta_{j-1} \leq \nu_t < \theta_j, \tag{2.19}$$

where l_t is the log daily trading volume of British Petroleum (BP) stock starting from 1/2/2009, $u_{t,j} = 1$ if $u_t = j$, $u_{t,j} = 0$ otherwise for $j = 1, 2, 3$, $\gamma = (0.2, -0.3, 0.5)'$, $\epsilon_t \overset{iid}{\sim} \mathrm{N}(0, 1)$, and $\boldsymbol{\theta} = (\theta_1, \theta_2)' = (2.23, 4.8)'$. Next, the observation process $\{y_t\}$ and the latent process $\{h_t\}$ are generated from

$$\begin{aligned} y_t &= \exp(h_t/2)\epsilon_t, \\ h_t &= \phi_1 + \phi_2 h_{t-1} + (\phi_3 + \phi_4 h_{t-1}) \frac{1}{1 + \exp\{-v_{t-1}\}} + \sigma_h \eta_t, \end{aligned}$$

with $(\phi_1, \phi_2, \phi_3, \phi_4)' = (-0.3, 0.6, -0.3, 0.35)'$, $\sigma_h^2 = 0.1, 0.5, 1$ and $\{\epsilon_t\}$ and $\{\eta_t\}$ are standard Gaussian white-noise. The values of the thresholds $\boldsymbol{\theta}$ and the covariate, log trading volume, are used to make the simulated data closer to the real data example in Section 2.4. When generating series of length 1000, we use the first 1000 log daily trading volume of BP starting from 1/2/2009 as the covariate l_t. When generating series of length 2000, we use the first 1000 log volumes sequentially, as the covariate.

By doing so, we focus on the effect of the sample size.

We generated 100 datasets as stated above and estimated the parameters $(\phi, \sigma^2; \gamma, \theta)$ via MCMC for each simulated dataset. In the MCMC algorithm, we specify the hyperparameters as $\sigma^2_{\phi_1} = 100$, $\sigma^2_{\phi_2} = 100$, $\delta^2_\gamma = 10,000$, $a_\sigma = 1$ and $b_\sigma = 1$. We run the MCMC simulation of $15,000$ draws for each simulated dataset and keep the last $10,000$ for inference in each MCMC sample chain. We assess the convergence of the MCMC algorithm through visual inspection of the trace plots of the sample chains and the Gelman-Rubin diagnostic (Brooks and Gelman, 1998). In particular, we randomly pick 10 of the 100 simulated datasets and generate three sample chains for each of the parameters and assess their convergence. Visual inspection of the trace plots of the sample chains does not show any evidence of lack of convergence for any of the parameters and the Gelman-Rubin statistic, $\widehat{R}$, is less than 1.01 for all parameters. A typical example of the estimated volatility compared with the true volatility is provided in Figure 2.1. The sampling properties of the Bayesian MCMC estimators are shown in Table 2.2. From this table, we see that when the sample size increases from 1000 to 2000, the frequentist coverage probability of the log volatility gets closer to the ideal value 0.9.

In the SV-LSTAR-Cv model, the intercept of the process $\{h_t\}$ varies between a_1 and a_2 and the slope of the process $\{h_t\}$ varies between b_1 and b_2 according to the value of ν_t. More specifically, in Regime 1, the intercept varies between a_1 and $a_1 + \phi_3/(1 + e^{-\theta_1})$ and the slope varies between b_1 and $b_1 + \phi_4/(1 + e^{-\theta_1})$; in Regime 2, the intercept varies between $a_1 + \phi_3/(1 + e^{-\theta_1})$ and $a_1 + \phi_3/(1 + e^{-\theta_2})$ and the slope varies between $b_1 + \phi_4/(1 + e^{-\theta_1})$ and $b_1 + \phi_4/(1 + e^{-\theta_2})$; in Regime 3, the intercept varies between $a_1 + \phi_3/(1 + e^{-\theta_2})$ and a_2 and the slope varies between

Figure 2.1: Comparison of the estimated log-volatility (black line) for one simulated dataset from the SV-LSTAR-Cv model and the true log-volatility (red line). The grey region is the 90% credible interval. Note, the estimated log-volatility is $\{\widehat{h}_t\}$. The first 200 of the series are enlarged to show the details.

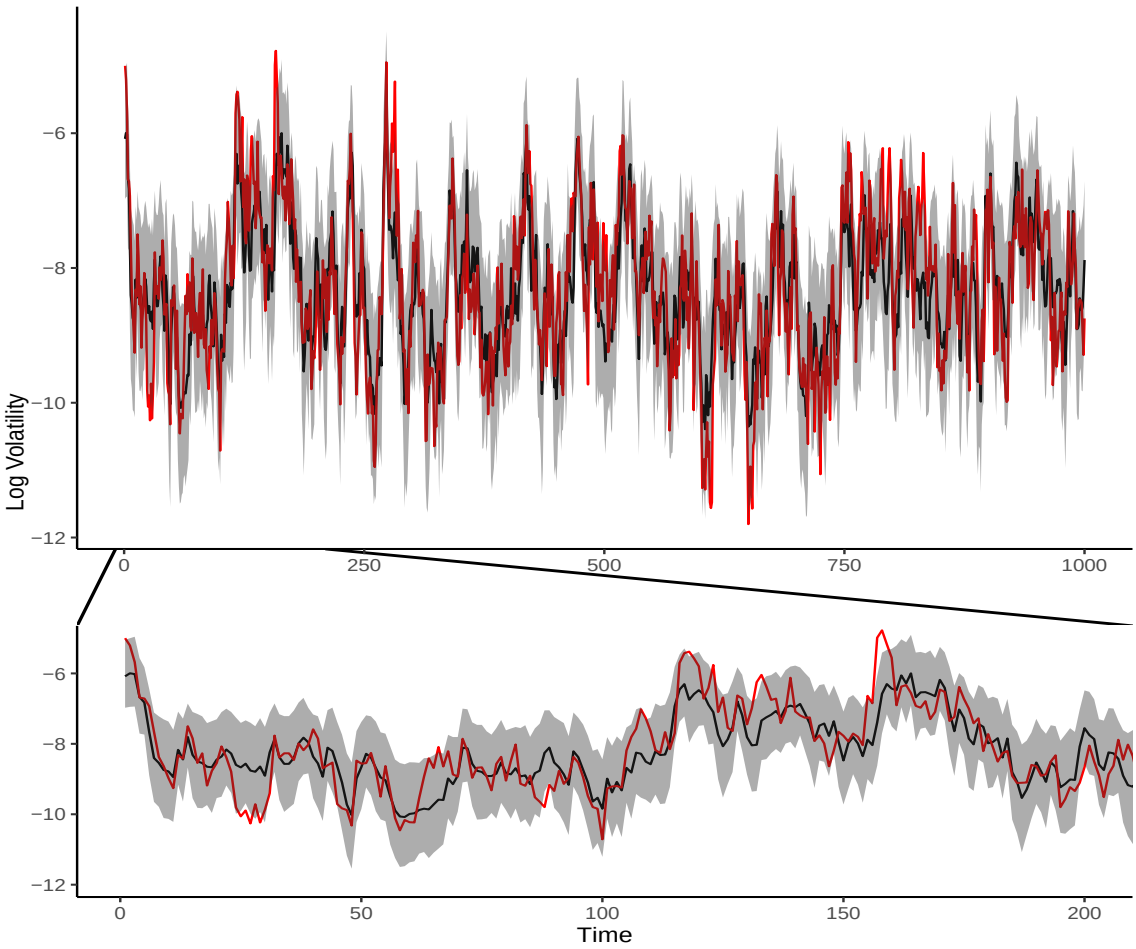


Table 2.2: Summary of parameter estimation for the SV-LSTAR-Cv model based on 100 simulated datasets. We report the posterior mean, standard deviation (s.d.) and mean square error (MSE) of the Bayesian estimates for the series length $T = 1000, 2000$ in the first simulation study (Section 2.3). The CP in the parenthesis stands for the coverage probability of 90% credible interval of the log volatility.

T			a_1	b_1	a_2	b_2	θ_1	θ_2	γ_1	γ_2	γ_3	σ_h^2
(CP)	σ_h^2		-0.300	0.600	-0.600	0.950	2.230	4.800	0.200	-0.300	0.500	
1000	1	Mean	-0.519	0.477	-0.643	0.946	2.232	4.811	0.199	-0.294	0.513	0.975
		s.d.	0.745	0.275	0.123	0.013	0.690	0.695	0.042	0.170	0.138	0.099
(0.889)		MSE	0.598	0.089	0.017	0.00018	0.471	0.478	0.00175	0.029	0.0191	0.0104
1000	0.5	Mean	-0.599	0.455	-0.679	0.942	2.278	4.856	0.202	-0.293	0.516	0.547
		s.d.	0.705	0.270	0.132	0.015	0.672	0.686	0.042	0.167	0.136	0.072
(0.906)		MSE	0.581	0.093	0.024	0.00028	0.449	0.469	0.00174	0.028	0.0186	0.0073
1000	0.1	Mean	-0.658	0.420	-0.871	0.921	2.314	4.893	0.204	-0.294	0.516	0.178
		s.d.	0.576	0.285	0.200	0.021	0.655	0.654	0.040	0.170	0.138	0.036
(0.912)		MSE	0.457	0.113	0.113	0.00127	0.432	0.433	0.00159	0.029	0.0192	0.0074
2000	1	Mean	-0.306	0.596	-0.613	0.947	2.191	4.762	0.198	-0.313	0.487	0.954
		s.d.	0.710	0.172	0.120	0.011	0.452	0.450	0.029	0.124	0.094	0.085
(0.892)		MSE	0.499	0.029	0.014	0.00013	0.204	0.202	0.00082	0.015	0.0089	0.0092
2000	0.5	Mean	-0.472	0.567	-0.593	0.946	2.161	4.731	0.196	-0.312	0.488	0.518
		s.d.	0.686	0.187	0.127	0.012	0.457	0.459	0.029	0.122	0.094	0.056
(0.903)		MSE	0.496	0.036	0.016	0.00016	0.212	0.212	0.00083	0.015	0.0088	0.0036
2000	0.1	Mean	-0.491	0.534	-0.657	0.935	2.216	4.787	0.200	-0.313	0.488	0.142
		s.d.	0.558	0.174	0.186	0.014	0.437	0.445	0.028	0.121	0.093	0.025
(0.909)		MSE	0.345	0.035	0.037	0.00042	0.189	0.196	0.00075	0.015	0.0088	0.0024

Note: $a_1 = \phi_1$, $a_2 = \phi_1 + \phi_3$, $b_1 = \phi_2$ and $b_2 = \phi_2 + \phi_4$.

$b_1 + \phi_4/(1 + e^{-\theta_2})$ and b_2. In the simulation examples, based on the true values of parameters, we expect to see that in Regime 1 the intercept varies between -0.3 and -0.571 and the slope varies between 0.6 and 0.916; in Regime 2, the intercept varies between -0.571 and -0.598 and the slope varies between 0.916 and 0.947; in Regime 3, the intercept varies between -0.598 and -0.6 and the slope varies between 0.947 and 0.95. As expected, the increase of the sample size helps to reduce the bias and standard deviations of the Bayesian estimates. When we increase the true value of σ_h^2, that is, the signal-to-noise ratio, the mean square error (MSE) of each parameter in the latent process of y_t decreases. When σ_h^2 becomes too small, its estimation becomes significantly biased. Nevertheless, as the sample size T increases, the biases and the standard deviations of these estimators still decrease. In SV models it is common to moderately overestimate σ_h^2. In this simulation, it appears that the overestimation of σ_h^2 leads to the underestimation of b_2 as well as the overestimation of a_2.

2.3.2 Simulation 2: SV-MRSTAR-C

Similar to the first simulation, we generate 100 datasets of $\{u_t\}$, $\{\nu_t\}$, $\{h_t\}$, and $\{y_t\}$ of lengths $T = 1000, 2000$ using a three-regime SV-MRSTAR-C model. The only difference is that the transition function is $M(s_t; \lambda, \theta_j) = 1/[1+\exp\{-\lambda(\mathbf{z}'_{t-1}\boldsymbol{\gamma}-\theta_j)\}]$. Therefore,

$$
\begin{aligned}
y_t &= \exp(h_t/2)\epsilon_t, \\
h_t &= \phi_{1,1} + \phi_{1,2}h_{t-1} + \frac{\phi_{2,1} + \phi_{2,2}h_{t-1}}{1 + \exp\{-\lambda(\mathbf{z}'_{t-1}\boldsymbol{\gamma} - \theta_1)\}} + \frac{\phi_{3,1} + \phi_{3,2}h_{t-1}}{1 + \exp\{-\lambda(\mathbf{z}'_{t-1}\boldsymbol{\gamma} - \theta_2)\}} + \sigma_h\eta_t,
\end{aligned}
$$

where $\Phi_1 = (\phi_{1,1}, \phi_{1,2})' = (0, 0.7)'$, $\Phi_2 = (\phi_{2,1}, \phi_{2,2})' = (-0.5, 0.1)'$, $\Phi_3 = (\phi_{3,1}, \phi_{3,2})' = (-1, 0.15)'$, $\lambda = 20, 50$, $\sigma_h = 1$, and $\{\epsilon_t\}$, $\{\eta_t\}$ are standard Gaussian white-noise. The ordinal process $\{\nu_t\}$ has three categories and is generated using (2.18) with $\boldsymbol{\gamma} = (1, -1.5, 2.5)'$. The categorical process $\{u_t\}$ is defined using (2.19) with $\boldsymbol{\theta} = (-1, 1)'$. We estimate the parameters $(\phi, \sigma^2; \gamma, \boldsymbol{\theta}, \lambda)$ via MCMC for each simulated dataset. In the MCMC, we specify the hyperparameters as $\sigma^2_{\phi_1} = 100$, $\sigma^2_{\phi_2} = 100$, $\delta^2_\gamma = 10,000$, $a_\sigma = 1$, $b_\sigma = 1$. For each simulated dataset, we run an MCMC simulation of $15,000$ draws and keep the last $10,000$ for inference in each MCMC sample chain. We assess the MCMC algorithm's convergence using both trace plots of the sample chains and the Gelman-Rubin diagnostic. We randomly pick 10 of the 100 simulated datasets and generate three sample chains for each of the parameters and assess their convergence. Visual inspection of the trace plots of the sample chains does not show any evidence of lack of convergence for any of the parameters, and the Gelman-Rubin statistic, $\widehat{R}$, is less than 1.01 for all parameters. The sampling properties of the Bayesian estimators are given in Table 2.3. In this simulation, we set $\boldsymbol{\theta} = (-1, 1)'$ and this makes the transition very smooth when $\lambda < 5$. Thus, we set $\lambda = 20, 50$ to investigate the performance of SV-MRSTAR-C model when the transition is not as smooth. This simulation shows that when λ is large enough, it is hard to distinguish between two different large λs because large values of λ make the transition function nearly a step function. However, a large λ value leads to a larger separation between regimes so that the smooth transition parameters are easier to estimate. In other words, large values of λ lead to better estimations of the smooth transition parameters in terms of the mean and the standard deviation, while smaller values of λ are easier to be estimated. Moreover, this simulation also suggests a larger sample size improves the estimation of

small values of λ; i.e., under large sample sizes, the average of the posterier means of every parameter is closer to the true value and the standard deviation of the posterior means of every parameter is smaller.

2.3.3 Simulation 3: Semiparametric SV-LSTAR-Cv

Similar to SV-LSTAR-Cv, we generate 100 datasets of $\{u_t\}$, $\{\nu_t\}$, $\{h_t\}$, and $\{y_t\}$ processes of length $T = 1000$ with $M(s_t; \lambda, \theta_j) = M(\nu_{t-1}; 1, 0) = 1/\{1 + \exp(-\nu_{t-1})\}$. The observation process $\{y_t\}$ and the latent process $\{h_t\}$ are generated as

$$
\begin{aligned}
y_t &= \exp(h_t/2)\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathrm{St}(1/3, 3), \\
h_t &= \phi_1 h_{t-1} + \phi_2 h_{t-1} \frac{1}{1 + \exp\{-\nu_{t-1}\}} + \sigma_h \eta_t,
\end{aligned}
$$

where $\phi_1 = 0.8$, $\phi_2 = 0.15$, $\sigma_h = 1$, and $\{\eta_t\}$ is Gaussian white-noise and $\mathrm{St}\{(v-2)/v, v\}$ denotes a Student-t density standardized to variance 1 and v degree of freedom. Under simulation, we compare the estimation of the SV-LSTAR-Cv and the semiparametric SV-LSTAR-Cv models. For each simulated dataset, we run the MCMC simulation with $15,000$ draws and keep the last $10,000$ for inference in each MCMC sample chain. We assess the MCMC algorithm's convergence using both trace plots of the sample chains and the Gelman-Rubin diagnostic. We randomly pick 10 of the 100 simulated datasets and generate three sample chains for each of the parameters and assess their convergence. Visual inspection of the trace plots of the sample chains does not show any evidence of lack of convergence for any of the parameters and the Gelman-Rubin statistic, $\widehat{R}$, is less than 1.02 for all parameters. The convergence of the nonparametric parameters of the DPM-P model can not be directly

Table 2.3: Summary of parameter estimation for the SV-MRSTAR-C model based on 100 simulated datasets. We report the mean and standard deviation (s.d.) of the posterior means for the series length $T = 1000, 2000$ and $\lambda = 20, 50$ in the second simulation study (Section 2.3.2).

$T = 1000$	a_1	b_1	a_2	b_2	a_3	b_3	θ_1	θ_2
True	0	0.7	-0.5	0.8	-1.5	0.95	-1	1
Mean	0.0019	0.7020	-0.4960	0.8447	-1.5510	0.9475	-1.0185	1.0402
s.d.	0.1924	0.0360	0.2486	0.0337	0.1912	0.0223	0.0730	0.0644
$\lambda = 20$	γ_1	γ_2	γ_3	σ_h^2	λ			
True	1	-1.5	2.5	1	20			
Mean	0.9748	-1.5032	2.5277	1.0594	26.2819			
s.d.	0.0705	0.0953	0.1193	0.1313	5.3307			
$T = 2000$	a_1	b_1	a_2	b_2	a_3	b_3	θ_1	θ_2
True	0	0.7	-0.5	0.8	-1.5	0.95	-1	1
Mean	0.0016	0.6972	-0.5012	0.8451	-1.4802	0.9481	-1.0213	1.0327
s.d.	0.1582	0.0304	0.1861	0.0276	0.1595	0.0207	0.0570	0.0490
$\lambda = 20$	γ_1	γ_2	γ_3	σ_h^2	λ			
True	1	-1.5	2.5	1	20			
Mean	0.9805	-1.5061	2.5302	1.0538	24.1877			
s.d.	0.0612	0.0731	0.0891	0.1074	4.6723			
$T = 1000$	a_1	b_1	a_2	b_2	a_3	b_3	θ_1	θ_2
True	0	0.7	-0.5	0.8	-1.5	0.95	-1	1
Mean	0.0016	0.7017	-0.4965	0.8443	-1.5570	0.9479	-1.0193	1.0322
s.d.	0.1854	0.03643	0.2444	0.0302	0.1864	0.0221	0.0704	0.0633
$\lambda = 50$	γ_1	γ_2	γ_3	σ_h^2	λ			
True	1	-1.5	2.5	1	50			
Mean	0.9901	-1.5034	2.5138	1.0565	92.9030			
s.d.	0.0700	0.0927	0.1120	0.1288	16.0126			
$T = 2000$	a_1	b_1	a_2	b_2	a_3	b_3	θ_1	θ_2
True	0	0.7	-0.5	0.8	-1.5	0.95	-1	1
Mean	0.0014	0.7013	-0.4968	0.8376	-1.5331	0.9485	-1.0134	1.0308
s.d.	0.1574	0.0305	0.1783	0.0269	0.1455	0.0203	0.0542	0.0472
$\lambda = 50$	γ_1	γ_2	γ_3	σ_h^2	λ			
True	1	-1.5	2.5	1	50			
Mean	0.9884	-1.5036	2.4921	1.0550	96.1298			
s.d.	0.0566	0.0727	0.0883	0.1079	15.2207			

Note: a_i denotes $\sum_{j=1}^{i} \phi_{2j-1} = \phi_1 + \phi_3 + \cdots + \phi_{2i-1}$
and b_i denotes $\sum_{j=1}^{i} \phi_{2j} = \phi_2 + \phi_4 + \cdots + \phi_{2i}$.

Table 2.4: Summary of parameter estimation for regular SV-LSTAR-Cv and semiparametric SV-LSTAR-Cv based on 100 simulated stochastic volatility datasets with Student-t innovations. We report the posterior mean and standard deviation (s.d.) of the Bayesian estimates in the third simulation study. Note that the semiparametric SV-LSTAR-Cv model has no intercept in the volatility process (i.e., no "a_{s1}" or "a_{s2}") and some parameters have different interpretations in the two models and therefore are not directly comparable.

Parametric	a_{p1}	b_{p1}	a_{p2}	b_{p2}	θ_1	θ_2	γ_1	γ_2	γ_3	σ_h^2
True	0.0	0.8	0.0	0.95	-1	1	0.2	-0.3	0.5	1
Mean	0.110	0.747	0.079	0.849	-1.013	0.996	0.178	-0.297	0.494	1.914
s.d.	0.223	0.080	0.181	0.067	0.056	0.059	0.054	0.134	0.058	0.298
Semiparametric		b_{s1}		b_{s2}	θ_1	θ_2	γ_1	γ_2	γ_3	σ_h^2
True		0.8		0.95	-1	1	0.2	-0.3	0.5	1
Mean		0.787		0.919	-1.030	1.002	0.189	-0.269	0.492	1.334
s.d.		0.044		0.033	0.051	0.062	0.045	0.123	0.063	0.232

Note: $a_1 = \phi_1$, $a_2 = \phi_1 + \phi_3$, $b_1 = \phi_2$ and $b_2 = \phi_2 + \phi_4$.

assessed as these mixture parameters cannot be identified (Jensen and Maheu, 2010).

The properties of the Bayesian MCMC estimators are shown in Table 2.4. Specifically, when the innovations of the returns have heavy-tailed distributions, the assumption of Gaussian innovations for the returns makes the SV-LSTAR-Cv model overestimate the innovation variance of the volatility and significantly underestimate the slope of the autoregressive process of the volatility, whereas the bias of slope estimated by semiparametric SV-LSTAR-Cv is substantially less (Table 2.4). Thus, the flexible semiparametric prior can better characterize the variations of the volatility when the innovations of the volatility are significantly non-Gaussian.

2.4 Application and Comparison with SV-TAR: British Petroleum Stock Returns

We use an application of the proposed model to evaluate the performance of each model to real data. Also, a comparison between the proposed models and SV-TAR model shows the necessity of introducing a smooth transition into the regime-switching in some cases.

2.4.1 Application of Multi-Regime Smooth Transition Stochastic Volatility Models

We analyze the British Petroleum (BP) stock series listed on the New York Stock Exchange (NYSE). The variables of interest are the return of its daily closing prices and its trading volume, both of which are corrected by dividends and stock splits. The analysis period of the BP stock return series begins 1/2/2009, and ends on 9/30/2016. The first 1850 returns, which end on 5/10/2016, are used as the in-sample period. The one-step-ahead rolling prediction of returns starts from 5/11/2016, and ends on 9/30/2016. A one-step-ahead prediction forecasts the response variable at the next step (next day in our cases) based on the observed data. In our stock return cases, one-step-ahead rolling predictions are the predicted returns of successive days, and each of the predictions is based on the observed data until that day.

An exploratory analysis is first conducted to examine the behavior of the returns in different regimes. We utilize daily trading volume as one of the covariates in the multi-regime process and consider three regimes: high volume (top five daily trading volume over the previous 50 days), low volume (bottom five daily trading volume over the previous 50 days), and medium volume (otherwise). Thus, every day is classified

into one of the three regimes according to their rank of trading volumes among the previous 50 days (including itself). This classification was previously used in Gervais et al. (2001) to characterize the different behavior of returns in different regimes. Thus, we believe that these three regimes may be useful in our case.

In our analysis, we apply the classic SV-AR(1) (Kim et al., 1998), SV-LSTAR-Cv, SV-MRSTAR-C and semiparametric SV-LSTAR-Cv models to the BP data. Using SV-MRSTAR-C model, the sample chain of the smoothness parameter in the MCMC algorithm does not achieve convergence. As previously discussed, depending on the smoothness of transitioning between regimes, this parameter can be difficult to estimate. Therefore, the SV-MRSTAR-C model is removed from consideration and we compare the remaining models. We use the proposed three regimes and choose $\mathbf{z}_t = (l_t, u_{t-1,1}, u_{t-1,2})'$, where l_t is the logarithm of daily trading volume of the BP stock, and $u_{t,j} = 1$ if $u_t = j$, $u_{t,j} = 0$ otherwise for $j = 1, 2, 3$. We run the MCMC algorithm of $15,000$ draws and keep the last $10,000$ for inference in each MCMC sample chain. We assess the MCMC algorithm's convergence using both trace plots of the sample chains and the Gelman-Rubin diagnostic. We generate three sample chains for each of the parameters and assess their convergence. Visual inspection of the trace plots of the sample chains does not show any evidence of lack of convergence for any of the parameters, and the Gelman-Rubin statistic, $\widehat{R}$, is less than 1.02 for all parameters. The model comparison in Table 2.5 contains parameter estimates, log-likelihood, the deviance information criterion (DIC), and the Bayesian p-value. DIC proposed by Spiegelhalter et al. (2002) is a criterion used in Bayesian model selection, and models with smaller DIC values are preferred. The Bayesian p-value is used as a check of goodness-of-fit (Gelman et al., 1996). A Bayesian p-value close

to 0.5 indicates a good model fit, while a Bayesian p-value close to one or zero indicates overfitting or underfitting, respectively. In this case, we use the deviance as the discrepancy measure, i.e., $-2\log\, p(\mathbf{y}|\cdot)$. The Bayesian p-values of the SV-AR(1), the SV-LSTAR-Cv and the semiparametric SV-LSTAR-Cv are 0.413, 0.507 and 0.773, respectively. This suggests that both the SV-AR(1) and the SV-LSTAR-Cv do not exhibit lack of fit, while the semiparametric SV-LSTAR-Cv may not fit the data well. It is very likely that the innovations of the BP returns do not substantially deviate enough from a Gaussian distribution so that the semiparametric SV-LSTAR-Cv model can benefit from the flexible semiparametric prior. Moreover, the DIC is hard to obtain for the semiparametric SV-LSTAR-Cv model. We focus on the comparison of the SV-AR(1) and the SV-LSTAR-Cv models. The DICs of the SV-AR(1) and the SV-LSTAR-Cv are 5210.219 and 5257.157, respectively, which can be viewed as similar. Although, the SV-AR(1) has slightly lower DIC, the SV-LSTAR-Cv provides additional regime specific inferences.

Parameter estimates of the SV-LSTAR-Cv are shown in Table 2.6. Particularly, the posterior mean (95% credible interval) of the intercept parameters a_1 and a_2 are -0.2116 $(-0.4540, 0.0969)$ and -0.3772 $(-0.5811, -0.1299)$, respectively. The posterior mean (95% credible interval) of the slope parameters b_1 and b_2 are 0.6312 $(0.5161, 0.8742)$ and 0.9547 $(0.9171, 0.9811)$, respectively. Based on non-overlapping 95% CIs, these Bayesian estimates show a significant regime-switching in the slope of the volatility process. This result suggests that when the trading volume is relatively high over the previous 50 trading days, there is a higher persistence in the volatility level. When the trading volume is relatively low, higher persistence in the volatility level is low. That is, when the trading volume grows from a small value to a large

Table 2.5: Model Comparison of SV-AR(1), SV-LSTAR-Cv and semiparametric SV-LSTAR-Cv for BP returns in terms of log likelihood, DIC, and Bayesian p-value based on a deviance discrepency measure.

	SV-AR(1)	SV-LSTAR-Cv
Log likelihood	5151.599	5179.191
DIC	5210.219	5257.157
Bayesian p-value	0.413	0.507

value, the expected value of the persistence of the volatility moves from 0.6312 towards 0.9547.

The estimated volatility using SV-LSTAR-Cv is shown together with returns and daily trading volumes in Figure 2.2. From this plot, we can see that the returns, the trading volume, and the estimated volatility all show a significant peak between April 2010 and July 2010 and a minor peak around August 2011. The BP oil spill happened in April 2010; that is, the oil rig explosion occurred on the night of 4/20/2010. This event is noted in Figure 2.2 by the left vertical dashed line. Since this event, the BP stock returns exhibited high volatility through July 2010. The right dashed line in Figure 2.2 denotes 8/1/2011, around which the oil slick was reported for a second time and raised the volatility of the BP stock returns once again. In short, using daily trading volumes, the SV-LSTAR-Cv model successfully identifies regime-switching and provides inference on the volatility behavior under different regimes, which is something the SV-AR(1) model is unable to achieve. Finally, the estimated volatility corroborates historical events relevant to the BP stock returns.

Figure 2.2: The top plot shows the absolute values of returns (grey line) and the estimated volatility, $\exp(h_t/2)$ (black line) from 1/2/2009 to 5/9/2016. The bottom plot shows the log daily trading volume from 1/2/2009 to 5/9/2016. The left vertical dashed line denotes 4/20/2010, the day of the BP oil rig explosion. The right dashed line denotes 8/1/2011, the date around which the oil slicks were reported a second time.

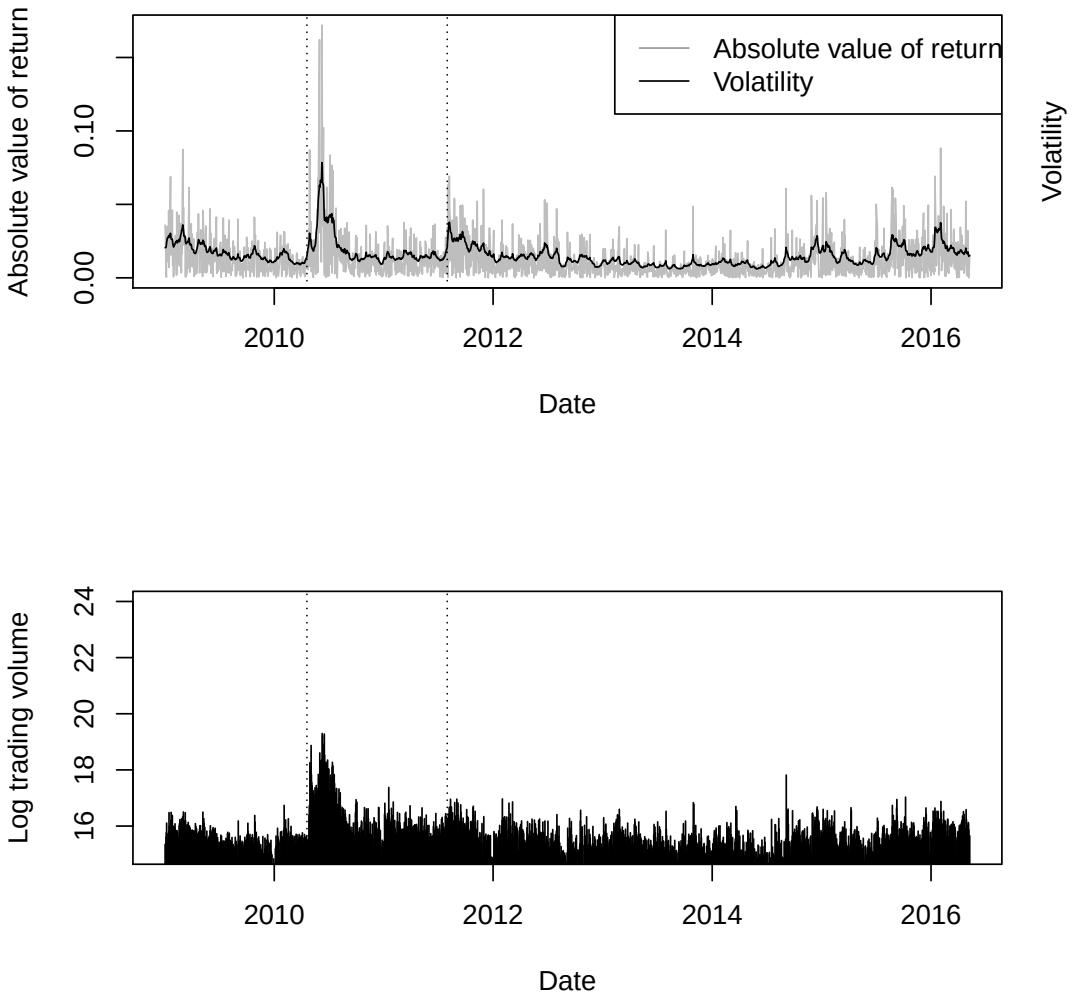


Table 2.6: Posterior mean and 95% credible interval of each parameter of SV-LSTAR-Cv model for BP returns. Recall that a_1 and a_2 are the lower limit and upper limit of autoregressive intercept, repectively and b_1 and b_2 are the lower limit and upper limit of autoregressive slope, repectively.

	a_1	b_1	a_2	b_2	θ_1	θ_2
Posterior mean	-0.2116	0.6312	-0.3772	0.9547	-1.8111	3.6494
Lower limit of 95% CI	-0.4540	0.5161	-0.5811	0.9171	-2.4987	3.9590
Upper limit of 95% CI	0.0969	0.8742	-0.1299	0.9811	-1.2843	4.0599
	γ_1	γ_2	γ_3	σ_t^2		
Posterior mean	0.6404	-2.0085	-1.0152	0.0561		
Lower limit of 95% CI	0.6232	-2.3790	-1.3178	0.0320		
Upper limit of 95% CI	0.6674	-1.7018	-0.7590	0.0803		

Note: $a_1 = \phi_1$, $a_2 = \phi_1 + \phi_3$, $b_1 = \phi_2$ and $b_2 = \phi_2 + \phi_4$.

2.4.2 Comparison with the SV-TAR model

To assess the necessity of smooth transition in the latent volatility process, we compare SV-LSTAR-Cv model with a reduced model without smooth transition model (SV-TAR). This model is defined as follows

$$y_t = \exp(h_t/2)\epsilon_t \tag{2.20}$$

$$h_t = a_{u_{t-1}} + b_{u_{t-1}} h_{t-1} + \sigma_h \eta_t, \tag{2.21}$$

where $\{a_t\}$ and $\{b_t\}$ are fixed intercepts and slopes of autoregressive process under different known regimes. The comparison is conducted using the same BP stock data as the application example so that the BP stock returns are y_t for $t = 1, \ldots, T$. We use the same regimes u_t, $t = 1, \ldots, T$, defined as in the application of SV-LSTAR-Cv. As such, the comparison is focused on the introduction of a smooth transition.

The inference of parameters in the SV-TAR is conducted through MCMC. Similar

to SV-LSTAR-Cv model, we run the MCMC algorithm of $15,000$ draws and keep the last $10,000$ for inference in each MCMC sample chain. We assess convergence of the MCMC using both trace plots of the sample chains and the Gelman-Rubin diagnostic. We generate three sample chains for each of the parameters and assess their convergence. Visual inspection of the trace plots of the sample chains does not indicate any lack of convergence for any of the parameters, and the Gelman-Rubin statistic, $\widehat{R}$, is less than 1.01 for all parameters. The parameter estimate of SV-TAR (Table 2.7) shows that the constant regime-wise intercepts and slopes do not explain a large proportion of the volatility variations. Instead, the variations are attributed to the innovation variance of the volatility. Large estimated innovation variance leads to an unstable prediction of returns. One-step-ahead rolling prediction MSE is used to assess the performance of each model. Among the one-step-ahead predictions of the 100 returns from time $t = 1001$ to 1100, many MCMC samples of the predicted of y_t are extremely large. So, the posterior mean of the samples may not be a good predictor. Thus, we use both posterior mean and median as predictors. The prediction MSE of SV-TAR is the worst (see Table 2.8).

From the above result, we can see the SV-TAR model does not fit this BP stock return data well. By considering a smooth transition in regime-switching, we can explain the variations of the volatility better and obtain a better prediction of the volatility.

Table 2.7: Posterior mean and s.d. of each parameter of SV-TAR model, BP stock return data. Each a, b are the constant regime-wise intercept and slope.

	a_1	b_1	a_2	b_2	a_3	b_3	σ_h^2
Posterior mean	-7.0746	0.1075	-7.2204	0.0863	-6.4383	0.0815	30.3238
S.D.	1.6949	0.1922	0.6998	0.0850	1.5053	0.2230	1.9281

Table 2.8: The one-step-ahead prediction MSE of each model, BP stock return data. This MSE is defined as $\sum(\tilde{y}_t^2 - y_t^2)^2$. SV-TAR (mean) uses the posterior mean as predictor and the SV-TAR (median) uses the posterior median as the predictor.

Model	MSE
SV-LSTAR-Cv	$3.0850e-08$
SV-TAR (mean)	$2.2725e+04$
SV-TAR (median)	$1.1213e+00$
Regular SV-AR(1)	$2.8755e-08$

2.5 Discussion

In this chapter, we proposed a new Bayesian smooth transition stochastic volatility model with ordered categorical variables. This model allows us to utilize other auxiliary variables to characterize the regime-switching behavior of the volatility. Additionally, a flexible Dirichlet mixture prior was introduced in the model for the case when the innovations show significant non-Gaussian behavior. Similar to the SV-AR(1) model, in our examples, the proposed multi-regime smooth transition model provided an adequate fit to the return data. In contrast, the multi-regime smooth transition models can provide an additional layer of inference on the volatility behavior under each regime. The comparison with SV-TAR shows that the smooth transition between different regimes of the volatility is necessary to explain the variations of the volatility.

For the different financial assets, the choice of categorical variables can vary, e.g., stock index (SP500, Dow Jones Industrial Average), crude oil prices, or any other information affecting the volatility. Moreover, the proposed smooth transition stochastic volatility models with ordered categorical variables applies to different types of time series with time-varying variance, e.g., environmental data and ecological data, among others.

2.6 Appendix: the MCMC Algorithm

For convenience, we make $\tilde{X}(\mathbf{s};\lambda,\boldsymbol{\theta})$ and $\mathbf{Z}$ denote the design matrices, such that

$$\tilde{X}(\mathbf{s};\lambda,\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{x}_1' & \mathbf{x}_1' M(s_1;\lambda,\theta_1) & \cdots & \mathbf{x}_1' M(s_1;\lambda,\theta_q) \\ \mathbf{x}_2' & \mathbf{x}_2' M(s_2;\lambda,\theta_1) & \cdots & \mathbf{x}_2' M(s_2;\lambda,\theta_q) \\ \vdots & \vdots & & \vdots \\ \mathbf{x}_T' & \mathbf{x}_T' M(s_T;\lambda,\theta_1) & \cdots & \mathbf{x}_T' M(s_T;\lambda,\theta_q) \end{bmatrix}, \quad Z = \begin{bmatrix} \mathbf{z}_1' \\ \mathbf{z}_2' \\ \vdots \\ \mathbf{z}_T' \end{bmatrix},$$

where and $\mathbf{x}_t' = (1, h_{t-1})'$, are respectively the design matrices of the autoregressive process of $\{h_t\}$ and the transition process $\{\nu_t\}$. The transition function $M(\cdot)$ will be specified in each of the following models.

2.6.1 Detailed Sampling Algorithms for SV-LSTAR-Cv

In the kth iteration of the MCMC simulation, do the following steps and move onto next iteration.

1) Generate $\{h_t^{(k)}\}$ from $p(h_1, h_2, \ldots, h_T|\cdot)$ using FFBS (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). Conditional on all the other parameters, $\{h_t\}$ are the

latent variable in the following dynamic linear model

$$
\begin{aligned}
y_t^* &= \log(y_t^2) = h_t + \epsilon_t^*, \quad \epsilon_t^* \sim \mathrm{N}(m_{r_t}, w_{r_t}^2), r_t \in \{1, \ldots, 10\} \text{ with known } \{m_{r_t}, w_{r_t}^2\} \\
h_t &= \left(\phi_1 + \frac{\phi_3}{1+\exp\{-\nu_{t-1}\}}\right) + \left(\phi_2 + \frac{\phi_4}{1+\exp\{-\nu_{t-1}\}}\right) h_{t-1} + \eta_t \\
&= C_{t-1} + D_{t-1} h_{t-1} + \eta_t \\
h_0 &\sim \mathrm{N}\left(0, \frac{\sigma^2}{1-(\phi_2+\phi_4)^2}\right),
\end{aligned}
\tag{2.22}
$$

2) Sample $\sigma_h^{2(k)}$ from $\sigma_h^2|\mathbf{h}, \mu, \boldsymbol{\phi} \sim \mathrm{IG}(a^*, b^*)$, where $a^* = T/2 + a_\sigma$ and $b^* = \frac{1}{2}\left((\mathbf{h} - \tilde{X}^{(k-1)}\boldsymbol{\phi}^{(k-1)})'(\mathbf{h} - \tilde{X}^{(k-1)}\boldsymbol{\phi}^{(k-1)}) + h_0/(1-\phi_2-\phi_4)^2\right) + b_\sigma$, where $\tilde{X}^{(k-1)}$ denotes $\tilde{X}^{(k-1)}(\boldsymbol{\nu}^{(k-1)}; \mathbf{1}, \mathbf{0})$.

3) Sample $\boldsymbol{\phi}^{(k)}$ from its full conditional distribution

$$N\left((\tilde{X}'^{(k-1)}\tilde{X}^{(k-1)} + \sigma_h^{2(k)}\delta^2 I)^{-1}\tilde{X}'^{(k-1)}\mathbf{h}, (\tilde{X}'^{(k-1)}\tilde{X}^{(k-1)} + \sigma_h^{2(k)}\delta^2 I)^{-1}\right).$$

4) Sample $\gamma^{(k)}$ from its full conditional distribution

$$\mathrm{N}\Big((Z'Z + \sigma_h^{2(k)}\delta^2 I)^{-1}Z'\nu^{(k-1)}, \sigma_h^{2(k)}(Z'Z + \sigma_h^{2(k)}\delta^2 I)^{-1}\Big).$$

5) Draw samples of θ_j sequentially from the following truncated normal proposal distribution for $j = 1, \ldots, q$, $J(\theta_j^{(k)}|\theta_{(j-1)}, \theta_j^{(k-1)}, \theta_{(j+1)}^{(k-1)}) = \mathrm{N}_{[\theta_{(j-1)}^{(k)}, \theta_{(j+1)}^{(k-1)}]}(\theta_j^{(k-1)}, \sigma_\theta^2)$, where σ_θ^2 is a tuning parameter and controls the acceptance rate and $\mathrm{N}_{[a,b]}(\mu, \sigma^2)$ denotes a truncated normal distribution with mean μ and variance σ^2 and a truncated

range $[a, b]$. Next, calculate

$$R_\theta = \prod_{j=1}^{q} \frac{\Phi\left(\frac{\theta_{(j+1)}^{(k-1)} - \theta_{(j)}^{(k-1)}}{\sigma_\theta}\right) - \Phi\left(\frac{\theta_{(j)}^{(k)} - \theta_{(j-1)}^{(k-1)}}{\sigma_\theta}\right)}{\Phi\left(\frac{\theta_{(j+1)}^{(k)} - \theta_{(j)}^{(k)}}{\sigma_\theta}\right) - \Phi\left(\frac{\theta_{(j)}^{(k-1)} - \theta_{(j-1)}^{(k)}}{\sigma_\theta}\right)}$$
$$\times \prod_{j=1}^{N} \sum_{j=1}^{N} \left\{ I_{u_t=j} \frac{\Phi\left(\theta_j^{(k)} - z_t'\boldsymbol{\gamma}^{(k)}\right) - \Phi\left(\theta_j^{(k)} - z_t'\boldsymbol{\gamma}^{(k)}\right)}{\Phi\left(\theta_j^{(k-1)} - z_t'\boldsymbol{\gamma}^{(k)}\right) - \Phi\left(\theta_j^{(k-1)} - z_t'\boldsymbol{\gamma}^{(k)}\right)} \right\}$$

and accept the new draw $\theta^{(k)}$ with probability $\min(R_\theta, 1)$.

6) Sample $\{\nu_t\}$ from its full conditional distribution

$$p(\nu_t|.) \propto \exp\left\{-\frac{1}{2}(\nu_t - z_t'\boldsymbol{\gamma})^2\right\}$$
$$\times \exp\left\{-\frac{1}{2}\left(h_{t+1} - C_t - D_t h_t\right)^2\right\} I_{(\theta_{j-1} \le \nu_t < \theta_j)},$$

where C_t and D_t are defined in (2.22). Using a Metropolis-Hastings step, we choose a truncated normal proposal distribution for each ν_t as follows

$$J(\nu_t^{(k)}|\nu_t^{(k-1)}) = \mathrm{N}(\nu_t^{(k-1),\sigma_\nu^2}) I_{[\theta_{j-1}^{(k)}, \theta_j^{(k)})}$$
$$= \frac{1}{\Phi\left(\frac{\theta_j^{(k)} - \nu_t^{(k-1)}}{\sigma_\nu}\right) - \Phi\left(\frac{\theta_{j-1}^{(k)} - \nu_t^{(k-1)}}{\sigma_\nu}\right)}$$
$$\times \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\left\{-\frac{1}{2}\left(\frac{\nu_t^{(k)} - \nu_t^{(k-1)}}{\sigma_\nu}\right)^2\right\} I_{(\theta_{j-1}^{(k)} \le \nu_t^{(k)} < \theta_j^{(k)})}.$$

Finally, we calculate the ratio

$$R_\nu = \min\left\{\frac{p(\nu_t^{(k)}|\cdot)J(\nu_t^{(k-1)}|\nu_t^{(k)})}{p(\nu_t^{(k-1)}|\cdot)J(\nu_t^{(k)}|\nu_t^{(k-1)})}, 1\right\}$$

and accept the new draw $\nu_t^{(k)}$ with probability R_ν.

7) Observe that $y_t^* - h_t = \epsilon_t^*$ with $\epsilon_t|r_t^* \sim \mathrm{N}(m_{r_t}, w_{r_t}^2)$, one can easily obtain the posterior probability $P(r_t = k|\cdot)$ for $k \in \{1, \ldots, 10\}$ and $t \in \{1, \ldots, T\}$ according to

$$P(r_t = k|\cdot) \propto P(r_t = k)\frac{1}{w_k}\exp\left\{-\frac{(\epsilon_t^* - m_k)^2}{2w_k^2}\right\},$$

where $P(r_t = k)$ denotes the mixture weights of the kth components (Omori et al., 2007).

2.6.2 Detailed Sampling Algorithms for SV-MRSTAR-C

In the kth iteration of MCMC, do the following steps and move onto next iteration.
1) Generate $\{h_t^{(k)}\}$ from $p(h_1, h_2, \ldots, h_T|\cdot)$ using FFBS analogous to SV-LSTAR-Cv.

2) Sample $\sigma_h^{2(k)}$ from $\sigma_h^2|\mathbf{h}, \mu, \boldsymbol{\phi} \sim \mathrm{IG}(a^*, b^*)$, where $a^* = T/2 + a_\sigma$ and
$b^* = \frac{1}{2}\left((\mathbf{h} - \tilde{X}^{(k-1)}\boldsymbol{\phi}^{(k-1)})'(\mathbf{h} - \tilde{X}^{(k-1)}\boldsymbol{\phi}^{(k-1)}) + h_0/(1-\phi_2-\phi_4)^2\right) + b_\sigma$, where $\tilde{X}^{(k-1)}$ denotes $\tilde{X}^{(k-1)}(\mathbf{z}'_{t-1}\boldsymbol{\gamma}^{(k-1)}; \lambda, \boldsymbol{\theta})$.

3) Sample $\phi^{(k)}$ from its full conditional distribution
$N\{(\tilde{X}'^{(k-1)}\tilde{X}^{(k-1)} + \sigma_h^{2(k)}\delta^2 I)^{-1}\tilde{X}'^{(k-1)}\mathbf{h}, (\tilde{X}'^{(k-1)}\tilde{X}^{(k-1)} + \sigma_h^{2(k)}\delta^2 I)^{-1}\}$.

4) Sample $\gamma^{(k)}$ from its full conditional distribution

$N\{(Z'Z+\sigma_h^{2(k)}\delta^2 I)^{-1}Z'\nu^{(k-1)}, \sigma_h^{2(k)}(Z'Z+\sigma_h^{2(k)}\delta^2 I)^{-1}\}$.

5) Draw samples of θ_j sequentially from the following truncated normal proposal distribution for $j=1,\ldots,q$, $J(\theta_j^{(k)}|\theta_{(j-1)},\theta_j^{(k-1)},\theta_{(j+1)}^{(k-1)}) = \mathrm{N}_{[\theta_{(j-1)}^{(k)},\theta_{(j+1)}^{(k-1)}]}(\theta_j^{(k-1)},\sigma_\theta^2)$. Then, use a normal proposal distribution for the logarithm of λ such that $J(\log(\lambda^{(k)})|\log(\lambda^{(k-1)})) = \mathrm{N}(\log(\lambda^{(k-1)}),\sigma_\lambda^2)$ and a normal proposal distribution for γ such that $J(\gamma^{(k)}|\gamma^{(k-1)}) = \mathrm{N}(\gamma^{(k-1)},\Sigma_\gamma)$, where σ_θ^2, σ_λ^2 and Σ_γ are tuning parameters and control the acceptance rate. Finally, calculate

$$
\begin{aligned}
R_\theta = & \prod_{j=1}^{q} \frac{\Phi\left(\frac{\theta_{(j+1)}^{(k-1)}-\theta_{(j)}^{(k-1)}}{\sigma_\theta}\right)-\Phi\left(\frac{\theta_{(j)}^{(k)}-\theta_{(j-1)}^{(k-1)}}{\sigma_\theta}\right)}{\Phi\left(\frac{\theta_{(j+1)}^{(k)}-\theta_{(j)}^{(k)}}{\sigma_\theta}\right)-\Phi\left(\frac{\theta_{(j)}^{(k-1)}-\theta_{(j-1)}^{(k)}}{\sigma_\theta}\right)} \\
& \times \prod_{t=1}^{T}\sum_{j=1}^{q}\left\{ I_{u_t=j}\frac{\Phi(\theta_j^{(k)}-z_t'\gamma^{(k)})-\Phi(\theta_j^{(k)}-z_t'\gamma^{(k)})}{\Phi(\theta_j^{(k-1)}-z_t'\gamma^{(k)})-\Phi(\theta_j^{(k-1)}-z_t'\gamma^{(k)})}\right\} \\
& \times \frac{\exp\left(-\frac{1}{2\sigma_h^{2(k)}}(\mathbf{h}-\tilde{X}^{(k)}\phi^{(k)})'(\mathbf{h}-\tilde{X}^{(k)}\phi^{(k)})\right)}{\exp\left(-\frac{1}{2\sigma_h^{2(k)}}(\mathbf{h}-\tilde{X}^{(k-1)}\phi^{(k)})'(\mathbf{h}-\tilde{X}^{(k-1)}\phi^{(k)})\right)} \times \frac{\mathrm{IG}(\lambda^{(k)};A_\lambda,B_\lambda)}{\mathrm{IG}(\lambda^{(k-1)};A_\lambda,B_\lambda)}
\end{aligned}
$$

and accept the new draw $(\theta^{(k)},\lambda^{(k)},\gamma^{(k)})$ with probability $\min(R_\theta,1)$. In cases where the acceptance rate is low, the sampling can be speeded up by iteratively drawing samples of $(\theta^{(k)},\lambda^{(k)},\gamma^{(k)})$ for multiple times and taking the sample of the last iteration. In practice, we find 20 typically yields satisfactory acceptance rates.

6) Observing that $y_t^* - h_t = \epsilon_t^*$ with $\epsilon_t^*|r_t \sim \mathrm{N}(m_{r_t},w_{r_t}^2)$, we obtain the posterior

probabilities $P(r_t = k|\cdot)$ for $k \in \{1, \ldots, 10\}$ and $t \in \{1, \ldots, T\}$ according to

$$P(r_t = k|\cdot) \propto P(r_t = k)\frac{1}{w_k}\exp\left\{-\frac{(\epsilon_t^* - m_k)^2}{2w_k^2}\right\},$$

where $P(r_t = k)$ denotes the mixture weights of the kth components (Omori et al., 2007) shown in Table 2.1.

2.6.3 Detailed Sampling Algorithms for Semiparametric SV-LSTAR-Cv

In the kth iteration of MCMC, do the following steps and move onto next iteration.
1) Generate $\sigma_h^{2(k)}$ from $\text{IG}\Big(T/2 + a_\sigma, (\mathbf{h} - \tilde{X}^{(k-1)}\phi^{(k-1)})'(\mathbf{h} - \tilde{X}^{(k-1)}\phi^{(k-1)})/2 + b_\sigma\Big)$, where $\tilde{X}^{(k-1)}$ denotes $\tilde{X}^{(k-1)}(\nu^{(k-1)}; \mathbf{1}, \mathbf{0})$ and $\mathbf{x}_t' = (h_{t-1})'$.

2) Sample $\boldsymbol{\phi}^{(k)}$ from its full conditional distribution
$N\left((\tilde{X}'^{(k-1)}\tilde{X}^{(k-1)} + \sigma^{2(k)}\delta^2 I)^{-1}\tilde{X}'^{(k-1)}\mathbf{h}, (\tilde{X}'^{(k-1)}\tilde{X}^{(k-1)} + \sigma^{2(k)}\delta^2 I)^{-1}\right)$.

3) Sample $\gamma^{(k)}$ from its full conditional distribution
$N\left((Z'Z + \sigma^{2(k)}\delta^2 I)^{-1}Z'\nu^{(k-1)}, \sigma^{2(k)}(Z'Z + \sigma^{2(k)}\delta^2 I)^{-1}\right)$.

4) Draw samples of θ_j sequentially from the following truncated normal proposal distribution for $j = 1, \ldots, q$, $J(\theta_j^{(k)}|\theta_{(j-1)}, \theta_j^{(k-1)}, \theta_{(j+1)}^{(k-1)}) = \text{N}_{(\theta_{(j-1)}^{(k)}, \theta_{(j+1)}^{(k-1)})}(\theta_j^{(k-1)}, \sigma_\theta^2)$,

where σ_θ^2 is a tuning parameter and controls the acceptance rate. Next, calculate

$$R_\theta = \prod_{j=1}^{q} \frac{\Phi\left(\frac{\theta_{(j+1)}^{(k-1)} - \theta_{(j)}^{(k-1)}}{\sigma_\theta}\right) - \Phi\left(\frac{\theta_{(j)}^{(k)} - \theta_{(j-1)}^{(k-1)}}{\sigma_\theta}\right)}{\Phi\left(\frac{\theta_{(j+1)}^{(k)} - \theta_{(j)}^{(k)}}{\sigma_\theta}\right) - \Phi\left(\frac{\theta_{(j)}^{(k-1)} - \theta_{(j-1)}^{(k)}}{\sigma_\theta}\right)}$$
$$\times \prod_{t=1}^{T} \sum_{j=1}^{q} \left\{ I_{u_t=j} \frac{\Phi\left(\theta_j^{(k)} - z_t'\boldsymbol{\gamma}^{(k)}\right) - \Phi\left(\theta_j^{(k)} - z_t'\boldsymbol{\gamma}^{(k)}\right)}{\Phi\left(\theta_j^{(k-1)} - z_t'\boldsymbol{\gamma}^{(k)}\right) - \Phi\left(\theta_j^{(k-1)} - z_t'\boldsymbol{\gamma}^{(k)}\right)} \right\}$$

and accept the new draw $\theta^{(k)}$ with probability $\min(R_\theta, 1)$.

5) Given that the full conditional distribution of latent variable ν_t is

$$p(\nu_t|.) \propto \exp\left\{-\frac{1}{2}(\nu_t - z_t'\boldsymbol{\gamma})^2\right\} \exp\left\{-\frac{1}{2}(h_{t+1} - a_t - b_t h_t)^2\right\} I_{(\theta_{j-1} \le \nu_t < \theta_j)},$$

it is sampled through a Metropolis-Hastings (MH) step. We use truncated normal proposal distribution for each ν_t as follows

$$\begin{aligned} J(\nu_t^{(k)}|\nu_t^{(k-1)}) =& \mathrm{N}(\nu_t^{(k-1),\sigma_\nu^2}) I_{[\theta_{j-1}^{(k)}, \theta_j^{(k)})} \\ =& \frac{1}{\Phi\left(\frac{\theta_j^{(k)} - \nu_t^{(k-1)}}{\sigma_\nu}\right) - \Phi\left(\frac{\theta_{j-1}^{(k)} - \nu_t^{(k-1)}}{\sigma_\nu}\right)} \\ & \times \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\left\{-\frac{1}{2}\left(\frac{\nu_t^{(k)} - \nu_t^{(k-1)}}{\sigma_\nu}\right)^2\right\} I_{(\theta_{j-1}^{(k)} \le \nu_t^{(k)} < \theta_j^{(k)})}. \end{aligned}$$

Finally, we calculate the ratio

$$R_\nu = \min\left\{ \frac{p(\nu_t^{(k)}|\cdot) J(\nu_t^{(k-1)}|\nu_t^{(k)})}{p(\nu_t^{(k-1)}|\cdot) J(\nu_t^{(k)}|\nu_t^{(k-1)})}, 1 \right\},$$

and accept the new draw $\nu_t^{(k)}$ with probability R_ν.

6) Sampling h_t by random length blocks using Metropolis-Hastings algorithm as mentioned in Section 2.2.2. Let $h_{(t,\tau)} = (h_t, h_{t+1}, \ldots, , h_\tau)', 1 \leq t \leq \tau \leq T$ and its length $l_t = \tau - t + 1$ is one plus a random draw from a Poisson distribution with the hyperparameter $\kappa_h = 3$; i.e., $E[l_t] = 4$ (κ_h was selected by Jensen and Maheu (2010) to minimize the numerical inefficiency of model parameters). Following the approximation of Jensen and Maheu (2010) the proposal density of $h_{(t,\tau)}$ is obtained as

$$f_{St}(h_{(t,\tau)}|, \zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu) \propto \left[1 + h_{(t,\tau)} - \zeta'_{(t,\tau)}\Sigma^{-1}_{(t,\tau)}(h_{(t,\tau)} - \zeta_{(t,\tau)})/\nu\right]^{-(l_t+\nu)/2},$$

where $f_{St}(h_{(t,\tau)}|, \zeta_{(t,\tau)}\Sigma_{(t,\tau)}, \nu)$ is the density of a l_t-variate Student-t distribution with mean $\zeta_{(t,\tau)} = m_{(t,\tau)} - 0.5\Sigma_{(t,\tau)}(\mathbf{1}_{l_t} - D_{(t,\tau)}\tilde{y}^2_{(t,\tau)})$, covariance $\Sigma_{(t,\tau)}\nu/(\nu-2)$, and ν degrees of freedom. For the endpoints h_1 and h_n, we generate them individually using the same proposal density. Given the previous draw of $h_{(t,\tau)}$, the candidate draw from $\hat{h}_{(t,\tau)} \sim St(\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)$, will be accepted with the probability as given below

$$\min\left\{\frac{f(y_{(t,\tau)}|\phi_{(t,\tau)}, \hat{h}_{(t,\tau)})\pi(\hat{h}_{(t,\tau)}|h_{t-1}, h_{t+1}, \psi)f_{St}(h_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)}{f(y_{(t,\tau)}|\phi_{(t,\tau)}, h_{(t,\tau)})\pi(h_{(t,\tau)}|h_{t-1}, h_{t+1}, \psi)f_{St}(\hat{h}_{(t,\tau)}|\zeta_{(t,\tau)}, \Sigma_{(t,\tau)}, \nu)}, 1\right\},$$

where $f(y_{(t,\tau)}|\phi_{(t,\tau)}, h_{(t,\tau)}) = \prod_{j=t}^{\tau} f_N(y_j|\eta_j, \omega_j^{-2}\exp\{h_j\})$ and $\pi(h_{(t,\tau)}|h_{t-1}, h_{t+1}, \phi) = \prod_{j=t}^{\tau+1}\exp\left\{-\frac{(h_j - \phi_1 h_{j-1}G(s_{j-1};\nu_{j-1},\theta_1) - \cdots - \phi_q G(s_{j-1};\nu_{j-1},\theta_q))^2}{2\sigma_\nu^2}\right\}$.

7) Sampling θ_ω and s_ω with a two-step procedure:

Step 1. Sample s_ω and k_ω by drawing ω_t^2 for $t = 1, \ldots, T$ from

$$\omega_t | y_t^*, \omega^{(t)}, s_\omega^{(t)} \sim c \frac{\alpha}{\alpha + T - 1} g(y_t) G(d\omega_t | y_t^*) + \frac{c}{\alpha + T - 1} \sum_{j-1}^{\kappa^{(t)}} n_j^{(t)} f(y_t | \mu, \exp\{h_t\} \omega_t^{-2}) \delta_{\omega_j^2}(d\omega_t^2),$$

where $g(y_t) = f_{St}(y_t | \mu, \exp\{h_t\}\nu_0/s_0, \nu_0)$, and $G(d\omega_t | y_t)$ is the distribution $\Gamma(\bar{\nu}, \bar{s}_t)$ with $\bar{\nu} = \nu_0 + 1/2$ and $\bar{s}_t = s_0 + (y_t - \mu)^2/(2\exp\{h_t\})$, c is the integrating constant and $n_j^{(t)}$ is the number of times that the jth cluster occurs at tth iteration.

Step 2. Given the s_ω and k_ω from Step 1, sample ω_j^2 for $j = 1, \ldots, k_\omega$, from

$$\omega_j^2 | \{y_t : s_{\omega t} = j\} \propto \prod_{t: s_{\omega t} = j} f_N(y_t | \mu, \exp\{-h_t/2\}, \omega_j^{-2}) K_0(d\omega_j),$$

which is the $\Gamma(\bar{\nu}_j, \bar{s}_j)$ distribution with $\bar{\nu}_j = \nu_0 + n_j/2$ and $\bar{s}_j = s_0 + \sum_{t: s_{\omega t} = j} (y_t - \mu)^2/(2\exp\{h_t\})$.

8) Sample α from $\pi(\alpha | k_\omega)$ by first sampling ξ, s.t., $\xi | \alpha, k_\omega \sim \text{Beta}(\alpha + 1, T)$, and next sampling α from the mixture distribution, s.t., $\alpha | \xi, k_\omega \sim \pi_\xi \Gamma(a + k_\omega, b - \log\xi) + (1 - \pi_\xi)\Gamma(a + k_\omega - 1, b - \log\xi)$, where $\pi_\xi/(1 - \pi_\xi) = (a + k_\omega - 1)/[n(b - \log\xi)]$.

9) Sampling μ from $\mathrm{N}(\bar{\mu}, \bar{\tau})$ where

$$\bar{\mu} = \frac{m/\tau + \sum_t y_t \exp\{-h_t\}\omega_t^2}{1/\tau + \sum_t \exp\{-h_t\}\omega_t^2},$$

$$\bar{\tau} = \left(1/\tau + \sum_t \exp\{-h_t\}\omega_t^2\right)^{-1}.$$

Chapter 3

Bayesian Circular Lattice Filters for Multivariate Time Series Analysis

3.1 Introduction

Multivariate time series data are measured and recorded for inquiries of interest in subject-matter disciplines such as biology, ecology, economics, finance, and medicine. For example, multivariate nonstationary time series models work well on correlated economic indicators not only to use time-varying parameters to evaluate the effect of policy changes and the resulting private sector behavior changes but also to understand the effect of the changes in policy on other factors of the economy (Huerta and Lopes, 2000; Primiceri, 2005; Nakajima et al., 2011; Hunter et al., 2017). Another example is that multi-channel electroencephalography (EEG) data are analyzed through their time-frequency representation to reveal how the neuronal activity in one area of

the human brain may influence another (Ombao et al., 2005; Zhao and Prado, 2020). In the big data era, the number of series of multivariate time series data and their length have been increasing due to technological advances. Such an expansion is also increasing data complexity. Thus, developing efficient methods that scale to such an enormous amount of time series data is imperative.

Parametric and nonparametric approaches for analyzing multivariate time series have been developed to reveal features both between multiple time series and within a single time series itself. Some parametric methods, e.g., vector autoregressive (VAR) models, are designed for stationary time series. Time-varying vector autoregressive (TV-VAR) models and many other parametric models (Ombao et al., 2005; Kowal et al., 2017), deal with multivariate nonstationary time series. There are also nonparametric methods, e.g., multivariate time-dependent spectral analysis (Guo and Dai, 2006), multivariate polynomial regression (Masry, 1996; Fan and Yao, 2008).

Parametric models of nonstationary multivariate time series play an increasingly important role when modern techniques make high-dimensional data available for analysis. They have several advantages: (1) easy to make forecasts, (2) straightforward assessment of uncertainty, (3) concise descriptions of the underlying model scheme. The TV-VAR model is arguably the most widely used model among the parametric methods. In the time domain, such models have been developed and applied to correlated economic and financial data variables (Primiceri, 2005; Del Negro and Primiceri, 2015; Nakajima et al., 2011; Nakajima and West, 2013). In contrast, the frequency domain helps to reveal features of the spectrum of each time series as well as the coherence between them. For example, the applications of such approaches to multi-channel EEG data can help us understand the brain's connective activities

(Ombao et al., 2005; Zhao and Prado, 2020), human speech signals (Ombao et al., 2001) and multi-location climate data (Zhao and Prado, 2020) by estimating their spectrum and coherence.

According to the assumptions often imposed on the innovation variance of TV-VAR models, some of these methods belong to the class of stochastic volatility models (Shephard, 2005), e.g., Primiceri (2005); Del Negro and Primiceri (2015); Nakajima et al. (2011). In the Bayesian settings, these methods typically require Markov chain Monte Carlo (MCMC) methods for estimation and, therefore, are computationally expensive. Besides the class of models, some other methods for TV-VAR models assume the innovation variance follows a random walk process. Gersch and Stone (1995) proposed a frequentist method to estimate the (univariate) time-varying autoregressive (TV-AR) model by imposing smoothing priors on the time-varying coefficients. These methods used a two-stage estimation approach and assumed a constant innovation covariance. The multivariate dynamic linear models (MDLMs) by West and Harrison (1997) imposed random walk priors on the parameters. This method is less stable and has a high computational cost. Based on MDLMs, Zhao and Prado (2020) proposed multivariate Bayesian Lattice Filters (BLFs), that extend the approach of Yang et al. (2016) for TV-VAR models and assumed a constant innovation covariance to reduce the computation cost. The constant innovation covariance was estimated by an approximate estimator (Triantafyllopoulos, 2007). Although the use of BLFs dramatically reduces the computation cost, the computation time still increases exponentially with the increase of the dimension of the data. Moreover, the assumption of constant innovation covariance may not be reasonable for many types of data. Several methods (e.g., Primiceri (2005); Del Negro and Primiceri (2015); Nakajima et al.

(2011)) successfully assumed and estimated a time-varying innovation covariance for the multivariate economic index data and wind data.

To overcome the drawbacks of the high computation cost and the constant innovation covariance assumption, we propose an approach that uses a "one channel at-a-time" scheme (Pagano, 1978) to transform the multivariate time series model into a periodic univariate process and estimates it using a circular lattice filter approach. This "one channel at-a-time" algorithm efficiently reduces the computation cost and makes parallel computing possible for high-dimensional data. Moreover, this approach uses dynamic linear models (DLMs) to facilitate the estimation in each stage of the lattice structure, allowing both the coefficients and the innovation variances to vary with time. By modeling the time-varying innovation covariance, the resulting models are more broadly applicable.

The remainder of the chapter is organized as follows. First, we introduce the Bayesian circular lattice filters (BCLFs) in Section 3.2 and evaluate the method via simulation studies in Section 3.3. In Section 3.4, we apply the method to several real datasets. Finally, Section 3.5 contains conclusion and discussion. We include the detailed procedures and algorithms for parameter estimation, model selection, and forecasting in the Appendix.

3.2 Methodology

In this chapter, we propose BCLFs to estimate the order and parameters of a TV-VAR model. This approach is computationally efficient for high-dimensional multivariate time series data due to its "one channel at-a-time" scheme and Bayesian lattice struc-

ture. To introduce this approach, we begin with a general description of TV-VAR models.

3.2.1 Time-varying Vector Autoregressive Model and Bayesian Inference

Suppose an observed K-variate time series is defined as $\mathbf{x}_t = (x_{1,t}, \ldots, x_{K,t})'$ for $t = 1, \ldots, T$. For this time series, we define the time-varying vector autoregressive model with order P (TV-VAR(P)) as follows

$$\mathbf{x}_t = \sum_{p=1}^{P} \mathbf{\Phi}_{p,t} \mathbf{x}_{t-p} + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(0, \mathbf{\Sigma}_t), \tag{3.1}$$

where $\mathbf{\Phi}_{p,t}$ and $\mathbf{u}_t$ are the $K \times K$ autoregressive coefficient matrix for time lag p and the $K \times 1$ innovation vector at time t, respectively. The innovations, $\mathbf{u}_t$, at time t are assumed to follow the multivariate Gaussian distribution with zero-mean and time-dependent covariance matrix, $\mathbf{\Sigma}_t$. Consequently, the TV-VAR model exemplifies nonstationary vector autoregressive models by the coefficient and innovation covariance matrices varying over time.

Among many Bayesian inference approaches to TV-AR models, $\mathbf{\Phi}_{p,t}$ are often assumed to follow a random walk process, but the assumptions on $\mathbf{u}_t$ vary. The stochastic volatility class of TV-VAR models (Primiceri, 2005; Del Negro and Primiceri, 2015; Nakajima et al., 2011) assume the logarithm of the innovation covariance follows a random walk or an autoregressive process. As such, these models can capture abruptly changing innovation variance. Alternatively, MDLMs provide full posterior inference for TV-VAR models and assume the innovation covariance follows a ran-

dom walk process. This assumption allows us to use the idea of BLFs to improve the efficiency of the estimation. The inference of MDLMs involves potentially high-dimensional matrix computation. Thus, the usage is limited to those time series of a small number of series and TV-VAR models of low orders. Zhao and Prado (2020) proposed a multivariate BLF in which the computational cost increases linearly with the model order. Nevertheless, this approach is still computationally expensive because the computational cost still increases exponentially with the number of series. To completely address this issue, we propose a way to avoid cumbersome matrix calculations in the Bayesian BLFs.

3.2.2 Time-varying Periodic Time Series

Pagano (1978) proposed a "one-channel-at-a-time" modeling approach to break the vector autoregressive (VAR) model into scalar periodic AR processes. This approach makes the computational cost increase linearly rather than exponentially with the model order. Using this idea, for a K-variate TV-VAR model as (3.1), we consider an LDL decomposition (also called modified Cholesky decomposition), such that, $\mathbf{\Sigma}_t = L_t W_t L_t'$, where L_t is a lower unit triangular matrix and W_t is a diagonal matrix. With both sides of (3.1) premultiplied by the inverse of L_t, the covariance matrix reduces to a diagonal matrix W_t, such that $W_t = \text{diag}(\sigma_{1,t}^2, \ldots, \sigma_{K,t}^2)$, we obtain the instantaneous response-orthogonal innovations model (Gersch and Stone, 1994, 1995; Kitagawa and Gersch, 1996)

$$L_t^{-1}\mathbf{x}_t = \sum_{p=1}^{P} A_{p,t}\mathbf{x}_{t-l} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(0, W_t), \tag{3.2}$$

with $\boldsymbol{\varepsilon}_t$ is transformed from $\mathbf{u}_t$ and $A_{p,t}$ is transformed from $\boldsymbol{\Phi}_{p,t}$, such that $A_{p,t} = L_t^{-1}\boldsymbol{\Phi}_{p,t}$. (3.2) is the model we actually estimate. The parameter estimation of (3.1) is achieved through a transformation from that of (3.2). By interlacing the multiple series in this instantaneous response-orthogonal innovations model, we obtain an equivalent periodic TV-AR model, in which it is defined that $y_{k+Kt} = x_{k,t+1}$, $k = 1, \ldots, K$, $t = 0, \ldots, T-1$. This periodic TV-AR model is given as

$$y_{k+Kt} = \sum_{m=1}^{M_k} a_{m,k+Kt} y_{k+Kt-m} + \epsilon_{k+Kt}, \quad k = 1, \ldots, K, \quad t = 1, \ldots, T, \tag{3.3}$$

where ϵ_{k+Kt} is the innovation with variance $\sigma^2_{k+Kt} = \sigma^2_{k,t}$ and $M_k = KP + k - 1$ is the order of the kth series , y_{k+Kt} for $t = 0, \ldots, T-1$. As such, a TV-VAR model given by (3.1) is rewritten as a periodic TV-AR model as in (3.3). As an example, for $K = 3$, $P = 2$, (3.3) can be written in matrix form as

$$\begin{aligned}
\begin{bmatrix} 1 & 0 & 0 \\ -a_{1,3t+2} & 1 & 0 \\ -a_{2,3t+3} & -a_{1,3t+3} & 1 \end{bmatrix} \begin{bmatrix} y_{3t+1} \\ y_{3t+2} \\ y_{3t+3} \end{bmatrix} &= \begin{bmatrix} a_{3,3t+1} & a_{2,3t+1} & a_{1,3t+1} \\ a_{4,3t+2} & a_{3,3t+2} & a_{2,3t+2} \\ a_{5,3t+3} & a_{4,3t+3} & a_{3,3t+3} \end{bmatrix} \begin{bmatrix} y_{3t-2} \\ y_{3t-1} \\ y_{3t} \end{bmatrix} + \\
&\begin{bmatrix} a_{6,3t+1} & a_{5,3t+1} & a_{4,3t+1} \\ a_{7,3t+2} & a_{6,3t+2} & a_{5,3t+2} \\ a_{8,3t+3} & a_{7,3t+3} & a_{6,3t+3} \end{bmatrix} \begin{bmatrix} y_{3t-5} \\ y_{3t-4} \\ y_{3t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_{3t+1} \\ \epsilon_{3t+2} \\ \epsilon_{3t+3} \end{bmatrix}, \quad t = 1, \ldots, T.
\end{aligned} \tag{3.4}$$

The parameters in (3.1) and the parameters in the corresponding periodic TV-AR

model as in (3.3) have the relationship:

$$\mathbf{\Phi}_{p,t} = -L_t A_{p,t} \tag{3.5}$$

$$\Sigma_t = L_t W_t L_t'. \tag{3.6}$$

This relationship allows us to rewrite a TV-VAR model as a periodic TV-AR model, and vice versa.

3.2.3 Circular Lattice Filter

For multivariate autoregressive models, Sakai (1982) developed a circular lattice structure to estimate the autoregressive coefficients iteratively. This approach allows us to estimate the autoregressive coefficients one series at a time and, therefore, minimizes the need to calculate big matrices. This strategy makes the computational cost increase linearly, rather than exponentially, with the model order. Gersch and Stone (1995) applied this circular lattice to the time-varying multivariate AR model with smoothness prior (Kitagawa and Gersch, 1996) on the autoregressive coefficients. This method uses a two-stage estimation approach for the coefficients and innovation covariance and only allows coefficients to be time-varying. The assumption of constant innovation covariance limits application for many types of data.

To better model multivariate nonstationary time series data, we propose a BCLF model, which can handle both time-varying coefficients and time-varying innovation covariance. According to the Levinson–Durbin algorithm, there exists a unique correspondence between the PARCOR coefficients and the AR coefficients, (Shumway and Stoffer, 2006; Kitagawa, 2010; Yang et al., 2016). This lattice structure provides

a direct way of using the observed time series to estimate the PARCOR coefficients, which are associated with particular AR model (see Hayes (1996) and the Supplementary Appendix Yang et al. (2016)). Here, we modify the above lattice filter for TV-VAR models. By interlacing the K-variate time series $\{x_t\}$ into a scalar series $\{y_t\}$ as in (3.3), a TV-VAR(P) is transformed into a time-varying periodic AR model, which is equivalent to K independent TV-AR models of order M_k ($M_k = KP + k - 1$) (Gersch and Stone, 1994, 1995; Kitagawa and Gersch, 1996). Each season is fitted by a TV-AR model. Efficient estimation of a TV-AR model can be conducted by associating the PARCOR coefficients with the AR coefficients through the circular lattice structure. Before going through the details of the circular lattice structure, recall that $y_{k+(t-1)K} = x_{k,t}$. We denote $f^{(M_k)}_{k+(t-1)K}$ and $b^{(M_k)}_{k+(t-1)K}$ to be the prediction error of kth series at time t for the forward and backward time-varying AR(M_k) models, respectively, where

$$f^{(M_k)}_{k+(t-1)K} = y_{k+(t-1)K} - \sum_{m=1}^{M_k} a^{(M_k)}_{m,k+(t-1)K} y_{k+(t-1)K-m}$$
$$b^{(M_k)}_{k+(t-1)K} = y_{k+(t-1)K} - \sum_{m=1}^{M_k} d^{(M_k)}_{m,k+(t-1)K} y_{k+(t-1)K+m}$$

and $a^{(M_k)}_{m,k+(t-1)K}$ and $b^{(M_k)}_{m,k+(t-1)K}$ are the forward and backward autoregressive coefficients of the corresponding time-varying periodic AR(M_k) models. Then, in the mth stage of the lattice filter for $m = 1, \ldots, M_k$, the forward and the backward coefficients

and the forward and the backward prediction errors have the relationship

$$f^{(m)}_{k+(t-1)K} = f^{(m-1)}_{k+(t-1)K} - \alpha^{(m)}_{f,m,k+(t-1)K} b^{(m-1)}_{k-1+tK}, \tag{3.7}$$

$$b^{(m)}_{k+(t-1)K} = b^{(m-1)}_{k-1+tK} - \alpha^{(m)}_{b,m,k+(t-1)K} f^{(m-1)}_{k+(t-1)K}, \tag{3.8}$$

with the initial condition, $f^{(0)}_{k+(t-1)K} = b^{(0)}_{k+(t-1)K} = y_{k+(t-1)K}$, and where $\alpha^{(m)}_{m,k+(t-1)K}$ and $\beta^{(m)}_{b,m,k+(t-1)K}$ are the lag m forward and backward PARCOR coefficients of the kth series at time t, respectively. At the mth stage of the lattice structure, we fit a time-varying AR(1) model for the kth series and there are m time-varying autoregressive coefficients in this model. The kth series jth lag forward and backward autoregressive coefficients at time t, $a^{(m)}_{j,k+(t-1)K}$ and $d^{(m)}_{j,k+(t-1)K}$, can be obtained according to the following equations

$$\begin{aligned} a^{(m)}_{j,k+(t-1)K} &= a^{(m-1)}_{j,k+(t-1)K} - a^{(m)}_{m,k+(t-1)K} d^{(m-1)}_{m-j,k+(t-1)K}, \\ d^{(m)}_{j,k+(t-1)K} &= d^{(m-1)}_{j,k+(t-1)K} - d^{(m)}_{m,k+(t-1)K} a^{(m-1)}_{m-j,k+(t-1)K}, \end{aligned} \tag{3.9}$$

with $j = 1, \ldots, m-1$, $a^{(m)}_{m,k+(t-1)K} = \alpha^{(m)}_{m,k+(t-1)K}$ and $d^{(m)}_{m,k+(t-1)K} = \alpha^{(m)}_{b,m,k+(t-1)K}$. Going from the first stage through the end, $\{a^{(M_k)}_{m,k+(t-1)K}\}$, $t = 1, \ldots, T$, $m = 1, \ldots, M_k$, are the coefficients in (3.3). The estimated autoregressive coefficients are then obtained through (3.5) and (3.6). For kth series, when the true process is TV-AR(M_k), the forward innovation variance at stage M_k, $\{\sigma^{2(M_k)}_{f,k+(t-1)K}\}$ is equal to the innovation variance $\{\sigma^2_{k,t}\}$ for $t = 1, \ldots, T$. The estimation algorithm of the coefficients and innovation covariance is given in the Appendix.

3.2.4 Model Specification and Bayesian Inference

To give time-varying structures to the forward and backward PARCOR coefficients and innovation variances in (3.7) and (3.8), we define (3.7) and (3.8) as DLMs with time-varying parameters (West and Harrison, 1997; West et al., 1999). We consider random walks for the PARCOR coefficients and multiplicative random walks for the innovation variance. The PARCOR coefficients are modeled as

$$\alpha^{(m)}_{f,m,k+tK} = \alpha^{(m)}_{f,m,k+(t-1)K} + \epsilon_{f,m,k+tK}, \quad \epsilon_{f,m,k+tK} \sim N(0, \zeta_{f,m,k+tK}),$$
$$\alpha^{(m)}_{b,m,k+tK} = \alpha^{(m)}_{b,m,k+(t-1)K} + \epsilon_{b,m,k+tK}, \quad \epsilon_{b,m,k+tK} \sim N(0, \zeta_{b,m,k+tK}),$$

where $\zeta_{f,m,k+tK}$ and $\zeta_{b,m,k+tK}$ are time dependent evolution variance. These evolution variances are defined via the *discount factors* $\gamma_{f,k,m}$ and $\gamma_{b,k,m}$ within the range $(0, 1)$, respectively (see Appendix, Yang et al. (2016) and West and Harrison (1997) for details). The discount factor γ actually controls the smoothness of PARCOR coefficients. Here, we assume $\gamma_{f,k,m} = \gamma_{b,k,m} = \gamma_{k,m}$ at each stage m and select their value through a grid-search based the likelihood of the fitted TV-AR model at each stage. Similarly, the forward and backward innovation variances are modeled as

$$\sigma^2_{f,m,k+tK} = \sigma^2_{f,m,k+(t-1)K}(\delta_{f,m}/\eta_{f,m,k+tK}), \quad \eta_{f,m,k+tK} \sim Beta(g_{f,m,k+tK}, h_{f,m,k+tK}),$$
$$\sigma^2_{b,m,k+tK} = \sigma^2_{b,m,k+(t-1)K}(\delta_{b,m}/\eta_{b,m,k+tK}), \quad \eta_{b,m,k+tK} \sim Beta(g_{b,m,k+tK}, h_{f,m,k+tK}),$$

where $\delta_{f,m}$ and $\delta_{b,m}$ are also discount factors in the range (0,1), and the multiplicative innovations, $\eta_{f,m,k+tK}$ and $\eta_{b,m,k+tK}$ follow beta distributions with parameters $(g_{f,m,k+tK}, h_{f,m,k+tK})$ and $(g_{b,m,k+tK}, h_{b,m,k+tK})$, respectively (see Appendix, Yang

et al. (2016) and West and Harrison (1997) for details). The smoothness of innovation variance is controlled by both γ and δ. Similar to the PARCOR coefficients, we assume $\delta_{f,k,m} = \delta_{b,k,m} = \delta_{k,m}$ at each stage. Note that $\epsilon_{b,m,k+tK}$, $\epsilon_{b,m,k+tK}$, $\eta_{f,m,k+tK}$ and $\eta_{b,m,k+tK}$ are mutually independent and are also independent of any other variables in the model.

We use conjugate normal priors for the forward and backward PARCOR coefficients, so that

$$p(\alpha_{f,m,k-K}|D_{f,m,k-K}) \sim N(\mu_{f,m,k-K}, C_{f,m,k-K}),$$
$$p(\alpha_{b,m,k-K}|D_{b,m,k-K}) \sim N(\mu_{b,m,k-K}, C_{b,m,k-K}),$$

where $m = 1, \ldots, M_k$, $k = 1, \ldots, K$, $D_{f,m,k-K}$ and $D_{b,m,k-K}$ denotes the information available at the initial time t = 0, $\mu_{f,m,k-K}$ and $C_{f,m,k-K}$ are the mean and the variance of the normal prior distribution. We also specify conjugate initial priors for the forward and backward innovation variance, so that

$$p(\sigma^2_{f,m,k-K}|D_{f,m,k-K}) \sim G(\nu_{f,m,k-K}/2, \kappa_{f,m,k-K}/2),$$
$$p(\sigma^2_{b,m,k-K}|D_{b,m,k-K}) \sim G(\nu_{b,m,k-K}/2, \kappa_{b,m,k-K}/2),$$

where $G(\cdot, \cdot)$ is the gamma distribution, and $\nu_{f,m,k-K}/2$ and $\kappa_{f,m,k-K}/2$ are the shape and rate parameters of the gamma prior distribution. Typically, we suppose these starting values as constants over all stages. In order to reduce the effect of the prior distribution, we choose $\mu_{f,m,k-K}/2$ and $C_{f,m,k-K}$ to be zero and one, respectively and fix $\nu_{f,m,0} = 1$ and set $\kappa_{f,m,k-K}$ to equal to the sample variance of the initial part of each series. Following these settings, we can obtain the DLM sequential filtering

and smoothing algorithms (West and Harrison, 1997) to derive the marginal posterior distributions of the forward and backward parameter parameters in (3.7) and (3.8). The detailed algorithm of the sequential filtering and smoothing are given in the Appendix.

3.2.5 Model Selection

To apply the BCLF, we need to find not only the optimal order but also the optimal discount factors. For selection an optimal model order, we can do it visually using scree plots or numerically using the model selection criterion shown below. We start with setting up a maximal order, P_{max}, and start from the first stage of the lattice filter through the $M_{k,max}$-th stage ($M_{k,max} = KP_{max} + k - 1$) for each series. At the mth stage, for the kth series, we search a group of pre-specified sets of discount factors $\{\gamma_{k,m}, \delta_{k,m}\}$ to find the set which maximizes the likelihood of the fitted DLM, and use the corresponding estimated parameters as the result. Having all the optimal discount factors and the corresponding estimated parameters, the model selection criterion for the order P can be computed based on the estimation obtained from the first through the $(PK + K - 1)$th stage. According to the model selection criterion values, the optimal order is selected. The details are shown in the Appendix. For the model order selection, we consider several criteria as follows.

Scree Plot

Yang et al. (2016) used a scree plot method to visually (and analytically) select the model order by plotting the log-likelihood against the order p. If the observed time series indeed follows a multivariate time-varying autoregressive model, the log-likelihood

will increase with the model order until an optimal number, and this number is then selected as the best order. A numerical way to determine where the likelihood stops increasing is to compute the percent of change in the log-likelihood from one order to next. Additionally, for each model order, a grid search of the discount factors is conducted, and when the percent change is small enough, we regard the likelihood stops increasing. The discount factors that achieve the largest log-likelihood is selected and the corresponding likelihood is used in model selection.

In our case, by plotting the log-likelihood against the order, we can find the order after which the log-likelihood stops increasing. The stopping point can be evaluated by the percent change. If the percent change is less than a prespecified threshold, this point is identified as the best order.

Bayesian Information Criterion (BIC)

We consider an approach based on BIC to select the statistically "best" model order. For a TV-VAR(P) model with parameters denoted as $\boldsymbol{\theta}$, BIC is defined as

$$BIC(P) = -2\mathcal{L} + n_\Theta log(KT).$$

BIC is computed for each of the fitted models of order $P = 1, \ldots, P_{max}$ to a K-dimensional time series of length T. $\mathcal{L}$ is the log likelihood based on multivariate time-varying autoregressive model of order P as (3.1). n_Θ is the number of parameters estimated in the initial state of the state space model. In our case, $n_\Theta = 2PK^2 + (K-1)K$, which is the total number of estimated time-varying variables in all stages. We use BIC as the model selection criterion throughout this chapter. BIC works well

on identifying the correct order on simulated examples. The penalty term of BIC considers the length of the series. For a given time series, the longer series leads to a larger likelihood. Therefore, we need a penalty term changing with the series length.

Deviance Information Criterion (DIC)

We can also use the deviance information criterion (DIC) to choose the model order (see Gelman et al. (2013) and references therein). With parameters denoted as $\boldsymbol{\theta}$ and data denoted as $\mathbf{y}$, the DIC of a model is defined as

$$DIC = -2\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}_{Bayes}) + 2p_{DIC},$$

where $\hat{\boldsymbol{\theta}}_{Bayes}$ is the Bayes estimator of $\boldsymbol{\theta}$ and p_{DIC} is the effective number of parameters. The effective number of parameters is given by

$$p_{DIC} = 2\left[\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}_{Bayes}) - E_{post}(\log p(\mathbf{y}|\boldsymbol{\theta}))\right],$$

where E_{post} is the expectation under the posterior distribution. In our case, the terms above are estimated through MC samples $\boldsymbol{\theta}^{(s)}$ as

$$p_{DIC} = 2\left[\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}_{Bayes}) - \frac{1}{S}\sum_{s=1}^{S}(\log p(\mathbf{y}|\boldsymbol{\theta}^{(s)}))\right].$$

where samples $\boldsymbol{\theta}^{(s)}$, $s = 1, \ldots, S$, are generated from $p(\boldsymbol{\theta}|\mathbf{y})$, the posterior distribution of $\boldsymbol{\theta}$ (See the Appendix for details).

Widely Applicable Akaike Information Criterion (WAIC)

WAIC is a generalized version of AIC and is more appropriate for Bayesian hierarchical models (Watanabe, 2010). For a model with parameters denoted as $\boldsymbol{\theta}$, considering data $\mathbf{y} = (y_1, \ldots, y_T)'$, the WAIC is defined as

$$WAIC = -2\text{log}p(\mathbf{y}|\hat{\boldsymbol{\theta}}_{Bayes}) + 2p_{WAIC},$$

where

$$p_{WAIC} = 2\sum_{t=1}^{T} \left(\text{log}(E_{post}p(y_t|\boldsymbol{\theta}) - E_{post}(\text{log}p(y_t|\boldsymbol{\theta}))\right).$$

WAIC is estimated by using MC samples from the S posterior draws $\boldsymbol{\theta}^{(s)}$ as

$$\hat{p}_{WAIC} = 2\sum_{t=1}^{T} \left(\text{log}(\frac{1}{S}\sum_{s=1}^{S} p(y_t|\boldsymbol{\theta}^{(s)}) - \frac{1}{S}\sum_{s=1}^{S}(\text{log}p(y_t|\boldsymbol{\theta}^{(s)}))\right).$$

$\boldsymbol{\theta}^{(s)}$ is generated from $p(\boldsymbol{\theta}|\mathbf{y})$, the posterior distribution of $\boldsymbol{\theta}$. See the Appendix for details.

To achieve a best model selection, we first search for an order which achieves the smallest BIC, DIC, WAIC. If the numerical criteria, BIC, DIC, WAIC, do not agree with each other, we use the visual scree plot to assist the model selection. In practice, for BCLF, BIC works best for selecting the model order among the numerical criteria in most of the simulation cases we conducted.

3.2.6 Forecasting

Having estimated all parameters, we consider 1-step ahead forecasts of the TV-VAR(P) model. The 1-step ahead predictive distribution of the PARCOR coefficients can be obtained according to West and Harrison (1997). We generate J samples for each of the PARCOR coefficients can be drawn from their 1-step-ahead predictive distribution as $\{\alpha^{(m)(j)}_{f,m,k+TK}, \alpha^{(m)(j)}_{b,m,k+TK}$ for all $k, m\}$, $j = 1, \ldots, J$. The samples of the 1-step ahead prediction of TV-VAR parameters, $\Phi_{p,T+1}$ can be obtained as $\{\mathbf{\Phi}^{(j)}_{p,T+1}, j = 1, \ldots, J\}$, for $p = 1, \ldots, P$, by transforming the samples of the PARCOR coefficients according to (3.5) and (3.6). Finally, the 1-step ahead forecast is given as

$$\mathbf{x}^{(j)}_{T+1} = \sum_{p=1}^{P} \mathbf{\Phi}^{(j)}_{p,T+1} \mathbf{x}_{T+1-p}. \tag{3.10}$$

We use the posterior mean of $\mathbf{x}^{(j)}_{T+1}$ $(j = 1, \ldots, J)$ obtained through the samples in (3.10) as the 1-step ahead forecast. This forecast can be easily extended to h-steps. The details of forecasting up to h-steps can be found in the Appendix.

3.3 Simulation Studies

To assess the effectiveness of our method, we conducted similar simulation studies to Zhao and Prado (2020) and compared our results with results obtained using their approach and a TV-VAR(P) model. The comparison were on spectral analysis and forecasting. In spectral analysis. the results are assessed using the average squared error (ASE) (Ombao et al., 2001) between the true spectral density and the estimated spectral density. In forecasting, the mean squared predictive error (MSPE) is used to

evaluate the forecasting performance.

3.3.1 Simulation 1: Bivariate TV-VAR(2) Process

We considered 500 bivariate time series of length $T = 1034$ simulated from a TV-VAR model (Zhao and Prado, 2020) as follows: $\mathbf{x}_t = \mathbf{\Phi}_{1,t}\mathbf{x}_{t-1} + \mathbf{\Phi}_{2,t}\mathbf{x}_{t-2} + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t)$ with

$$\mathbf{\Phi}_{1,t} = \begin{pmatrix} r_{1,t}\cos(\frac{2\pi}{\lambda_{1,t}}) & \phi_{1,1,2,t} \\ 0 & r_{2,t}\cos(\frac{2\pi}{\lambda_{2,t}}) \end{pmatrix} \quad \text{and} \quad \mathbf{\Phi}_{2,t} = \begin{pmatrix} -r_{1,t}^2 & \phi_{2,1,2,t} \\ 0 & -r_{2,t}^2 \end{pmatrix}$$

where $r_{1,t} = \frac{0.1}{T}t + 0.85$, $r_{2,t} = -\frac{0.1}{T}t + 0.95$, $r_{3,t} = \frac{0.2}{T}t - 0.9$, $r_{4,t} = \frac{0.2}{T}t + 0.7$, $\lambda_{1,t} = \frac{15}{T}t + 5$, and $\lambda_{2,t} = -\frac{10}{T}t + 15$. We considered three different cases for the values of $\phi_{1,1,2,t}$ and $\phi_{2,1,2,t}$, namely (i) $\phi_{1,1,2} = 0$, $\phi_{2,1,2} = 0$; (ii) $\phi_{1,1,2} = -0.8$, $\phi_{2,1,2} = 0$; (iii) $\phi_{1,1,2} = r_{3,t}$, $\phi_{2,1,2} = r_{4,t}$. These three cases have the covariance $\mathbf{\Sigma}_t = I_2, 2I_2, 3I_2$, where I_2 denotes a 2×2 identity matrix. We also consider cases (iv), (v), (vi) with the same $\mathbf{\Phi}_t$ as (i), (ii), (ii), respectively, but with different covariance $\mathbf{\Sigma}_t$, such that $\Sigma_{1,1,t} = \Sigma_{2,2,t} = 1 + \frac{t}{T}$, $\Sigma_{1,2,t} = \Sigma_{2,1,t} = 0$ for $t = 1, \ldots, T$. The bivariate spectral matrix of this process can be obtained by

$$\mathbf{g}(t, \omega) = \mathbf{\Psi}(t, \omega)^{-1} \times \mathbf{\Sigma}_t \times \mathbf{\Psi}^*(t, \omega)^{-1},$$

where $\mathbf{\Psi}(t, \omega) = \mathbf{I}_2 - \sum_{p=1}^{P} \mathbf{\Phi}_{p,t}\exp\{-2p\pi i\omega\}$, $\mathbf{\Sigma}_t$ is time-varying innovation covariance and $*$ stands for the Hermitian transpose of a matrix. The spectral matrix $\mathbf{g}(t, \omega)$ is symmetric and consists of series $g_{11}(t, \omega)$, $g_{22}(t, \omega)$ and $g_{12}(t, \omega)$, representing the spectrum of the first series, the spectrum of the second series and the cross-spectrum

between the first and the second series, respectively. The squared coherence between the first and the second series is defined as

$$\rho_{12}^2(t,\omega) = \frac{|g_{12}(t,\omega)|^2}{g_{11}(t,\omega)g_{22}(t,\omega)}.$$

The spectral matrix can be estimated by

$$\hat{\mathbf{g}}(t,\omega) = \hat{\mathbf{\Psi}}(t,\omega)^{-1} \times \hat{\mathbf{\Sigma}}_t \times \hat{\mathbf{\Psi}}^*(t,\omega)^{-1},$$

where $\hat{\mathbf{\Psi}}(t,\omega) = |\mathbf{I}_2 - \sum_{p=1}^{P} \hat{\mathbf{\Phi}}_{p,t} \exp\{-2p\pi i\omega\}|$ and $\hat{\mathbf{\Sigma}}_t$ are estimated values. The estimated squared coherence $\hat{\rho}_{12}^2(t,w)$ can be obtained accordingly.

To evaluate the performance in estimating the time-frequency representations, we compared different models by the mean and standard deviations of the ASEs. We calculate ASE for each realization as follows (Ombao et al., 2001):

$$\text{ASE}_n = (TL)^{-1} \sum_{t=1}^{T} \sum_{l=1}^{L} \left(\log(\hat{\mathbf{g}}(t,\omega_l)) - \log(\mathbf{g}(t,\omega_l))\right)^2, \tag{3.11}$$

where $n = 1,\ldots,500$, w_l is the lth frequency, and L is the number of frequencies in the time-frequency representation. We denote the average over all realizations as $\overline{ASE} = 1/500 \sum_{n=1}^{500} ASE_n$.

To evaluate the performance of our proposed method (BCLF), we choose a $P_{max} = 5$ and fit the simulated datasets with the TV-VAR order adaptively selected based on BIC as we suggest. The other model selection criteria were also obtained and are shown in Table 3.1. BIC suggests that TV-VAR(2) is the best model for most simulated datasets. In practice, BIC works extremely well for model selection of

the BCLF and significantly better than DIC and WAIC. Consequently, we use BIC for model order selection of the BCLF throughout this chapter. Our method is compared with TV-VPARCOR and TV-VAR (a method denoted by TV-VAR in the simulations of Zhao and Prado (2020)), which are the methods used in Zhao and Prado (2020). For every dataset, TV-VPARCOR and TV-VAR have the model orders selected based on DIC while BCLF has the model order selected based on BIC. All model selections have a maximum order of 5. The discount factors in all of BCLF, TV-VPARCOR and TV-VAR are chosen from a grid of values in $[0.99, 1]$ based on the likelihood. We select this range to make our results comparable to those of Zhao and Prado (2020). Note that TV-VAR and BCLF consider time-varying covariance while TV-PARCOR assumes constant covariance. Figures 3.1 and 3.2 are the boxplots, which summarize the ASE based on ASEn for six cases compared with TV-VAR and TV-VPARCOR as Zhao and Prado (2020). Figures 3.3, 3.4, and 3.5 summarize the posterior inference by the BCLF approach on the 500 simulated datasets. These figures show the average of the estimated spectrum of the two series and the average of the estimated coherence between them over the 500 simulated datasets. Moreover, we compare the forecast of BCLF with TV-PARCOR in the following ways. The forecast of TV-VPARCOR can be conducted in two ways. The first way (TV-VPARCOR 1) is that we obtain predicted values of PARCOR coefficients from their forecast distribution, transform them into predicted values of the autoregressive coefficients, and then compute the predicted values. The other way (TV-VPARCOR 2) is that we draw MC samples of PARCOR coefficients from their forecast distribution and transform these samples into the samples of autoregressive coefficients, and, finally, obtain the samples of predicted observations. The posterior

mean of the samples of predicted observations are used to be the predicted values of the observations. We introduce another forecasting method for the BCLF (BCLF 1) to compare with the forecasts of TV-VPARCOR. Specifically, for BCLF 1, we first obtain predicted values of PARCOR coefficients from their forecast distribution, transform them into predicted values of the autoregressive coefficients, and then compute the predicted values of the observations. The formal forecast method for the BCLF is given in Sections 3.2.6 and 3.6.4 (BCLF 2). Using the four forecasting approaches, we conduct one-step ahead rolling prediction for $t = 1025{:}1034$ for the first 3 simulation cases. For each case and each method, 100 series of length 1034 are generated, the last 10 observations of the series are held out for prediction. Table 3.2 shows the mean and the standard deviation of the 100 MSPEs for each case and each method. TV-VPARCOR 1 and BCLF 1 are not fully Bayesian predictions while TV-PARCOR 2 and BCLF 2 are Bayesian predictions, which may be more appropriate for uncertainty quantification. BCLF 2 one-step ahead prediction works slightly better than TV-PARCOR 2 in this study. Table 3.3 shows the computation time of the simulation studies. The time of Simulation 1 is the time used for fitting one dataset of case (i), including the selection of discount factors, the selection of the model order and the estimation of parameters. Note that the TV-VPARCOR method is conducted using Rcpp and parallel computing while the BCLF is coded in pure R. If TV-VPARCOR is conducted using pure R, it would take significantly longer computation time than that shown in this table. Moreover, BCLF can be accelerated by using Rcpp and parallel computing. Therefore, the computation time will be greatly reduced if BCLF is also conducted using Rcpp and parallel computing. All of the six cases take approximately the same time and therefore, we use case (i) as an example.

Table 3.4 shows BCLF has smaller ASEs in all cases except Case 1, which is described further below. Each of the methods TV-VAR, TV-VPARCOR and BCLF has some drawbacks that lower their performance. TV-VAR and TV-VPARCOR use an approximate estimator to estimate the covariance, which does not work as well as our time-varying covariance estimator in some cases. Additionally, TV-VAR uses one discount factor for all coefficients. This produces larger ASEs for TV-VAR in most cases. BCLF uses the modified Cholesky decomposition due to the "one-channel-at-a-time" algorithm. This results in some small loss in ASEs.

Table 3.4 shows the ASEs of each method for different true coefficients and innovation covariance when the innovation covariance is time-invariant. It suggests that the BCLF performs better for the spectrum estimation in most cases. In other words, BCLF is more robust to the level of the covariance. That is, when the true covariance increases from I_2 to $2I_2$ and to $3I_2$, the ASEs of BCLF stay at approximately the same level, while the ASEs of TV-VAR and TV-VPARCOR become significantly larger. Moreover, Table 3.5 shows that when the innovation covariance is time-varying, BCLF works better than the constant covariance models TV-VAR and TV-VPARCOR as expected.

Table 3.1: Model selection of TV-VAR model for 500 simulated datasets from bivariate TV-VAR(2) process. Each column gives the model order P and the percentage of the datasets that are selected to this order according to BIC, DIC, and WAIC.

Case	P	1	2	3	4	5
1	BIC	0%	100%	0%	0%	0%
1	DIC	0%	20%	10%	23%	47%
1	WAIC	90%	10%	0%	0%	0%
2	BIC	0%	100%	0%	0%	1%
2	DIC	0%	92%	1%	4%	3%
2	WAIC	99%	1%	0%	0%	0%
3	BIC	0%	100%	0%	0%	0%
3	DIC	0%	87%	8%	0%	5%
3	WAIC	100%	0%	0%	0%	0%
4	BIC	0%	100%	0%	0%	0%
4	DIC	0%	18%	11%	19%	52%
4	WAIC	100%	0%	0%	0%	0%
5	BIC	0%	100%	0%	0%	0%
5	DIC	0%	88%	4%	0%	8%
5	WAIC	100%	0%	0%	0%	0%

Table 3.2: MSPE values for the 1-step ahead rolling forecast (t = 1025:1034) and corresponding standard deviations (in parentheses) obtained from BCLF and TV-VPARCOR methods for the TV-VAR(2) simulated data. All four methods are described in the simulation example.

Model	Case (i)	Case (ii)	Case (iii)
TV-VPARCOR 1	1.029(0.315)	1.036(0.318)	1.046(0.322)
TV-VPARCOR 2	1.046(0.332)	1.061(0.327)	1.103(0.371)
BCLF 1	1.069(0.327)	1.046(0.325)	1.100(0.434)
BCLF 2	1.038(0.317)	1.046(0.325)	1.085(0.354)

Figure 3.1: Boxplots of ASEs by three methods in cases (i), (ii), (iii) when $\mathbf{\Sigma}_t = I_2$. In each plot, the index 1, 2, and 3 denotes TV-VAR, TV-, and BCLF, respectively.

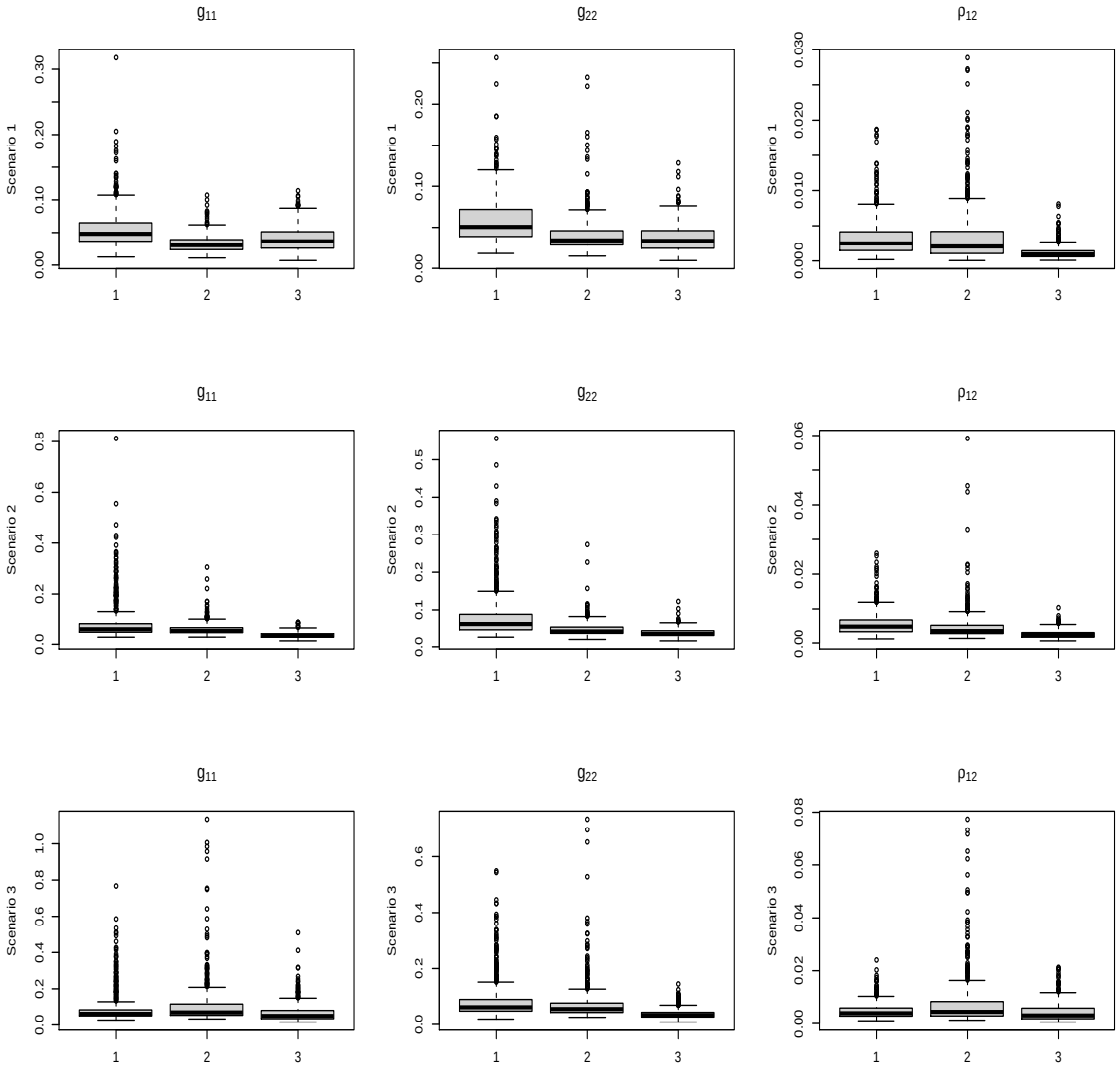


Figure 3.2: Boxplots of ASEs by three methods in cases (iv), (v), (vi). In each plot, the index 1, 2, and 3 denotes TV-VAR, TV-PARCOR, and BCLF, respectively.

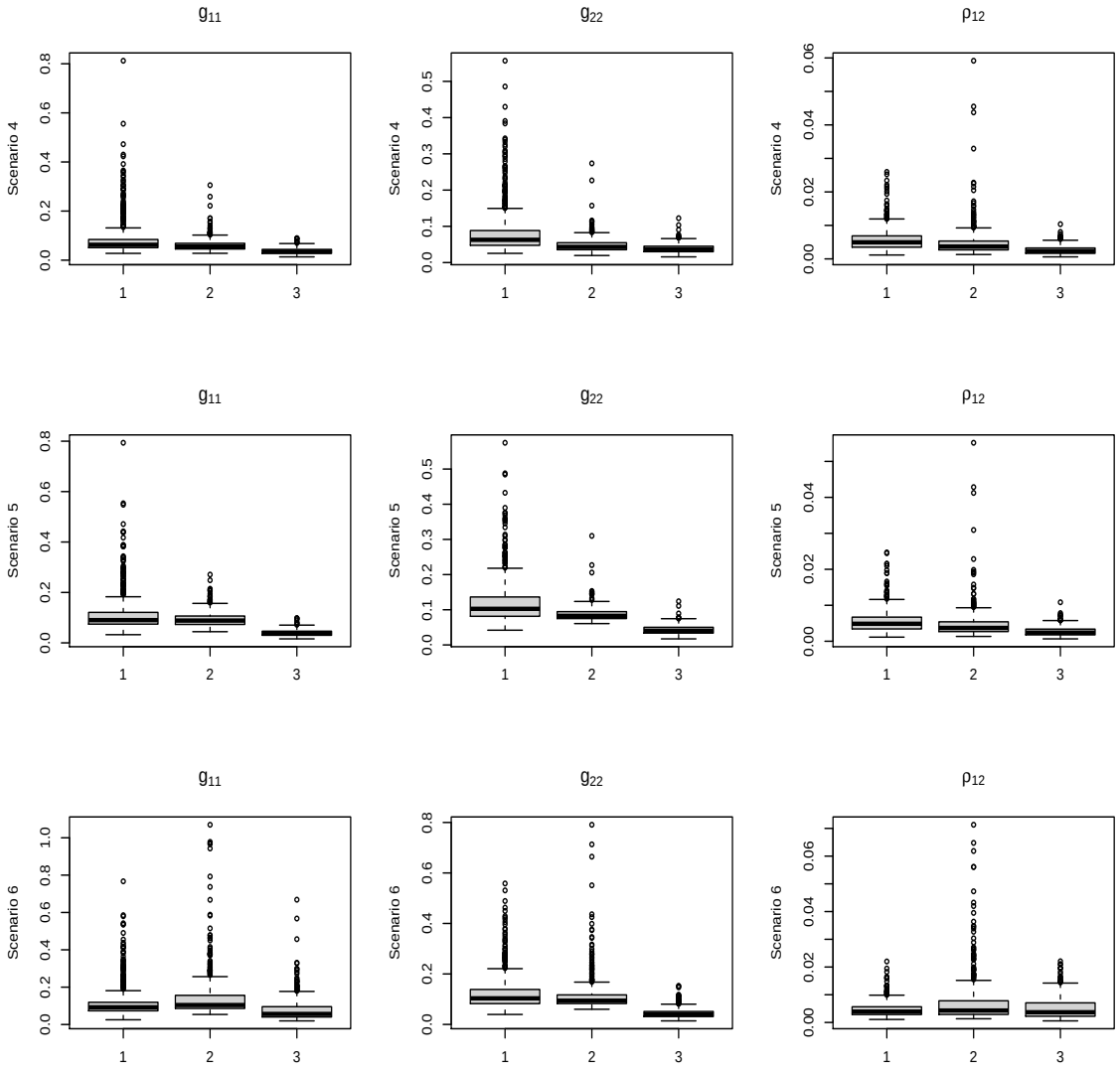


Figure 3.3: Case with $\phi_{1,1,2,t} = 0$ and $\phi_{2,1,2,t} = 0$. Top: True log spectral density $g_{11}(t,\omega)$ (left), true log spectral density $g_{22}(t,\omega)$ (middle), true squared coherence $\rho^2_{12}(t,\omega)$ (right). Bottom: Average of the estimated $\hat{g}_{11}(t,\omega)$ (left), average of the estimated $\hat{g}_{22}(t,\omega)$ (middle), average of the estimated $\hat{\rho}^2_{12}(t,\omega)$ (right) using BCLF.

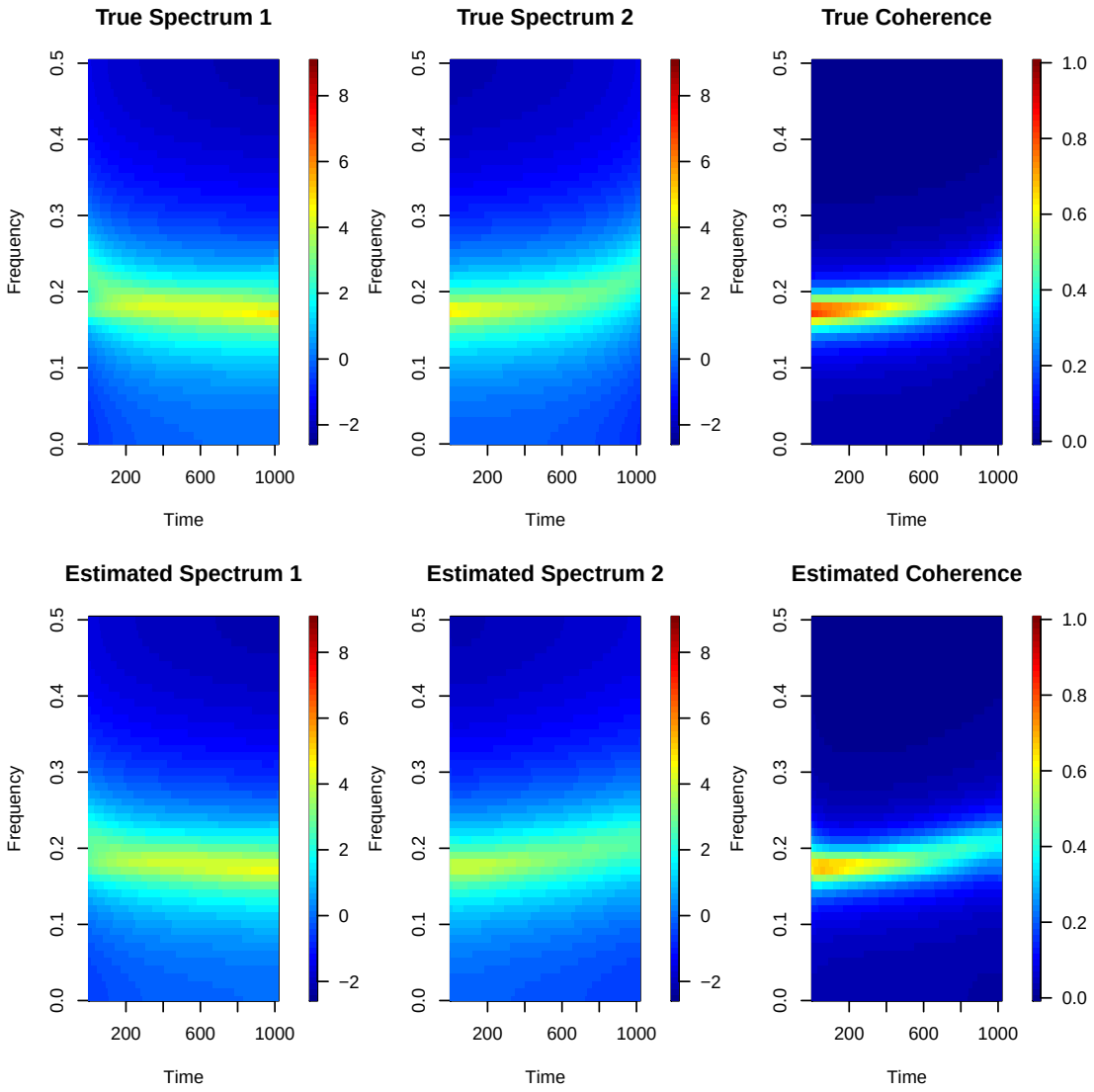


Figure 3.4: Case with $\phi_{1,1,2,t} = -0.8$ and $\phi_{2,1,2,t} = 0$. Top: True log spectral density $g_{11}(t,\omega)$ (left), true log spectral density $g_{22}(t,\omega)$ (middle), true squared coherence $\rho^2_{12}(t,\omega)$ (right). Bottom: Average of the estimated $\hat{g}_{11}(t,\omega)$ (left), average of the estimated $\hat{g}_{22}(t,\omega)$ (middle), average of the estimated $\hat{\rho}^2_{12}(t,\omega)$ (right) using BCLF.

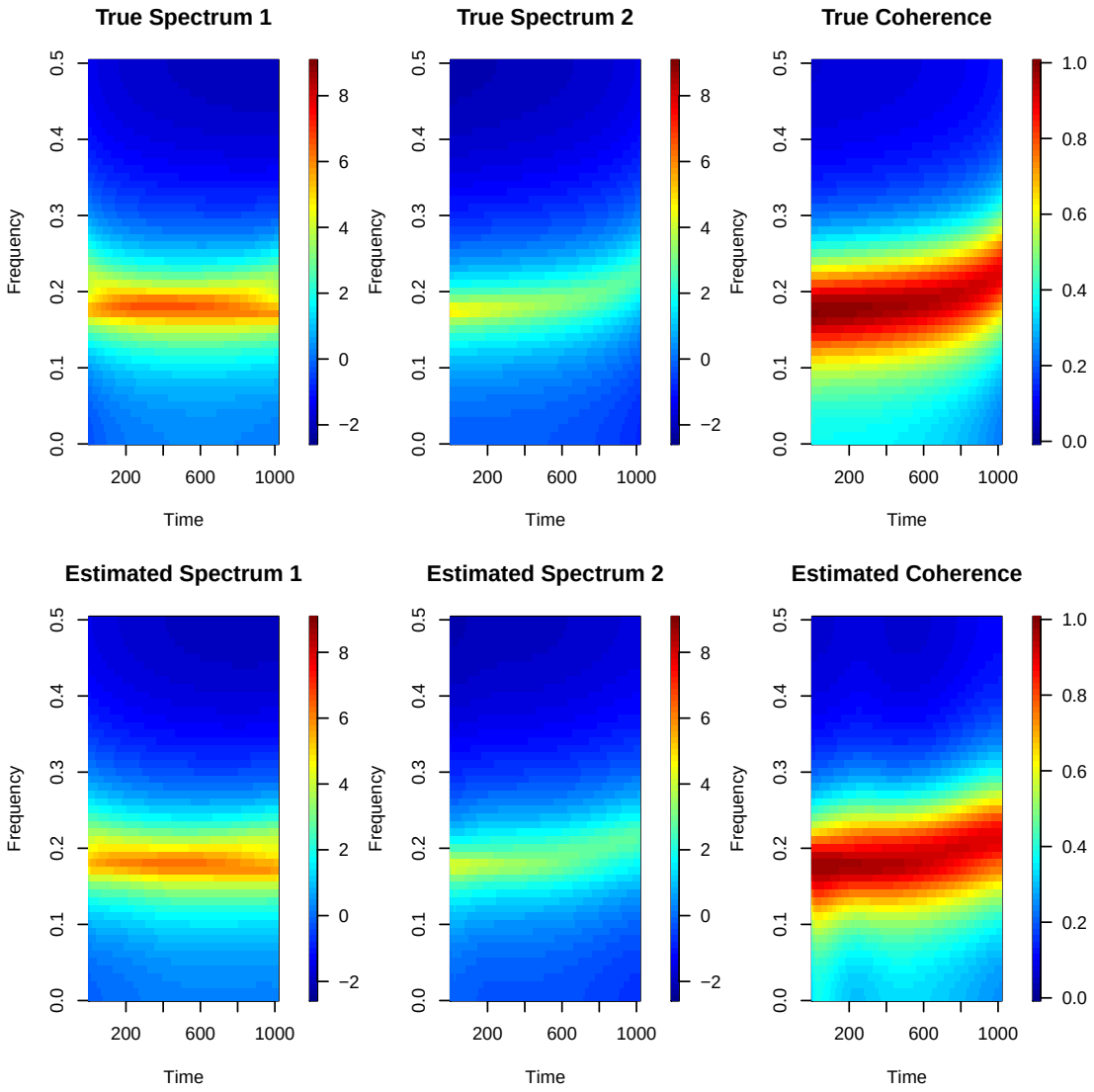


Figure 3.5: Case with $\phi_{1,1,2,t} = r_{3,t}$ and $\phi_{2,1,2,t} = r_{4,t}$. Top: True log spectral density $g_{11}(t,\omega)$ (left), true log spectral density $g_{22}(t,\omega)$ (middle), true squared coherence $\rho^2_{12}(t,\omega)$ (right). Bottom: Average of the estimated $\hat{g}_{11}(t,\omega)$ (left), average of the estimated $\hat{g}_{22}(t,\omega)$ (middle), average of the estimated $\hat{\rho}^2_{12}(t,\omega)$ (right) using BCLF.

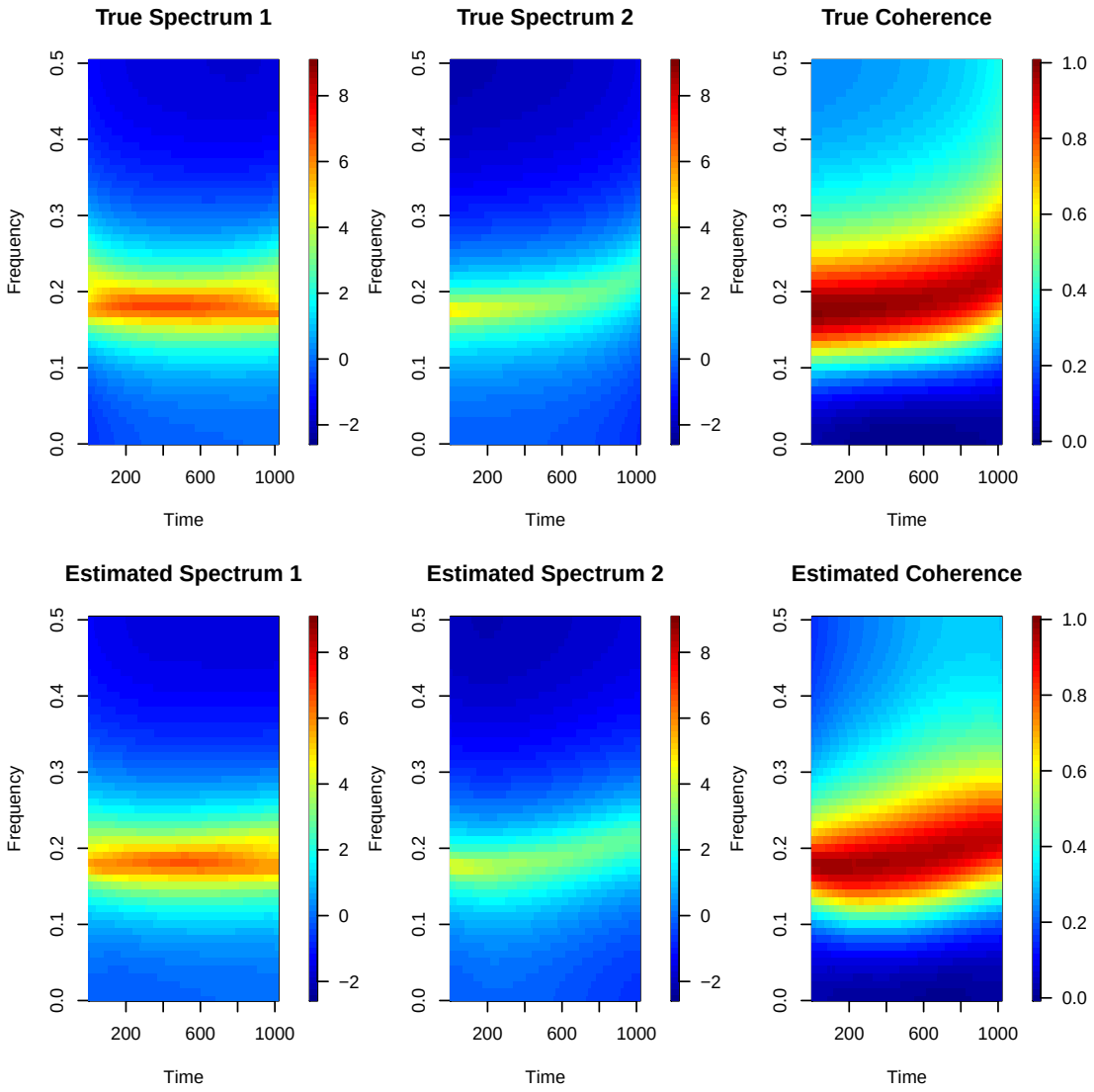


Table 3.3: Computation times (in seconds) for TV-VPARCOR and BCLF models.

Model	Simulation 1	Simulation 2
TV-VPARCOR (Rcpp)	10.83 seconds	17874.36 seconds
BCLF (R)	9.86 seconds	279.76 seconds

3.3.2 Simulation 2: 20-Dimensional TV-VAR(1)

We consider a simulation in Zhao and Prado (2020) and generate data from a 20-dimensional nonstationary TV-VAR(1) process of length T = 300. The elements $\Phi_{i,j,t}$ of the autoregressive coefficients at time t, Φ_t, are given as follows:

$$\phi_{i,j} = \begin{cases} 0.7 + \frac{0.2}{299} \times t & \text{for all} \quad i = j, i = 1, \ldots, 10, \\ 0.7 + \frac{0.2}{299} \times t & \text{for all} \quad i = j, i = 1, \ldots, 10, \\ 0.9 & \text{for} \quad (i,j) \in [(1,5),(2,15)] \\ -0.9 & \text{for} \quad (i,j) \in [(6,12),(15,20)] \\ 0 & \text{otherwise,} \end{cases}$$

for $t = 1, \ldots, 300$. Additionally, the innovation covariance is specified as $\Sigma = 0.1\mathbf{I}$

Similar to Zhao and Prado (2020), we choose a $P_{max} = 3$ and fit the simulated datasets with the TV-VAR order adaptively selected based on BIC as we suggest. The other model selection criteria were also obtained. All of BIC, DIC, and WAIC suggest that TV-VAR(1) is the best model for the simulated datasets. The discount factors are chosen from a grid of values in $[0.99, 1]$ based on the likelihood of the fitted TV-AR models in each stage. Figures 3.6 and 3.7 summarize the posterior inference obtained from the BCLF approach on the 100 simulated datasets. These figures

Table 3.4: Mean ASEs and corresponding standard deviations (in parentheses) using different methods for 500 simulated datasets from TV-VAR(2).

Model	g_{11}	g_{22}	ρ_{12}^2
Case (i): $\phi_{1,1,2,t} = 0, \phi_{2,1,2,t} = 0, \Sigma_t = I_2$			
TV-VAR	0.05393 (0.02349)	0.05497 (0.04214)	0.00292 (0.00194)
TV-VPARCOR	0.03302 (0.01313)	0.04062 (0.02288)	0.00353 (0.00424)
BCLF	0.04037 (0.01905)	0.03643 (0.01713)	0.00123 (0.00107)
Case (i): $\phi_{1,1,2,t} = 0, \phi_{2,1,2,t} = 0, \Sigma_t = 2I_2$			
TV-VAR	0.07538 (0.03663)	0.07748 (0.0852)	0.00484 (0.00357)
TV-VPARCOR	0.04144 (0.02542)	0.04960 (0.0305)	0.00582 (0.00595)
BCLF	0.03509 (0.01680)	0.03622 (0.0169)	0.00124 (0.00112)
Case (i): $\phi_{1,1,2,t} = 0, \phi_{2,1,2,t} = 0, \Sigma_t = 3I_2$			
TV-VAR	0.10443 (0.05255)	0.10575 (0.05543)	0.00643 (0.00574)
TV-VPARCOR	0.05114 (0.04774)	0.06347 (0.05972)	0.00804 (0.00877)
BCLF	0.03513 (0.01634)	0.03623 (0.01699)	0.00124 (0.00111)
Case (ii): $\phi_{1,1,2,t} = -0.8, \phi_{2,1,2,t} = 0, \Sigma_t = I_2$			
TV-VAR	0.09164 (0.08122)	0.0870 (0.070740)	0.00583 (0.00364)
TV-VPARCOR	0.06124 (0.02644)	0.0475 (0.021017)	0.00491 (0.00480)
BCLF	0.03712 (0.01255)	0.0384 (0.012639)	0.00262 (0.00133)
Case (ii): $\phi_{1,1,2,t} = -0.8, \phi_{2,1,2,t} = 0, \Sigma_t = 2I_2$			
TV-VAR	0.10632 (0.07053)	0.09847 (0.07628)	0.00733 (0.00551)
TV-VPARCOR	0.19222 (0.30678)	0.12383 (0.19629)	0.01293 (0.01698)
BCLF	0.03833 (0.01303)	0.03937 (0.01285)	0.00259 (0.00125)
Case (ii): $\phi_{1,1,2,t} = -0.8, \phi_{2,1,2,t} = 0, \Sigma_t = 3I_2$			
TV-VAR	0.26823 (0.96065)	0.16537 (0.28518)	0.00978 (0.00817)
TV-VPARCOR	0.28447 (0.50113)	0.17805 (0.33872)	0.01791 (0.02220)
BCLF	0.03836 (0.01304)	0.03937 (0.01285)	0.00258 (0.00125)
Case (iii): $\phi_{1,1,2,t} = r_{3,t}, \phi_{2,1,2,t} = r_{4,t}, \Sigma_t = I_2$			
TV-VAR	0.09722 (0.09334)	0.09132 (0.07920)	0.00493 (0.00310)
TV-VPARCOR	0.11324 (0.12927)	0.07579 (0.07308)	0.00807 (0.01014)
BCLF	0.06804 (0.05361)	0.03902 (0.01943)	0.00442 (0.00372)
Case (iii): $\phi_{1,1,2,t} = r_{3,t}, \phi_{2,1,2,t} = r_{4,t}, \Sigma_t = 2I_2$			
TV-VAR	0.10583 (0.07045)	0.10023 (0.08221)	0.00622 (0.00492)
TV-VPARCOR	0.19213 (0.30681)	0.12378 (0.19631)	0.01292 (0.01698)
BCLF	0.06968 (0.05467)	0.03996 (0.01970)	0.00446 (0.00377)
Case (iii): $\phi_{1,1,2,t} = r_{3,t}, \phi_{2,1,2,t} = r_{4,t}, \Sigma_t = 3I_2$			
TV-VAR	0.19117 (0.68771)	0.15403 (0.28426)	0.00836 (0.00907)
TV-VPARCOR	0.28374 (0.49940)	0.17709 (0.33651)	0.01791 (0.02219)
BCLF	0.06980 (0.05447)	0.04005 (0.01978)	0.00445 (0.00378)

Table 3.5: Mean ASE values and corresponding standard deviations (in parentheses) for 500 simulated datasets from TV-VAR(2).

Case (iv): $\phi_{1,1,2,t} = 0, \phi_{2,1,2,t} = 0$			
Model	g_{11}	g_{22}	ρ_{12}^2
TV-VAR	0.0916 (0.0812)	0.0870 (0.0707)	0.0058 (0.0036)
TV-VPARCOR	0.0614 (0.0264)	0.0475 (0.0210)	0.0049 (0.0048)
BCLF	0.0371 (0.0126)	0.0384 (0.0126)	0.0026 (0.0013)
Case (v): $\phi_{1,1,2,t} = -0.8, \phi_{2,1,2,t} = 0$			
Model	g_{11}	g_{22}	ρ_{12}^2
TV-VAR	0.1174 (0.0819)	0.1232 (0.0701)	0.0056 (0.0034)
TV-VPARCOR	0.0938 (0.0296)	0.0871 (0.0204)	0.0048 (0.0045)
BCLF	0.0396 (0.0129)	0.0427 (0.0130)	0.0027 (0.0012)
Case (vi): $\phi_{1,1,2,t} = r_{3,t}, \phi_{2,1,2,t} = r_{4,t}$			
Model	g_{11}	g_{22}	ρ_{12}^2
TV-VAR	0.1228 (0.0934)	0.1281 (0.0782)	0.0047 (0.0029)
TV-VPARCOR	0.1465 (0.1270)	0.1159 (0.0736)	0.0073 (0.0087)
BCLF	0.0788 (0.0660)	0.0449 (0.0206)	0.0051 (0.0040)

show the average of the estimated spectrum of the two series and the average of the estimated coherence between them over the 100 simulated datasets. Table 3.3 shows the computation time of the simulation studies, including the selection of discount factors, the selection of the model order and the estimation of parameters.

3.4 Case Studies

3.4.1 U.S. Economy Data

We analyze a U.S. economic time series from the first quarter of 1960 to the fourth quarter of 2001, which contains three variables: inflation rate and unemployment rate and a short-term nominal interest rate. This dataset is also used by Primiceri

Figure 3.6: Top: True log spectral densities of Components 1, 2, 8, and 15. Bottom: estimated log spectral densities of Components 1, 2, 8, and 15 using the BCLF.

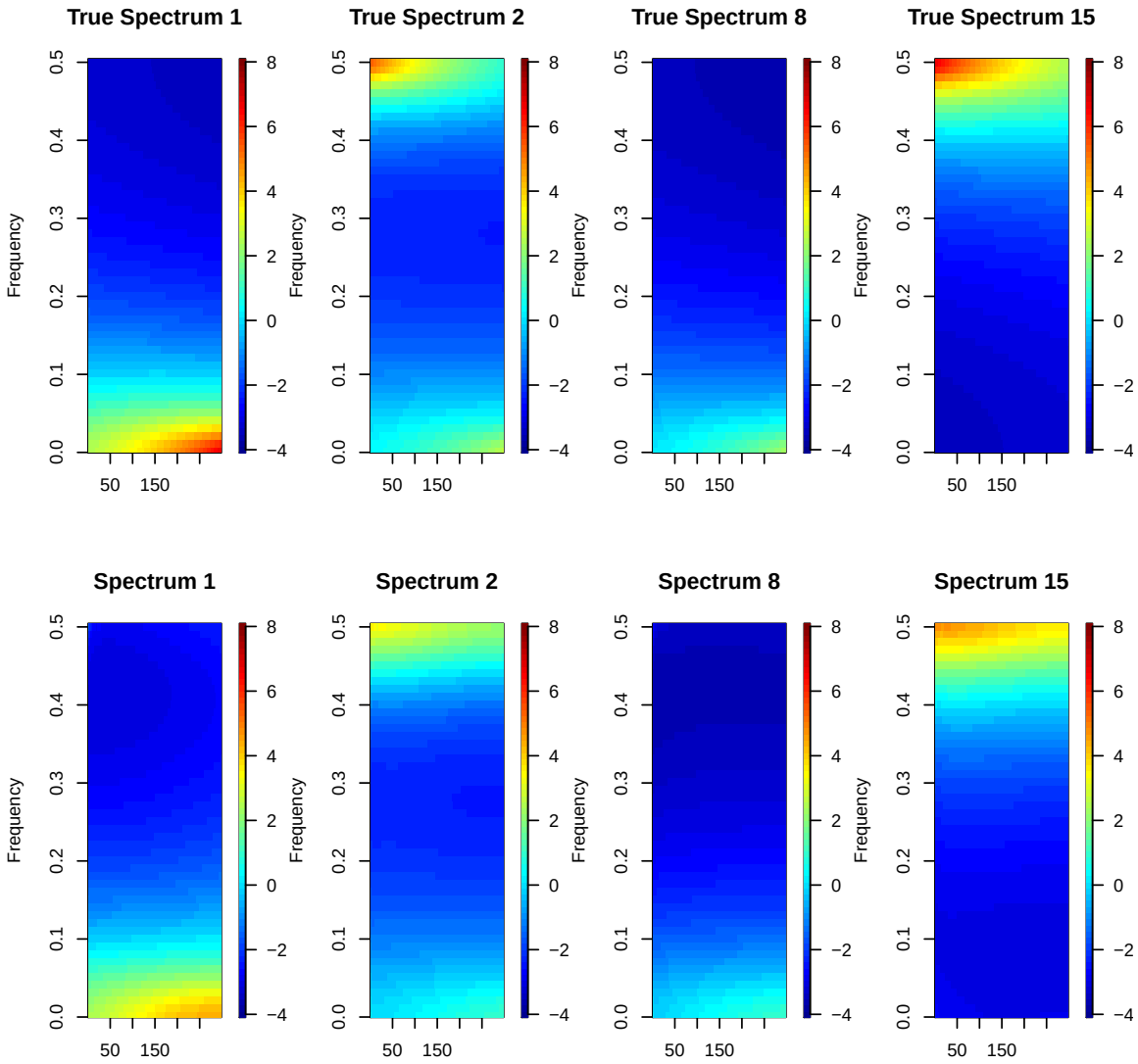
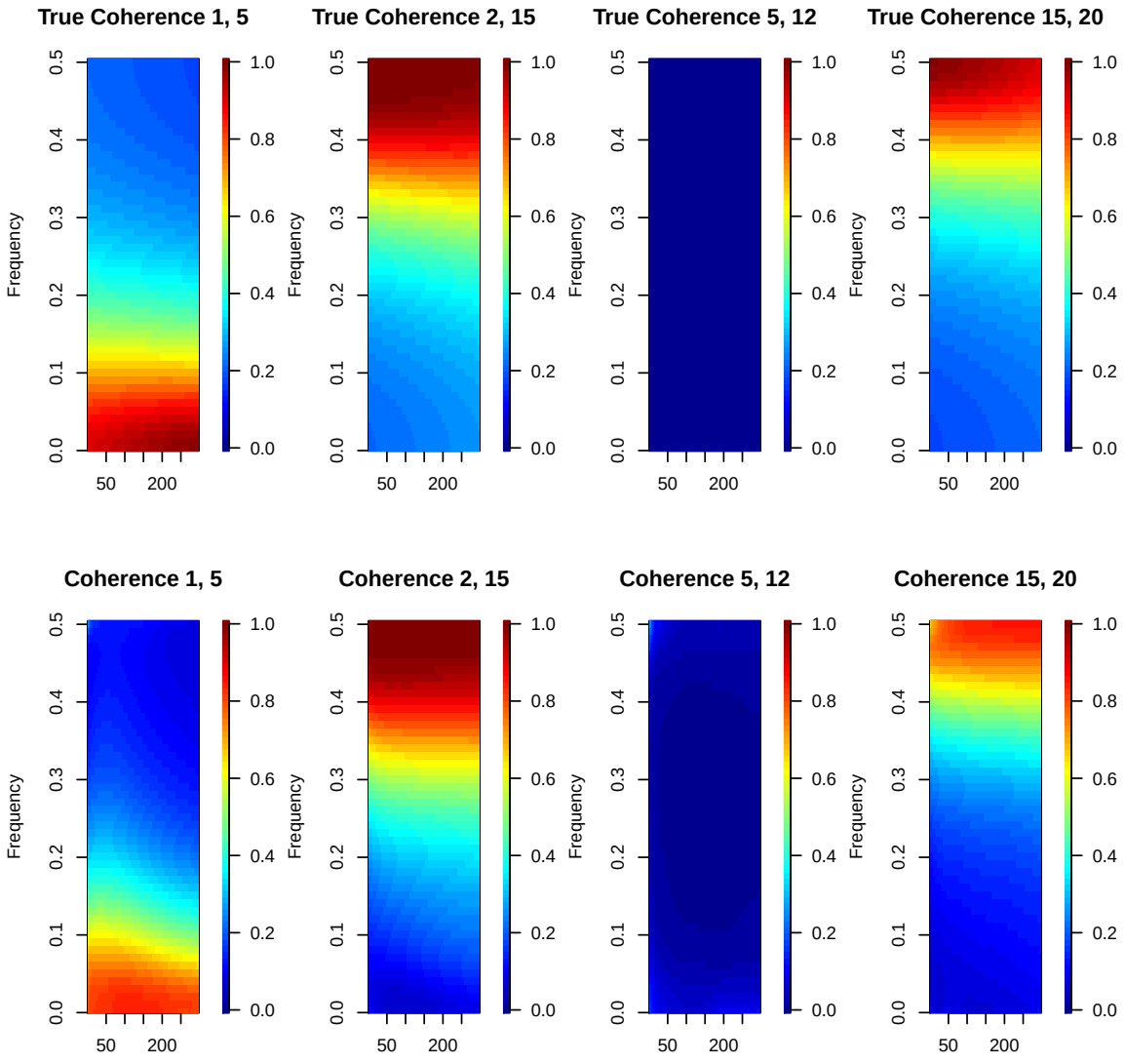


Figure 3.7: Top: True coherence between Components 1 and 5, 2 and 15, 5 and 12, and 15 and 20. Bottom: Estimated coherences between Components 1 and 5, 2 and 15, 5 and 12, and 15 and 20 using the BCLF.



(2005) to analyze the effect of the monetary policy to the high unemployment and inflation in 1970s and early 1980s. We apply the BCLF to 168 quarters of the three series, each of them has been mean centered (using the series sample mean). We set the maximal order $P_{max} = 10$ and assume each variable has its own optimal discount factor values. According to the simulation studies and some experiments on real data, we find that a discount factor close to 1 would make the estimated parameters more smooth and significantly biased. Therefore, we select $[0.90, 0.995]$ as the range of the discount factor. The series plots of the three variable are given as Figure 3.8. The model selection criteria (see Table 3.6) suggest that the best model order is 2 and for the order 1 is nearly as good. Considering the order 2 is used in Primiceri (2005), we pick order 2 for this data example.

The two major differences between our model and Primiceri (2005): (1) our model does not have intercept, (2) our method is a fully Bayesian approach, which selects the smoothness of time-varying parameters adaptively as part of the MCMC while Primiceri (2005) uses an empirical Bayes approach with the priors estimated from an exploratory analysis of the data. Therefore, the application of our method should not give the same results as the application in Primiceri (2005). The estimated impulse responses at three different dates are shown in Figures 3.9 and 3.10. These dates are selected as Primiceri (2005) to represent three Federal Reserve chairmanships. These impulse responses move with the change of monetary policy (short-term nominal interest rate). The short-term interest rate was from low to high in 1975:I, 1996:I and 1981:III. This order matches the plots that the curves of 1975:I and 1981:III are at two ends and 1996:I in the middle. Figure 3.11 shows the estimated innovation covariance and their 90% credible interval. Some of the credible intervals are extremely narrow

Table 3.6: Model selection of TV-VAR model for quarterly economic data. This table displays the BIC for model orders 1 through 10.

Model Order	1	2	3	4	5
BIC	-169.16	-169.36	-83.12	-99.62	-138.79
Model Order	6	7	8	9	10
BIC	-68.34	32.08	125.04	195.89	310.11

and, therefore, difficult to see in the plot. We can see that the innovation variance of the short-term interest rate is significantly higher in the period 1979-1983. It coincided with Paul Volcker's first term of chairmanship of the Federal Reserve. The changing pattern of the interest rate's response to inflation and unemployment and the interest's effect on inflation and unemployment confirm with the results of Primiceri (2005). Providing similar results, our method is much more efficient in the sense that the estimation does not require MCMC simulation and the model selection is conducted simultaneously for all orders under the maximal order. Moreover, the smoothness of the time-varying variables are automatically tuned in the model fitting and does not need to estimate the smoothness through an exploratory analysis of the data. These features make this method easy to use and allow us to extend similar analyses to much longer series and more variables.

3.4.2 Wind Data

We analyze the wind data used by Zhao and Prado (2020), which contains the median wind speed and direction measurements collected every 4 hours from 6/1/2010 to 8/15/2020 in Monterey, Salinas and Watsonville, 3 stations located near Monterey Bay, Northern California. These data were obtained from a publicly avail-

Figure 3.8: U.S. quarterly economic data from 1960 to 2001. The first series is the inflation rate. The second series is the unemployment rate. The third series is the nominal short-term interest rate.

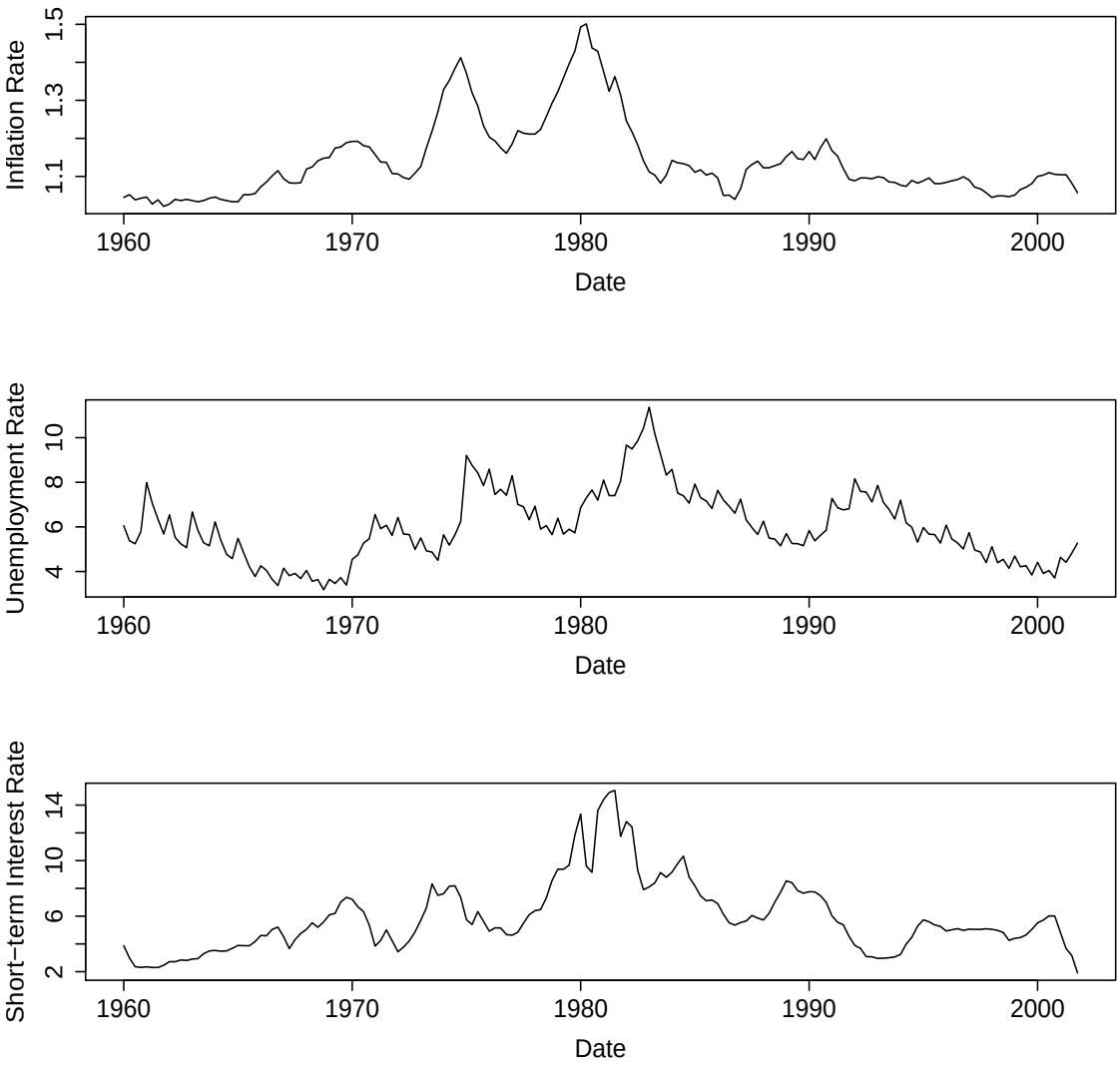


Figure 3.9: Impulse responses of inflation to monetary policy (short-term nominal interest rate) shocks in 1975:I, 1981:III and 1996:I and their 95% credible intervals.

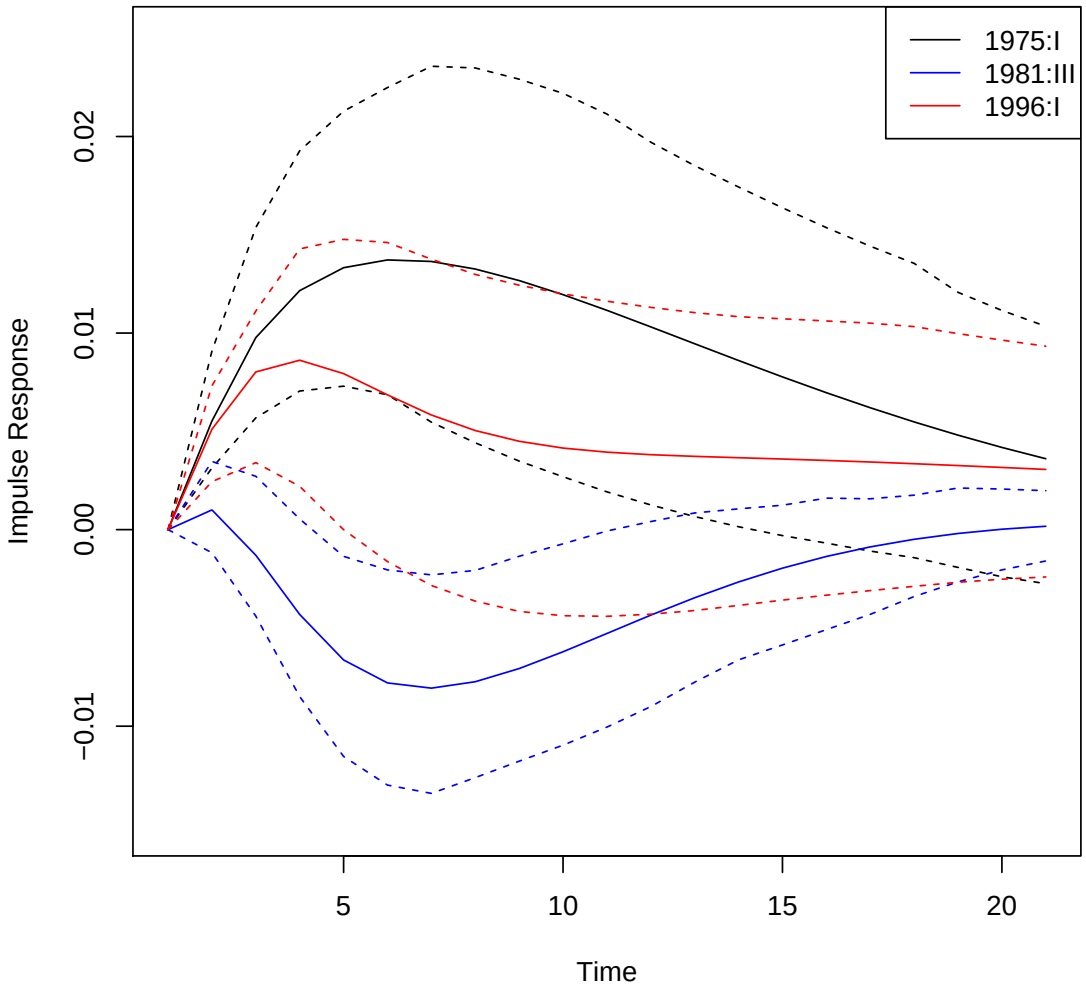


Figure 3.10: Impulse responses of unemployment to monetary policy (short-term nominal interest rate) shocks in 1975:I, 1981:III and 1996:I and their 95% credible intervals.

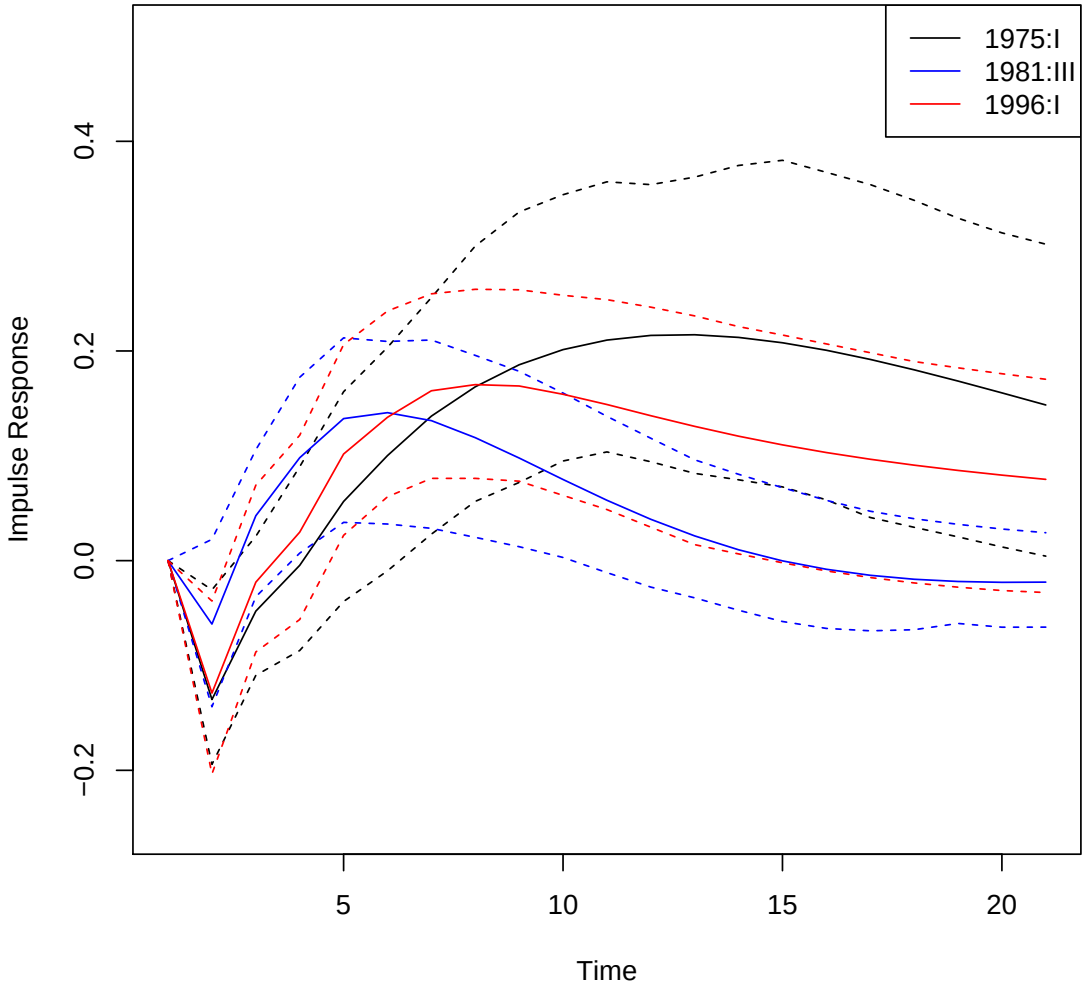
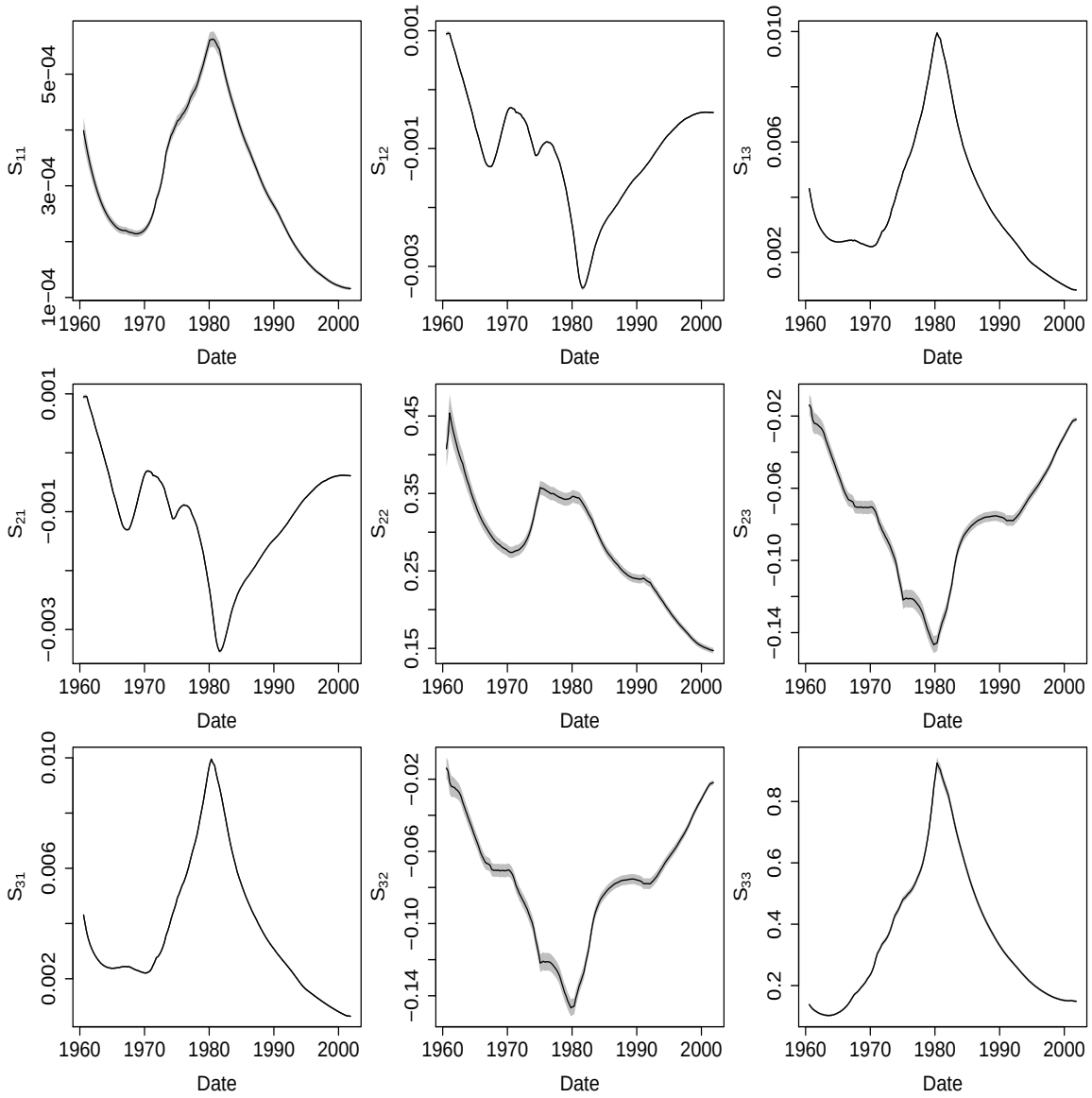


Figure 3.11: The estimated time-varying innovation covariance of a TV-VAR(2) model on U.S. quarterly economic data. The grey region shows the credible intervals of each variable.



able database, the Iowa Environmental Mesonet (IEM) (see http://mesonet.agron.iastate.edu/ASOS/). ASOS stations are located at airports and take observations and basic reports from the National Weather Service (NWS), the Federal Aviation Administration (FAA), and the Department of Defense (DOD). We use the BCLF approach for a multivariate analysis of the six-dimensional time series of the wind speed component and wind direction component for the 3 stations. We set $P_{max} = 10$ and consider discount factor values on a grid in the $(0.95, 0.99]$ range and assume that each station has their own optimal discount factor values. The BIC selects the model order 6 for the TV-VAR model. After applying BCLF to fit the TV-VAR(6) model to the six-dimensional data, we obtain the estimate of its time-frequency representation. Figure 3.12 shows the estimated log spectral densities of the X (East-West) component and the Y (North-South) component for each location. Figure 3.13 shows the estimated squared coherence between each pair of wind components across the three locations. Generally, the results reveal a 24-hour cycle in X components, Y components and their coherence and the magnitude of the cycles vary with time.

3.5 Discussion

We propose a computationally efficient model-based parametric approach for nonstationary multivariate time series. We use a response-orthogonal reparameterization to transform a TV-VAR model into a scalar periodic AR model and further use the Bayesian Lattice filter together with PARCOR to speed up the estimation. Our approach can simultaneously estimate the time-varying coefficients and the time-varying innovation covariance. The "one channel at-a-time" modeling avoids matrix inver-

Figure 3.12: Top Row: Estimated log-spectral densities of the X (East-West) components for Monterey (Spectrum 1), Salinas (Spectrum 2), and Watsonville (Spectrum 3). Bottom Row: Estimated log-spectral densities of the Y (North-South) components for Monterey (Spectrum 4), Salinas (Spectrum 5), and Watsonville (Spectrum 6). The unit of time in the plots is 4 hours.

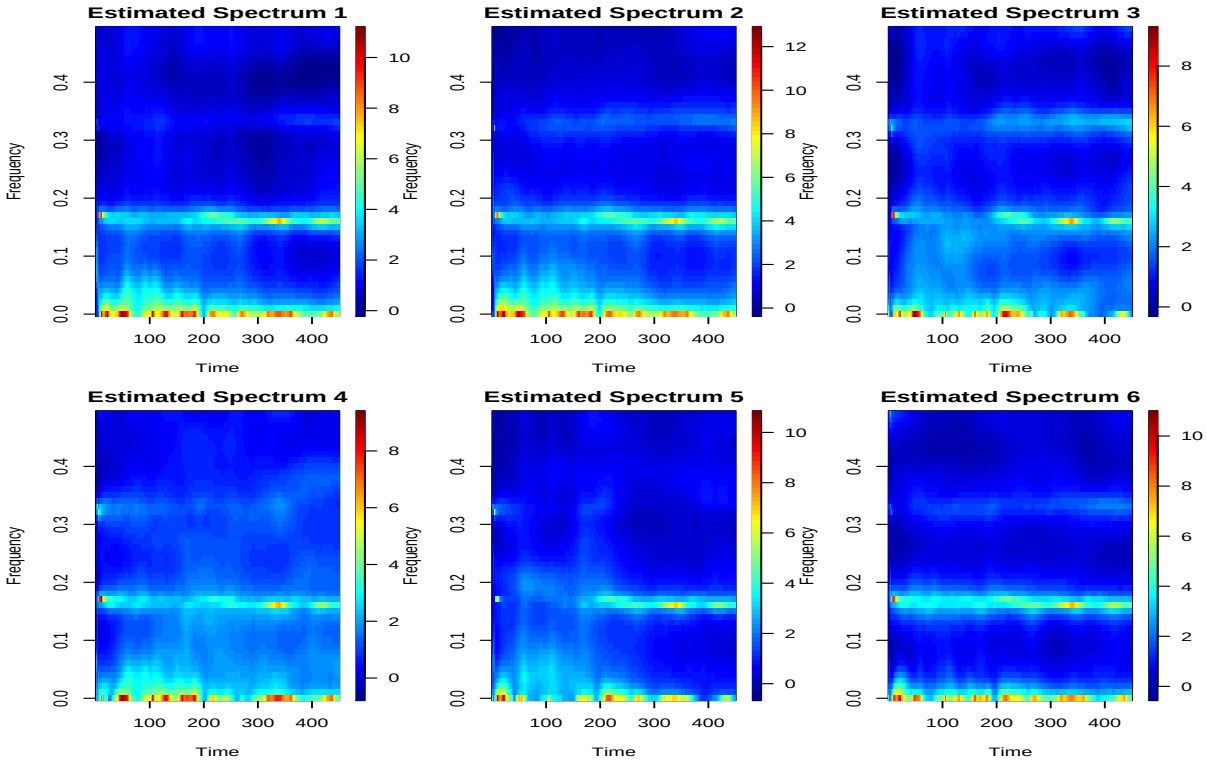
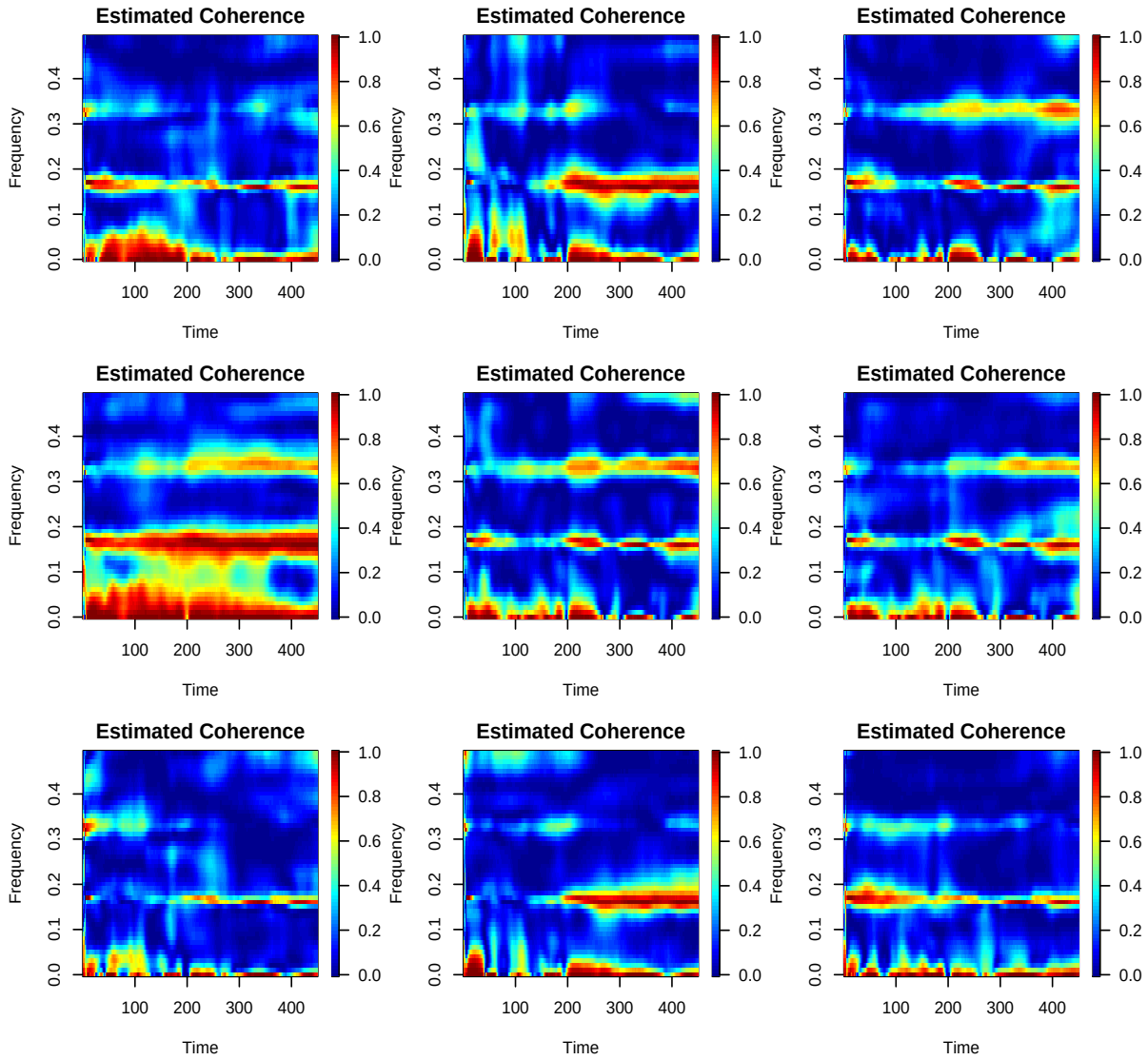


Figure 3.13: Top Row: Estimated squared coherences between the X (East-West) component and Y (North-South) component in Monterey, Salinas, and Watsonville. Middle Row: Estimated squared coherences between the X (East-West) components of Monterey and Salinas, Monterey and Watsonville, and Salinas and Watsonville. Bottom Row: Estimated squared coherences between the Y (North-South) components of Monterey and Salinas, Monterey and Watsonville, and Salinas and Watsonville. The unit of time in the plots is 4 hours.



sion computations in the estimation, which significantly reduces the computation cost and makes the computation time increase linearly, rather than exponentially, with the data dimension. Additionally, this "one channel at-a-time" modeling makes it possible for our approach to benefit from parallel computing in the future. Moreover, the use of Bayesian lattice filters makes the computation time increase linearly, rather than exponentially, with the model order. We also provide a model selection method to choose the discount factors and the optimal model order. The simulation cases show that the parameter estimation is fairly accurate. The real data example shows that the time-varying coefficients and innovation covariance can be effectively used to reveal the time-varying structure.

Our method benefits from the LDL decomposition (modified Cholesky decomposition) of the innovation covariance as (3.6). This assumption has a drawback as stated in Primiceri (2005). In particular, the ordering of the multiple variables matters because any permutation of the ordering can lead to different prior distributions for the autoregressive coefficients. The prior distributions are firstly assigned to the PARCOR coefficients after making a transformation through the Cholesky decomposition. When transforming back to the original model space, these priors are transformed to the priors of the autoregressive coefficients. This transformation depends on the order of the variables in the multivariate model. This difference in the prior distributions does not have a significant effect on the time-invariant parameter models. However, it may show a more noticeable difference in the time-varying parameter models when using different ordering. In Simulation 1, we see this phenomenon. However, in both our real data application and in Primiceri (2005), different orders of the variables led to very similar results. This issue also received attention from other research (Kastner

et al. (2017), among others). A solution to this issue in the context of the BCLF is a subject of future research.

3.6 Appendix

3.6.1 Algorithm for Fitting Multivariate Time Series

We summarize the algorithm of our approach to fitting a TV-VAR model as follows:

- Step 1. Interlace the multivariate times series as in (3.1) into a periodic time series as in (3.3).
- Step 2. Set up a value for order P and a set of values for $\gamma_{k,m}$, $\delta_{k,m}$ for $m = 1, \ldots, M_k$, $k = 1, \ldots, K$, where $M_k = kP_{max} + K - 1$, as well as the initial values of parameters at $t = 0$.
- Step 3. Repeat Step 4 for stage $m = 1, \ldots, M_k$, $k = 1, \ldots, K$.
- Step 4. Apply the sequential filtering and smoothing algorithm to the prediction errors of last stage, $f_{k,t}^{(m-1)}$ and $b_{k,t}^{(m-1)}$ to obtain $\alpha_t^{(m)} = \mu_{t,T}^{(m)}$ and $\sigma_t^{2(m)} = s_{t,T}^{(m)}$ of the forward and backward equations, and the forward and backward prediction errors, $f_{k,t}^{(m)}$ and $b_{k,t}^{(m)}$ for $t = 1, \ldots, T$.
- Step 5. To compute the model selection criterion of TV-VAR(P), collect the computed parameters up to order $kP + K - 1$ for the kth series, $k = 1, 2, \ldots, K$. Transform the computed parameters to obtain the estimated values of the parameters $\{\mathbf{\Phi}_{p,t}\}$ and $\{\mathbf{\Sigma}_t\}$ in (3.1). Compute the criterion following Section 3.2.5. Pick the best order according to the model selection criterion.

- Step 6. (Optional) Compute the spectrum and cross-spectrum based on the parameter estimation following Section 3.3.1.

- Step 7. (Optional) Make h-step ahead forecasting following Section 3.2.6.

3.6.2 Sequential Filtering and Smoothing Algorithm

Suppose we need to fit (3.7) and (3.8) to estimate the PARCOR coefficients and innovation variance. Their posterior distributions can be obtained by following a sequential filtering and smoothing algorithm. The algorithm given here is for the forward autoregression case as in (3.7). Filtering and smoothing algorithm can be obtained for the backward case in a similar manner. For any series, any stage, we denote the posterior distribution at time t as $(\alpha_t|D_t) \sim T_{\nu_t}(\mu_t, C_t)$ a multivariate T distribution with ν_t df, location parameter μ_t, and scale matrix C_t, and $(\sigma_t^{-2}|D_t) \sim G(\nu_t/2, \kappa_t/2)$, a gamma distribution with shape parameter $\nu_t/2$ and scale parameter $\kappa_t/2$. These parameters can be computed for all t using the filtering equations below. Note we use $s_t = \kappa_t/\nu_t$ to denote the usual point estimate of σ_t^2. f_t in the equation is the forward prediction error. For $t = 2, \ldots, T$, we have

$$\mu_t = \mu_{t-1} + z_t e_t,$$
$$C_t = (R_t - z_t z_t' q_t)(s_t/s_{t-1}),$$

and

$$\nu_t = \delta\nu_{t-1} + 1,$$
$$\kappa_t = \delta\kappa_{t-1} + s_{t-1}e_t^2/q_t,$$

where

$$e_t = f_t - z'_{t-1}m_{t-1},$$
$$q_t = z'_{t-1}R_t z'_{t-1} + s_{t-1},$$

and

$$z_t = R_t f_{t-1}/q_t,$$
$$R_t = C_{t-1} + G_t,$$
$$G_t = C_t(1-\beta)/\beta.$$

After the filtering equations up to T, we compute the full marginal posterior distribution $(\alpha_t|D_T) \sim T_{\nu_t}(\mu_{t,T}, C_t)$ and $(\sigma_t^{-2}|D_T) \sim G(\nu_{t,T}/2, \kappa_{t,T}/2)$ through the smoothing equations

$$\mu_{t,T} = (1-\beta)\mu_t + \beta\mu_{t+1,T}$$
$$C_{t,T} = [(1-\beta)C_t + \beta^2 C_{t+1,T}](s_{t,T}/s_t)$$
$$\nu_{t,T} = (1-\delta)\nu_t + \delta\nu_{t+1,T}$$
$$1/s_{t,T} = (1-\delta)/s_t + \delta s_{t+1,T}$$

and $\kappa_{t,T} = \nu_{t,T} s_{t,T}$ for $t = T-1, \ldots, 1$.

3.6.3 Generating MC Samples of $\boldsymbol{\theta}$ for Model Selection Criteria

First, generate samples $\{\alpha_t\}$ from $T_{\nu_t}(\mu_t, C_t)$ and samples $\{\sigma_t^{-2}\}$ from $G(\nu_t/2, \kappa_t/2)$ for $t = 1, \ldots, T$ as Section 3.6.2. Then, transform them into TV-VAR parameters, $\{\boldsymbol{\Phi}_{p,t}\}$ and $\{\boldsymbol{\Sigma}_t\}$, by (3.5) and (3.6). Finally, generate S sets of samples $\{\boldsymbol{\theta} = (\boldsymbol{\Phi}_{t,p}^{(s)}, \boldsymbol{\Sigma}_t^{(s)}), s = 1, \ldots, S\}$, which are used to compute DIC and WAIC.

3.6.4 Forecasting

We can make h-step ahead forecasting by following these steps.

- For stage $m = 1, \ldots, M_k$ and series $k = 1, \ldots, K$, compute the h-step ahead posterior predictive distribution of the PARCOR coefficients following West and Harrison (1997): $(\alpha^{(m)}_{f,m,k+(T+h-1)K}|D_T) \sim N(\mu^{(m)}_{f,m,k+(T-1)K}(h), C^{(m)}_{f,m,k+(T-1)K}(h))$ where $\mu_{f,m,k+(T-1)K}(h) = \mu^{(m)}_{f,m,k+(T+h-1)K}$ and $C^{(m)}_{f,m,k+(T-1)K}(h) = C^{(m)}_{f,m,k+(T-1)K} + hG^{(m)}_{f,m,k+TK}$ with $G^{(m)}_{f,m,k+TK} = C^{(m)}_{k+(T+h-1)K}(1-\beta)/\beta$.

- For stage $m = 1, \ldots, M_k$ and series $k = 1, \ldots, K$, draw J samples for each of the h-step ahead of the AR coefficients $\{a^{(M_k)}_{m,k+(T+h-1)K}, m = 1, \ldots, M_k\}$ from the samples of $\{\alpha^{(m)}_{f,m,k+(T+h-1)K}\}$ and $\{\alpha^{(m)}_{b,m,k+(T+h-1)K}\}$ for series $k = 1, \ldots, K$.

- Transform the samples of $\{a^{(M_k)}_{m,k+(T+h-1)K}, m = 1, \ldots, M_k\}$ into the samples of TV-VAR parameters $\{\boldsymbol{\Phi}_{p,T+h}, p = 1, \ldots, P\}$.

- The samples of the h-step ahead forecast is obtained as

$$\mathbf{x}_{T+h}^{(j)} = \sum_{p=1}^{P} \mathbf{\Phi}_{p,T+h}^{(j)} \mathbf{x}_{T+h-p}^{(j)}, \tag{3.12}$$

  where $\mathbf{x}_{T+h-p}^{(j)} = \mathbf{x}_{T+h-p}$ if $h - p \leq 0$.

- We use the posterior mean of $\mathbf{x}_{T+h}$ obtained through the samples in (3.12) as the h-step ahead forecast.

Chapter 4

Computationally Efficient Time-varying Poisson Autoregressive Models

4.1 Introduction

Modeling count time series is essential in many applications such as pandemic incidences, insurance claims, and integer financial data such as transactions. Different types of count time series models have been broadly developed within the classes of observation-driven and parameter-driven models. The observation-driven models include the INGARCH model (Ferland et al., 2006; Fokianos et al., 2009), integer-valued autoregressive model, also called Poisson autoregressive model (Al-Osh and Alzaid, 1987), generalized linear ARMA (GLARMA) model (Zeger, 1988; Dunsmuir, 2016) and Poisson AR model (Brandt and Williams, 2001), among others (see Davis et al. (2016) for a comprehensive overview). The parameter-driven models include the

Poisson state space model (Smith and Miller, 1986), Poisson exponentially weighted moving average (PEWMA) model (Brandt et al., 2000), and dynamic count mixture model (Berry and West, 2019), among others. Some of this research is proceeds under a Bayesian framework (Berry and West, 2019; Bradley et al., 2020).

For nonstationary count time series with changing trends, e.g., daily new COVID cases data, many traditional methods (Ferland et al., 2006; Fokianos et al., 2009; Brandt et al., 2000; Brandt and Williams, 2001) may not capture local trends or give good multi-step forecasts. The motivation for our study is to propose an efficient method to capture the time-varying pattern of the means for such nonstationary count time series and, therefore, make better forecasts than traditional methods. To capture the evolutionary properties, a parameter-driven (process-driven) model with a time-varying coefficient latent process provides one good choice. Moreover, a latent process with appropriately modeled innovations can address the over-dispersion issue, which is common in count time series modeling.

We propose a time-varying Poisson autoregressive (TV-Pois-AR) model for nonstationary count time series. We use a time-varying autoregressive (TV-AR) latent process to model the nonstationary intensity of the Poisson process. This flexible model can capture the latent dynamics of the time series and, therefore, make superior forecasts. The estimation of such a TV-AR process is greatly sped up by using the Bayesian Lattice Filter (BLF, Yang et al. (2016)). Moreover, in our model and many other models, the intensity parameter does not exhibit a closed-form full conditional distribution. To sample the intensity parameter, the No-U-Turn Sampler (NUTS, Hoffman and Gelman (2014)) is used, instead of a random walk Metropolis-Hastings algorithm, for faster mixing of the sample chains. Martino et al. (2015) investigated

the use of the Hamilton Monte Carlo (HMC) method inside the Gibbs sampling. According to this paper, as a self-tuned extension to HMC, NUTS should work well as a univariate sampler inside the Gibbs sampling. Benefiting from the joint use of the Bayesian lattice filter and No-U-Turn Sampler, estimation of the TV-Pois-AR model is efficient and fast, especially for higher model orders or longer length time series.

The rest of the chapter is organized as follows. Section 4.2 describes the proposed Bayesian method. Section 4.3 shows the results of several simulation studies. In Section 4.4, our proposed model is demonstrated through an example of COVID-19 spread for New York State. In Section 4.5, we summarize the proposed method and discuss possible future research.

4.2 Methodology

4.2.1 TV-Pois-AR(P) Model

We propose a Poisson model with latent Gaussian process (LGP). For a univariate count-valued series, z_t, $t = 1, \ldots, T$, a TV-Pois-AR model of order P (TV-Pois-AR(P)) is defined as

$$\begin{aligned} z_t|\mu, y_t &\sim \text{Pois}(\exp(\mu + y_t)), \\ y_t &= \sum_{j=1}^{P} a_{j,t} y_{t-j} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma_t^2), \end{aligned} \tag{4.1}$$

where μ is the mean level of the logarithm of the Poisson intensity and y_t is the autoregressive component of the logarithm of the Poisson intensity which follows a

TV-AR process. It can also be interpreted that $\exp(\mu)$ is the scale parameter of the intensity. We define $a_{j,t}$ and σ_t^2 to be time-varying AR coefficients of TV-AR(P) associated with time lag j at time t and the innovation variance at time t, respectively. The innovations, ξ_t, are defined to be independent Gaussian errors. For this Bayesian model, we use Gibbs sampling to estimate the parameters and latent variables. The later sections show the derivation of the full conditional distributions,

4.2.2 The Lattice Structure of the TV-AR Process

Conditional on the latent variable y_t for $t = 1, \ldots, T$, the conditional distribution parameters in the latent TV-AR process of y_t can be inferred through the BLF Yang et al. (2016). According to the Durbin-Levinson algorithm, there exists a unique correspondence between the PARCOR coefficients and the AR coefficients, (Shumway and Stoffer, 2006; Kitagawa, 2010; Yang et al., 2016). This lattice structure provides a direct way of using the observed time series to estimate the PARCOR coefficients, which are associated with particular AR model (see Hayes (1996) and the Supplementary Appendix Yang et al. (2016)). The efficient estimation of this TV-AR process can be conducted through the following P-stage lattice filter. We denote $f_t^{(P)}$ and $b_t^{(P)}$ to be the prediction error at time t for the forward and backward TV-AR(P) models, respectively, where

$$\begin{aligned} f_t^{(P)} &= y_t - \sum_{j=1}^{P} a_{j,t}^{(P)} y_{t-j} \\ b_t^{(P)} &= y_t - \sum_{j=1}^{P} d_{j,t}^{(P)} y_{t+j}, \end{aligned} \tag{4.2}$$

and $a_{j,t}^{(P)}$ and $b_{j,t}^{(P)}$ are the forward and backward autoregressive coefficients of the corresponding TV-AR(P) models. Then, in the jth stage of the lattice filter for $j = 1, \ldots, P$, the forward and the backward coefficients and the forward and backward prediction errors have the following relationship

$$\begin{aligned} f_t^{(j)} &= f_t^{(j-1)} - \alpha_{j,t}^{(j)} b_{t-1}^{(j-1)}, \\ b_t^{(j)} &= b_{t-1}^{(j-1)} - \beta_{j,t}^{(j)} f_t^{(j-1)}, \end{aligned} \tag{4.3}$$

where $\alpha_{j,t}^{(j)}$ and $\beta_{j,t}^{(j)}$ are the lag j forward and backward PARCOR coefficients at time t, respectively. The initial condition, $f_t^{(0)} = b_t^{(0)} = y_t$ can be obtained from the definition by (4.3). It means that the samples of y_t are plugged in as the initial values of $f_t^{(0)}$ and $b_t^{(0)}$ in the Gibbs sampling. At the jth stage of the lattice structure, we fit time-varying AR(1) models to estimate $\alpha_{j,t}^{(j)}$ and $\beta_{j,t}^{(j)}$. The corresponding forward and backward autoregressive coefficients at time t, $a_{j,t}^{(j)}$ and $d_{j,t}^{(j)}$ can be obtained according to the following equations:

$$\begin{aligned} a_{i,t}^{(j)} &= a_{i,t}^{(j-1)} - a_{j,t}^{(j)} d_{j-i,t}^{(j-1)}, \\ d_{i,t}^{(j)} &= d_{i,t}^{(j-1)} - d_{j,t}^{(j)} a_{j-i,t}^{(j-1)}, \end{aligned} \tag{4.4}$$

with $i = 1, \ldots, j-1$, $a_{j,t}^{(j)} = \alpha_{j,t}^{(j)}$ and $d_{j,t}^{(j)} = \beta_{j,t}^{(j)}$. Finally, the distribution of $a_{j,t}^{(P)}$ and ξ_t for $j = 1, \ldots, P$ are obtained. These distributions are used as conditional distributions of $a_{j,t}$ and ξ_t in the Gibbs sampling.

4.2.3 Model Specification and Bayesian Inference

We assume that the mean level μ follows a normal distribution, i.e., $\mu \sim N(\mu_0, \tau_0^2)$ and the initial state of the latent variable y_0 follows a normal distribution, s.t., $y_0 \sim N(m_0, s_0^2)$. In Gibbs sampling, μ is sampled efficiently by NUTS and this speeds up the mixing of the sample chains. NUTS is an extension to Hamilton Monte Carlo (HMC, Neal (1994); Neal et al. (2011); Duane et al. (1987)) algorithm through automatic tuning. Compared to Metropolis-Hastings algorithm, which uses a Gaussian random walk as proposal distribution, HMC proposes moves using the leapfrog integration and, therefore, generates samples converging to the target distribution. The target distribution of y_t, for $t = 1, \ldots, T$, is its conditional distribution with the density function

$$\begin{aligned} p(y_t|z_t, \mathbf{y}_{-t}, \boldsymbol{\theta}) &= p(z_t|\mu, y_t)p(y_t|\mathbf{y}_{-t}, \boldsymbol{\theta}) \\ &\propto p(z_t|\mu + y_t)p(y_t|\mathbf{y}_{t-1}, \boldsymbol{\theta})p(y_{t+1}|\mathbf{y}_t, \boldsymbol{\theta}) \\ &= p_{\text{Pois}}(z_t|\exp(\mu + y_t))p_N(y_t|\mathbf{a}_t\mathbf{y}_{t-1}, \sigma_t^2)p_N(y_{t+1}|\mathbf{a}_t\mathbf{y}_t, \sigma_{t+1}^2), \end{aligned}$$

where $\mathbf{y}_t = (y_t, \ldots, y_{t-P+1})'$, $\mathbf{a}_t = (a_{1,t}, \ldots, a_{P,t})'$, $\boldsymbol{\theta}$ denotes $a_{j,t}$ and σ_t^2 for all t and $\mathbf{y}_{-t}$ denotes y_i for all i but t. Having the target distribution, NUTS can adaptively draw samples of y_t conditional on all other variables for all t.

To use the BLF to derive the conditional distributions of the parameters in the latent autoregressive process of y_t, we define the distribution of its coefficients $a_{j,t}^{(P)}$ by defining the distributions of the forward and backward PARCOR coefficients in (4.3). To give time-varying structures to the forward and backward PARCOR coefficients, we consider random walks for the PARCOR coefficients. The PARCOR coefficients

are modeled as

$$\alpha_{j,t}^{(j)} = \alpha_{j,t-1}^{(j)} + \epsilon_{\alpha,j,t}, \quad \epsilon_{\alpha,j,t} \sim N(0, \omega_{\alpha,j,t}),$$
$$\beta_{j,t}^{(j)} = \beta_{j,t-1}^{(j)} + \epsilon_{\beta,j,t}, \quad \epsilon_{\beta,j,t} \sim N(0, \omega_{\beta,j,t}),$$

where $\omega_{\alpha,j,t}$ and $\omega_{\beta,j,t}$ are time dependent evolution variance. These evolution variances are defined via the standard discount method in terms of the *discount factors* $\gamma_{f,j}$ and $\gamma_{b,j}$ within the range $(0, 1)$, respectively (see Appendix and West and Harrison (1997) for details). The discount factor γ actually controls the smoothness of PARCOR coefficients. Here, we assume $\gamma_{f,j} = \gamma_{b,j} = \gamma_j$ at each stage j. Similarly, the innovation variances are assumed to follow multiplicative random walks and modeled as

$$\sigma_{f,j,t}^2 = \sigma_{f,j,t-1}^2(\delta_{f,j}/\eta_{f,j,t}), \quad \eta_{f,j,t} \sim Beta(g_{f,j,t}, h_{f,j,t}),$$
$$\sigma_{b,j,t}^2 = \sigma_{b,j,t-1}^2(\delta_{b,j}/\eta_{b,j,t}), \quad \eta_{b,j,t} \sim Beta(g_{b,j,t}, h_{f,j,t}),$$

where $\delta_{f,j}$ and $\delta_{b,j}$ are also discount factors in the range (0,1), and the multiplicative innovations, $\eta_{f,j,t}$ and $\eta_{b,j,t}$ follow beta distributions (see Appendix and West and Harrison (1997) for details). The smoothness of innovation variance is controlled by both γ and δ. Similar to the PARCOR coefficients, we assume $\delta_{f,j} = \delta_{b,j} = \delta_j$ at each stage. Note that $\epsilon_{\beta,j,t}$, $\epsilon_{\beta,j,t}$, $\eta_{f,j,t}$ and $\eta_{b,j,t}$ are mutually independent and are also independent of any other variables in the model. The discount factors γ and δ are selected adaptively through a grid-search based on the likelihood (see the Appendix for details) in each iteration of MCMC.

We specify conjugate initial priors for the forward and backward PARCOR coef-

ficients, so that

$$p(\alpha_{j,0}|D_{f,j,0}, \sigma_{f,j,0}) \sim N(\mu_{f,j,0}, C_{f,j,0}),$$
$$p(\beta_{j,0}|D_{b,j,0}, \sigma_{b,j,0}) \sim N(\mu_{b,j,0}, C_{b,j,0}),$$

where $p = 1, \ldots, P$, $D_{f,j,0}$ and $D_{b,j,0}$ denotes the information available at the initial time $t = 0$, $\mu_{f,j,0}$ and $C_{f,j,0}$ are the mean and the variance of the normal prior distribution. We also specify conjugate initial priors for the forward and backward innovation variance, so that

$$p(\sigma^2_{f,j,0}|D_{f,j,0}) \sim G(\nu_{f,j,0}/2, \kappa_{f,j,0}/2),$$
$$p(\sigma^2_{b,j,0}|D_{b,j,0}) \sim G(\nu_{b,j,0}/2, \kappa_{b,j,0}/2),$$

where $G(\cdot, \cdot)$ is the gamma distribution, and $\nu_{f,j,0}/2$ and $\kappa_{f,j,0}/2$ are the shape and rate parameters for the gamma prior distribution. Usually, we treat these starting values as constants over all stages. In order to reduce the effect of the prior distribution, we choose $\mu_{f,j,0}/2$ and $C_{f,j,0}$ to be zero and one, respectively and fix $\nu_{f,j,0} = 1$ and set $\kappa_{f,j,0}$ to equal to the sample variance of the initial part of each series. In such prior settings, we can use the DLM sequential filtering and smoothing algorithms (West and Harrison, 1997) to derive the joint conditional posterior distributions of the forward and backward PARCOR coefficients and innovation variances in (4.3). Conditional on other the variables and the data, the full conditional distribution of the latent variable y_t can be easily obtained individually. To efficiently draw samples from individual y_t's full conditional distribution, we use the NUTS algorithm (Hoffman and Gelman, 2014) instead of a traditional random walk Metropolis. The detailed algorithm of the

sequential filtering and smoothing are given in the Appendix.

4.2.4 Model Selection

In order to determine the model order, we set a maximal order P_{max} and fit TV-Pois-AR(P) for $P = 1, \ldots, P_{max}$. The model selection criteria is computed one by one for any specified order. By comparing model selection criteria, we can select the best model order. Since Bayesian inference for the TV-Pois-AR model is conducted through MCMC simulations, we choose DIC (Spiegelhalter et al., 2002; Gelman et al., 2013) and WAIC (Watanabe, 2010, 2013) introduced in Section 3.2.5 as the model selection criteria.

4.2.5 Forecasting

Having estimated all parameters, we consider 1-step ahead forecasts of the TV-Pois-AR(P) model. Then, the 1-step ahead predictive posterior distribution of the PARCOR coefficients and innovation variance can be obtained according to West and Harrison (1997). The samples of the PARCOR coefficients and innovation variance can be drawn from their predictive distribution. The samples of the 1-step ahead prediction of the parameters $\mathbf{a}_{T+1} = (a_{1,T+1}, \ldots, a_{P,T+1})'$ can be obtained through the Durbin-Levinson algorithm from the samples of the PARCOR coefficients. After drawing the samples of innovation variance σ^2_{T+1} from its predictive distribution, the samples of y_{T+1} are drawn from its predictive distribution, such that,

$$y^{(j)}_{T+1} | y_{1:T}, \mathbf{a}^{(j)}_{T+1}, \sigma^{2(j)}_{T+1} \sim \mathrm{N}(\sum_{p=1}^{P} a^{(j)}_{p,T+1} y_{T+1-p}, \sigma^{2(j)}_{T+1}), \quad j = 1, \ldots, J.$$

With the samples of μ from its posterior distribution, the samples of the 1-step ahead forecast are given as

$$z_{T+1}^{(j)}|y_{T+1}^{(j)}, \mu^{(j)} \sim \text{Pois}(\exp(y_{T+1}^{(j)} + \mu^{(j)})), \quad j =, \ldots, J. \tag{4.5}$$

We use the posterior mean or median of z_{T+1} obtained through the samples in (4.5) as the 1-step ahead forecast. This forecast can be easily extended to h-steps ahead. The details of forecasting up to h-steps ahead can be found in the Appendix.

4.3 Simulation Study

In this section, first, we simulate the nonstationary Poisson time series from the exact TV-Pois-AR(P) model to evaluate the parameter estimation of the latent TV-AR process. Second, we generate a nonstationary Poisson time series based on a known time-dependent intensity parameter in order to compare our TV-Pois-AR model with other models.

4.3.1 Simulation 1

We simulated 100 time series for each of the lengths $T = 200, 300, 400$ from the following Poisson TV-AR(6) model, for $t = 1, \ldots, T$,

$$\begin{aligned} z_t|\mu, y_t &\sim \text{Pois}(\exp(\mu + y_t)), \\ y_t &= \sum_{j=1}^{6} a_{j,t} y_{t-j} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, 1), \end{aligned}$$

where $\mu = 3$. The latent process of y_t is the same time-varying TV-AR(6) process as in Rosen et al. (2009). This TV-AR(6) process can be defined as $a_t(B)y_t = \xi_t$, $t = 1, \ldots, T$, through a characteristic polynomial function $a_t(B)$, with B is the backshift operator (i.e., $B^p y_t = y_{t-p}$). In this TV-AR(6) process, the characteristic polynomial function is factorized as

$$a_t(B) = (1 - a_{t,1}B)(1 - a^*_{t,1}B)(1 - a_{t,2}B)(1 - a^*_{t,2}B)(1 - a_{t,3}B)(1 - a^*_{t,3}B),$$

where the superscript $*$ denotes the complex conjugate of a numbers. Also, let $a^{-1}_{t,j} = A_p \exp(2\pi i d_{t,j})$ for $p = 1, 2, 3$, where the $d_{t,j}s$ are defined by $d_{t,1} = 0.05 + (0.1/(T-1))t$, $d_{t,2} = 0.25$, and $d_{t,3} = 0.45 - (0.1/(T-1))t$ and the values of A_1, A_2 and A_3 equal to 1.1, 1.12, and 1.1, respectively. Based on DIC and WAIC, a model order of 6 is correctly selected. According to DIC and WAIC, 98% of the simulated datasets are identified to follow an order 2 model (TV-Pois-AR). Figure 4.1 and 4.2 shows the MSEs of 6 coefficients, the innovation variance, the mean level μ and the latent variable $\mathbf{y}$ over 100 simulated datasets. As expected, when the length of series increases, the TV-Pois-AR model gives more accurate estimation of each parameter.

4.3.2 Simulation 2 - An Empirical Example

In this study, we simulated the signals in a way, so that they exhibit similar properties as the COVID-19 data in New York State (see Section 4.4 for a complete discussion). We generated 100 time series of length $T = 278$ from the following Poisson process,

$$z_t | \lambda_t \sim \text{Pois}(\lambda_t), \quad t = 1, \ldots, T,$$

Figure 4.1: Boxplots of the MSEs of each of the six time-varying coefficient $a_{1,t}$ through $a_{6,t}$ for 100 simulated datasets of different length: 200, 300, 400.

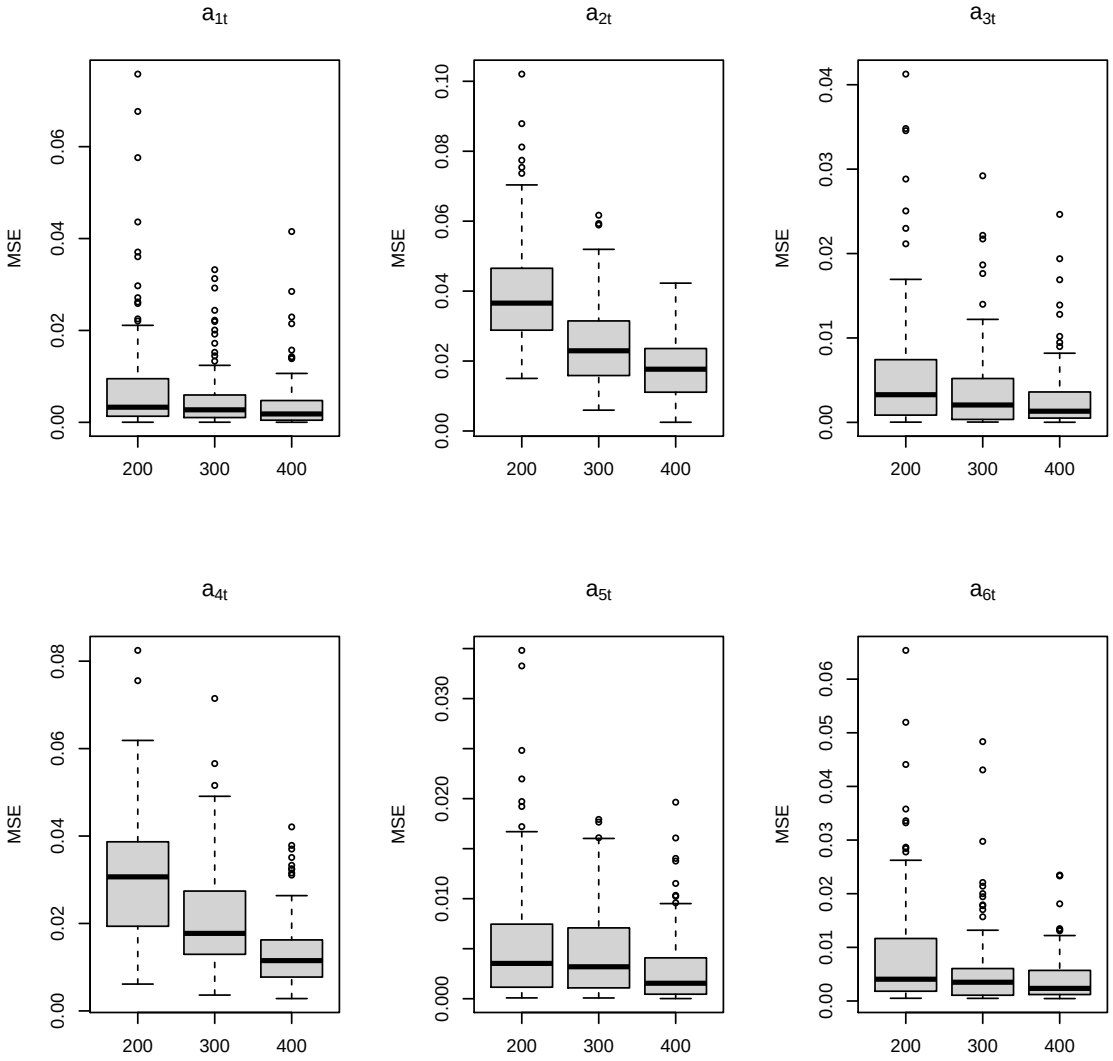


Figure 4.2: Boxplots of the MSEs of the intensity and the parameters in the latent process for 100 simulated datasets of different length. For each of them, three boxplots of length $200, 300, 400$ are put side by side from left to right. The left plot shows the MSEs of the innovation variance. The middle plot shows the MSEs of the mean level μ. The right plot shows the MSEs of the latent varaible $\mathbf{y}$.

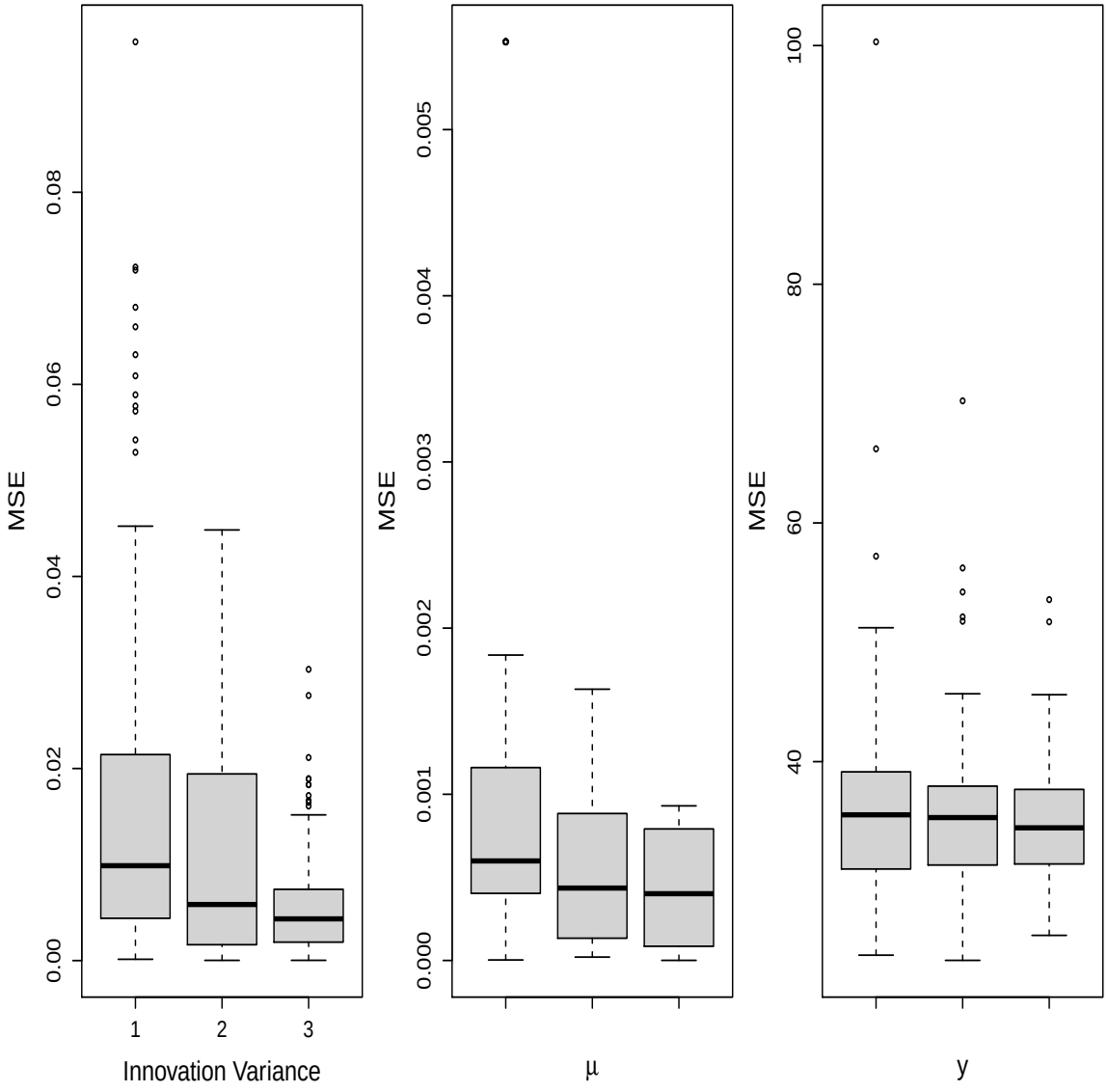
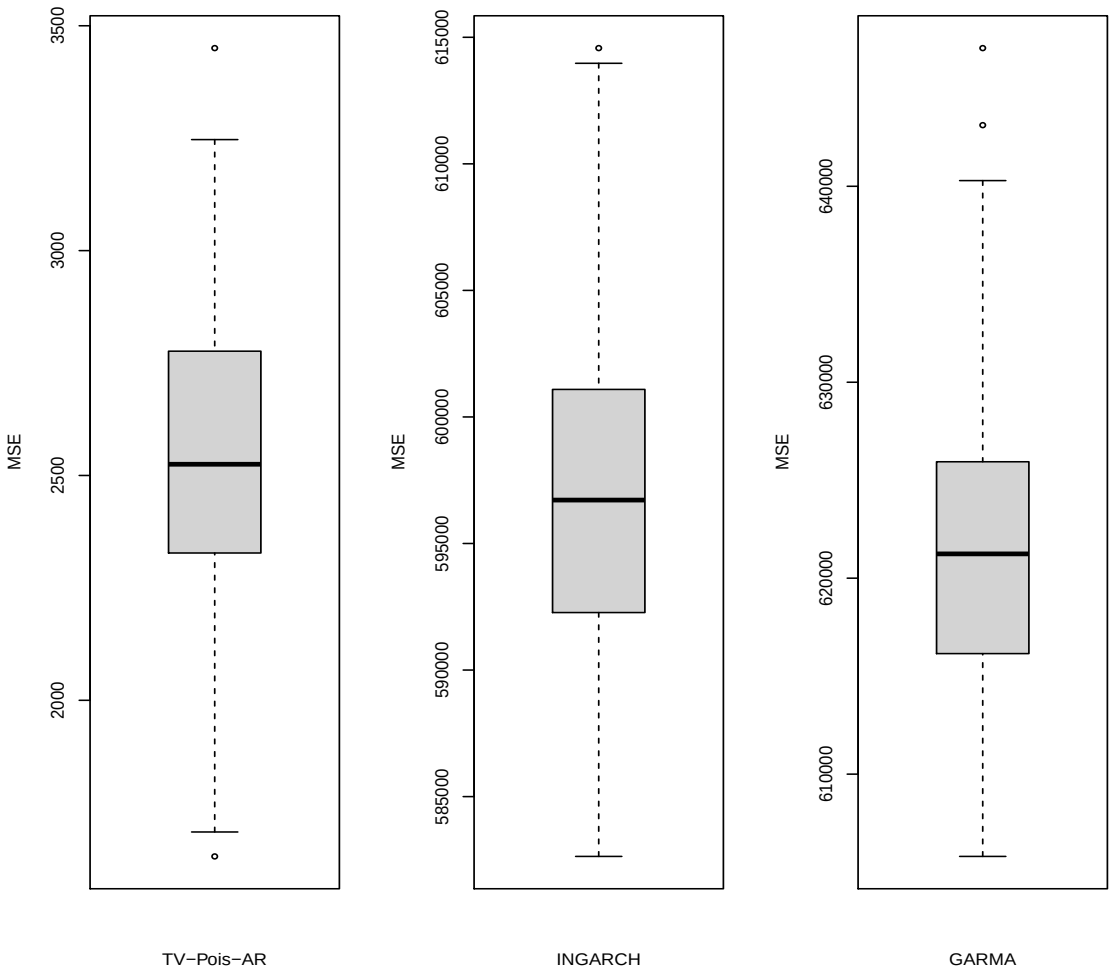


Table 4.1: Average Mean Squared Error (AMSE) of the estimated intensity over the 100 simulated datasets using different methods.

P	INGARCH	GARMA	TV-Pois-AR
AMSE	597041.3	621732.5	2568.6

where $\lambda_t = x_t + 1$ with x_t was daily new COVID-19 cases in New York State from 3/3/2020 to 12/5/2020. With this type of nonstationary signal, different models are compared by the estimation of the known time-varying intensity parameter λ_t, including the generalized autoregressive conditional heteroscedasticity model (INGARCH) and generalized linear autoregressive moving average (GARMA) model. The INGARCH and GARMA models are conducted by *tsglm* from R package *tscount*. Using the Akaike information criterion (AIC) and the quasi information criterion (QIC) Pan (2001), INGARCH(1,0) and GARMA(1,0) are selected. Both DIC and WAIC indicate that TV-Pois-AR(1) is the best model on these simulated datasets (see details in Section 4.2.4). To compare the parameter estimation of frequentist and Bayesian models, the average MSE (AMSE) of the Poisson intensity parameter is computed and shown in Table 4.1, where AMSE $= \frac{1}{100T}\sum_{s=1}^{100}\sum_{t=1}^{T}(\lambda_t - \hat{\lambda}_t)^2$. The boxplots in Figure 4.3 summarize the MSEs of 100 simulated datasets. As shown, TV-Pois-AR makes better forecasts on these simulated datasets. We expect the TV-Pois-AR model to have better performance than INGARCH and GARMA on similar pandemic data and other nonstationary count time series that shows similar characteristics.

Figure 4.3: Boxplots of MSE of the estimated intensity over 100 simulated datasets using different methods. Note that the scales of the three boxplots are different.



4.4 Case Study: COVID-19 in New York State

We collect the number of daily new confirmed COVID-19 cases in New York State from 3/3/2020 to 12/5/2020 from The COVID Tracking Project (https://covidtracking.com). We pick New York State data, as New York city remained an epicenter in the U.S. for about a month. Our research is motivated by the time-varying nature of the COVID-19 data. Inferences on the trend of the data may give us some insight into the spread of the COVID-19 and possibly insight into the effect of government interventions.

A TV-Pois-AR model is applied with the model order selected based on DIC and WAIC (see details in Section 3.2.5). By setting a maximum order of 5, order 2 is considered as the best. The difference between the estimated exponential of the intensity parameter, $\exp(\lambda_t)$ and the observed series are shown in Figure 4.4. Figure 4.5 shows the estimated parameters. Table 4.2 shows the model selection results. A series of restrictions in New York State began on 3/12/2020 and a state-wide stay-at-home order was declared on 3/20/2020. The number of new cases reached its peak about two week after the lockdown. In Figure 4.5, the first dashed line in the first two plots denotes 3/20/2020 the time when state-wide stay-at-home order was declared. We can see the estimated autoregressive coefficients keep changing significantly after this date. This change in the autoregressive coefficients coincides with the lockdown process. The second dashed line in the first two plots denotes 9/26/2020. On that day, the number showed an uptick in cases, with more than 1,000 daily COVID-19 cases, which was the first time since early June. About two weeks before this date, the coefficients show some evidence of pattern change. This may be an indication that the lockdown affected the spread of the COVID-19. The

innovation variance of the intensity is getting smaller and smaller probably due to the improvement in testing and reporting. The dashed line in the third plot denotes the date when the peak number of new cases occurred. Since then, the innovation variance was stabilized to a low level.

To evaluate the performance of forecasting, we make rolling 20-day-ahead prediction and compare the mean squared prediction error (MSPE). We pick a starting date and make 20-day-ahead prediction based on the data up to this date. Then, we move to the next day and make 20-day-ahead prediction based on the data up to the new date. By repeating this until 20 days before the last day, we obtain the rolling 20-day-ahead prediction.

We compare the forecast performance of four methods, where Naive denotes the naive forecast, that is, using the previous period to forecast for the next period (carry-forward). Table 4.3 shows the percentage of TV-Pois-AR model's 20-day-ahead forecast performance beating other methods in terms of MSPE over the rolling observed COVID data. Tables 4.4 and 4.5 show the same percentage when using the posterior mean and the posterior median as forecast values. We pick two starting dates for the rolling predictions: 7/19/2020 and 9/19/2020. The first date is picked to give at least 100 observations of fitting data and 7/19/2020 is two weeks after the New York state entered Phase 3 of reopening. Since, at the time, the common COVID-19 quarantine period was 14 days, we use it as a starting date to conduct prediciton. In other words, we suppose the reopening took 14 days to show an impact. The second date allows at least 200 observations for model fitting. Our simulations studies showed the model fitting and the model selection works ideally for at least 200 observations. The rolling prediction is conducted from the selected start day to

Table 4.2: Model selection of Poisson TV-AR model for the daily new COVID-19 cases in New York State. Each column gives the model order P and the value of the model selection criteria.

P	1	2	3	4	5
DIC	1738.540	1735.748	1783.484	1779.946	1796.255
WAIC	303.004	295.531	317.326	316.626	320.476

Table 4.3: Percentage of TV-Pois-AR(2) giving better forecasts of 20-step-ahead rolling prediction on COVID-19 data in New York State from 3/3/2020 to 12/5/2020. The posterior means of the intensity are used as the forecast values. There are two start dates of the rolling predictions.

Start Date	7/19/2020	9/19/2020
GARMA(1,0)	94.2%	89.8%
INGARCH(12,2)	83.5%	78.0%
Naive	56.2%	71.2%

the last available day. The average MSPE over these days is used for comparison. We evaluate three different forecast values (Tables 4.3 through 4.5). The rolling prediction results suggest that the posterior median of the future observations gives a better forecasts than the posterior mean of the future observations and the posterior mean of the intensity (see Appendix for additional forecasting details). Moreover, the posterior median of the future observations outperforms the two existing models and the Naive forecasting. Figure 4.6 shows an example of 20-day forecast of the daily new COVID-19 cases from 10/18/2020 to 11/6/2020 in New York State using the posterior mean of the intensity as the forecast values. Figures 4.7 and 4.8 show the same example with the posterior mean and the posterior median as forecast values. The example demonstrates how the TV-Pois-AR model captures the time-varying trend.

Figure 4.4: The difference between daily new COVID-19 cases in New York State and the estimated expected values. The black line is the difference and the grey region shows the corresponding 90% credible intervals. The top plot shows the difference in the original scale and the bottom plot show the difference in log scale.

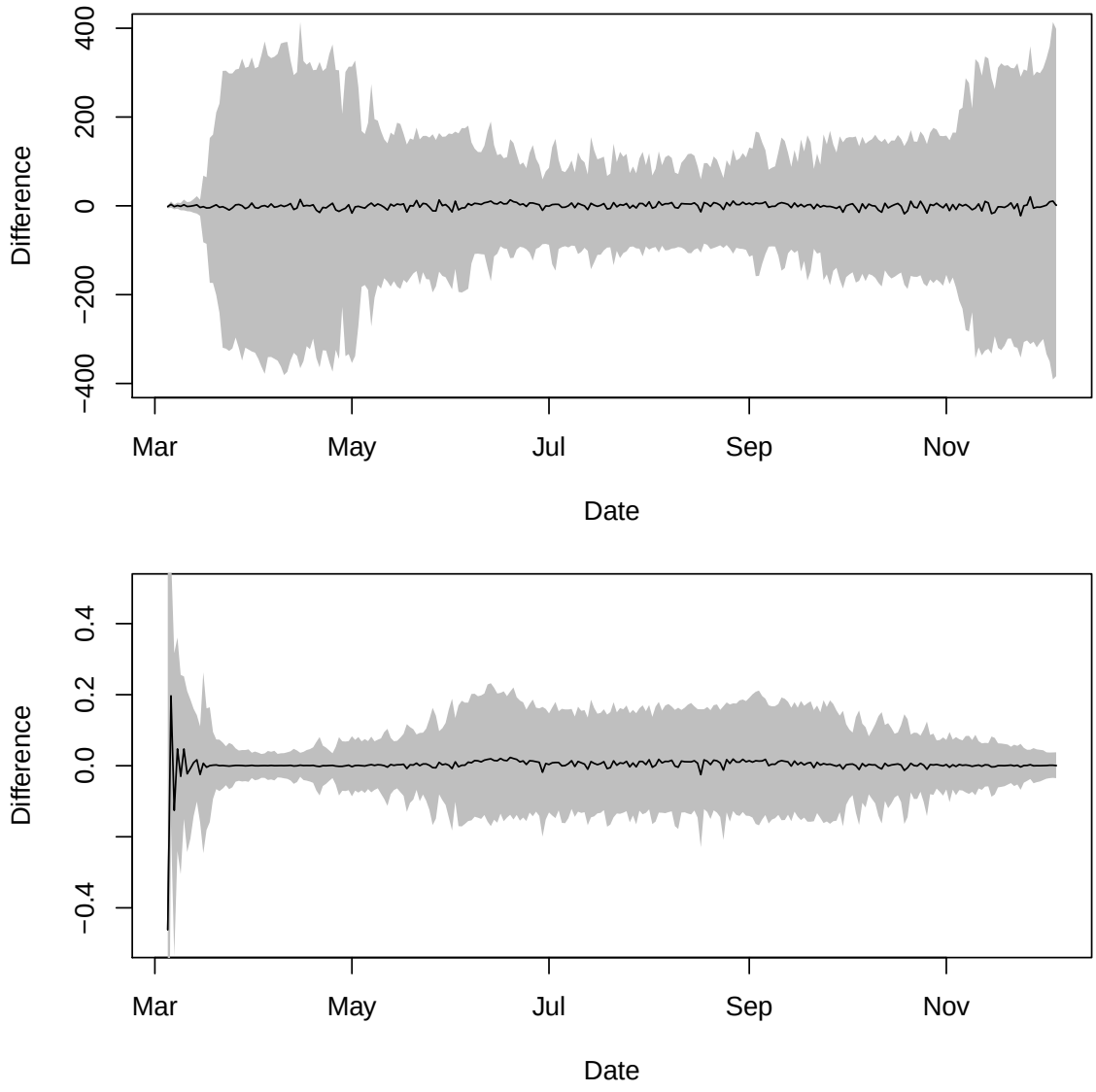


Figure 4.5: The estimated $a_{1,t}$, $a_{2,t}$, and σ_t^2 of Poisson TV-VAR(2) model applied to daily new COVID-19 cases in New York State from top to bottom, respectively. The grey region shows the corresponding 90% credible intervals.

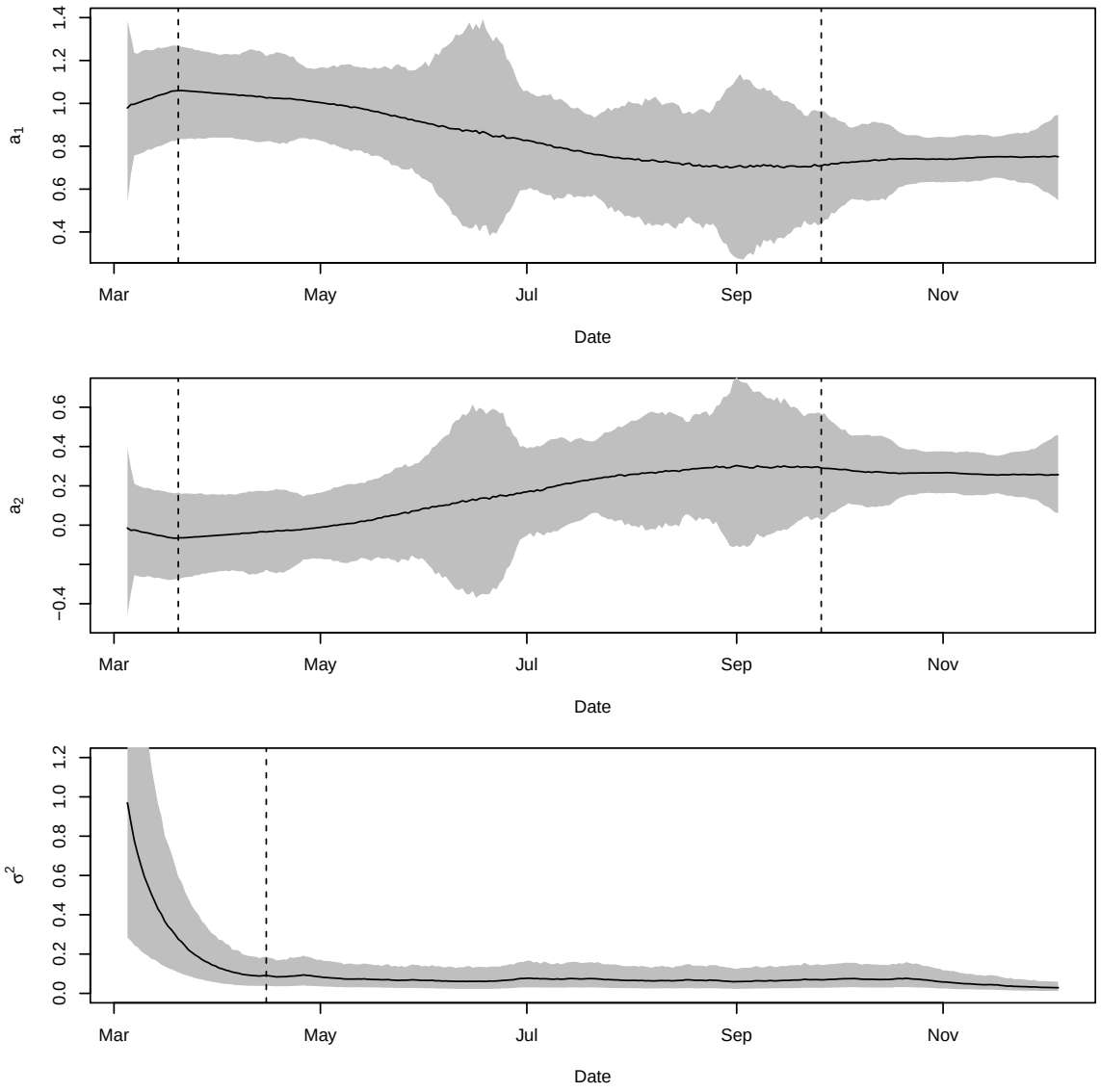


Figure 4.6: The 20-day forecast of the daily new COVID-19 cases of the last 20 days in New York State. The posterior means of the intensity are used as the forecast values. The black overplotted points and lines are the observed daily new cases used for model fitting from 3/3/2020 to 10/17/2020. The black dots are the true daily new cases in the forecast region from 10/18/2020 to 11/6/2020. The blue line shows the 20-day forecast. The light blue region is the 90% prediction interval.

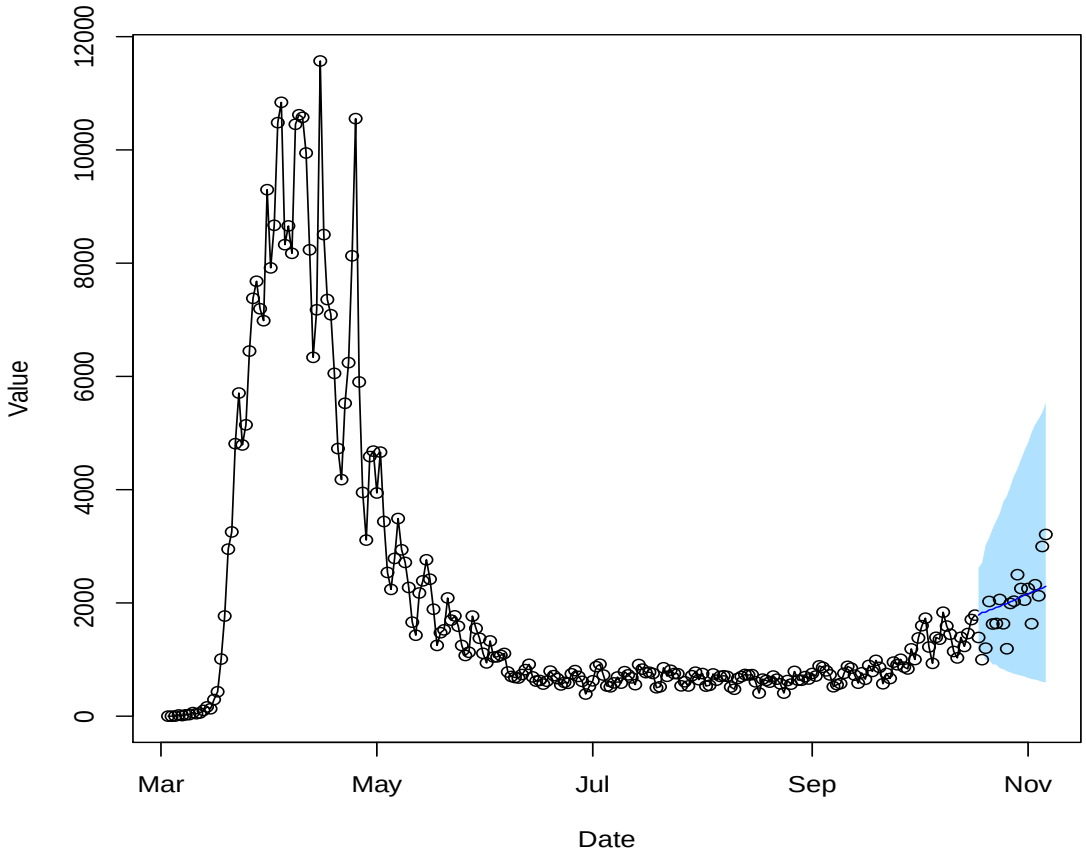


Figure 4.7: The 20-day forecast of the daily new COVID-19 cases of the last 20 days in New York State. The posterior means of the future observations are used as the forecast values. The black overplotted points and lines are the observed daily new cases used for model fitting from 3/3/2020 to 10/17/2020. The black dots are the true daily new cases in the forecast region from 10/18/2020 to 11/6/2020. The blue line shows the 20-day forecast. The light blue region is the 90% prediction interval.

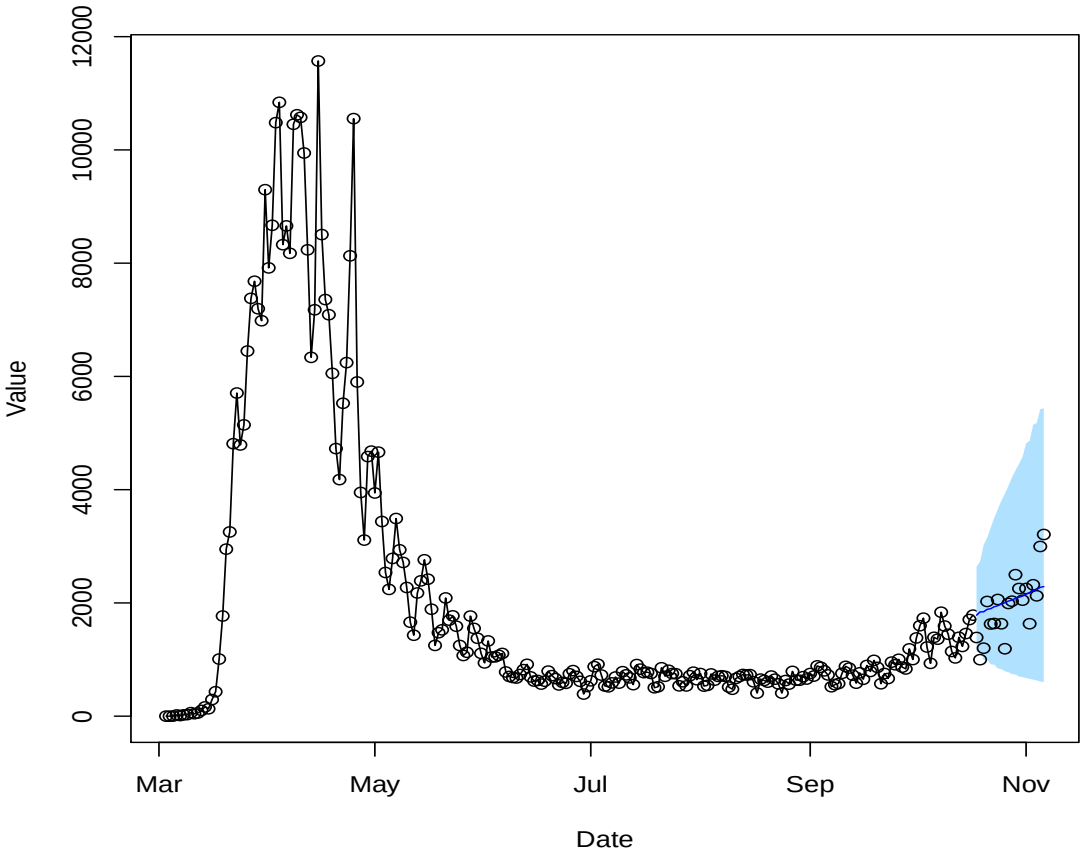


Figure 4.8: The 20-day forecast of the daily new COVID-19 cases of the last 20 days in New York State. The posterior medians of the future observations are used as the forecast values. The black overplotted points and lines are the observed daily new cases used for model fitting from 3/3/2020 to 10/17/2020. The black dots are the true daily new cases in the forecast region from 10/18/2020 to 11/6/2020. The blue line shows the 20-day forecast. The light blue region is the 90% prediction interval.

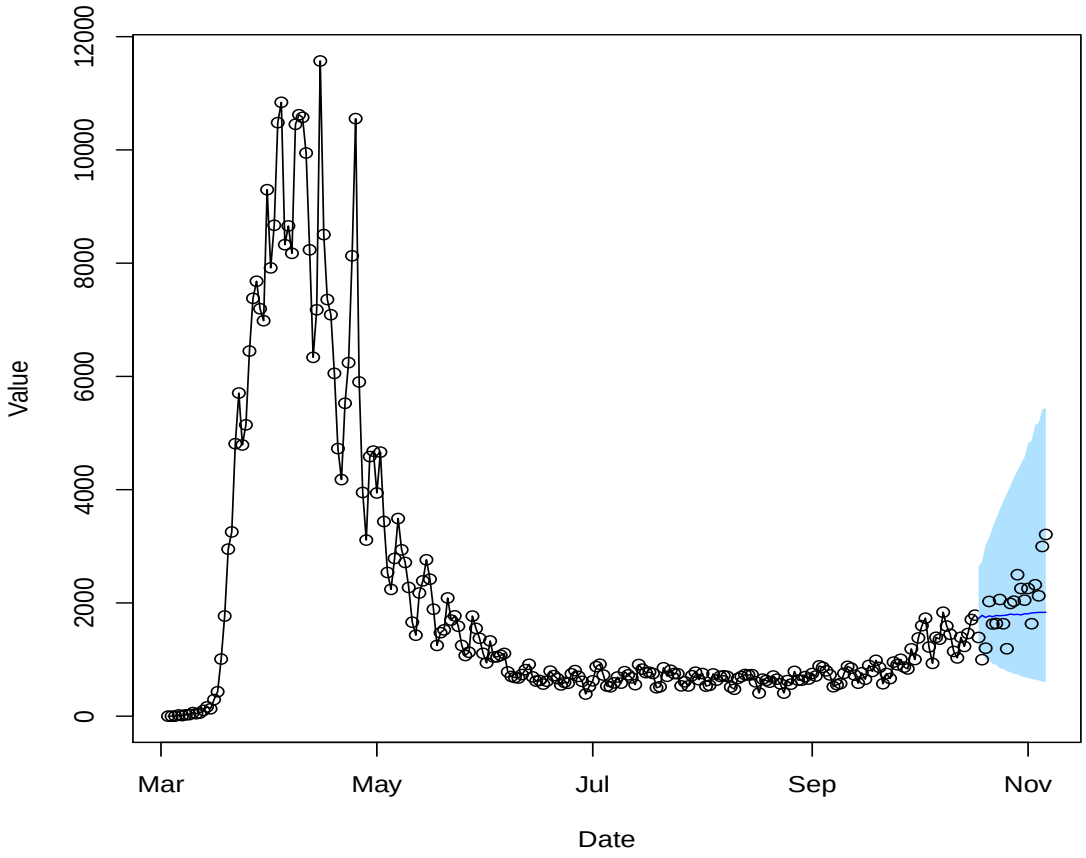


Table 4.4: Percentage of TV-Pois-AR(2) giving better forecasts of 20-step-ahead rolling prediction on COVID-19 data in New York State from 3/3/2020 to 12/5/2020. The posterior means of the future observations are used as the forecast values. There are two start dates of the rolling predictions. Posterior mean used as forecasts.

Start Date	7/19/2020	9/19/2020
GARMA(1,0)	56.2%	94.9%
INGARCH(12,2)	65.3%	86.4%
Naive	50.41%	89.8%

Table 4.5: Percentage of TV-Pois-AR(2) giving better forecasts of 20-step-ahead rolling prediction on COVID-19 data in New York State from 3/3/2020 to 12/5/2020. The posterior medians of the future observations are used as the forecast values. There are two start dates of the rolling predictions. Posterior median used as forecasts.

Start Date	7/19/2020	9/19/2020
GARMA(1,0)	82.6%	100%
INGARCH(12,2)	96.6%	96.6%
Naive	57.9%	100%

4.5 Discussion

We develop a novel hierarchical Bayesian model to model the nonstationary count time-varying models and propose an efficient estimation approach using a MCMC sampling scheme with embedded NUTS algorithm. We also provide a model selection method to choose the discount factors and the optimal model order. The simulation cases show that the estimation of parameters is extremely accurate. The real data example shows that the time-varying coefficients and innovation covariance can reveal the changing pattern over time. The proposed method can be applied to not only the confirmed cases of COVID-19 but also the number of deaths, number of recovered cases, number of critical cases, and many others for different diseases. Such studies may provide important insights into the spread and measures required.

One subject of future research is an extension to multivariate and/or spatio-temporal cases by adding some region-specific effects and modeling the series in multiple regions jointly. Modeling multivariate count-valued time series data is an important research topic in ecology and climatology. Moreover, regarding univariate applications on epidemic disease data, we can consider different government intervention as covariates, which usually has an essential impact on the spread of any infectious disease.

4.6 Appendix

4.6.1 Algorithm of Fitting Poisson TV-AR Time Series

To fit a TV-Pois-AR model We use Gibbs sampling to generate samples from the full conditional distribution of each of the parameters and the latent variables iteratively. After burn-in, the sample distributions of these parameters and latent variables are the estimated posterior distributions. These samples are generated by a Gibbs sampling with the following steps:

- Draw samples of $\mathbf{y} = (y_1, \ldots, y_T)'$ from the full conditional distritbuion $p(\mathbf{y}|\mathbf{z}, \mu, \mathbf{a}, \boldsymbol{\sigma}^2) \propto p(\mathbf{z}|\mathbf{y}, \mu)p(\mathbf{y}|\mathbf{a}, \boldsymbol{\sigma}^2)$ using a No-U-Turn Sampler (Hoffman and Gelman, 2014).
- Draw samples of $\mathbf{a}$ and $\boldsymbol{\sigma}^2$ from the the joint full conditional distribution $p(\mathbf{a}, \boldsymbol{\sigma}^2|\mathbf{y})$ obtained from the BLF, where $\mathbf{a} = (a_1, \ldots, a_T)'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_T^2)'$.
- Draw samples of μ from the full conditional distribution $p(\mu|\mathbf{z}, \mathbf{y}) \propto p(\mathbf{z}|\mathbf{y}, \mu)p(\mu)$ using a a No-U-Turn Sampler.

4.6.2 Bayesian Lattice Filter

- Step 1. Repeat Step 2 for stage $p = 1, \ldots, P$.
- Step 2. Apply the sequential filtering and smoothing algorithm (see 3.6.2) to the prediction errors of last stage, $f_t^{(p-1)}$ and $b_t^{(p-1)}$ to obtain $\hat{\alpha}_t^{(p)} = \mu_t^{(p)}$ and $\hat{\sigma^2}_t^{(p)} = s_t^{(p)}$ of the forward and backward equations, and the forward and backward prediction errors, $f_t^{(p)}$ and $b_t^{(p)}$ for $t = 1, \ldots, T$.

- Step 3. Compute the full conditional distribution of $\mathbf{a}$ and σ^2 based on the results of Step 2.

4.6.3 Forecasting

We can make h-step ahead forecasting by following these steps.

- For stage p, compute the h-step ahead predictive distribution of the PARCOR coefficients following West and Harrison (1997): $(\alpha_{p,T+h}^{(p)}|D_T) \sim N(\mu_T^{(p)}(h), C_T^{(p)}(h))$ where $\mu_T(h) = \mu_T^{(p)}$ and $C_T^{(p)}(h) = C_T^{(p)} + hG_{T+h}^{(p)}$ with $G_{T+1}^{(p)} = C_t^{(p)}(1-\beta)/\beta$.

- Draw J samples of $\{\alpha_{p,T+h}^{(p)}, p = 1, \ldots, P\}$ from their predictive distribution.

- For stage p, compute the h-step ahead predictive distribution of innovation variance following West and Harrison (1997): $(\sigma_{T+h}^{2(p)}|D_T) \sim G(\nu_T^{(p)}(h)/2, \kappa_T^{(p)}(h)/2)$ where $\nu_T^{(p)}(h) = \delta^h \nu_T^{(p)}$ and $\kappa_T^{(p)}(h) = \delta^h \kappa_T^{(p)}$.

- For stage p, draw J samples of $\sigma_{T+h}^{2(p)}$ from the its predictive distribution $G(\nu_{k+(T-1)K}^{(p)}(h)/2, \kappa_{k+(T-1)K}^{(p)}(h)/2)$.

- Compute the samples of the AR coefficients $\{a_{p,T+h}^{(P)}, p = 1, \ldots, P\}$ through the Durbin-Levinson algorithm from the samples of $\{\alpha_{p,T+h}^{(P)}, p = 1, \ldots, P\}$.

- The samples of y_{T+h} are generated from its predictive distribution, such that,

$$y_{T+h}^{(j)}|y_{1:T}, y_{T:(T+h-1)}^{(j)}, \mathbf{a}_{T+h}^{(j)}, \sigma_{T+h}^{2(j)} \sim \mathrm{N}(\sum_{p=1}^{P} a_{p,T+h}^{(j)} y_{T+h-p}^{(j)}, \sigma_{T+h}^{2(j)}), \quad j = 1, \ldots, J,$$

where $y_{T+h-p}^{(j)} = y_{T+h-p}$ if $h - p \leq 0$.

- With the samples of μ from its posterior distribution, the samples of the h-step ahead forecast are drawn as

$$z_{T+h}^{(j)} | y_{T+h}^{(j)}, \mu^{(j)} \sim \text{Pois}(\exp(y_{T+h}^{(j)} + \mu^{(j)})), \quad j =, \ldots, J. \tag{4.6}$$

- We use the posterior mean or median of z_{T+h} obtained through the samples in (4.6) as the h-step ahead forecast.

Bibliography

Abanto-Valle, C. A., Migon, H. S., and Lopes, H. F. (2010). "Bayesian modeling of financial returns: A relationship between volatility and trading volume." *Applied Stochastic Models in Business and Industry*, 26, 2, 172–193.

Aktekin, T., Polson, N., Soyer, R., et al. (2018). "Sequential Bayesian analysis of multivariate count data." *Bayesian Analysis*, 13, 2, 385–409.

Al-Osh, M. and Alzaid, A. A. (1987). "First-order integer-valued autoregressive (INAR (1)) process." *Journal of Time Series Analysis*, 8, 3, 261–275.

Berry, L. R. and West, M. (2019). "Bayesian forecasting of many count-valued time series." *Journal of Business & Economic Statistics*, 1–15.

Black, F. (1976). "Studies of stock market volatility changes." *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 177––181.

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2020). "Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural

exponential family." *Journal of the American Statistical Association*, 115, 532, 2037–2052.

Bradley, J. R., Holan, S. H., Wikle, C. K., et al. (2018). "Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion)." *Bayesian Analysis*, 13, 1, 253–310.

Brandt, P. T. and Williams, J. T. (2001). "A linear Poisson autoregressive model: The Poisson AR (p) model." *Political Analysis*, 164–184.

Brandt, P. T., Williams, J. T., Fordham, B. O., and Pollins, B. (2000). "Dynamic modeling for persistent event-count time series." *American Journal of Political Science*, 823–843.

Brooks, S. P. and Gelman, A. (1998). "General methods for monitoring convergence of iterative simulations." *Journal of Computational and Graphical Statistics*, 7, 4, 434–455.

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81, 3, 541–553.

Carvalho, C. M. and Lopes, H. F. (2007). "Simulation-based sequential analysis of Markov switching stochastic volatility models." *Computational Statistics & Data Analysis*, 51, 9, 4526–4542.

Christie, A. A. (1982). "The stochastic behavior of common stock variances: Value, leverage and interest rate effects." *Journal of Financial Economics*, 10, 4, 407–432.

Clark, P. K. (1973). "A subordinated stochastic process model with finite variance for speculative prices." *Econometrica: Journal of the Econometric Society*, 135–155.

Davis, G. M. and Ensor, K. B. (2007). "Multivariate Time-Series Analysis With Categorical and Continuous Variables in an LSTR Model." *Journal of Time Series Analysis*, 28, 6, 867–885.

Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N., eds. (2016). *Handbook of Discrete-Valued Time Series*. Chapman & Hall/CRC.

Del Negro, M. and Primiceri, G. E. (2015). "Time varying structural vector autoregressions and monetary policy: a corrigendum." *The Review of Economic Studies*, 82, 4, 1342–1345.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). "Hybrid Monte Carlo." *Physics Letters B*, 195, 2, 216–222.

Dunsmuir, W. T. (2016). "Generalized Linear Autoregressive Moving Average Models." *Handbook of Discrete-Valued Time Series*, 51.

Durham, G. B. (2006). "Monte Carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models." *Journal of Econometrics*, 133, 1, 273–305.

Engle, R. F. and Ng, V. K. (1993). "Measuring and testing the impact of news on volatility." *The Journal of Finance*, 48, 5, 1749–1778.

Escobar, M. D. and West, M. (1995). "Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90, 430, 577–588.

Fan, J. and Yao, Q. (2008). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Science & Business Media.

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 209–230.

Ferland, R., Latour, A., and Oraichi, D. (2006). "Integer-valued GARCH process." *Journal of Time Series Analysis*, 27, 6, 923–942.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). "Poisson autoregression." *Journal of the American Statistical Association*, 104, 488, 1430–1439.

Frühwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models." *Journal of Time Series Analysis*, 15, 2, 183–202.

— (2006). *Finite Mixture and Markov Switching Models*. Springer Science & Business Media.

Gallant, A. R., Hsieh, D., and Tauchen, G. (1997). "Estimation of stochastic volatility models with diagnostics." *Journal of Econometrics*, 81, 1, 159–192.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC.

Gelman, A., Meng, X.-L., and Stern, H. (1996). "Posterior predictive assessment of model fitness via realized discrepancies." *Statistica Sinica*, 6, 4, 733–760.

Gersch, W. and Stone, D. (1994). "One channel at-a-time multichannel autoregressive modeling of stationary and nonstationary time series." In *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach: Volume 3 Engineering and Scientific Applications*, eds. H. Bozdogan, S. L. Sclove, A. K. Gupta, D. Haughton, G. Kitagawa, T. Ozaki, and K. Tanabe, 165–192. Springer.

— (1995). "Multivariate autoregressive time semes modeling: one scalar autoregressive model at-a-time." *Communications in Statistics-Theory and Methods*, 24, 11, 2715–2733.

Gervais, S., Kaniel, R., and Mingelgrin, D. H. (2001). "The High-Volume Return Premium." *The Journal of Finance*, 56, 3, 877–919.

Geweke, J. (2007). "Interpretation and inference in mixture models: Simple MCMC works." *Computational Statistics & Data Analysis*, 51, 7, 3529–3550.

Guo, W. and Dai, M. (2006). "Multivariate time-dependent spectral analysis using Cholesky decomposition." *Statistica Sinica*, 16, 3, 825.

Hamilton, J. D. and Susmel, R. (1994). "Autoregressive conditional heteroskedasticity and changes in regime." *Journal of Econometrics*, 64, 1, 307–333.

Hayes, M. H. (1996). *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons.

Hoffman, M. D. and Gelman, A. (2014). "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.*, 15, 1, 1593–1623.

Huerta, G. and Lopes, H. F. (2000). "Bayesian forecasting and inference in latent structure for the brazilian industrial production index." *Brazilian Review of Econometrics*, 20, 1, 1–26.

Hunter, J., Burke, S. P., and Canepa, A. (2017). *Multivariate Modelling of Non-Stationary Economic Time Series*. Springer.

Jacquier, E., Polson, N., and Rossi, P. (1994). "Bayesian analysis of stochastic volatility models (with discussion)." *Journal Bussines of Econometrics Statistics, 12 (4)*, 371–417.

Jensen, M. J. and Maheu, J. M. (2010). "Bayesian semiparametric stochastic volatility modeling." *Journal of Econometrics*, 157, 2, 306–316.

— (2014). "Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture." *Journal of Econometrics*, 178, 523–538.

Kalimipalli, M. and Susmel, R. (2004). "Regime-switching stochastic volatility and short-term interest rates." *Journal of Empirical Finance*, 11, 3, 309–329.

Karpoff, J. M. (1987). "The relation between price changes and trading volume: A survey." *Journal of Financial and Quantitative Analysis*, 22, 1, 109–126.

Kastner, G., Frühwirth-Schnatter, S., and Lopes, H. F. (2017). "Efficient Bayesian inference for multivariate factor stochastic volatility models." *Journal of Computational and Graphical Statistics*, 26, 4, 905–917.

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic volatility: likelihood inference and comparison with ARCH models." *The Review of Economic Studies*, 65, 3, 361–393.

Kitagawa, G. (2010). *Introduction to Time Series Modeling*. Chapman & Hall/CRC.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*. Springer.

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017). "A Bayesian multivariate functional dynamic linear model." *Journal of the American Statistical Association*, 112, 518, 733–744.

Lam, K.-c., Li, W., and Wong, P. (1990). "Price changes and trading volume relationship in the Hong Kong stock market." *Asia Pacific Journal of Management*, 7, 2, 25–42.

Li, W. and Lam, K. (1995). "Modelling asymmetry in stock returns by a threshold autoregressive conditional heteroscedastic model." *The Statistician*, 333–341.

Liesenfeld, R. and Jung, R. (2000). "Stochastic volatility models: conditional normality versus heavy-tailed distributions." *Journal of Applied Econometrics*, 5, 137–160.

Livingston Jr, G. and Nur, D. (2017). "Bayesian inference for smooth transition autoregressive (STAR) model: A prior sensitivity analysis." *Communications in Statistics-Simulation and Computation*, 46, 7, 5440–5461.

Lopes, H. F. and Carvalho, C. M. (2007). "Factor stochastic volatility with time varying loadings and Markov switching regimes." *Journal of Statistical Planning and Inference*, 137, 10, 3082–3091.

Lopes, H. F. and Salazar, E. (2006). "Bayesian model uncertainty in smooth transition autoregressions." *Journal of Time Series Analysis*, 27, 1, 99–117.

Lubrano, M. (2000). "Bayesian analysis of nonlinear time series models with a threshold." In *Nonlinear Econometric Modeling in Time Series: Proceedings of the*

Eleventh International Symposium in Economic Theory, vol. 11, 79. Cambridge University Press.

Mahieu, R. J. and Schotman, P. C. (1998). "An empirical application of stochastic volatility models." *Journal of Applied Econometrics*, 13, 4, 333–360.

Martino, L., Yang, H., Luengo, D., Kanniainen, J., and Corander, J. (2015). "A fast universal self-tuned sampler within Gibbs sampling." *Digital Signal Processing*, 47, 68–83.

Masry, E. (1996). "Multivariate local polynomial regression for time series: uniform strong consistency and rates." *Journal of Time Series Analysis*, 17, 6, 571–599.

Nakajima, J., Kasuya, M., and Watanabe, T. (2011). "Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy." *Journal of the Japanese and International Economies*, 25, 3, 225–245.

Nakajima, J. and West, M. (2013). "Bayesian analysis of latent threshold dynamic models." *Journal of Business & Economic Statistics*, 31, 2, 151–164.

Neal, R. M. (1994). "An improved acceptance procedure for the hybrid Monte Carlo algorithm." *Journal of Computational Physics*, 111, 1, 194–203.

Neal, R. M. et al. (2011). "MCMC using Hamiltonian dynamics." *Handbook of Markov Chain Monte Carlo*, 2, 11, 2.

Ombao, H., Raz, J., Von Sachs, R., and Malow, B. (2001). "Automatic statistical analysis of bivariate nonstationary time series." *Journal of the American Statistical Association*, 96, 454, 543–560.

Ombao, H., Von Sachs, R., and Guo, W. (2005). "SLEX analysis of multivariate nonstationary time series." *Journal of the American Statistical Association*, 100, 470, 519–531.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference." *Journal of Econometrics*, 140, 2, 425–449.

Pagano, M. (1978). "On periodic and multiple autoregressions." *The Annals of Statistics*, 6, 6, 1310–1317.

Pan, W. (2001). "Akaike's information criterion in generalized estimating equations." *Biometrics*, 57, 1, 120–125.

Primiceri, G. E. (2005). "Time varying structural vector autoregressions and monetary policy." *The Review of Economic Studies*, 72, 3, 821–852.

Rosen, O., Stoffer, D. S., and Wood, S. (2009). "Local spectral analysis via a Bayesian mixture of smoothing splines." *Journal of the American Statistical Association*, 104, 485, 249–262.

Sakai, H. (1982). "Circular lattice filtering using Pagano's method." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30, 2, 279–287.

Shephard, N. (1994). "Partial non-Gaussian state space." *Biometrika*, 81, 1, 115–131.

— (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press on Demand.

Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and Its Applications (2nd ed)*. Springer.

Smith, R. and Miller, J. (1986). "A non-Gaussian state space model and application to prediction of records." *Journal of the Royal Statistical Society: Series B (Methodological)*, 48, 1, 79–88.

So, M. E. P., Lam, K., and Li, W. K. (1998). "A stochastic volatility model with Markov switching." *Journal of Business & Economic Statistics*, 16, 2, 244–253.

So, M. K. and Choi, C. (2009). "A threshold factor multivariate stochastic volatility model." *Journal of Forecasting*, 28, 8, 712–735.

So, M. K., Li, W., and Lam, K. (2002). "A threshold stochastic volatility model." *Journal of Forecasting*, 21, 7, 473–500.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 4, 583–639.

Triantafyllopoulos, K. (2007). "Covariance estimation for multivariate conditionally Gaussian dynamic linear models." *Journal of Forecasting*, 26, 8, 551–569.

Virbickaitė, A., Lopes, H. F., Ausín, M. C., and Galeano, P. (2019a). "Particle learning for Bayesian semi-parametric stochastic volatility model." *Econometric Reviews*.

Virbickaitė, A., Lopes, H. F., et al. (2019b). "Bayesian semiparametric Markov switching stochastic volatility model." *Applied Stochastic Models in Business and Industry*, 35, 4, 978–997.

Wang, J. C. and Holan, S. H. (2012). "Bayesian multi-regime smooth transition regression with ordered categorical variables." *Computational Statistics and Data Analysis*.

Watanabe, S. (2010). "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory." *Journal of Machine Learning Research*, 11, Dec, 3571–3594.

— (2013). "A widely applicable Bayesian information criterion." *Journal of Machine Learning Research*, 14, Mar, 867–897.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (2nd)*. Springer.

West, M., Prado, R., and Krystal, A. D. (1999). "Evaluation and comparison of EEG traces: latent structure in nonstationary time series." *Journal of the American Statistical Association*, 94, 446, 375–387.

Yang, W.-H., Holan, S. H., and Wikle, C. K. (2016). "Bayesian lattice filters for time–varying autoregression and time–frequency analysis." *Bayesian Analysis*, 11, 4, 977–1003.

Yu, J. (2012). "A semiparametric stochastic volatility model." *Journal of Econometrics*, 167, 2, 473–482.

Zeger, S. L. (1988). "A regression model for time series of counts." *Biometrika*, 75, 4, 621–629.

Zhao, W. and Prado, R. (2020). "Effcient Bayesian PARCOR approaches for dynamic modeling of multivariate time series." *Journal of Time Series Analysis*.

VITA

Yuelei Sui was born in Shandong, China in 1984. After graduating with a Bachelor of Statistics degree from Shanghai University of Finance & Economics in 2007, he attended University of Missouri - Kansas City and obtained a degree of Master of Mathematics and Statistics with concentration in Statistics in 2010. He then entered the University of Missouri and began research on this project with Professor Scott H. Holan since 2011.