Bayesian Markov–Switching Vector Autoregressive Process

Battulga Gankhuu*

Abstract

This study introduces marginal density functions of the general Bayesian Markov–Switching Vector Autoregressive (MS–VAR) process. In the case of the Bayesian MS–VAR process, we provide closed–form density functions and Monte–Carlo simulation algorithms, including the importance sampling method. The Monte–Carlo simulation method departs from the previous simulation methods because it removes the duplication in a regime vector.

Keywords: Bayesian MS–VAR process, Monte–Carlo simulation method, Stochastic DDM.

1 Introduction

Classic Vector Autoregressive (VAR) process was proposed by Sims (1980) who criticize large–scale macro–econometric models, which are designed to model interdependencies of economic variables. Besides Sims (1980), there are some other important works on multiple time series modeling, see, e.g., Tiao and Box (1981), where a class of vector autoregressive moving average models was studied. For the VAR process, a variable in the process is modeled by its past values and the past values of other variables in the process. After the work of Sims (1980), VARs have been used for macroeconomic forecasting and policy analysis. However, if the number of variables in the system increases or the time lag is chosen high, then too many parameters need to be estimated. This will reduce the degrees of freedom of the model and entail a risk of over–parametrization.

Therefore, to reduce the number of parameters in a high–dimensional VAR process, Litterman (1979) introduced probability distributions for coefficients that are centered at the desired restrictions but that have a small and nonzero variance. Those probability distributions are known as Minnesota prior in Bayesian VAR (BVAR) literature, which is widely used in practice. Due to over–parametrization, the generally accepted result is that the forecast of the BVAR model is better than the VAR model estimated by the frequentist technique. Research works have shown that BVAR is an appropriate tool for modeling large data sets; for example, see Bańbura, Giannone, and Reichlin (2010).

Sudden and dramatic changes in the financial market and economy are caused by events such as wars, market panics, or significant changes in government policies. To model those events, some authors used regime–switching models. The regime–switching model was introduced by seminal works of Hamilton (1989, 1990) (see also books of Hamilton (1994) and Krolzig (1997)), and the model is hidden Markov model with dependencies; see Zucchini, MacDonald, and Langrock (2016). However, Markov regime–switching models have been introduced before Hamilton (1989), see, Goldfeld and Quandt (1973), Quandt (1958), and Tong (1983). The regime–switching model assumes that a discrete unobservable Markov process randomly switches among a finite set of regimes and that a particular parameter set defines each regime. The model fits some financial data well and has become popular in financial modeling, including equity options, bond prices, and others.

*Department of Applied Mathematics, National University of Mongolia; E-mail: battulgag@num.edu.mn; Phone Number: 976–99246036

A model that considers all of the above is the Bayesian Markov–Switching VAR (MS–VAR) process. Its applications in finance can be found in Battulga (2023b), Battulga (2024a), and Battulga (2024b). In some existing option pricing models, the underlying asset price is governed by some stochastic process, and economic variables such as GDP, inflation, unemployment rate, and so on are not taken into account. For this reason, the author has developed option pricing models, depending on economic variables. Applying the Bayesian MS–VAR process, with direct calculation and change of probability measure for some frequently used options, Battulga (2024b) derived pricing formulas. Also, the author used the Bayesian MS–VAR process to price equity–linked life insurance products and rainbow options, see Battulga (2024a) and Battulga (2023b).

Monte–Carlo simulation methods using the Gibbs sampling algorithm for Bayesian MS–VAR process are proposed by some authors. In particular, the Monte–Carlo simulation method of the Bayesian MS–AR(p) process is provided by Albert and Chib (1993), and its multidimensional extension is given by Krolzig (1997). In this paper, we introduce a new Monte–Carlo simulation method that removes duplication in a regime vector. We also introduce importance sampling method to estimate probability of rare event, which corresponds to endogenous variables. Importance sampling is an effective variance reduction technique for studying the rare events. Glasserman, Heidelberger, and Shahabuddin (2000) used the importance sampling method to model portfolio loss random variable by using approximation. Also, Glasserman and Li (2005) study a loss random variable of credit portfolio applying the method, see also McNeil, Frey, and Embrechts (2005).

Dividend discount models (DDMs), first introduced by Williams (1938), are common methods for stock valuation. The basic idea is that a stock price of a firm is equal to a sum of dividend paid by the firm and the stock price of the firm, which correspond to the next period and which are discounted at required rate of return on stock. As the outcome of DDMs depends crucially on dividend payment forecasts, most research in the last few decades has been around the proper estimations of dividend development. To model the dividends of a firm, Battulga, Jacob, Altangerel, and Horsch (2022) used the compound Poisson process. Also, parameter estimation of DDMs is a challenging task. Battulga et al. (2022) introduced parameter estimation methods for practically popular DDMs. Battulga (2023a) provided parameter estimation methods of the required rate of returns for public and private companies. Under the normal framework, Battulga (2022) obtained pricing and hedging formulas for the European options and equity–linked life insurance products by introducing a DDM with regime–switching process. A review of some existing DDMs, including deterministic and stochastic models can be found in D'Amico and De Blasis (2020).

The rest of the paper is organized as follows: In Section 2, for the general Bayesian MS–VAR process, we obtain some conditional density functions, which are helpful for general Monte–Carlo simulation. Section 3 is dedicated to studying a special case of the process, where we obtain closed-form conditional density functions of our model's random components. Some of the conditional density functions have not been explored before. In Section 3, we provide Monte–Carlo simulation methods, including the importance sampling method. Section 4 gives numerical results on three companies, listed in the S&P 500 index. Finally, Section 5 concludes the study.

2 Bayesian MS–VAR(p) process

Let $(\Omega, \mathcal{H}_T, \mathbb{P})$ be a complete probability space, where $\mathbb{P}$ is a given physical or real–world probability measure. Other elements of the probability space will be defined below. To introduce a regime–switching, we assume that $\{s_t\}_{t=1}^T$ is a homogeneous Markov chain with N state and $\mathsf{P} := \{p_{ij}\}_{i=0,j=1}^N$ is a random transition probability matrix, including an initial probability vector, where $\{p_{0j}\}_{j=1}^N$ is the initial probability vector. We consider a Bayesian Markov–Switching Vector Autoregressive process of p order (MS–VAR(p)), which is given by the following equation

$$y_t = A_{0,s_t}\psi_t + A_{1,s_t}y_{t-1} + \dots + A_{p,s_t}y_{t-p} + \xi_t,\ t = 1, \dots, T, \tag{1}$$

where $y_t = (y_{1,t}, \dots, y_{n,t})'$ is an $(n \times 1)$ vector of endogenous variables, $\psi_t = (1, \psi_{2,t}, \dots, \psi_{l,t})'$ is an $(l \times 1)$ vector of exogenous variables, $\xi_t = (\xi_{1,t}, \dots, \xi_{n,t})'$ is an $(n \times 1)$ residual process, A_{0,s_t} is an $(n \times l)$ random coefficient matrix at regime s_t, corresponding to the vector of exogenous variables, for $i = 1, \dots, p$, A_{i,s_t} are $(n \times n)$ random coefficient matrices at regime s_t, corresponding to $y_{t-1}, \dots, y_{t-p}$. Equation (1) can be written by

$$y_t = \Pi_{s_t} \mathsf{Y}_t + \xi_t, \ t = 1, \dots, T, \tag{2}$$

where $\Pi_{s_t} := [A_{0,s_t} : A_{1,s_t} : \dots : A_{p,s_t}]$ is an $(n \times d)$ random coefficient matrix with $d := l + np$ at regime s_t, which consist of all the random coefficient matrices and $\mathsf{Y}_t := (\psi_t', y_{t-1}', \dots, y_{t-p}')'$ is a $(d \times 1)$ vector, which consist of exogenous variable ψ_t and last p lagged values of the process y_t. The process Y_t is measurable with respect to a σ–field $\mathcal{F}_{t-1}$, which is defined below.

For the residual process ξ_t, we assume that it has $\xi_t := \Sigma_{s_t}^{1/2} \varepsilon_t$, $t = 1, \dots, T$ representation, see Lütkepohl (2005) and McNeil et al. (2005), where $\Sigma_{s_t}^{1/2}$ is a Cholesky factor of a positive definite $(n \times n)$ random matrix Σ_{s_t}, which is measurable with respect to σ–field $\mathcal{H}_{t-1}$, defined below and depends on $(n_* \times d_*)$ random coefficient matrix $\Gamma_{s_t} := [B_{0,s_t} : B_{1,s_t} : \dots : B_{p_*+q_*,s_t}]$ with $d_* := l_* + n_*(p_* + q_*)$. Here B_{0,s_t} is an $(n_* \times l_*)$ random matrix, for $i = 1, \dots, p_* + q_*$, B_{i,s_t} are $(n_* \times n_*)$ random matrices, and $\varepsilon_1, \dots, \varepsilon_T$ is a random sequence of independent identically multivariate normally distributed random vectors with means of 0 and covariance matrices of n dimensional identity matrix I_n. Then, in particular, for multivariate GARCH process of (p_*, q_*) order, dependence of $\Sigma_{s_t}^{1/2}$ on Γ_{s_t} is given by

$$\operatorname{vech}\big(\Sigma_{s_t}\big) = B_{0,s_t} + \sum_{i=1}^{p_*} B_{i,s_t} \operatorname{vech}\big(\xi_{t-i}\xi_{t-i}'\big) + \sum_{j=1}^{q_*} B_{p_*+j,s_t} \operatorname{vech}(\Sigma_{s_{t-j}}), \tag{3}$$

where B_{0,s_t} and B_{i,s_t} for $i = 1, \dots, p_* + q_*$ are suitable $([n(n+1)/2] \times 1)$ random vector and suitable $([n(n+1)/2] \times [n(n+1)/2])$ matrices, respectively, and the vech is an operator that stacks elements on and below a main diagonal of a square matrix.

Let us introduce stacked vectors and matrices: $y := (y_1', \dots, y_T')'$, $s := (s_1, \dots, s_T)'$, $\Pi_s := [\Pi_{s_1} : \dots : \Pi_{s_T}]$, and $\Gamma_s := [\Gamma_{s_1} : \dots : \Gamma_{s_T}]$. We also assume that the strong white noise process $\{\varepsilon_t\}_{t=1}^T$ is independent of the random coefficient matrices Π_s and Γ_s, random transition matrix P, and regime vector s conditional on initial information $\mathcal{F}_0 := \sigma(y_{1-p}', \dots, y_0', \psi_1, \dots, \psi_T, \Sigma_{1-q_*}, \dots, \Sigma_0)$. Here for a generic random vector X, $\sigma(X)$ denotes a σ–field generated by the random vector X, $\Sigma_{1-q_*}, \dots, \Sigma_0$ is initial values of the random matrix process Σ_{s_t}, $\psi_1, \dots, \psi_T$ are values of exogenous variables and they are known at time zero. We further suppose that the transition probability matrix P is independent of the random coefficient matrices Π_s and Γ_s given initial information $\mathcal{F}_0$ and regime vector s.

To ease of notations, for a generic vector $o = (o_1', \dots, o_T')'$, we denote its first t and last $T - t$ sub vectors by $\bar{o}_t$ and $\bar{o}_t^c$, respectively, that is, $\bar{o}_t := (o_1', \dots, o_t')'$ and $\bar{o}_t^c := (o_{t+1}', \dots, o_T')'$. We define σ–fields: for $t = 0, \dots, T$, $\mathcal{F}_t := \mathcal{F}_0 \vee \sigma(\bar{y}_t)$ and $\mathcal{H}_t := \mathcal{F}_t \vee \sigma(\Pi_s) \vee \sigma(\Gamma_s) \vee \sigma(s) \vee \sigma(\mathsf{P})$ where for generic sigma fields $\mathcal{O}_1$ and $\mathcal{O}_2$, $\mathcal{O}_1 \vee \mathcal{O}_2$ is the minimal σ–field containing the σ–fields $\mathcal{O}_1$ and $\mathcal{O}_2$. For the first–order Markov chain, a conditional probability that the regime at time $t+1$, s_{t+1} equals some particular value conditional on the past regimes $\bar{s}_t$, transition probability matrix P, and initial information $\mathcal{F}_0$ depends only through the most recent regime at time t, s_t, transition probability matrix P, and initial information $\mathcal{F}_0$, that is,

$$p_{s_t s_{t+1}} := \mathbb{P}[s_{t+1} = s_{t+1} | s_t = s_t, \mathsf{P}, \mathcal{F}_0] = \mathbb{P}\big[s_{t+1} = s_{t+1} | \bar{s}_t = \bar{s}_t, \mathsf{P}, \mathcal{F}_0\big] \tag{4}$$

for $t = 0, \dots, T-1$, where $p_{s_1} := p_{s_0 s_1} = \mathbb{P}[s_1 = s_1 | \mathsf{P}, \mathcal{F}_0]$ is an initial probability. A distribution of a residual random vector $\xi := (\xi_1', \dots, \xi_T')'$ is given by

$$\xi = (\xi_1', \dots, \xi_T')' \mid \mathcal{H}_0 \sim \mathcal{N}(0, \Sigma_s), \tag{5}$$

where $\Sigma_s := \operatorname{diag}\{\Sigma_{s_1}, \dots, \Sigma_{s_T}\}$ is a block diagonal matrix.

To remove duplicates in the random coefficient matrix (Π_s, Γ_s), for a generic regime vector with length k, $o = (o_1, \dots, o_k)'$, we define sets

$$\mathcal{A}_{\bar{o}_t} := \mathcal{A}_{\bar{o}_{t-1}} \cup \big\{o_t \in \{o_1, \dots, o_k\} \big| o_t \notin \mathcal{A}_{\bar{o}_{t-1}}\big\}, \quad t = 1, \dots, k, \tag{6}$$

where for $t = 1, \dots, k$, $o_t \in \{1, \dots, N\}$ and an initial set is empty set, i.e., $\mathcal{A}_{\bar{o}_0} = \emptyset$. The final set $\mathcal{A}_o = \mathcal{A}_{\bar{o}_k}$ consists of different regimes in regime vector $o = \bar{o}_k$ and $|\mathcal{A}_o|$ represents a number of different regimes in the regime vector o.

Let us assume that elements of sets $\mathcal{A}_s$, $\mathcal{A}_{\bar{s}_t}$, $\mathcal{A}_{\bar{s}_t^c}$, intersection set of the sets $\mathcal{A}_{\bar{s}_t}$ and $\mathcal{A}_{\bar{s}_t^c}$, and difference sets between the sets $\mathcal{A}_{\bar{s}_t^c}$ and $\mathcal{A}_{\bar{s}_t}$ are given by $\mathcal{A}_s = \{\hat{s}_1, \dots, \hat{s}_{r_{\hat{s}}}\}$, $\mathcal{A}_{\bar{s}_t} = \{\alpha_1, \dots, \alpha_{r_\alpha}\}$, $\mathcal{A}_{\bar{s}_t^c} = \{\beta_1, \dots, \beta_{r_\beta}\}$, $\mathcal{A}_{\bar{s}_t} \cap \mathcal{A}_{\bar{s}_t^c} = \{\gamma_1, \dots, \gamma_{r_\gamma}\}$, $\mathcal{A}_{\bar{s}_t^c} \backslash \mathcal{A}_{\bar{s}_t} = \{\delta_1, \dots, \delta_{r_\delta}\}$, and $\mathcal{A}_{\bar{s}_t} \backslash \mathcal{A}_{\bar{s}_t^c} = \{\epsilon_1, \dots, \epsilon_{r_\epsilon}\}$, respectively, where $r_{\hat{s}} := |\mathcal{A}_s|$, $r_\alpha := |\mathcal{A}_{\bar{s}_t}|$, $r_\beta := |\mathcal{A}_{\bar{s}_t^c}|$, $r_\gamma := |\mathcal{A}_{\bar{s}_t} \cap \mathcal{A}_{\bar{s}_t^c}|$, $r_\delta := |\mathcal{A}_{\bar{s}_t^c} \backslash \mathcal{A}_{\bar{s}_t}|$, and $r_\epsilon := |\mathcal{A}_{\bar{s}_t} \backslash \mathcal{A}_{\bar{s}_t^c}|$ are numbers of elements of the sets, respectively. Note that

$$\mathcal{A}_{\bar{s}_t} = (\mathcal{A}_{\bar{s}_t} \backslash \mathcal{A}_{\bar{s}_t^c}) \cup (\mathcal{A}_{\bar{s}_t} \cap \mathcal{A}_{\bar{s}_t^c}), \tag{7}$$

$$\mathcal{A}_{\bar{s}_t^c} = (\mathcal{A}_{\bar{s}_t} \cap \mathcal{A}_{\bar{s}_t^c}) \cup (\mathcal{A}_{\bar{s}_t^c} \backslash \mathcal{A}_{\bar{s}_t}), \tag{8}$$

and

$$\mathcal{A}_s = \mathcal{A}_{\bar{s}_t} \cup \mathcal{A}_{\bar{s}_t^c} = (\mathcal{A}_{\bar{s}_t} \backslash \mathcal{A}_{\bar{s}_t^c}) \cup \mathcal{A}_{\bar{s}_t^c} = \mathcal{A}_{\bar{s}_t} \cup (\mathcal{A}_{\bar{s}_t^c} \backslash \mathcal{A}_{\bar{s}_t}) \tag{9}$$

and intersection sets of the sets of right hand sides of equations (7) and (8), and (9) are empty sets. We introduce the following regime vectors: $\hat{s} := (\hat{s}_1, \dots, \hat{s}_{r_{\hat{s}}})'$ is an $(r_{\hat{s}} \times 1)$ vector, $\alpha := (\alpha_1, \dots, \alpha_{r_\alpha})'$ is an $(r_\alpha \times 1)$ vector, $\beta = (\beta_1, \dots, \beta_{r_\beta})'$ is an $(r_\beta \times 1)$ vector, $\gamma = (\gamma_1, \dots, \gamma_{r_\gamma})'$ is an $(r_\gamma \times 1)$ vector, $\delta = (\delta_1, \dots, \delta_{r_\delta})'$ is an $(r_\delta \times 1)$ vector, and $\epsilon = (\epsilon_1, \dots, \epsilon_{r_\epsilon})'$ is an $(r_\epsilon \times 1)$ vector. For the regime vector $a = (a_1, \dots, a_{r_a})' \in \{\hat{s}, \alpha, \beta, \gamma, \delta, \epsilon\}$, we also introduce duplication removed random coefficient matrices, whose block matrices are different: $\Pi_a = [\Pi_{a_1} : \dots : \Pi_{a_{r_a}}]$ is an $(n \times [dr_a])$ matrix, $\Gamma_a = [\Gamma_{a_1} : \dots : \Gamma_{a_{r_a}}]$ is an $(n_* \times [d_* r_a])$ matrix, and (Π_a, Γ_a).

We assume that for given duplication removed regime vector $\hat{s}$ and initial information $\mathcal{F}_0$, the coefficient matrices $(\Pi_{\hat{s}_1}, \Gamma_{\hat{s}_1}), \dots, (\Pi_{\hat{s}_{r_{\hat{s}}}}, \Gamma_{\hat{s}_{r_{\hat{s}}}})$ are independent. Under the last assumption, a joint density function of the random coefficient random matrix $(\Pi_{\hat{s}}, \Gamma_{\hat{s}})$ is represented by

$$f\big(\Pi_{\hat{s}}, \Gamma_{\hat{s}} \big| \hat{s}, \mathcal{F}_0\big) = \prod_{t=1}^{r_{\hat{s}}} f\big(\Pi_{\hat{s}_t}, \Gamma_{\hat{s}_t} \big| \hat{s}_t, \mathcal{F}_0\big), \tag{10}$$

where for a generic random vector X, we denote its density function by $f(X)$. Throughout the paper we fix $t = 1, \dots, T-1$. For the regime vectors α and β, the above joint density function can be written by

$$f\big(\Pi_{\hat{s}}, \Gamma_{\hat{s}} \big| \hat{s}, \mathcal{F}_0\big) = f\big(\Pi_\alpha, \Gamma_\alpha \big| \alpha, \mathcal{F}_0\big) f_*\big(\Pi_\delta, \Gamma_\delta \big| \delta, \mathcal{F}_0\big) \tag{11}$$

where the density function $f_*\big(\Pi_\delta, \Gamma_\delta \big| \delta, \mathcal{F}_0\big)$ equals

$$f_*\big(\Pi_\delta, \Gamma_\delta \big| \delta, \mathcal{F}_0\big) := \begin{cases} f\big(\Pi_\delta, \Gamma_\delta \big| \delta, \mathcal{F}_0\big), & \text{if} \quad r_\delta \neq 0, \\ 1, & \text{if} \quad r_\delta = 0. \end{cases} \tag{12}$$

Then, the following Proposition, which is useful for Monte–Carlo simulation holds, see below.

Proposition 1. *Conditional on initial information $\mathcal{F}_0$, a joint density function of the random vectors $\bar{y}_t$ and s and random matrices $\Pi_{\hat{s}}$, $\Gamma_{\hat{s}}$, and P is given by*

$$f(\bar{y}_t, \Pi_{\hat{s}}, \Gamma_{\hat{s}}, s, \mathsf{P} | \mathcal{F}_0) = f\big(\bar{y}_t, \Pi_\alpha, \Gamma_\alpha, \bar{s}_t \big| \mathcal{F}_0\big) f_*\big(\Pi_\delta, \Gamma_\delta \big| \delta, \mathcal{F}_0\big) f(s, \mathsf{P} | \mathcal{F}_0) \big/ f(\bar{s}_t | \mathcal{F}_0). \tag{13}$$

In particular, the following relationships holds

$$f(\bar{s}_t^c | \bar{s}_t, \mathcal{F}_t) = f(\bar{s}_t^c | \bar{s}_t, \mathcal{F}_0), \tag{14}$$

$$f\big(\Pi_\delta,\Gamma_\delta\big|\Pi_\alpha,\Gamma_\alpha,s,\mathcal{F}_t\big)=f_*\big(\Pi_\delta,\Gamma_\delta\big|\delta,\mathcal{F}_0\big), \tag{15}$$

$$f\big(\Pi_\beta,\Gamma_\beta\big|s,\mathcal{F}_t\big)=f\big(\Pi_\gamma,\Gamma_\gamma\big|\bar{s}_t,\mathcal{F}_t\big)f_*\big(\Pi_\delta,\Gamma_\delta\big|\delta,\mathcal{F}_0\big), \tag{16}$$

$$f(\mathsf{P}|\bar{s}_t,\mathcal{F}_t)=f(\mathsf{P}|\bar{s}_t,\mathcal{F}_0), \tag{17}$$

$$f(\bar{s}_t^c|\bar{s}_t,\mathsf{P},\mathcal{F}_t)=f(\bar{s}_t^c|\bar{s}_t,\mathsf{P},\mathcal{F}_0), \tag{18}$$

and

$$f(\Pi_{\hat{s}},\Gamma_{\hat{s}}|s,\mathsf{P},\mathcal{F}_t)=f(\Pi_{\hat{s}},\Gamma_{\hat{s}}|s,\mathcal{F}_t). \tag{19}$$

Proof. Let us consider a joint density function of the random vectors $\bar{y}_t$ and s and random matrices $\Pi_{\hat{s}}$, $\Gamma_{\hat{s}}$, and P for given initial information $\mathcal{F}_0$. According to the conditional probability formula, one gets that

$$f(\bar{y}_t,\Pi_{\hat{s}},\Gamma_{\hat{s}},s,\mathsf{P}|\mathcal{F}_0)=f(\bar{y}_t|\Pi_{\hat{s}},\Gamma_{\hat{s}},s,\mathsf{P},\mathcal{F}_0)f(\Pi_{\hat{s}},\Gamma_{\hat{s}}|s,\mathsf{P},\mathcal{F}_0)f(s|\mathsf{P},\mathcal{F}_0)f(\mathsf{P}|\mathcal{F}_0). \tag{20}$$

Since the random vector $\bar{y}_t$ depends on Π_α, Γ_α, $\bar{s}_t$, the first joint density function of the right–hand side of the above equation equals $f(\bar{y}_t|\Pi_\alpha,\Gamma_\alpha,\bar{s}_t,\mathcal{F}_0)$. As the random coefficient matrix $(\Pi_{\hat{s}},\Gamma_{\hat{s}})$ is independent of the transition probability matrix P conditional on s and $\mathcal{F}_0$, the second joint density function of the right–hand side can be represented by equation (11). According to the Markov property (4), the third joint density function equals $f(\bar{s}_t|\mathsf{P},\mathcal{F}_0)f(\bar{s}_t^c|\bar{s}_t,\mathsf{P},\mathcal{F}_0)$. Consequently, one obtains equation (13). If we integrate equation (13) by P, then one finds that

$$f(\bar{y}_t,\Pi_{\hat{s}},\Gamma_{\hat{s}},s|\mathcal{F}_0)=f\big(\bar{y}_t,\Pi_\alpha,\Gamma_\alpha,\bar{s}_t\big|\mathcal{F}_0\big)f_*\big(\Pi_\delta,\Gamma_\delta\big|\delta,\mathcal{F}_0\big)f(\bar{s}_t^c|\bar{s}_t,\mathcal{F}_0). \tag{21}$$

By integrating the above equation by the random matrix $(\Pi_{\hat{s}},\Gamma_{\hat{s}})$, one obtains equation (14). Since $f(\bar{y}_t|\Pi_\alpha,\Gamma_\alpha,\bar{s}_t,\bar{s}_t^c,\mathcal{F}_0)=f(\bar{y}_t|\Pi_\alpha,\Gamma_\alpha,\bar{s}_t,\mathcal{F}_0)$ and $f(\Pi_\alpha,\Gamma_\alpha|\bar{s}_t,\bar{s}_t^c,\mathcal{F}_0)=f(\Pi_\alpha,\Gamma_\alpha|\bar{s}_t,\mathcal{F}_0)$, we have that

$$\begin{aligned} f(\bar{y}_t,\bar{s}_t^c|\bar{s}_t,\mathcal{F}_0) &= \int_{\Pi_\alpha,\Gamma_\alpha} f(\bar{y}_t|\Pi_\alpha,\Gamma_\alpha,\bar{s}_t,\bar{s}_t^c,\mathcal{F}_0)f(\Pi_\alpha,\Gamma_\alpha|\bar{s}_t,\bar{s}_t^c,\mathcal{F}_0)f(\bar{s}_t^c|\bar{s}_t\mathcal{F}_0)d\Pi_\alpha d\Gamma_\alpha \\ &= f(\bar{y}_t|\bar{s}_t,\mathcal{F}_0)f(\bar{s}_t^c|\bar{s}_t,\mathcal{F}_0). \end{aligned} \tag{22}$$

Thus, conditional on $\bar{s}_t$ and $\mathcal{F}_0$, the random vectors $\bar{y}_t$ and $\bar{s}_t^c$ are independent. Consequently, it holds that

$$f(\bar{y}_t,s|\mathcal{F}_0)=f(\bar{y}_t,\bar{s}_t|\mathcal{F}_0)f(\bar{s}_t^c|\bar{s}_t,\mathcal{F}_0). \tag{23}$$

If we divide equation (21) by the above equation, then one obtains

$$f\big(\Pi_{\hat{s}},\Gamma_{\hat{s}}\big|s,\mathcal{F}_t\big)=f\big(\Pi_\alpha,\Gamma_\alpha\big|\bar{s}_t,\mathcal{F}_t\big)f_*\big(\Pi_\delta,\Gamma_\delta\big|\delta,\mathcal{F}_0\big). \tag{24}$$

Since $f(\bar{y}_t,\Pi_\alpha,\Gamma_\alpha|s,\mathcal{F}_0)=f(\bar{y}_t,\Pi_\alpha,\Gamma_\alpha|\bar{s}_t,\mathcal{F}_0)$ and conditional on $\bar{s}_t$ and $\mathcal{F}_0$, the random vectors $\bar{y}_t$ and $\bar{s}_t^c$ are independent, we have $f\big(\Pi_\alpha,\Gamma_\alpha\big|s,\mathcal{F}_t\big)=f\big(\Pi_\alpha,\Gamma_\alpha\big|\bar{s}_t,\mathcal{F}_t\big)$. Consequently, due to the conditional probability formula, we have that

$$f\big(\Pi_{\hat{s}},\Gamma_{\hat{s}}\big|s,\mathcal{F}_t\big)=f\big(\Pi_\alpha,\Gamma_\alpha\big|\bar{s}_t,\mathcal{F}_t\big)f\big(\Pi_\delta,\Gamma_\delta\big|\Pi_\alpha,\Gamma_\alpha,s,\mathcal{F}_0\big). \tag{25}$$

Thus, equating equation (24) with the above equation, we get equation (15). If we integrate equation (24) by $(\Pi_\epsilon,\Gamma_\epsilon)$, then by equation (7) and (9), we obtain equation (16). Integrating equation (13) by $(\Pi_{\hat{s}},\Gamma_{\hat{s}})$, one obtains that

$$f(\bar{y}_t,s,\mathsf{P}|\mathcal{F}_0)=f\big(\bar{y}_t,\bar{s}_t\big|\mathcal{F}_0\big)f(s,\mathsf{P}|\mathcal{F}_0)\big/f(\bar{s}_t|\mathcal{F}_0). \tag{26}$$

Consequently if we integrate equation (26) by $\bar{s}_t^c$ and divide $f(\bar{y}_t, \bar{s}_t|\mathcal{F}_0)$, we get equation (17). Thus, conditional on $\bar{s}_t$ and $\mathcal{F}_0$, $\bar{y}_t$ and P are independent. Also, it follows from equation (26) that

$$f(\bar{s}_t^c, \mathsf{P}|\bar{s}_t, \mathcal{F}_t) = f(s, \mathsf{P}|\mathcal{F}_0)/f(\bar{s}_t|\mathcal{F}_0). \tag{27}$$

By dividing the above equation by equation (17), we obtain equation (18). To prove equation (19), let us consider a conditional density function $f(\Pi_{\hat{s}}, \Gamma_{\hat{s}}, \mathsf{P}|s, \mathcal{F}_t)$. Equation (20) can be written by

$$f(\bar{y}_t, \Pi_{\hat{s}}, \Gamma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0) = f(\bar{y}_t, \Pi_{\hat{s}}, \Gamma_{\hat{s}}|s, \mathcal{F}_0)f(s, \mathsf{P}|\mathcal{F}_0). \tag{28}$$

On the other hand, by the conditional probability formula, it holds

$$f(\bar{y}_t, \Pi_{\hat{s}}, \Gamma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0) = f(\bar{y}_t, \Pi_{\hat{s}}, \Gamma_{\hat{s}}|s, \mathsf{P}, \mathcal{F}_0)f(s, \mathsf{P}|\mathcal{F}_0). \tag{29}$$

Therefore, one conclude that conditional on s and $\mathcal{F}_0$, the random matrix $(\bar{y}_t, \Pi_{\hat{s}}, \Gamma_{\hat{s}})$ and transition probability matrix P are independent. Thus, equation (19) holds. That completes the proof. $\square$

It should be noted that according to the Markov property (4), it follows from equation (18) that the assumption for a Markov chain in the book of Hamilton (1994) always holds, namely,

$$f(s_{t+1}|\bar{s}_t, \mathsf{P}, \mathcal{F}_t) = f(s_{t+1}|s_t, \mathsf{P}, \mathcal{F}_0). \tag{30}$$

3 Special Case of Bayesian MS–VAR(p) process

In this section, we consider a special case of the Bayesian MS–VAR(p) process. The Bayesian MS–VAR(p) process can be written by the following equation

$$y_t = \Pi_{s_t}\mathsf{Y}_t + \xi_t = (\mathsf{Y}_t' \otimes I_n)\pi_{s_t} + \xi_t, \quad t = 1, \dots, T, \tag{31}$$

where $\otimes$ is the Kronecker product of two matrices and $\pi_{s_t} := \text{vec}(\Pi_{s_t})$ is an $(nd \times 1)$ vectorization of the random coefficient matrix Π_{s_t}. Now we define distributions of the random coefficient vector π_{s_t} and covariance matrix Σ_{s_t}. We assume that conditional on the regime s_t and initial information $\mathcal{F}_0$, a distribution of the random covariance matrix Σ_{s_t} is given by

$$\Sigma_{s_t} \mid s_t, \mathcal{F}_0 \sim \mathcal{IW}(\nu_{0,s_t}, V_{0,s_t}) \tag{32}$$

where the notation $\mathcal{IW}$ denotes the Inverse–Wishart distribution, $\nu_{0,s_t} > n-1$ is a degrees of freedom and V_{0,s_t} is a positive definite scale matrix and both are prior hyperparameters, corresponding to the regime s_t. Consequently, a distribution of the residual vector ξ_t equals

$$\xi_t \mid \Sigma_{s_t}, s_t, \mathcal{F}_0 \sim \mathcal{N}\Big(0, \Sigma_{s_t}\Big), \tag{33}$$

where $\mathcal{N}$ denotes the normal distribution. Also, we assume that conditional on the covariance matrix Σ_{s_t}, regime s_t, and initial information $\mathcal{F}_0$, a distribution of the random coefficient vector π_{s_t} is given by

$$\pi_{s_t} \mid \Sigma_{s_t}, s_t, \mathcal{F}_0 \sim \mathcal{N}\Big(\pi_{0,s_t}, \Lambda_{0,s_t} \otimes \Sigma_{s_t}\Big), \tag{34}$$

where π_{0,s_t} is an $(nd \times 1)$ prior hyperparameter vector at regime s_t and Λ_{0,s_t} is a symmetric positive definite $(d \times d)$ prior hyperparameter matrix at regime s_t.

3.1 Distributions

For the regime vector $\bar{s}_t$ and regime α_k, we define sets

$$S_{t,\alpha_k} := \big\{u \in \{1,\dots,t\}\big|s_u = \alpha_k,\ u = 1,\dots,t\big\}, \quad k = 1,\dots,r_\alpha. \tag{35}$$

For $k = 1,\dots,r_\alpha$, the set S_{t,α_k} consists of indexes of regimes in the regime vector $\bar{s}_t$ that equal the regime α_k. Let us suppose that $q_{t,\alpha_k} := |S_{t,\alpha_k}|$ is a number of regimes in the regime vector $\bar{s}_t$ that equal the regime α_k and elements of the set S_{t,α_k} are given by

$$S_{t,\alpha_k} = \{k_{t,1},\dots,k_{t,q_{t,\alpha_k}}\}, \quad k = 1,\dots,r_\alpha. \tag{36}$$

Further, we define indexes

$$o_t := \big\{k \in \{1,\dots,r_\alpha\}\big|s_t = \alpha_k,\ k = 1,\dots,r_\alpha\big\}, \quad t = 1,\dots,t. \tag{37}$$

The index o_t represents a position of the regime s_t in the regime vector α. Let $\pi_\alpha := \big(\pi_{\alpha_1}',\dots,\pi_{\alpha_{r_\alpha}}'\big)'$ be an $([ndr_\alpha]\times 1)$ random coefficient vector, whose sub–vectors are different and which corresponds to the regime vector $\bar{s}_t$, $y_{t,\alpha_k} := \Big(y_{k_{t,1}}',\dots,y_{k_{t,q_{t,\alpha_k}}}'\Big)'$ be an $([nq_{t,\alpha_k}]\times 1)$ vector of endogenous variables, corresponding to the regime α_k, and $\mathsf{Y}_{t,\alpha_k}^\circ := \big[\mathsf{Y}_{k_{t,1}} : \dots : \mathsf{Y}_{k_{t,q_{t,\alpha_k}}}\big]$ be a $(d\times q_{t,\alpha_k})$ matrix of exogenous and endogenous variables, corresponding to the regime α_k. By using a $(t\times r_\alpha)$ matrix $D_\alpha := [j_{o_1} : \dots : j_{o_t}]'$ one can revive the vector $\pi_{\bar{s}_t} := \mathrm{vec}(\Pi_{\bar{s}_t})$ from the vector π_α, that is, $\pi_{\bar{s}_t} = (D_\alpha \otimes I_{nd})\pi_\alpha$, where j_o is an $(r_\alpha \times 1)$ unit vector, whose o–th element equals one and others zero.

3.1.1 Conditional Distributions

It follows from equations (33) and (34) that distributions of the random vectors $\bar{\xi}_t$ and π_α are obtained by

$$\bar{\xi}_t \mid \Sigma_\alpha, \bar{s}_t, \mathcal{F}_0 \sim \mathcal{N}\big(0, \Sigma_{\bar{s}_t}\big), \tag{38}$$

and

$$\pi_\alpha \mid \Sigma_\alpha, \alpha, \mathcal{F}_0 \sim \mathcal{N}\Big(\pi_{0,\alpha}, \Sigma_{\pi_\alpha}\Big), \tag{39}$$

respectively, where $\Sigma_{\bar{s}_t} := \mathrm{diag}\{\Sigma_{s_1},\dots,\Sigma_{s_t}\}$ is an $([nt]\times[nt])$ block diagonal matrix, corresponding to the regime vector $\bar{s}_t$ and $\Sigma_\alpha := \big[\Sigma_{\alpha_1} : \dots : \Sigma_{\alpha_{r_\alpha}}\big]'$ is an $([nr_\alpha]\times n)$ matrix, $\pi_{0,\alpha} := \big(\pi_{0,\alpha_1}',\dots,\pi_{0,\alpha_{r_\alpha}}'\big)'$ is an $([ndr_\alpha]\times 1)$ prior hyperparameter vector, and $\Sigma_{\pi_\alpha} := \mathrm{diag}\big\{\Lambda_{0,\alpha_1}\otimes\Sigma_{\alpha_1},\dots,\Lambda_{0,\alpha_{r_\alpha}}\otimes\Sigma_{\alpha_{r_\alpha}}\big\}$ is an $([ndr_\alpha]\times[ndr_\alpha])$ block diagonal matrix, all of which correspond to the duplication removed regime vector α. A connection between the random matrices $\Sigma_{\bar{s}_t}$ and Σ_α is

$$\Sigma_{\bar{s}_t} = \mathrm{diag}\big\{\big((D_\alpha\otimes I_n)\Sigma_\alpha\big)_1,\dots,\big((D_\alpha\otimes I_n)\Sigma_\alpha\big)_t\big\}, \tag{40}$$

where the matrix $\big((D_\alpha\otimes I_n)\Sigma_\alpha\big)_j$ equals j–th block matrix of the matrix $(D_\alpha\otimes I_n)\Sigma_\alpha$. On the other hand, by following Battulga (2024b), a distribution of $([nT]\times 1)$ the random vector $y = (y_1',\dots,y_T')'$ is given by

$$y \mid \pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathcal{F}_0 \sim \mathcal{N}\Big(\Psi_s^{-1}\varphi_s, \Psi_s^{-1}\Sigma_s(\Psi_s^{-1})'\Big), \tag{41}$$

where the matrix Ψ_s and the vector δ_s are

$$\Psi_s := \begin{bmatrix} I_n & 0 & \dots & 0 & \dots & 0 & 0 \\ -A_{1,s_2} & I_n & \dots & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & -A_{p-1,s_{T-1}} & \dots & I_n & 0 \\ 0 & 0 & \dots & -A_{p,s_T} & \dots & -A_{1,s_T} & I_n \end{bmatrix} \tag{42}$$

and

$$\varphi_s := \begin{bmatrix} A_{0,s_1}\psi_1 + A_{1,s_1}y_0 + \dots + A_{p,s_1}y_{1-p} \\ A_{0,s_2}\psi_2 + A_{2,s_2}y_0 + \dots + A_{p,s_2}y_{2-p} \\ \vdots \\ A_{0,s_{T-1}}\psi_{T-1} \\ A_{0,s_T}\psi_T \end{bmatrix}, \tag{43}$$

respectively. To price default–free options, Battulga (2024b) used the conditional distribution of the random vector y. For a generic vector $o = (o_1', \dots, o_n')'$ with $(m \times 1)$ vector o_i, we introduce an $(m \times n)$ matrix notation $o^\circ := [o_1 : \dots : o_n]$. Then, the following Proposition holds.

Proposition 2. *Let for $t = 1, \dots, T-1$, $\pi_{s_t} \mid \Sigma_{s_t}, s_t, \mathcal{F}_0 \sim \mathcal{N}\big(\pi_{0,s_t}, \Lambda_{0,s_t} \otimes \Sigma_{s_t}\big)$, and $\Sigma_{s_t} \mid s_t, \mathcal{F}_0 \sim \mathcal{IW}(\nu_{0,s_t}, V_{0,s_t})$. Then, first, conditional on the regime vector $\bar{s}_t$ and initial information $\mathcal{F}_0$, a joint density function of the random vector $\bar{y}_t$ is given by*

$$f(\bar{y}_t|\bar{s}_t, \mathcal{F}_0) = \frac{1}{\pi^{nt/2}} \prod_{k=1}^{r_\alpha} \frac{|\Lambda_{0,\alpha_k}^{-1}|^{n/2}\Gamma_n\big((\nu_{0,\alpha_k} + q_{t,\alpha_k})/2\big)|V_{0,\alpha_k}|^{\nu_{0,\alpha_k}/2}}{|\Lambda_{0,\alpha_k|t}|^{n/2}\Gamma_n(\nu_{0,\alpha_k}/2)\big|\bar{B}_{t,\alpha_k} + V_{0,\alpha_k}\big|^{(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}, \tag{44}$$

where $\Gamma_n(\cdot)$ is the multivariate gamma function, $\Lambda_{0,\alpha_k|t} := \mathsf{Y}_{t,\alpha_k}^\circ(\mathsf{Y}_{t,\alpha_k}^\circ)' + \Lambda_{0,\alpha_k}^{-1}$ is a $(d \times d)$ matrix, and $\bar{B}_{t,\alpha_k}$ is an $(n \times n)$ positive semi–definite matrix and equals

$$\begin{aligned} \bar{B}_{t,\alpha_k} &:= y_{t,\alpha_k}^\circ(y_{t,\alpha_k}^\circ)' + \pi_{0,\alpha_k}^\circ\Lambda_{0,\alpha_k}^{-1}(\pi_{0,\alpha_k}^\circ)' - C_{t,\alpha_k}\Lambda_{0,\alpha_k|t}C_{t,\alpha_k}' \\ &= \big(y_{t,\alpha_k}^\circ - \pi_{0,\alpha_k}^\circ\mathsf{Y}_{t,\alpha_k}^\circ\big)\big(I_{q_{t,\alpha_k}} + (\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\mathsf{Y}_{t,\alpha_k}^\circ\big)^{-1}\big(y_{t,\alpha_k}^\circ - \pi_{0,\alpha_k}^\circ\mathsf{Y}_{t,\alpha_k}^\circ\big)' \end{aligned} \tag{45}$$

with $(n \times d)$ matrix $C_{t,\alpha_k} := \big(y_{t,\alpha_k}^\circ(\mathsf{Y}_{t,\alpha_k}^\circ)' + \pi_{0,\alpha_k}^\circ\Lambda_{0,\alpha_k}^{-1}\big)\Lambda_{0,\alpha_k|t}^{-1}$. Second, conditional on the random covariance matrix Σ_α, regime vector $\bar{s}_t$, and information $\mathcal{F}_t$, a joint density function of the random coefficient vector π_α is given by

$$\begin{aligned} & f(\pi_\alpha|\Sigma_\alpha, \bar{s}_t, \mathcal{F}_t) \\ &= \frac{1}{(2\pi)^{ndr_\alpha/2}\prod_{k=1}^{r_\alpha}|A_{\alpha_k}|^{1/2}}\exp\bigg\{-\frac{1}{2}\sum_{k=1}^{r_\alpha}\Big(\pi_{\alpha_k} - A_{\alpha_k}b_{\alpha_k}\Big)'A_{\alpha_k}^{-1}\Big(\pi_{\alpha_k} - A_{\alpha_k}b_{\alpha_k}\Big)\bigg\}, \end{aligned} \tag{46}$$

where for $k = 1, \dots, r_\alpha$, $A_{\alpha_k} := \big(\Lambda_{0,\alpha_k|t}^{-1} \otimes \Sigma_{\alpha_k}\big)$ is an $([nd] \times [nd])$ matrix and $b_{\alpha_k} := \big(\mathsf{Y}_{t,\alpha_k}^\circ \otimes \Sigma_{\alpha_k}^{-1}\big)y_{t,\alpha_k} + \big(\Lambda_{0,\alpha_k}^{-1} \otimes \Sigma_{\alpha_k}^{-1}\big)\pi_{0,\alpha_k}$ is an $([nd] \times 1)$ vector. Third, conditional on the regime vector $\bar{s}_t$ and information $\mathcal{F}_t$, a joint density function of the random coefficient matrix Σ_α is given by

$$\begin{aligned} f(\Sigma_\alpha|\bar{s}_t, \mathcal{F}_t) &= \prod_{k=1}^{r_\alpha}\frac{\big|\bar{B}_{t,\alpha_k} + V_{0,\alpha_k}\big|^{(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}{\Gamma_n\big((\nu_{0,\alpha_k} + q_{t,\alpha_k})/2\big)2^{n(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}|\Sigma_{\alpha_k}|^{-(\nu_{0,\alpha_k}+q_{t,\alpha_k}+n+1)/2} \\ &\times \exp\bigg\{-\frac{1}{2}\sum_{k=1}^{r_\alpha}\mathrm{tr}\Big(\big(\bar{B}_{t,\alpha_k} + V_{0,\alpha_k}\big)\Sigma_{\alpha_k}^{-1}\Big)\bigg\}. \end{aligned} \tag{47}$$

Finally, conditional on the regime vector s and information $\mathcal{F}_t$, a joint density function of the random coefficient matrix π_β° is given by

$$\begin{aligned} f(\pi_\beta^\circ|s, \mathcal{F}_t) &= \prod_{k=1}^{r_\gamma}\frac{|\Lambda_{0,\gamma_k|t}|^{n/2}\big|\bar{B}_{t,\gamma_k} + V_{0,\gamma_k}\big|^{-d/2}\Gamma_n\big((\nu_{0,\gamma_k} + q_{t,\gamma_k} + d)/2\big)}{\pi^{nd/2}\Gamma_n((\nu_{0,\gamma_k} + q_{t,\gamma_k})/2)} \\ &\times \big|I_n + (\bar{B}_{t,\gamma_k} + V_{0,\gamma_k})^{-1}(\pi_{\gamma_k}^\circ - C_{t,\gamma_k})\Lambda_{0,\gamma_k|t}(\pi_{\gamma_k}^\circ - C_{t,\gamma_k})'\big|^{-(\nu_{0,\gamma_k}+q_{t,\gamma_k}+d)/2} \\ &\times \prod_{\ell=1}^{r_\delta}\frac{|\Lambda_{0,\delta_\ell}|^{-n/2}|V_{0,\delta_\ell}|^{-d/2}\Gamma_n\big((\nu_{0,\delta_\ell} + d)/2\big)}{\pi^{nd/2}\Gamma_n(\nu_{0,\delta_\ell}/2)} \\ &\times \big|I_n + V_{0,\delta_\ell}^{-1}(\pi_{\delta_\ell}^\circ - \pi_{0,\delta_\ell}^\circ)\Lambda_{0,\delta_\ell}^{-1}(\pi_{\delta_\ell}^\circ - \pi_{0,\delta_\ell}^\circ)'\big|^{-(\nu_{0,\delta_\ell}+d)/2}. \end{aligned} \tag{48}$$

Proof. First, since $(\pi_{\alpha_1},\Sigma_{\alpha_1}),\dots,(\pi_{r_\alpha},\Sigma_{r_\alpha})$ are independent for given regime vector $\bar{s}_t$ and initial information $\mathcal{F}_0$, observe that conditional density functions of the random vectors y and $\pi_{\hat{s}}$ and random matrix $\Sigma_{\hat{s}}$ are given by

$$f(\bar{y}_t|\pi_\alpha,\Sigma_\alpha,\bar{s}_t,\mathcal{F}_0)=\frac{1}{(2\pi)^{nt/2}}\prod_{k=1}^{r_\alpha}|\Sigma_{\alpha_k}|^{-q_{t,\alpha_k}/2} \tag{49}$$
$$\times\exp\bigg\{-\frac{1}{2}\sum_{k=1}^{r_\alpha}\Big(y_{t,\alpha_k}-((\mathsf{Y}^\circ_{t,\alpha_k})'\otimes I_n)\pi_{\alpha_k}\Big)'\Big(I_{q_{t,\alpha_k}}\otimes\Sigma_{\alpha_k}^{-1}\Big)\Big(y_{t,\alpha_k}-((\mathsf{Y}^\circ_{t,\alpha_k})'\otimes I_n)\pi_{\alpha_k}\Big)\bigg\},$$

$$\begin{aligned}f(\pi_\alpha|\Sigma_\alpha,\bar{s}_t,\mathcal{F}_0) &= \frac{1}{(2\pi)^{nr_\alpha d/2}\prod_{k=1}^{r_\alpha}|\Lambda_{0,\alpha_k}|^{n/2}}\prod_{k=1}^{r_\alpha}|\Sigma_{\alpha_k}|^{-d/2} \\ &\times \exp\bigg\{-\frac{1}{2}\sum_{k=1}^{r_\alpha}\big(\pi_{\alpha_k}-\pi_{0,\alpha_k}\big)'\big(\Lambda_{0,\alpha_k}^{-1}\otimes\Sigma_{\alpha_k}^{-1}\big)\big(\pi_{\alpha_k}-\pi_{0,\alpha_k}\big)\bigg\},\end{aligned} \tag{50}$$

and

$$f(\Sigma_\alpha|\bar{s}_t,\mathcal{F}_0)=\prod_{k=1}^{r_\alpha}\frac{|V_{0,\alpha_k}|^{\nu_{0,\alpha_k}/2}}{\Gamma_n(\nu_{0,\alpha_k}/2)2^{n\nu_{0,\alpha_k}/2}}|\Sigma_{\alpha_k}|^{-(\nu_{0,\alpha_k}+n+1)/2}\exp\bigg\{-\frac{1}{2}\mathrm{tr}\Big(V_{0,\alpha_k}\Sigma_{\alpha_k}^{-1}\Big)\bigg\}, \tag{51}$$

respectively. Consequently, by the completing square method, a joint conditional density function of the random vectors y and $\pi_{\hat{s}}$ is

$$\begin{aligned}&f(\bar{y}_t,\pi_\alpha|\Sigma_\alpha,\bar{s}_t,\mathcal{F}_0)\\ &=c_1\prod_{k=1}^{r_\alpha}|\Sigma_{\alpha_k}|^{-(q_{t,\alpha_k}+d)/2}\exp\bigg\{-\frac{1}{2}\sum_{k=1}^{r_\alpha}\Big(\pi_{\alpha_k}-A_{\alpha_k}b_{\alpha_k}\Big)'A_{\alpha_k}^{-1}\Big(\pi_{\alpha_k}-A_{\alpha_k}b_{\alpha_k}\Big)\bigg\}\\ &\times\exp\bigg\{-\frac{1}{2}\sum_{k=1}^{r_\alpha}\Big(y_{t,\alpha_k}'\big(I_{q_{t,\alpha_k}}\otimes\Sigma_{\alpha_k}^{-1}\big)y_{t,\alpha_k}+\pi_{0,\alpha_k}'\big(\Lambda_{0,\alpha_k}^{-1}\otimes\Sigma_{\alpha_k}^{-1}\big)\pi_{0,\alpha_k}-b_{\alpha_k}'A_{\alpha_k}b_{\alpha_k}\Big)\bigg\},\end{aligned} \tag{52}$$

where normalizing coefficient equals

$$c_1:=\frac{1}{(2\pi)^{n(t+r_\alpha d)/2}\prod_{k=1}^{r_\alpha}|\Lambda_{0,\alpha_k}|^{n/2}}. \tag{53}$$

If we integrate from the above joint density function with respect to the vector π_α, then an integral, corresponding to the first exponential is proportional to $\prod_{k=1}^{r_\alpha}|A_{\alpha_k}|^{1/2}=\prod_{k=1}^{r_\alpha}|(\mathsf{Y}_{t,\alpha_k}\mathsf{Y}_{t,\alpha_k}'+\Lambda_{0,\alpha_k}^{-1})|^{-n/2}|\Sigma_{\alpha_k}|^{d/2}$. Therefore, we have that

$$f(\bar{y}_t|\Sigma_\alpha,\bar{s}_t,\mathcal{F}_0)=c_2\prod_{k=1}^{r_\alpha}|\Sigma_{\alpha_k}|^{-q_{t,\alpha_k}/2}$$
$$\times\exp\bigg\{-\frac{1}{2}\sum_{k=1}^{r_\alpha}\Big(y_{t,\alpha_k}'\big(I_{q_{t,\alpha_k}}\otimes\Sigma_{\alpha_k}^{-1}\big)y_{t,\alpha_k}+\pi_{0,\alpha_k}'\big(\Lambda_{0,\alpha_k}^{-1}\otimes\Sigma_{\alpha_k}^{-1}\big)\pi_{0,\alpha_k}-b_{\alpha_k}'A_{\alpha_k}b_{\alpha_k}\Big)\bigg\}, \tag{54}$$

where the normalizing coefficient equals

$$c_2:=\frac{1}{(2\pi)^{nt/2}\prod_{k=1}^{r_\alpha}|\Lambda_{0,\alpha_k}|^{n/2}|\Lambda_{0,\alpha_k|t}|^{n/2}}. \tag{55}$$

Hence, according to the well–known formula that for suitable matrices A,B,C,D,

$$\mathrm{vec}(A)'(B\otimes C)\mathrm{vec}(D)=\mathrm{tr}(DB'A'C), \tag{56}$$

we find that

$$f(\bar{y}_t|\Sigma_\alpha, \bar{s}_t, \mathcal{F}_0) = c_2 \prod_{k=1}^{r_\alpha} |\Sigma_{\alpha_k}|^{-q_{t,\alpha_k}/2} \exp\bigg\{ -\frac{1}{2}\sum_{k=1}^{r_\alpha} \mathrm{tr}\big(\bar{B}_{t,\alpha_k}\Sigma_{\alpha_k}^{-1}\big)\bigg\}. \tag{57}$$

Thus, it follows from equations (51) and (57) that a joint conditional density of the random vector y and random matrix $\Sigma_{\hat{s}}$ is

$$f(\bar{y}_t, \Sigma_\alpha|\bar{s}_t, \mathcal{F}_0) = c_3 \prod_{k=1}^{r_\alpha} |\Sigma_{\alpha_k}|^{-(q_{t,\alpha_k}+\nu_{0,\alpha_k}+n+1)/2} \exp\bigg\{ -\frac{1}{2}\sum_{k=1}^{r_\alpha} \mathrm{tr}\Big(\big(\bar{B}_{t,\alpha_k} + V_{0,\alpha_k}\big)\Sigma_{\alpha_k}^{-1}\Big)\bigg\}, \tag{58}$$

where the normalizing coefficient equals

$$c_3 := \frac{1}{(2\pi)^{nt/2}} \prod_{k=1}^{r_\alpha} \frac{|V_{0,\alpha_k}|^{\nu_{0,\alpha_k}/2}}{|\Lambda_{0,\alpha_k}|^{n/2}|\Lambda_{0,\alpha_k|t}|^{n/2}\Gamma_n(\nu_{0,\alpha_k}/2)2^{n\nu_{0,\alpha_k}/2}}. \tag{59}$$

Consequently, a prior density of the random vector y is given by

$$\begin{aligned} f(\bar{y}_t|\bar{s}_t, \mathcal{F}_0) &= \int_{\Sigma_{\hat{s}_1},\dots,\Sigma_{\hat{s}_{r_\alpha}}>0} f(\bar{y}_t, \hat{\Sigma}_{\bar{s}_t}|\bar{s}_t, \mathcal{F}_0)d\Sigma_{\hat{s}_1}\dots d\Sigma_{\hat{s}_{r_\alpha}} \\ &= c_3 \prod_{k=1}^{r_\alpha} \frac{\Gamma_n\big((\nu_{0,\alpha_k}+q_{t,\alpha_k})/2\big)2^{n(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}{\big|\bar{B}_{t,\alpha_k}+V_{0,\alpha_k}\big|^{(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}. \end{aligned} \tag{60}$$

If we divide equations (52) and (58) by equations (54) and (60), respectively, then one obtains equations (46) and (47). According to equation (16), we have that

$$f(\pi_\beta, \Sigma_\beta|s, \mathcal{F}_t) = f(\pi_\gamma, \Sigma_\gamma|\bar{s}_t, \mathcal{F}_t)f_*(\pi_\delta, \Sigma_\delta|\delta, \mathcal{F}_0). \tag{61}$$

We consider the first joint conditional density of the right–hand side of the above equation. By integrating a product of density functions (46) and (47) by $(\pi_\epsilon, \Sigma_\epsilon)$ and taking account that

$$\sum_{k=1}^{r_\gamma}\Big(\pi_{\gamma_k} - A_{\gamma_k}b_{\gamma_k}\Big)'A_{\gamma_k}^{-1}\Big(\pi_{\gamma_k} - A_{\gamma_k}b_{\gamma_k}\Big) = \sum_{k=1}^{r_\gamma}\mathrm{tr}\Big((\pi_{\gamma_k}^\circ - C_{t,\gamma_k})\Lambda_{0,\gamma_k|t}(\pi_{\gamma_k}^\circ - C_{t,\gamma_k})'\Sigma_{\gamma_k}^{-1}\Big) \tag{62}$$

the joint conditional density function is

$$\begin{aligned} f(\pi_\gamma, \Sigma_\gamma|\bar{s}_t, \mathcal{F}_t) &= c_4 \prod_{k=1}^{r_\gamma} |\Sigma_{\gamma_k}|^{-(\nu_{0,\gamma_k}+q_{t,\gamma_k}+d+n+1)/2} \\ &\times \exp\bigg\{ -\frac{1}{2}\sum_{k=1}^{r_\gamma}\mathrm{tr}\Big(\big(\bar{B}_{t,\gamma_k} + V_{0,\gamma_k} + (\pi_{\gamma_k}^\circ - C_{t,\gamma_k})\Lambda_{0,\gamma_k|t}(\pi_{\gamma_k}^\circ - C_{t,\gamma_k})'\big)\Sigma_{\alpha_k}^{-1}\Big)\bigg\}, \end{aligned} \tag{63}$$

where the normalizing coefficient equals

$$c_4 := \frac{1}{(2\pi)^{ndr_\gamma/2}\prod_{k=1}^{r_\gamma}|\Lambda_{0,\gamma_k|t}|^{-n/2}} \prod_{k=1}^{r_\gamma} \frac{\big|\bar{B}_{t,\gamma_k}+V_{0,\gamma_k}\big|^{(\nu_{0,\gamma_k}+q_{t,\gamma_k})/2}}{\Gamma_n\big((\nu_{0,\gamma_k}+q_{t,\gamma_k})/2\big)2^{n(\nu_{0,\gamma_k}+q_{t,\gamma_k})/2}}. \tag{64}$$

If we integrate the above equation by Σ_γ, then one obtains that

$$\begin{aligned} f(\pi_\gamma^\circ|\bar{s}_t, \mathcal{F}_t) &= \prod_{k=1}^{r_\gamma} \frac{|\Lambda_{0,\gamma_k|t}|^{n/2}\big|\bar{B}_{t,\gamma_k}+V_{0,\gamma_k}\big|^{-d/2}\Gamma_n\big((\nu_{0,\gamma_k}+q_{t,\gamma_k}+d)/2\big)}{\pi^{nd/2}\Gamma_n\big((\nu_{0,\gamma_k}+q_{t,\gamma_k})/2\big)} \\ &\times \ \big|I_n + (\bar{B}_{t,\gamma_k}+V_{0,\gamma_k})^{-1}(\pi_{\gamma_k}^\circ - C_{t,\gamma_k})\Lambda_{0,\gamma_k|t}(\pi_{\gamma_k}^\circ - C_{t,\gamma_k})'\big|^{-(\nu_{0,\gamma_k}+q_{t,\gamma_k}+d)/2}. \end{aligned} \tag{65}$$

Similarly, if $r_\delta > 0$, it can be shown that

$$\begin{aligned} f(\pi_\delta^\circ|\delta,\mathcal{F}_0) &= \prod_{\ell=1}^{r_\delta} \frac{|\Lambda_{0,\delta_\ell}|^{-n/2}|V_{0,\delta_\ell}|^{-d/2}\Gamma_n\big((\nu_{0,\delta_\ell}+d)/2\big)}{\pi^{nd/2}\Gamma_n(\nu_{0,\delta_\ell}/2)} \\ &\times \big|I_n + V_{0,\delta_\ell}^{-1}(\pi_{\delta_\ell}^\circ - \pi_{0,\delta_\ell}^\circ)\Lambda_{0,\delta_\ell}^{-1}(\pi_{\delta_\ell}^\circ - \pi_{0,\delta_\ell}^\circ)'\big|^{-(\nu_{0,\delta_\ell}+d)/2}. \end{aligned} \tag{66}$$

Therefore, equation (48) holds. By the completing square method, the matrix $\bar{B}_{t,\alpha_k}$ can be written by

$$\begin{aligned} \bar{B}_{t,\alpha_k} &= \big(\tilde{y}_{t,\alpha_k}^\circ - \pi_{0,\alpha_k}^\circ\Lambda_{0,\alpha_k}^{-1}\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ\Phi_{t,\alpha_k}^{-1}\big)\Phi_{t,\alpha_k}^{-1}\big(\tilde{y}_{t,\alpha_k}^\circ - \pi_{0,\alpha_k}^\circ\Lambda_{0,\alpha_k}^{-1}\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ\Phi_{t,\alpha_k}^{-1}\big)' \\ &- \pi_{0,\alpha_k}^\circ\Lambda_{0,\alpha_k}^{-1}\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ\Phi_{t,\alpha_k}^{-1}(\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k|t}^{-1}\Lambda_{0,\alpha_k}^{-1}(\pi_{0,\alpha_k}^\circ)' \\ &+ \pi_{0,\alpha_k}^\circ\Lambda_{0,\alpha_k}^{-1}(\pi_{0,\alpha_k}^\circ)' - \pi_{0,\alpha_k}^\circ\Lambda_{0,\alpha_k}^{-1}\Lambda_{0,\alpha_k|t}^{-1}\Lambda_{0,\alpha_k}^{-1}(\pi_{0,\alpha_k}^\circ)', \end{aligned} \tag{67}$$

where $\Phi_{t,\alpha_k} := I_{q_{t,\alpha_k}} - (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ$ is a symmetric $(q_{t,\alpha_k}\times q_{t,\alpha_k})$ matrix. We consider the following product

$$I_{t,\alpha_k} := \big(I_{q_{t,\alpha_k}} + (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\Upsilon_{t,\alpha_k}^\circ\big)\big(I_{q_{t,\alpha_k}} - (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ\big). \tag{68}$$

It equals

$$\begin{aligned} I_{t,\alpha_k} &= I_{q_{t,\alpha_k}} + (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\Upsilon_{t,\alpha_k}^\circ - (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ \\ &- (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\Upsilon_{t,\alpha_k}^\circ(\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ. \end{aligned} \tag{69}$$

If we add and subtract the matrix $\Lambda_{0,\alpha_k}^{-1}$ into the term $\Upsilon_{t,\alpha_k}^\circ(\Upsilon_{t,\alpha_k}^\circ)'$ in the last line of the above equation, then the product matrix equals $I_{t,\alpha_k} = I_{q_{t,\alpha_k}}$. Consequently, the matrix $I_{q_{t,\alpha_k}} + (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\Upsilon_{t,\alpha_k}^\circ$ is an inverse matrix of the matrix Φ_{t,α_k}, that is,

$$\Phi_{t,\alpha_k}^{-1} = I_{q_{t,\alpha_k}} + (\Upsilon_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\Upsilon_{t,\alpha_k}^\circ. \tag{70}$$

Since it is a positive definite matrix, the matrix $\bar{B}_{t,\alpha_k}$ is a positive semi–definite matrix. Now, we consider the term $\Lambda_{0,\alpha_k}^{-1}\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ\Phi_{t,\alpha_k}^{-1}$ in the first line in equation (67). Similarly as before, by adding and subtracting $\Lambda_{0,\alpha_k}^{-1}$ into the term $\Upsilon_{t,\alpha_k}^\circ(\Upsilon_{t,\alpha_k}^\circ)'$, one obtains that

$$\Lambda_{0,\alpha_k}^{-1}\Lambda_{0,\alpha_k|t}^{-1}\Upsilon_{t,\alpha_k}^\circ\Phi_{t,\alpha_k}^{-1} = \Upsilon_{t,\alpha_k}^\circ. \tag{71}$$

Consequently, the sum of the second and third line of equation (67) equals zero. Let $\Lambda_{0,\alpha_k}^{1/2}$ be the Cholesky factor of the matrix Λ_{0,α_k}, i.e., $\Lambda_{0,\alpha_k} = \big(\Lambda_{0,\alpha_k}^{1/2}\big)'\Lambda_{0,\alpha_k}^{1/2}$. Then, according to the Sylvester's determinant theorem, see Lütkepohl (2005), a determinant of the matrix Φ_{t,α_k}^{-1} is

$$|\Phi_{t,\alpha_k}^{-1}| = \big|I_{q_{t,\alpha_k}} + \big(\Lambda_{0,s_t}^{1/2}\Upsilon_{t,\alpha_k}^\circ\big)'\Lambda_{0,s_t}^{1/2}\Upsilon_{t,\alpha_k}^\circ\big| = \big|I_d + \Lambda_{0,s_t}^{1/2}\Upsilon_{t,\alpha_k}^\circ\big(\Upsilon_{t,\alpha_k}^\circ\big)'\big(\Lambda_{0,s_t}^{1/2}\big)'\big|. \tag{72}$$

That completes the proof of the Proposition. □

It follows from equations (46) and (47) that sub coefficient vectors and sub covariance matrices are conditional independent. Note that the conditional independence is consistent with the assumption (10). From equations (46) and (47) one deduces that for $k = 1,\dots,r_\alpha$, the conditional density functions of the coefficient vector π_{α_k} and the covariance matrix Σ_{α_k} are given by

$$f(\pi_{\alpha_k}|\Sigma_{\alpha_k},\alpha_k,\mathcal{F}_t) = \frac{1}{(2\pi)^{nd/2}|A_{\alpha_k}|^{1/2}}\exp\bigg\{-\frac{1}{2}\Big(\pi_{\alpha_k} - A_{\alpha_k}b_{\alpha_k}\Big)'A_{\alpha_k}^{-1}\Big(\pi_{\alpha_k} - A_{\alpha_k}b_{\alpha_k}\Big)\bigg\} \tag{73}$$

and

$$
\begin{aligned}
f(\Sigma_{\alpha_k}|\alpha_k,\mathcal{F}_t) &= \frac{\big|\bar{B}_{t,\alpha_k}+V_{0,\alpha_k}\big|^{(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}{\Gamma_n\big((\nu_{0,\alpha_k}+q_{t,\alpha_k})/2\big)2^{n(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}|\Sigma_{\alpha_k}|^{-(\nu_{0,\alpha_k}+q_{t,\alpha_k}+n+1)/2} \\
&\times \ \exp\bigg\{-\frac{1}{2}\mathrm{tr}\Big(\big(\bar{B}_{t,\alpha_k}+V_{0,\alpha_k}\big)\Sigma_{\alpha_k}^{-1}\Big)\bigg\}, \qquad (74)
\end{aligned}
$$

respectively. Thus, the conditional distribution functions of the coefficient vector π_{α_k} and the covariance matrix Σ_{α_k} are multivariate normal and inverse Wishart, respectively. Also, it follows from equation (48) that the conditional density function of the random coefficient matrix π_β° equals products of matrix variate student t density functions. The conditional density function of the random matrix π_β° can be used to impulse response analysis. Because marginal density functions of the random coefficient matrix π_β° are the matrix variate student t, their means are given by

$$
\mathbb{E}\big[\pi_{\gamma_k}^\circ\big|\gamma_k,\mathcal{F}_t\big] = C_{t,\gamma_k}, \quad k=1,\dots,r_\gamma \qquad (75)
$$

and

$$
\mathbb{E}\big[\pi_{\delta_k}^\circ\big|\delta_k,\mathcal{F}_0\big] = \pi_{0,\delta_k}, \quad k=1,\dots,r_\delta. \qquad (76)
$$

Because, according to equation (45), density function (44) has a form of the matrix variate student t distribution, we refer to the density function as a conditional matrix variate student t density function. Furthermore, it follows from equation (9) and (44) that conditional on the regime vector s and information $\mathcal{F}_t$, a density function of future values of the vector of endogenous variables is given by the following equation

$$
\begin{aligned}
f(\bar{y}_t^c|s,\mathcal{F}_t) &= \frac{1}{\pi^{n(T-t)/2}}\prod_{k=1}^{r_\alpha}\frac{|\Lambda_{0,\alpha_k|t}|^{n/2}\Gamma_n\big((\nu_{0,\alpha_k}+q_{T,\alpha_k})/2\big)\big|\bar{B}_{t,\alpha_k}+V_{0,\alpha_k}\big|^{(\nu_{0,\alpha_k}+q_{t,\alpha_k})/2}}{|\Lambda_{0,\alpha_k|T}|^{n/2}\Gamma_n((\nu_{0,\alpha_k}+q_{t,\alpha_k}/2)\big|\bar{B}_{T,\alpha_k}+V_{0,\alpha_k}\big|^{(\nu_{0,\alpha_k}+q_{T,\alpha_k})/2}} \\
&\times \ \prod_{\ell=1}^{r_\delta}\frac{|\Lambda_{0,\delta_\ell}^{-1}|^{n/2}\Gamma_n\big((\nu_{0,\delta_\ell}+q_{T,\delta_\ell})/2\big)|V_{0,\delta_\ell}|^{\nu_{0,\delta_\ell}/2}}{|\Lambda_{0,\delta_\ell|T}|^{n/2}\Gamma_n(\nu_{0,\delta_\ell}/2)\big|\bar{B}_{T,\delta_\ell}+V_{0,\delta_\ell}\big|^{(\nu_{0,\delta_\ell}+q_{T,\delta_\ell})/2}}. \qquad (77)
\end{aligned}
$$

Let us consider the following matrix

$$
\Theta_{t,\alpha_k} := \Lambda_{0,\alpha_k} - \Lambda_{0,\alpha_k}\mathsf{Y}_{t,\alpha_k}^\circ\Phi_{t,\alpha_k}(\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}. \qquad (78)
$$

Since $\Phi_{t,\alpha_k} = I_{q_{t,\alpha_k}} - (\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k|t}^{-1}\mathsf{Y}_{t,\alpha_k}^\circ$, we have that

$$
\Theta_{t,\alpha_k} = \Lambda_{0,\alpha_k} - \Lambda_{0,\alpha_k}\mathsf{Y}_{t,\alpha_k}^\circ(\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k} + \Lambda_{0,\alpha_k}\mathsf{Y}_{t,\alpha_k}^\circ(\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k|t}^{-1}\mathsf{Y}_{t,\alpha_k}^\circ(\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}. \qquad (79)
$$

If we substitute the matrix $\mathsf{Y}_{t,\alpha_k}^\circ(\mathsf{Y}_{t,\alpha_k}^\circ)' = \Lambda_{0,\alpha_k|t} - \Lambda_{0,\alpha_k}^{-1}$ into the above equation, one obtains $\Theta_{t,\alpha_k} = \Lambda_{0,\alpha_k|t}^{-1}$. Consequently, by the formula of partitioned matrix's inverse (e.g., see Lütkepohl (2005)) it can be shown that the matrix $\bar{B}_{T,\alpha_k}+V_{0,\alpha_k}$ equals

$$
\begin{aligned}
\bar{B}_{T,\alpha_k}+V_{0,\alpha_k} &= \bar{B}_{t,\alpha_k}+V_{0,\alpha_k} \\
&+ \ \big(y_{t,\alpha_k}^* - \pi_{0,\alpha_k}^\circ\mathsf{Y}_{t,\alpha_k}^* - (y_{t,\alpha_k}^\circ - \pi_{0,\alpha_k}^\circ\mathsf{Y}_{t,\alpha_k}^\circ)\Phi_{t,\alpha_k}(\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\mathsf{Y}_{t,\alpha_k}^*\big) \\
&\times \ (I_{q_{t,\alpha}^*} + \mathsf{Y}_{t,\alpha_k}^*\Lambda_{0,\alpha_k|t}^{-1}\mathsf{Y}_{t,\alpha_k}^*)^{-1} \\
&\times \ \big(y_{t,\alpha_k}^* - \pi_{0,\alpha_k}^\circ\mathsf{Y}_{t,\alpha_k}^* - (y_{t,\alpha_k}^\circ - \pi_{0,\alpha_k}^\circ\mathsf{Y}_{t,\alpha_k}^\circ)\Phi_{t,\alpha_k}(\mathsf{Y}_{t,\alpha_k}^\circ)'\Lambda_{0,\alpha_k}\mathsf{Y}_{t,\alpha_k}^*\big)', \qquad (80)
\end{aligned}
$$

where the matrices y_{t,α_k}^* and $\mathsf{Y}_{t,\alpha_k}^*$ are come from the matrices $y_{T,\alpha_k}^\circ = [y_{t,\alpha_k}^\circ : y_{t,\alpha_k}^*]$ and $\mathsf{Y}_{T,\alpha_k}^\circ = [\mathsf{Y}_{t,\alpha_k}^\circ : \mathsf{Y}_{t,\alpha_k}^*]$, corresponding to the random vectors $\bar{y}_t$ and $\bar{y}_t^c$, and $q_{t,\alpha_k}^* = q_{T,\alpha_k} - q_{t,\alpha_k}$. As a result, density function (77) is represented by a product of the conditional matrix variate student t density functions. Using the idea of proof of Proposition 2, one obtains conditional density function of the random vector y_t for given s_t and $\mathcal{F}_t$

$$f(y_t|s_t,\mathcal{F}_t)=\frac{1}{\pi^{n/2}}\frac{|\Lambda_{0,s_t}|^{-n/2}\Gamma_n\big((\nu_{0,s_t}+1)/2\big)|V_{0,s_t}|^{\nu_{0,s_t}/2}}{\big(1+\mathsf{Y}_t'\Lambda_{0,s_t}\mathsf{Y}_t\big)^{n/2}\Gamma_n(\nu_{0,s_t}/2)\big|B_{t,s_t}+V_{0,s_t}\big|^{(\nu_{0,s_t}+1)/2}}, \tag{81}$$

where

$$B_{t,s_t}:=\frac{1}{1+\mathsf{Y}_t'\Lambda_{0,s_t}\mathsf{Y}_t}\big(y_t-\pi_{0,s_t}^{\circ}\mathsf{Y}_t\big)\big(y_t-\pi_{0,s_t}^{\circ}\mathsf{Y}_t\big)' \tag{82}$$

is a symmetric positive semi–definite $(n\times n)$ matrix. Note that the conditional density function only depends on the regime s_t and does not depend on the other regimes. To calculate smoothed probabilities, the conditional density function will be used, see below.

Let us assume that the prior density functions of each row of the transition probability matrix P follow Dirichlet distribution and they are mutually independent. Under the assumption, a joint density function of them is given by

$$f(\mathsf{P}|\mathcal{F}_0)=\prod_{i=0}^{N}\frac{\Gamma\big(\sum_{j=1}^{N}\alpha_{ij}\big)}{\prod_{j=1}^{N}\Gamma(\alpha_{ij})}\prod_{j=1}^{N}p_{ij}^{\alpha_{ij}-1} \tag{83}$$

where $\Gamma(\cdot)$ is the gamma function and the parameters of Dirichlet distribution satisfy $\alpha_{ij}>0$ for $i=0,\dots,N$ and $j=1,\dots,N$. Let us denote i–th row of the random transition probability matrix P by P_i, corresponding prior hyperparameter by $\alpha_i:=(\alpha_{i1},\dots,\alpha_{iN})'$, and Dirichlet distribution by $\mathrm{Dir}(\alpha_i)$. Then, the following Lemma holds.

Proposition 3. *Let for $i=0,\dots,N$, $\mathsf{P}_i\sim\mathrm{Dir}(\alpha_i)$ and they are mutually independent. Then, the followings are hold*

(i) for $t=1,\dots,T$, conditional on the information $\mathcal{F}_0$, a density function of regime vector $\bar{s}_t$ is given by

$$f(\bar{s}_t|\mathcal{F}_0)=\prod_{i=0}^{N}\frac{\Gamma\big(\sum_{j=1}^{N}\alpha_{ij}\big)}{\prod_{j=1}^{N}\Gamma(\alpha_{ij})}\frac{\prod_{j=1}^{N}\Gamma(\alpha_{ij}+n_{ij}(\bar{s}_t))}{\Gamma\big(\sum_{j=1}^{N}(\alpha_{ij}+n_{ij}(\bar{s}_t))\big)} \tag{84}$$

(ii) and for $t=2,\dots,T$, conditional on the regime vector $\bar{s}_t$ and information $\mathcal{F}_0$, a density function of regime s_t is given by

$$f(s_t|\bar{s}_{t-1},\mathcal{F}_0)=\frac{\alpha_{s_{t-1}s_t}+n_{s_{t-1}s_t}(\bar{s}_{t-1})}{\sum_{s_t=1}^{N}\big(\alpha_{s_{t-1}s_t}+n_{s_{t-1}s_t}(\bar{s}_{t-1})\big)}, \tag{85}$$

where the random variable $n_{ij}(\bar{s}_t)$ equals

$$n_{ij}(\bar{s}_t):=\#\big\{m\in\{0,1,\dots,t-1\}\big|s_{m-1}=i,s_m=j,\ m=2,\dots,t\big\} \tag{86}$$

for $t=2,\dots,T$, $i=1,\dots,N$, and $j=1,\dots,N$ and

$$n_{ij}(s_1):=\begin{cases}1 & \text{if}\quad i=0,\ s_1=j\\ 0 & \text{if}\quad \text{otherwise}\end{cases} \tag{87}$$

for $i=0,\dots,N$ and $j=1,\dots,N$.

Proof. Since for $t = 2,\dots,T$, the random variable $n_{ij}(\bar{s}_t)$ represents a number of consequential elements, which equals (i,j) of the regime vector $\bar{s}_t$, we have that

$$f(\bar{s}_t|\mathsf{P},\mathcal{F}_0) = \prod_{m=1}^{t} p_{s_{m-1}s_m} = \prod_{i=0}^{N}\prod_{j=1}^{N} p_{ij}^{n_{ij}(\bar{s}_t)}. \tag{88}$$

Consequently, it follows from the joint density function of the random transition probability matrix P, which is given by equation (75) that

$$f(\bar{s}_t|\mathcal{F}_0) = \int_{\mathsf{P}} f(\bar{s}_t|\mathsf{P},\mathcal{F}_0)f(\mathsf{P}|\mathcal{F}_0)d\mathsf{P} = \prod_{i=0}^{N}\frac{\Gamma\big(\sum_{j=1}^{N}\alpha_{ij}\big)}{\prod_{j=1}^{N}\Gamma(\alpha_{ij})}\frac{\prod_{j=1}^{N}\Gamma(\alpha_{ij}+n_{ij}(\bar{s}_t))}{\Gamma\big(\sum_{j=1}^{N}(\alpha_{ij}+n_{ij}(\bar{s}_t))\big)}. \tag{89}$$

On the other hand, as $f(s_t|\bar{s}_t,\mathcal{F}_0) = f(\bar{s}_t|\mathcal{F}_0)/f(\bar{s}_{t-1}|\mathcal{F}_0)$, one obtains

$$f(s_t|\bar{s}_{t-1},\mathcal{F}_0) = \prod_{i=0}^{N}\frac{\prod_{j=1}^{N}\Gamma(\alpha_{ij}+n_{ij}(\bar{s}_t))}{\Gamma\big(\sum_{j=1}^{N}(\alpha_{ij}+n_{ij}(\bar{s}_t))\big)}\frac{\Gamma\big(\sum_{j=1}^{N}(\alpha_{ij}+n_{ij}(\bar{s}_{t-1}))\big)}{\prod_{j=1}^{N}\Gamma(\alpha_{ij}+n_{ij}(\bar{s}_{t-1}))}. \tag{90}$$

Consequently, since $n_{ij}(\bar{s}_t) = n_{ij}(\bar{s}_{t-1}) + \delta_{ij}(s_t)$ for $t = 2,\dots,T$ and $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$, we find that

$$f(s_t|\bar{s}_{t-1},\mathcal{F}_0) = \frac{\alpha_{s_{t-1}s_t} + n_{s_{t-1}s_t}(\bar{s}_{t-1})}{\sum_{s_t=1}^{N}\big(\alpha_{s_{t-1}s_t} + n_{s_{t-1}s_t}(\bar{s}_{t-1})\big)}, \tag{91}$$

where the random variable $\delta_{ij}(s_t)$ equals

$$\delta_{ij}(s_t) = \begin{cases} 1 & \text{if} \quad s_{t-1} = i,\ s_t = j, \\ 0 & \text{if} \quad \text{otherwise}. \end{cases} \tag{92}$$

That completes the proof. $\square$

It is worth mentioning that according to equation (85) in the above Proposition, conditional on $\mathcal{F}_0$, the regime–switching process s_t is not a Markov chain.

3.1.2 Characteristic Function

By equation (62), for $k = 1,\dots,r_\gamma$, equation (73) can be written by

$$\begin{aligned} f(\pi^\circ_{\gamma_k}|\Sigma_{\gamma_k},\gamma_k,\mathcal{F}_t) &= \frac{1}{(2\pi)^{nd/2}|\Lambda_{0,\gamma_k|t}|^{-n/2}|\Sigma_{\gamma_k}|^{d/2}} \\ &\times\ \exp\bigg\{-\frac{1}{2}\mathrm{tr}\Big(\big(\pi^\circ_{\gamma_k} - C_{t,\gamma_k}\big)\Lambda_{0,\gamma_k|t}\big(\pi^\circ_{\gamma_k} - C_{t,\gamma_k}\big)'\Sigma^{-1}_{\gamma_k}\Big)\bigg\}. \end{aligned} \tag{93}$$

To obtain characteristic function of the random coefficient matrix $\pi^\circ_{\gamma_k}$ for given regime γ_k and information $\mathcal{F}_t$, we use the matrix generalized inverse Gaussian (MGIG) distribution. For a positive definite $(n \times n)$ matrix Σ, the density function of the MGIG distribution is given by

$$f(\Sigma) = \frac{2^{n\lambda}}{|\mathsf{A}|^{\lambda}\mathcal{B}_\lambda\big(\frac{1}{4}\mathsf{BA}\big)}|\Sigma|^{\lambda-(n+1)/2}\exp\bigg\{-\frac{1}{2}\mathrm{tr}\big(\mathsf{A}\Sigma^{-1} + \mathsf{B}\Sigma\big)\bigg\}, \tag{94}$$

where $\mathcal{B}_\lambda(\cdot)$ is the matrix argument modified Bessel function of the second kind with index λ, which is defined by

$$\mathcal{B}_\lambda\Big(\frac{1}{4}\mathsf{BA}\Big) := \Big|\frac{1}{2}\mathsf{B}\Big|^{-\lambda}\int_{\Sigma>0}|\Sigma|^{-\lambda-(n+1)/2}\exp\bigg\{-\frac{1}{2}\mathrm{tr}\big(\mathsf{A}\Sigma^{-1} + \mathsf{B}\Sigma\big)\bigg\}d\Sigma \tag{95}$$

and for $n \geq 2$, the index $\lambda \in \mathbb{R}$ and the $(n \times n)$ matrices A and B satisfy

$$\begin{cases} \mathsf{A} \geq 0, \ \mathsf{B} > 0 & \text{if} \quad \lambda \geq \frac{1}{2} \\ \mathsf{A} > 0, \ \mathsf{B} > 0 & \text{if} \quad -\frac{1}{2}(n-1) \leq \lambda < \frac{1}{2} \\ \mathsf{A} > 0, \ \mathsf{B} \geq 0 & \text{if} \quad \lambda < -\frac{1}{2}(n-1) \end{cases}, \tag{96}$$

see Butler (1998). Its one dimensional version is called generalized inverse Gaussian distribution and it is widely used to model returns of financial assets, see McNeil et al. (2005). It is the well–known fact that a characteristic function of the random coefficient matrix $\pi_{\gamma_k}^\circ$ for given regime γ_k, covariance matrix Σ_{γ_k}, and information $\mathcal{F}_t$ is given by

$$\begin{aligned} \varphi(Z_{\gamma_k}|\Sigma_{\gamma_k}, \gamma_k, \mathcal{F}_t) &= \mathbb{E}\big[\exp\big\{\operatorname{tr}\big(\mathrm{i}Z_{\gamma_k}'\pi_{\gamma_k}^\circ\big)\big\}\big|\Sigma_{\gamma_k}, \gamma_k, \mathcal{F}_t\big] \\ &= \exp\bigg\{\operatorname{tr}\Big(\mathrm{i}Z_{\gamma_k}'C_{t,\gamma_k} - \frac{1}{2}Z_{\gamma_k}'\Lambda_{0,\gamma_k|t}^{-1}Z_{\gamma_k}\Sigma_{\gamma_k}\Big)\bigg\}, \end{aligned} \tag{97}$$

where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit, Z_{γ_k} is an $(n \times d)$ matrix, corresponding to the regime γ_k. Consequently, by the iterated expectation formula, conditional density function (74), and the above characteristic function of the matrix normal distribution, conditional on the regime γ_k and information $\mathcal{F}_t$, a characteristic function of the random coefficient matrix $\pi_{\gamma_k}^\circ$ is obtained by

$$\begin{aligned} \varphi(Z_{\gamma_k}|\gamma_k, \mathcal{F}_t) &= \mathbb{E}\big[\exp\big\{\operatorname{tr}\big(\mathrm{i}Z_{\gamma_k}'\pi_{\gamma_k}^\circ\big)\big\}\big|\gamma_k, \mathcal{F}_t\big] \\ &= \frac{\exp\big\{\operatorname{tr}\big(\mathrm{i}Z_{\gamma_k}'C_{t,\gamma_k}\big)\big\}\big|\bar{B}_{t,\gamma_k} + V_{0,\gamma_k}\big|^{\nu_{0,\gamma_k}+q_{t,\gamma_k}}|Z_{\gamma_k}'\Lambda_{0,\gamma_k|t}^{-1}Z_{\gamma_k}|^{\nu_{0,\gamma_k}+q_{t,\gamma_k}}}{\Gamma_n\big((\nu_{0,\gamma_k}+q_{t,\gamma_k})/2\big)2^{n(\nu_{0,\gamma_k}+q_{t,\gamma_k})}} \\ &\times \ \mathcal{B}_{\nu_{0,\gamma_k}+q_{t,\gamma_k}}\big(Z_{\gamma_k}'\Lambda_{0,\gamma_k|t}^{-1}Z_{\gamma_k}\big(\bar{B}_{t,\gamma_k} + V_{0,\gamma_k}\big)/4\big). \end{aligned} \tag{98}$$

Similarly, it can be shown that

$$\begin{aligned} \varphi(Z_{\delta_k}|\delta_k, \mathcal{F}_0) &= \mathbb{E}\big[\exp\big\{\operatorname{tr}\big(\mathrm{i}Z_{\delta_k}'\pi_{\delta_k}^\circ\big)\big\}\big|\delta_k, \mathcal{F}_0\big] \\ &= \frac{\exp\big\{\operatorname{tr}\big(\mathrm{i}Z_{\delta_k}'\pi_{0,\delta_k}^\circ\big)\big\}\big|V_{0,\delta_k}\big|^{\nu_{0,\delta_k}}|Z_{\delta_k}'\Lambda_{0,\delta_k}Z_{\delta_k}|^{\nu_{0,\delta_k}}}{\Gamma_n\big((\nu_{0,\delta_k})/2\big)2^{n\nu_{0,\delta_k}}} \\ &\times \ \mathcal{B}_{\nu_{0,\delta_k}}\big(Z_{\delta_k}'\Lambda_{0,\delta_k}Z_{\delta_k}V_{0,\delta_k}/4\big). \end{aligned} \tag{99}$$

The above characteristic functions can be used to obtain raw moments of the random coefficient matrix π_β° for given the regime vector s and information $\mathcal{F}_t$. For example, since conditional on s and $\mathcal{F}_t$, for $k = 1, \dots, r_\gamma$ and $\ell = 1, \dots, r_\delta$, coefficient matrices π_{β_k} and π_{δ_ℓ} are independent, we have that

$$\begin{aligned} &\mathbb{E}\Big[\big(\pi_{\gamma_1}^\circ\big)_{i_{\gamma_1},j_{\gamma_1}}^{m_{\gamma_1}} \dots \big(\pi_{\gamma_{r_\gamma}}^\circ\big)_{i_{\gamma_{r_\gamma}},j_{\gamma_{r_\gamma}}}^{m_{\gamma_{r_\gamma}}} \big(\pi_{\delta_1}^\circ\big)_{i_{\delta_1},j_{\delta_1}}^{m_{\delta_1}} \dots \big(\pi_{\delta_{r_\delta}}^\circ\big)_{i_{\delta_{r_\delta}},j_{\delta_{r_\delta}}}^{m_{\delta_{r_\delta}}}\Big|s, \mathcal{F}_t\Big] \\ &= \prod_{k=1}^{r_\gamma}\frac{1}{\mathrm{i}^{m_{\gamma_k}}}\frac{\partial^{m_{\gamma_k}}\varphi(Z_{\gamma_k}|\gamma_k, \mathcal{F}_t)}{\partial(Z_{\gamma_k})_{i_{\gamma_k},j_{\gamma_k}}^{m_{\gamma_k}}} \times \prod_{\ell=1}^{r_\delta}\frac{1}{\mathrm{i}^{m_{\delta_\ell}}}\frac{\partial^{m_{\delta_\ell}}\varphi(Z_{\delta_\ell}|\delta_\ell, \mathcal{F}_0)}{\partial(Z_{\gamma_k})_{i_{\delta_\ell},j_{\delta_\ell}}^{m_{\gamma_k}}}, \end{aligned} \tag{100}$$

where for a generic $(n \times m)$ matrix O, $(O)_{i,j}$ denotes an (i,j)–th element of the matrix O for $i = 1, \dots, n$ and $j = 1, \dots, m$. The partial derivatives can be calculated by the numerical methods. The raw moments may be used to obtain forecast of the vector of endogenous variables. In particular, conditional on $\bar{s}_{t+2}$, the optimal forecast, which minimizes the mean squared errors for forecast horizon 2 at forecast origin t equals an expectation of the vector of endogenous variables at time $(t+2)$ for given $\mathcal{F}_t$. Thus, the forecast is given by the following equation

$$\begin{aligned} \mathbb{E}\big[y_{t+2}\big|\bar{s}_{t+2}, \mathcal{F}_t\big] &= \mathbb{E}\big[A_{0,s_{t+2}}\big|\bar{s}_{t+2}, \mathcal{F}_t\big] + \mathbb{E}\big[A_{1,s_{t+2}}A_{0,s_{t+1}}\big|\bar{s}_{t+2}, \mathcal{F}_t\big]\psi_{t+1} \\ &+ \sum_{k=1}^{p}\mathbb{E}\big[A_{1,s_{t+2}}A_{k,s_{t+1}}|\bar{s}_{t+2}, \mathcal{F}_t\big]y_{t+1-k} + \sum_{k=2}^{p}\mathbb{E}\big[A_{k,s_{t+2}}|\bar{s}_{t+2}, \mathcal{F}_t\big]y_{t+2-k}, \end{aligned} \tag{101}$$

where the conditional expectations $\mathbb{E}\big[A_{k,s_{t+2}}\big|\bar{s}_{t+2},\mathcal{F}_t\big]$ for $k=0,2,\dots,p$ are calculated by equations (75) and (76) and the conditional expectations $\mathbb{E}\big[A_{1,s_{t+2}}A_{k,s_{t+1}}\big|\bar{s}_{t+2},\mathcal{F}_t\big]$ for $k=0,\dots,p$ are calculated equation (100). To illustrative purpose, we assume that $s_{t+1},s_{t+2}\in\mathcal{A}_{\bar{s}_t}\cap\mathcal{A}_{\bar{s}_t^c}$ and positions of the regimes s_{t+1} and s_{t+2} in the regime vector γ are k_{1*} and k_{2*}, respectively, that is,

$$k_{i*}=\{k\in\{1,\dots,r_\gamma\}|\gamma_k=s_{t+i},k=1,\dots,r_\gamma\}\tag{102}$$

for $i=1,2$. Then, we have that for $i=1,\dots,n$, $j=1,\dots,l$, and $k=0$,

$$\Big(\mathbb{E}\big[A_{1,s_{t+2}}A_{k,s_{t+1}}|\bar{s}_{t+2},\mathcal{F}_t\big]\Big)_{i,j}=-\sum_{\ell=1}^n\frac{\partial\varphi(Z_{\gamma_{k_{2*}}}|\gamma_{k_{2*}},\mathcal{F}_t)}{\partial(Z_{\gamma_{k_{2*}}})_{i,l+\ell}}\frac{\partial\varphi(Z_{\gamma_{k_{1*}}}|\gamma_{k_{1*}},\mathcal{F}_t)}{\partial(Z_{\gamma_{k_{1*}}})_{\ell,j}},\tag{103}$$

for $s_{t+1}=s_{t+2}$, $i=1,\dots,n$, and $k=1$,

$$\begin{aligned}&\Big(\mathbb{E}\big[A_{1,s_{t+2}}A_{k,s_{t+1}}|\bar{s}_{t+2},\mathcal{F}_t\big]\Big)_{i,i}\\&=-\sum_{\ell=1,\ell\neq i}^n\frac{\partial\varphi(Z_{\gamma_{k_{1*}}}|\gamma_{k_{1*}},\mathcal{F}_t)}{\partial(Z_{\gamma_{k_{1*}}})_{i,l+\ell}}\frac{\partial\varphi(Z_{\gamma_{k_{1*}}}|\gamma_{k_{1*}},\mathcal{F}_t)}{\partial(Z_{\gamma_{k_{1*}}})_{l+\ell,i}}-\frac{\partial^2\varphi(Z_{\gamma_{k_{1*}}}|\gamma_{k_{1*}},\mathcal{F}_t)}{\partial(Z_{\gamma_{k_{1*}}})_{i,l+i}^2},\end{aligned}\tag{104}$$

and for other cases ($i,j=1,\dots,n$ and $k=1,\dots,p$),

$$\Big(\mathbb{E}\big[A_{1,s_{t+2}}A_{k,s_{t+1}}|\bar{s}_{t+2},\mathcal{F}_t\big]\Big)_{i,j}=-\sum_{\ell=1}^n\frac{\partial\varphi(Z_{\gamma_{k_{2*}}}|\gamma_{k_{2*}},\mathcal{F}_t)}{\partial(Z_{\gamma_{k_{2*}}})_{i,l+\ell}}\frac{\partial\varphi(Z_{\gamma_{k_{1*}}}|\gamma_{k_{1*}},\mathcal{F}_t)}{\partial(Z_{\gamma_{k_{1*}}})_{l+(k-1)p+\ell,j}}.\tag{105}$$

Because the exact calculation of forecast of the process of endogenous variables is complicated, we consider an approximation, which is used to calculate the forecast of the endogenous variables in Bańbura et al. (2010). For $u=t+1,\dots,T$, by the iterated expectation formula, conditional on the regime vector s, the exact forecast is given by the following equation

$$\mathbb{E}\big[y_u\big|s,\mathcal{F}_t\big]=\mathbb{E}\big[\Pi_{s_u}\mathsf{Y}_u\big|s,\mathcal{F}_t\big]=\mathbb{E}\big[\mathbb{E}\big[\Pi_{s_u}\big|s,\mathcal{F}_{u-1}\big]\mathsf{Y}_u\big|s,\mathcal{F}_t\big].\tag{106}$$

Bańbura et al. (2010) approximate the last expression by $\mathbb{E}\big[\Pi_{s_u}\big|s,\mathcal{F}_{u-1}\big]\mathbb{E}\big[\mathsf{Y}_u\big|s,\mathcal{F}_t\big]$. Consequently, the forecast is approximated by

$$\mathbb{E}\big[y_u\big|s,\mathcal{F}_t\big]\approx\mathbb{E}\big[\Pi_{s_u}\big|s,\mathcal{F}_{u-1}\big]\mathbb{E}\big[\mathsf{Y}_u\big|s,\mathcal{F}_t\big].\tag{107}$$

For $t=u-1$, the approximation becomes exact, namely,

$$\mathbb{E}\big[y_u\big|s,\mathcal{F}_{u-1}\big]=\mathbb{E}\big[\Pi_{s_u}\big|s,\mathcal{F}_{u-1}\big]\mathsf{Y}_u.\tag{108}$$

However, for $u=t+2,\dots,T$, one should study the quality of the very simple approximation.

3.1.3 Minnesota Prior

In practice, one usually adopts Minnesota prior to estimating the parameters of the VAR(p) process. The first version of Minnesota prior was introduced by Litterman (1979). Also, Bańbura et al. (2010) used Minnesota prior for large Bayesian VAR and showed that the forecast of large Bayesian VAR is better than small Bayesian VAR. However, there are many different variants of the Minnesota prior, we consider a prior, which is included in Miranda-Agrippino and Ricco (2018). The idea of Minnesota prior is that it shrinks diagonal elements of the matrix A_1 toward ϕ_i and off–diagonal elements of A_{1,s_t} and all elements of other matrices $A_{0,s_t},A_{2,s_t},\dots,A_{p,s_t}$ toward 0, where ϕ_i is 0 for a stationary variable $y_{i,t}$ and 1 for a variable with unit root $y_{i,t}$. For the prior, it is assumed that conditional on

Σ_{s_t}, s_t, and $\mathcal{F}_0$, $A_0, A_1, \dots, A_p$ are jointly normally distributed, and for (i,j)–th element of the matrix A_{ℓ,s_t} $(\ell = 0, \dots, p)$, it holds

$$\mathbb{E}\big((A_{\ell,s_t})_{i,j}\big|\Sigma_{s_t}, s_t, \mathcal{F}_0\big) = \begin{cases} \phi_i & \text{if} \quad i = j, \ \ell = 1, \\ 0 & \text{if} \quad \text{otherwise} \end{cases}, \tag{109}$$

$$\mathrm{Var}\big((A_{0,s_t})_{i,j}\big|\Sigma_{s_t}, s_t, \mathcal{F}_0\big) = (\sigma_{i,s_t}/\varepsilon_{s_t})^2, \tag{110}$$

and

$$\mathrm{Var}\big((A_{\ell,s_t})_{ij}\big|\Sigma_{s_t}, s_t, \mathcal{F}_0\big) = \begin{cases} \left(\dfrac{\sigma_{i,s_t}}{\ell^{\lambda_{2,s_t}}\lambda_{1,s_t}\tau_{i,s_t}}\right)^2 & \text{if} \quad i = j, \\ \left(\dfrac{\sigma_{i,s_t}}{\ell^{\lambda_{2,s_t}}\lambda_{1,s_t}\tau_{j,s_t}}\right)^2 & \text{if} \quad \text{otherwise} \end{cases} \quad \text{for } \ell = 1, \dots, p. \tag{111}$$

The parameter $\varepsilon_{s_t}^2$ is a small number and it corresponds to an uninformative diffuse prior for $(A_{0,s_t})_{i,j}$, the parameter λ_{1,s_t} controls the overall tightness of the prior distribution, the parameter λ_{2,s_t} controls amount of information prior information at higher lags, and τ_{i,s_t} is a scaling parameter, see Miranda-Agrippino and Ricco (2018). Thus, the factor $1/\ell^{2\lambda_{2,s_t}}$ represents a rate at which prior variance decreases with increasing lag length.

According to Bańbura et al. (2010), it can be shown that the following equation satisfies the prior conditions (109), (110), and (111)

$$\hat{y}_{s_t}^\circ = \Pi_{s_t}\hat{\mathsf{Y}}_{s_t}^\circ + \hat{\xi}^\circ, \tag{112}$$

where $\hat{y}_{s_t}^\circ$ and $\hat{\mathsf{Y}}_{s_t}^\circ$ are $(n \times d)$ and $(d \times d)$ matrices of dummy variables and are defined by

$$\hat{y}_{s_t}^\circ := \big[0_{[n\times l]} : \lambda_{1,s_t}\mathrm{diag}\{\phi_1\tau_{1,s_t}, \dots, \phi_n\tau_{n,s_t}\} : 0_{[n\times(n-1)p]}\big] \tag{113}$$

and

$$\hat{\mathsf{Y}}_{s_t}^\circ := \begin{bmatrix} \varepsilon_{s_t}I_l & 0_{[l\times np]} \\ 0_{[np\times l]} & \lambda_{1,s_t}\big(J_{s_t} \otimes \mathrm{diag}\{\tau_{1,s_t}, \dots, \tau_{n,s_t}\}\big) \end{bmatrix} \tag{114}$$

with $J_{s_t} := \mathrm{diag}\{1^{\lambda_{2,s_t}}, \dots, p^{\lambda_{2,s_t}}\}$, respectively, and $\hat{\xi}^\circ := [\xi_1 : \dots : \xi_d]$ is an $(n \times d)$ matrix of residual process. Note that one can add constraints for elements of the coefficient matrix Π_{s_t} to the matrices of dummy variables, see Miranda-Agrippino and Ricco (2018). It is worth mentioning that the matrices of dummy variables $\hat{y}_{s_t}$ and $\hat{\mathsf{Y}}_{s_t}$ should not depend on the covariance matrix Σ_{s_t}. If the dummy variables depend on the covariance matrix, an OLS estimator $\hat{\pi}_{s_t}$, and matrix Λ_{0,s_t} depend on the covariance matrix Σ_{s_t}, see below. Consequently, in this case, one can not use the results of Proposition 2. For this reason, we choose the prior condition (111). Equation (112), can be written by

$$\hat{y}_{s_t} = \big((\hat{\mathsf{Y}}_{s_t}^\circ)' \otimes I_n\big)\pi_{s_t} + \hat{\xi}, \tag{115}$$

where $\hat{y}_{s_t}$ and $\hat{\xi}$ are $([nd] \times 1)$ vectors and are vectorizations of the matrix of dummy variables $\hat{y}_{s_t}^\circ$ and matrix of the residual process $\hat{\xi}^\circ$, respectively, i.e., $\hat{y}_{s_t} := \mathrm{vec}(\hat{y}_{s_t}^\circ)$ and $\hat{\xi} := \mathrm{vec}(\hat{\xi}^\circ)$. It follows from equation (115) that

$$\pi_{s_t} \stackrel{d}{=} \big(((\hat{\mathsf{Y}}_{s_t}^\circ(\hat{\mathsf{Y}}_{s_t}^\circ)')^{-1}\hat{\mathsf{Y}}_{s_t}^\circ) \otimes I_n\big)\hat{y}_{s_t} + \big(((\hat{\mathsf{Y}}_{s_t}^\circ(\hat{\mathsf{Y}}_{s_t}^\circ)')^{-1}\hat{\mathsf{Y}}_{s_t}^\circ) \otimes I_n\big)\hat{\xi}, \tag{116}$$

where d denotes equal distribution. It should be noted that the first term of the right–hand side of the above equation is a vecorization of the ordinary least square (OLS) estimator of the coefficient matrix Π_{s_t}, namely, $\pi_{0,s_t} := \mathrm{vec}\big(\hat{\Pi}_{s_t}\big) = \mathrm{vec}\big(\hat{y}_{s_t}^\circ(\hat{\mathsf{Y}}_{s_t}^\circ)'(\hat{\mathsf{Y}}_{s_t}^\circ(\hat{\mathsf{Y}}_{s_t}^\circ)')^{-1}\big)$. Consequently, conditional on Σ_{s_t}, s_t, and $\mathcal{F}_0$, a distribution of the coefficient vector π_{s_t} is given by

$$\pi_{s_t} \mid \Sigma_{s_t}, s_t, \mathcal{F}_0 \sim \mathcal{N}\Big(\pi_{0,s_t}, \big(\Lambda_{0,s_t} \otimes \Sigma_{s_t}\big)\Big). \tag{117}$$

where $\Lambda_{0,s_t} := (\hat{\mathsf{Y}}_{s_t}(\hat{\mathsf{Y}}_{s_t}^\circ)')^{-1}$ is a $(d \times d)$ diagonal matrix. Consequently, conditional on Σ_{s_t}, s_t, and $\mathcal{F}_0$, columns of the random coefficient matrix Π_{s_t} are independent. For the moment conditions (109), (110), and (111), we can use the results of Proposition 2.

3.2 Simulation Methods

By applying Propositions 2 and 3, one can obtain exact density function of the random vector $\bar{y}_t$, namely,

$$f(\bar{y}_t|\mathcal{F}_0) = \sum_{\bar{s}_t} f(\bar{y}_t|\bar{s}_t, \mathcal{F}_0) \prod_{i=0}^N \frac{\Gamma\big(\sum_{j=1}^N \alpha_{ij}\big)}{\prod_{j=1}^N \Gamma(\alpha_{ij})} \frac{\prod_{j=1}^N \Gamma(\alpha_{ij} + n_{ij}(\bar{s}_t))}{\Gamma\big(\sum_{j=1}^N (\alpha_{ij} + n_{ij}(\bar{s}_t))\big)}. \tag{118}$$

However, this exact density function has the following two main disadvantages:

(i) it is difficult to obtain characteristics (such as mean, quantile, marginal densities, distribution, and so on) of the mixture density function $f(\bar{y}_t|\mathcal{F}_0)$,

(ii) and it is difficult to calculate the sum with respect to $\bar{s}_t$. For example, if the length of the regime vector $\bar{s}_t$ equals 30 and the regime number equals 3, then we have to calculate $3^{30} \approx 2.06 \times 10^{14}$ summands.

If the dimensions increase, the disadvantages are seriously worsen. Therefore, from a practical point of view, we need to develop Monte–Carlo simulation method.

3.2.1 General Method

According to the conditional probability formula and Proposition 1, we have that

$$\begin{aligned} f(\bar{y}_t^c, \pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_t) &= f(\bar{y}_t^c|\pi_\beta, \Sigma_\beta, s, \mathcal{F}_t) f_*(\pi_\delta|\Sigma_\delta, \delta, \mathcal{F}_0) f_*(\Sigma_\delta|\delta, \mathcal{F}_0) \\ &\times f(\pi_\alpha|\Sigma_\alpha, \bar{s}_t, \mathcal{F}_t) f(\Sigma_\alpha|\bar{s}_t, \mathcal{F}_t) f(\bar{s}_t^c|s_t, \mathsf{P}, \mathcal{F}_0) f(\bar{s}_t, \mathsf{P}|\mathcal{F}_t). \end{aligned} \tag{119}$$

The above equation tells us that how to generate random sample from $(\bar{y}_t^c, \pi_{\hat{s}_t}, \Sigma_{\hat{s}_t}, s, \mathsf{P})$ for given information $\mathcal{F}_t$. The direction of our simulation method move toward from right to left for the above equation. To generate random samples, first, generate the regime vector and transition probability matrix $(\bar{s}_t, \mathsf{P})$ from the posterior density function $f(\bar{s}_t, \mathsf{P}|\mathcal{F}_t)$. Next, using the regime vector $\bar{s}_t$ and transition probability matrix P, generate regime vector $\bar{s}_t^c$ from the conditional density function $f(\bar{s}_t^c|\bar{s}_t, \mathsf{P}, \mathcal{F}_0)$, so on.

First, we consider a simulation method that generate the regime vector and transition probability matrix $(\bar{s}_t, \mathsf{P})$ from the posterior density function $f(\bar{s}_t, \mathsf{P}|\mathcal{F}_t)$. As mentioned above, for given $\mathcal{F}_0$, the regime–switching process s_t is not a Markov chain. Therefore, we develop the Gibbs sampling method to generate $(\bar{s}_t, \mathsf{P})$. In the Bayesian statistics, the Gibbs sampling is often used when the joint distribution is not known explicitly or is difficult to sample from directly, but the conditional distribution of each variable is known and is easy to sample from. Constructing the Gibbs sampler to approximate the joint posterior distribution $f(\bar{s}_t, \mathsf{P}|\mathcal{F}_t)$ is straightforward: New values $\big(\bar{s}_t(\ell), \mathsf{P}(\ell)\big)$, $\ell = 1, \dots, \mathcal{L}$ can be generated by

- generate $\mathsf{P}(\ell)$ from $f\left(\mathsf{P}|\bar{s}_t(\ell), \mathcal{F}_t\right)$,
- generate $\bar{s}_t(\ell)$ from $f\big(\bar{s}_t\big|\mathsf{P}(\ell), \mathcal{F}_t\big)$.

To generate $\mathsf{P}(\ell)$ from $f(\mathsf{P}|\bar{s}_t(\ell), \mathcal{F}_t)$, we apply equation (17) in Proposition 1, density function (83), equation (84) in Proposition 3, and the Bayesian formula. Then, we have that

$$f(\mathsf{P}|\bar{s}_t(\ell), \mathcal{F}_t) = \prod_{i=0}^N \frac{\Gamma\big(\sum_{j=1}^N (\alpha_{ij} + n_{ij}(\bar{s}_t(\ell)))\big)}{\prod_{j=1}^N \Gamma(\alpha_{ij} + n_{ij}(\bar{s}_t(\ell)))} \prod_{j=1}^N p_{ij}^{\alpha_{ij} + n_{ij}(\bar{s}_t(\ell)) - 1}. \tag{120}$$

Thus, one can deduce that conditional on $\bar{s}_t(\ell)$ and $\mathcal{F}_t$, for $i = 0, \dots, N$, each row of the transition probability matrix P are independent and has Dirichlet distribution with parameter $\alpha_i(\bar{s}_t(\ell)) := (\alpha_{i1} + n_{i1}(\bar{s}_t(\ell)), \dots, \alpha_{iN} + n_{iN}(\bar{s}_t(\ell)))'$. Consequently, the simulation method that generate $\mathsf{P}(\ell)$ for given $\bar{s}_t(\ell)$ and $\mathcal{F}_t$ as follows:

- generate $\mathsf{P}_i(\ell)$ from $\text{Dir}(\alpha_i(\bar{s}_t(\ell)))$ for $i = 0, \dots, N$.

Collect $\mathsf{P}_i(\ell)$ for $i = 0, \dots, N$ into an $([N+1] \times N)$ matrix $\mathsf{P}(\ell)$, that is, $\mathsf{P}(\ell) := [\mathsf{P}_0(\ell)' : \dots : \mathsf{P}_N(\ell)']'$. Let $\hat{\mathsf{P}}(\ell) := [\mathsf{P}_1(\ell)' : \dots : \mathsf{P}_N(\ell)']'$ be an $(N \times N)$ transition probability matrix, which omits first row of the matrix $\mathsf{P}(\ell)$.

Now we consider a sampling method that generate $\bar{s}_t(\ell)$ from $f\big(\bar{s}_t \big| \mathsf{P}(\ell), \mathcal{F}_t\big)$. Here we follow the book of Hamilton (1994), see also Battulga (2022) and Battulga (2023a). If we assume that the regime–switching process in regime j at time u, then according to equation (81), the conditional density function of the random vector y_u is given by the following equation

$$\begin{aligned} \eta_{u,j} &:= f(y_u | s_u = j, \mathcal{F}_{u-1}) \\ &= \frac{1}{\pi^{n/2}} \frac{|\Lambda_{0,j}|^{-n/2} \Gamma_n\big((\nu_{0,j}+1)/2\big) |V_{0,j}|^{\nu_{0,j}/2}}{\big(1 + \mathsf{Y}_u' \Lambda_{0,j} \mathsf{Y}_u\big)^{n/2} \Gamma_n(\nu_{0,j}/2) \big|B_{u,j} + V_{0,j}\big|^{(\nu_{0,j}+1)/2}}, \end{aligned} \tag{121}$$

for $u = 1, \dots, t$ and $j = 1, \dots, N$, where $B_{u,j}$ is given by equation (82). For all $u = 1, \dots, t$, we collect the conditional density functions of y_u into an $(n \times 1)$ vector η_t, that is, $\eta_u := (\eta_{u,1}, \dots, \eta_{u,N})'$. Let us denote a probabilistic inference about the value of the regime–switching process s_u equals to j, based on the information $\mathcal{F}_u$ and transition probability matrix $\mathsf{P}(\ell)$ by $\mathbb{P}(s_u = j | \mathsf{P}(\ell), \mathcal{F}_u)$. Collect these conditional probabilities $\mathbb{P}(s_u = j | \mathsf{P}(\ell), \mathcal{F}_u)$ for $j = 1, \dots, N$ into an $(N \times 1)$ vector $z_{u|u}(\ell)$, that is, $z_{u|u}(\ell) := \big(\mathbb{P}(s_u = 1 | \mathsf{P}(\ell), \mathcal{F}_u), \dots, \mathbb{P}(s_u = N | \mathsf{P}(\ell), \mathcal{F}_u)\big)'$. Also, we need a probabilistic forecast about the value of the regime–switching process at time $u+1$ equals j conditional on data up to and including time u and transition probability matrix $\mathsf{P}(\ell)$. Collect these forecasts into an $(N \times 1)$ vector $z_{u+1|u}(\ell)$, that is, $z_{u+1|u}(\ell) := \big(\mathbb{P}(s_{u+1} = 1 | \mathsf{P}(\ell), \mathcal{F}_u), \dots, \mathbb{P}(s_{u+1} = N | \mathsf{P}(\ell), \mathcal{F}_u)\big)'$.

The probabilistic inference and forecast for each time $u = 1, \dots, t$ can be found by iterating on the following pair of equations:

$$z_{u|u}(\ell) = \frac{(z_{u|u-1}(\ell) \odot \eta_u)}{i_N'(z_{u|u-1}(\ell) \odot \eta_u)} \quad \text{and} \quad z_{u+1|u}(\ell) = \hat{\mathsf{P}}(\ell)' z_{u|u}(\ell), \quad u = 1, \dots, t, \tag{122}$$

where $\odot$ is the Hadamard product of two vectors, η_u is the $(N \times 1)$ vector, whose j-th element is given by equation (121), $\hat{\mathsf{P}}(\ell)$ is the $(N \times N)$ transition probability matrix, and i_N is an $(N \times 1)$ vector, whose elements equal 1. Given a starting value $z_{1|0}(\ell) := \mathsf{P}_0(\ell)'$ one can iterate on (122) for $u = 1, \dots, t$ to calculate the values of $z_{u|u}(\ell)$ and $z_{u+1|u}(\ell)$. To obtain marginal distributions of the regime vector $\bar{s}_t$ conditional on the transition probability matrix $\mathsf{P}(\ell)$ and information $\mathcal{F}_t$, let us introduce $(N \times 1)$ smoothed inference vector $z_{u|t}(\ell) := \big(\mathbb{P}(s_u = 1 | \mathsf{P}(\ell), \mathcal{F}_t), \dots, \mathbb{P}(s_u = N | \mathsf{P}(\ell), \mathcal{F}_t)\big)'$ for $u = 1, \dots, t$. The smoothed inference vectors can be obtained by using the Kim (1994)'s smoothing algorithm:

$$z_{u|t}(\ell) = z_{u|u}(\ell) \odot \big\{\hat{\mathsf{P}}(\ell)'(z_{u+1|t}(\ell) \oslash z_{u+1|u}(\ell))\big\}, \quad u = t-1, \dots, 1, \tag{123}$$

where $\oslash$ is an element–wise division of two vectors. The smoothed probabilities $z_{u|t}(\ell)$ are found by iterating on (123) backward for $u = t-1, \dots, 1$. This iteration is started with $z_{t|t}(\ell)$, which is obtained from (122) for $u = t$. Thus, the simulation method that generate $\bar{s}_t(\ell)$ for given $\mathsf{P}(\ell)$ and $\mathcal{F}_t$ as follows:

- generate $s_u(\ell)$ from $z_{u|t}(\ell)$ for $u = 1, \dots, t$.

Collect $s_u(\ell)$ for $u=1,\dots,t$ into $(t\times 1)$ vector $\bar{s}_t(\ell)$, namely, $\bar{s}_t(\ell):=(s_1(\ell),\dots,s_t(\ell))'$.

Second, we consider a simulation method that generate the regime vector $\bar{s}_t^c(\ell)$ from the density function $f(\bar{s}_t^c|s_t(\ell),\mathsf{P}(\ell),\mathcal{F}_0)$. Note that conditional on the transition probability matrix $\mathsf{P}(\ell)$ and initial information $\mathcal{F}_0$, the regime–switching process s_t is a Markov chain. Thus, to sample the regime vector $\bar{s}_t^c(\ell)$ from the density function $f(\bar{s}_t^c|s_t(\ell),\mathsf{P}(\ell),\mathcal{F}_0)$, we can use the Markov property. That is, we have that

$$f(\bar{s}_t^c|s_t(\ell),\mathsf{P}(\ell),\mathcal{F}_0)=f(s_T|s_{T-1},\mathsf{P}(\ell),\mathcal{F}_0)\dots f(s_{t+1}|s_t(\ell),\mathsf{P}(\ell),\mathcal{F}_0). \tag{124}$$

Thus, a simulation method that generate $\bar{s}_t^c(\ell)$ for given $s_t(\ell)$, $\mathsf{P}(\ell)$, and $\mathcal{F}_0$ as follows:

- generate $s_{t+1}(\ell)$ from $f(s_{t+1}|s_t(\ell),\mathsf{P}(\ell),\mathcal{F}_0)$,
- generate $s_{t+2}(\ell)$ from $f(s_{t+2}|s_{t+1}(\ell),\mathsf{P}(\ell),\mathcal{F}_0)$,

  $\dots$
- generate $s_T(\ell)$ from $f(s_T|s_{T-1}(\ell),\mathsf{P}(\ell),\mathcal{F}_0)$.

Collect $s_u(\ell)$ for $u=t+1,\dots,T$ into $([T-t]\times 1)$ vector $\bar{s}_t^c(\ell)$, namely, $\bar{s}_t^c(\ell):=(s_{t+1}(\ell),\dots,s_T(\ell))'$.

Third, we consider a simulation method that generate the coefficient vector $\pi_{\bar{s}_t^c(\ell)}(\ell)$ and covariance matrix $\Sigma_{\bar{s}_t^c(\ell)}(\ell)$ from the density function $f\big(\pi_{\bar{s}_t^c(\ell)},\Sigma_{\bar{s}_t^c(\ell)}\big|s(\ell),\mathcal{F}_t\big)$. Let $\mathcal{A}_{\bar{s}_t(\ell)}=\big\{\alpha_1(\ell),\dots,\alpha_{r_{\alpha(\ell)}}(\ell)\big\}$ and $\mathcal{A}_{\bar{s}_t^c(\ell)}=\big\{\beta_1(\ell),\dots,\beta_{r_{\beta(\ell)}}(\ell)\big\}$ be the duplication removed sets, corresponding the regime vectors $\bar{s}_t(\ell)$ and $\bar{s}_t^c(\ell)$, respectively. To eliminate unnecessary simulations, instead of the regimes in the set $\mathcal{A}_{\bar{s}_t(\ell)}$, one should consider regimes in intersection set of the sets $\mathcal{A}_{\bar{s}_t(\ell)}$ and $\mathcal{A}_{\bar{s}_t^c(\ell)}$. Let us assume that the intersection set and a difference set of the sets are given by $\mathcal{A}_{\bar{s}_t(\ell)}\cap\mathcal{A}_{\bar{s}_t^c(\ell)}=\big\{\gamma_1(\ell),\dots,\gamma_{r_{\gamma(\ell)}}(\ell)\big\}$ and $\mathcal{A}_{\bar{s}_t^c(\ell)}\backslash\mathcal{A}_{\bar{s}_t(\ell)}=\big\{\delta_1(\ell),\dots,\delta_k(\ell)\big\}$, respectively. Then, according to equations (10), (73) and (74), a simulation method that generates $\big(\pi_{\gamma_k(\ell)}(\ell),\Sigma_{\gamma_k(\ell)}(\ell)\big)$ for $\ell=1,\dots,r_{\gamma(\ell)}$ as follows: for $k=1,\dots,r_{\gamma(\ell)}$,

- generate $\Sigma_{\gamma_k(\ell)}(\ell)$ from $\mathcal{IW}\Big(\nu_{0,\gamma_k(\ell)}+q_{t,\gamma_k(\ell)},\bar{B}_{t,\gamma_k(\ell)}+V_{0,\gamma_k(\ell)}\Big)$,
- generate $\pi_{\gamma_k(\ell)}(\ell)$ from $\mathcal{N}\Big(A_{\gamma_k(\ell)}b_{\gamma_k(\ell)},A_{\gamma_k(\ell)}\Big)$.

On the other hand, according to equations (10) and (15) and marginal density functions of equations (50) and (51), a simulation method that generates $\big(\pi_{\delta_k(\ell)}(\ell),\Sigma_{\delta_k(\ell)}(\ell)\big)$ for $\ell=1,\dots,r_{\delta(\ell)}$ as follows: if $r_{\delta(\ell)}>0$, then for $k=1,\dots,r_{\delta(\ell)}$,

- generate $\Sigma_{\delta_k(\ell)}(\ell)$ from $\mathcal{IW}\Big(\nu_{0,\delta_k(\ell)},V_{0,\delta_k(\ell)}\Big)$,
- generate $\pi_{\delta_k(\ell)}(\ell)$ from $\mathcal{N}\Big(\pi_{0,\delta_k(\ell)},\Lambda_{0,\delta_k(\ell)}\otimes\Sigma_{\delta_k(\ell)}(\ell)\Big)$.

Let $\hat{\beta}(\ell):=\big(\gamma_1(\ell),\dots,\gamma_{r_{\gamma(\ell)}}(\ell),\delta_1(\ell),\dots,\delta_{r_{\delta(\ell)}}(\ell)\big)'$ be an $(r_{\beta(\ell)}\times 1)$ regime vector. The regime vector $\hat{\beta}(\ell)$ has same elements as the duplication removed regime vector $\beta(\ell)$, but positions of the elements are different for the two regime vectors. Collect the realizations for $k=1,\dots,r_{\gamma(\ell)}$, $\Sigma_{\gamma_k(\ell)}(\ell)$ and $\pi_{\gamma_k(\ell)}(\ell)$ and for $k=1,\dots,r_{\delta(\ell)}$, $\Sigma_{\delta_k(\ell)}(\ell)$ and $\pi_{\delta_k(\ell)}(\ell)$ into $([ndr_{\beta(\ell)}]\times 1)$ vector $\pi_{\hat{\beta}(\ell)}(\ell)$ and $(n\times[ndr_{\beta(\ell)}])$ matrix $\Sigma_{\hat{\beta}(\ell)}(\ell)$, namely,

$$\pi_{\hat{\beta}(\ell)}(\ell):=\Big(\pi_{\gamma_1(\ell)}(\ell)',\dots,\pi_{\gamma_{r_{\gamma(\ell)}}(\ell)}(\ell)',\pi_{\delta_1(\ell)}(\ell)',\dots,\pi_{\delta_{r_{\delta(\ell)}}(\ell)}(\ell)'\Big)' \tag{125}$$

and

$$\Sigma_{\hat{\beta}(\ell)}(\ell):=\Big[\Sigma_{\gamma_1(\ell)}(\ell):\dots:\Sigma_{\gamma_{r_{\gamma(\ell)}}(\ell)}(\ell),\Sigma_{\delta_1(\ell)}(\ell):\dots:\Sigma_{\delta_{r_{\delta(\ell)}}(\ell)}(\ell)\Big]. \tag{126}$$

Similar to equation (37), for $u = t+1, \dots, T$, we denote a position of the regime $s_u(\ell)$ in the regime vector $\hat{\beta}(\ell)$ by $o_\ell(\ell)$. Let us define a matrix $D_{\hat{\beta}(\ell)} := \big[j_{o_{t+1}(\ell)} : \dots : j_{o_T(\ell)}\big]'$, where $j_o(\ell)$ is an $(r_{\beta(\ell)} \times 1)$ unit vector, whose o–th element 1 and others 0. Then, one revives the vector $\pi_{\bar{s}_t^c(\ell)} = \big(D_{\hat{\beta}(\ell)} \otimes I_{nd}\big)\pi_{\hat{\beta}(\ell)}$ and matrix

$$\Sigma_{\bar{s}_t^c(\ell)} = \operatorname{diag}\big\{\big((D_{\hat{\beta}(\ell)} \otimes I_n)\Sigma_{\hat{\beta}(\ell)}\big)_1, \dots, \big((D_{\hat{\beta}(\ell)} \otimes I_n)\Sigma_{\hat{\beta}(\ell)}\big)_T\big\}. \tag{127}$$

Fourth, we consider a simulation method that generate the regime vector $\bar{y}_t^c(\ell)$ from the density function $f(\bar{y}_t^c|\pi_\beta, \Sigma_\beta, s, \mathcal{F}_t)$. Let us assume that

$$\varphi_s = \begin{bmatrix} \varphi_{1,\bar{s}_t} \\ \varphi_{2,\bar{s}_t^c} \end{bmatrix} \quad \text{and} \quad \Psi_s = \begin{bmatrix} \Psi_{11,\bar{s}_t} & 0 \\ \Psi_{21,\bar{s}_t^c} & \Psi_{22,\bar{s}_t^c} \end{bmatrix} \tag{128}$$

are partitions of the vector φ_s and matrix Ψ_s, corresponding to random sub vectors $\bar{y}_t$ and $\bar{y}_t^c$ of the random vector y. Then, due to Battulga (2024b), a distribution of the random vector $\bar{y}_t^c$ is given by

$$\bar{y}_t^c \mid \pi_\beta, \Sigma_\beta, s, \mathcal{F}_t \ \sim \ \mathcal{N}\Big(\Psi_{22,\bar{s}_t^c}^{-1}\big(\varphi_{2,\bar{s}_t^c} - \Psi_{21,\bar{s}_t^c}\bar{y}_t\big), \Psi_{22,\bar{s}_t^c}^{-1}\Sigma_{\bar{s}_t^c}\big(\Psi_{22,\bar{s}_t^c}^{-1}\big)'\Big), \tag{129}$$

where $\Sigma_{\bar{s}_t^c} = \operatorname{diag}\{\Sigma_{s_{t+1}}, \dots, \Sigma_{s_T}\}$ is an $([n(T-t)] \times [n(T-t)])$ block diagonal matrix, corresponding to the regime vector $\bar{s}_t^c$. Thus, a simulation method that generate the vector of endogenous variables $\bar{y}_t^c(\ell)$ for given $\pi_{\bar{s}_t(\ell)}(\ell)$, $\Sigma_{\bar{s}_t(\ell)}(\ell)$, $s(\ell)$, and $\mathcal{F}_t$ as follows:

- generate $\bar{y}_t^c(\ell)$ from $\mathcal{N}\Big(\Psi_{22,\bar{s}_t^c(\ell)}^{-1}\big(\varphi_{2,\bar{s}_t^c(\ell)} - \Psi_{21,\bar{s}_t^c(\ell)}\bar{y}_t\big), \Psi_{22,\bar{s}_t^c(\ell)}^{-1}\Sigma_{\bar{s}_t^c(\ell)}(\ell)\big(\Psi_{22,\bar{s}_t^c(\ell)}^{-1}\big)'\Big)$,

where the matrix $\Psi_{22,\bar{s}_t^c(\ell)}$ and vector $\varphi_{\bar{s}_t^c(\ell)} - \Psi_{21,\bar{s}_t^c(\ell)}\bar{y}_t$ are given by

$$\Psi_{22,\bar{s}_t^c(\ell)} := \begin{bmatrix} I_n & 0 & \dots & 0 & \dots & 0 & 0 \\ -A_{1,s_{t+2}(\ell)} & I_n & \dots & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & -A_{p-1,s_{T-1}(\ell)} & \dots & I_n & 0 \\ 0 & 0 & \dots & -A_{p,s_T(\ell)} & \dots & -A_{1,s_T(\ell)} & I_n \end{bmatrix} \tag{130}$$

and

$$\varphi_{2,\bar{s}_t^c(\ell)} - \Psi_{21,\bar{s}_t^c(\ell)}\bar{y}_t := \begin{bmatrix} A_{0,s_{t+1}(\ell)}\psi_{t+1} + A_{1,s_{t+1}(\ell)}y_t + \dots + A_{p,s_{t+1}(\ell)}y_{t+1-p} \\ A_{0,s_{t+2}(\ell)}\psi_{t+2} + A_{2,s_{t+2}(\ell)}y_t + \dots + A_{p,s_{t+2}(\ell)}y_{t+2-p} \\ \vdots \\ A_{0,s_{T-1}(\ell)}\psi_{T-1} \\ A_{0,s_T(\ell)}\psi_T \end{bmatrix} \tag{131}$$

and they are obtained from the vector $\pi_{\bar{s}_t^c(\ell)}(\ell)$. It should be noted that traditional methods that generate the vector $\bar{y}_t^c$ are based on an iterative method for $y_{t+1}, \dots, y_T$ by generating $\xi_{t+1}, \dots, \xi_T$, see Karlsson (2013). As a result, if $T-t$ is large, the simulation method reduces the computational burden that generates the random vector $\bar{y}_t^c$ as compared to the traditional algorithms.

3.2.2 Importance Sampling Method

Now, we consider the importance sampling method for the Bayesian MS–VAR process. We estimate a probability of a rare event, corresponding to the endogenous variables by the important sampling method. In the importance sampling method, one changes the real probability measure $\mathbb{P}$. The new probability measure $\tilde{\mathbb{P}}$ must be chosen that the rare event more frequently comes from than the real probability measure $\mathbb{P}$. Let $f(y, \pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0)$ and $\tilde{f}(y, \pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0)$ be joint density functions under the real probability measure $\mathbb{P}$ and new probability measure $\tilde{\mathbb{P}}$, respectively, for given initial information $\mathcal{F}_0$ and

$$L = L(y, \pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0) = f(y, \pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0)/\tilde{f}(y, \pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0) \tag{132}$$

be the likelihood ratio. Let us choose density functions that corresponds to the new probability measure $\tilde{\mathbb{P}}$ by

$$\tilde{f}(\bar{y}_t, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0) = f(\bar{y}_t, \Sigma_{\hat{s}}, s, \mathsf{P}|\mathcal{F}_0), \tag{133}$$

if $s_u \in \gamma$, then

$$\begin{aligned}
&\tilde{f}(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t) = f(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t) = f(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathcal{F}_t)\\
&\Big/\frac{1}{(2\pi)^{nd}|A_{s_u}|^{1/2}}\exp\bigg\{-\frac{1}{2}\Big(\pi_{s_u} - A_{s_u}b_{s_u}\Big)'A_{s_u}^{-1}\Big(\pi_{s_u} - A_{s_u}b_{s_u}\Big)\bigg\}\\
&\times\frac{1}{(2\pi)^{nd}|A_{s_u}|^{1/2}}\exp\bigg\{-\frac{1}{2}\Big(\pi_{s_u} - A_{s_u}b_{s_u} - \theta_{s_u}A_{s_u}(\mathsf{Y}_u\otimes I_n)z_u\Big)'\\
&\times A_{s_u}^{-1}\Big(\pi_{s_u} - A_{s_u}b_{s_u} - \theta_{s_u}A_{s_u}(\mathsf{Y}_u\otimes I_n)z_u\Big)\bigg\}
\end{aligned} \tag{134}$$

if $s_u \in \delta$, then

$$\begin{aligned}
&\tilde{f}(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t) = f(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t) = f(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathcal{F}_t)\\
&\Big/\frac{1}{(2\pi)^{nd}|\Lambda_{0,s_u}|^{n/2}|\Sigma_{s_u}|^{d/2}}\exp\bigg\{-\frac{1}{2}\Big(\pi_{s_u} - \pi_{0,s_u}\Big)'\big(\Lambda_{0,s_u}^{-1}\otimes\Sigma_{s_u}^{-1}\big)\Big(\pi_{s_u} - \pi_{0,s_u}\Big)\bigg\}\\
&\times\frac{1}{(2\pi)^{nd}|\Lambda_{0,s_u}|^{n/2}|\Sigma_{s_u}|^{d/2}}\exp\bigg\{-\frac{1}{2}\Big(\pi_{s_u} - \pi_{0,s_u} - \theta_{s_u}(\Lambda_{0,s_u}\mathsf{Y}_u\otimes\Sigma_{s_u})z_u\Big)'\\
&\times\big(\Lambda_{0,s_u}^{-1}\otimes\Sigma_{s_u}^{-1}\big)\Big(\pi_{s_u} - \pi_{0,s_u} - \theta_{s_u}(\Lambda_{0,s_u}\mathsf{Y}_u\otimes\Sigma_{s_u})z_u\Big)\bigg\},
\end{aligned} \tag{135}$$

and

$$\begin{aligned}
&\tilde{f}(\bar{y}_t^c|\pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t) = f(\bar{y}_t^c|\pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t) = f(\bar{y}_t^c|\pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathcal{F}_t)\\
&\Big/\frac{1}{(2\pi)^{n}|\Sigma_{s_u}|^{1/2}}\exp\bigg\{-\frac{1}{2}\Big(y_u - \Pi_{s_u}\mathsf{Y}_u\Big)'\Sigma_{s_u}^{-1}\Big(y_u - \Pi_{s_u}\mathsf{Y}_u\Big)\bigg\}\\
&\times\frac{1}{(2\pi)^{n}|\Sigma_{s_u}|^{1/2}}\exp\bigg\{-\frac{1}{2}\Big(y_u - \Pi_{s_u}\mathsf{Y}_u - \theta_{s_u}\Sigma_{s_u}z_u\Big)'\Sigma_{s_u}^{-1}\Big(y_u - \Pi_{s_u}\mathsf{Y}_u - \theta_{s_u}\Sigma_{s_u}z_u\Big)\bigg\}
\end{aligned} \tag{136}$$

for $u = t+1, \dots, T$, where θ_{s_u} is a positive constant, depending on the random coefficient vector π_{s_u}, random covariance matrix Σ_{s_t}, and regime s_u and z_u is an $(n \times 1)$ vector, whose elements are known. Note that if $\theta_{s_u} = 0$, then the new probability measure $\tilde{\mathbb{P}}$ equals the real probability measure $\mathbb{P}$. If we compare the density functions $\tilde{f}(\bar{y}_t^c|\pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t)$ and $f(\bar{y}_t^c|\pi_{\hat{s}}, \Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t)$, one can conclude that conditional distribution of y_u changes from $y_u|\pi_{s_u}, s_u, \mathcal{F}_{u-1} \sim \mathcal{N}(\Pi_{s_u}\mathsf{Y}_u, \Sigma_{s_u})$ to $y_u|\pi_{s_u}, s_u, \mathcal{F}_{u-1} \sim \mathcal{N}(\Pi_{s_u}\mathsf{Y}_u + \theta_{s_u}\Sigma_{s_u}z_u, \Sigma_{s_u})$ and for each $v = t+1, \dots, T$ $(v \neq u)$, the conditional distribution of other sub random vector y_v of the random vector $\bar{y}_t^c$ does not change. The same explanation holds for the density functions $\tilde{f}(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t)$ and $f(\pi_{\hat{s}}|\Sigma_{\hat{s}}, s, \mathsf{P}, \mathcal{F}_t)$. Then, it can be shown that for $u = t+1, \dots, T$, the likelihood ratio is given by

$$L_u = \exp\big\{-\theta_{s_u}X_u + \psi(\theta_{s_u})\big\}. \tag{137}$$

where the random variable X_u is given by $X_u := z_u'y_u$ and the quadratic function $\psi(\theta_{s_u})$ for θ_{s_u} is given by

$$\psi(\theta_{s_u}) := \begin{cases}\theta_u z_u'C_{u,s_u}\mathsf{Y}_u + \frac{1}{2}\theta_u^2\big(1 + \mathsf{Y}_u'\Lambda_{0,s_u|u}^{-1}\mathsf{Y}_u\big)z_u'\Sigma_{s_u}z_u & \text{if} \quad s_u \in \gamma,\\ \theta_u z_u'\Pi_{s_u}\mathsf{Y}_u + \frac{1}{2}\theta_u^2\big(1 + \mathsf{Y}_u'\Lambda_{0,s_u}\mathsf{Y}_u\big)z_u'\Sigma_{s_u}z_u & \text{if} \quad s_u \in \delta.\end{cases} \tag{138}$$

For each $u = t+1, \dots, T$, by choosing z_u by the unit vector, one can extract elements of the vector of endogenous variables y_u. If the process y_t consists of returns of financial assets, then by choosing z_u by weight vector, one obtains portfolio return.

Now we consider a conditional probability $\mathbb{P}(X_u > x_u|\pi_\beta, \Sigma_\beta, s, \mathcal{F}_{u-1})$ for large $x_u \in \mathbb{R}$ and $u = t+1, \dots, T$. Since θ_{s_u} is the positive constant, for the conditional probability, by equation (137), the following inequality holds

$$\mathbb{P}(X_u > x_u|\pi_\beta, \Sigma_\beta, s, \mathcal{F}_{u-1}) = \tilde{\mathbb{E}}\big[1_{\{X_u > x_u\}} L_u \big|\pi_\beta, \Sigma_\beta, s, \mathcal{F}_{u-1}\big] \leq \exp\big\{-\theta_{s_u} x_u + \psi(\theta_{s_u})\big\} \tag{139}$$

for $u = t+1, \dots, T$, where for a generic event $A \in \mathcal{H}_T$, 1_A is the indicator function of the event A, see Glasserman et al. (2000). Also, for the second order moment, it holds

$$m_2(x_u, \theta_{s_u}) := \tilde{\mathbb{E}}\big[1_{\{X_u > x_u\}} L_u^2 \big|\pi_\beta, \Sigma_\beta, s, \mathcal{F}_{u-1}\big] \leq \exp\big\{-2\theta_{s_u} x_u + 2\psi(\theta_{s_u})\big\}. \tag{140}$$

To reduce a variance of the importance sampling, we need to keep the right hand side of the above inequality as low as possible. To minimize the right–hand side of the above equation, we minimize the exponent using the parameter θ_{s_t}. The minimizer of the right–hand side of the above equation is given by

$$\theta_{s_u}(x_u) := \begin{cases} \dfrac{x_u - z_u' C_{u,s_u} \mathsf{Y}_u}{\big(1 + \mathsf{Y}_u' \Lambda_{0,s_u|u}^{-1} \mathsf{Y}_u\big) z_u' \Sigma_{s_u} z_u} & \text{if} \quad s_u \in \gamma, \\ \dfrac{x_u - z_u' \Pi_{s_u} \mathsf{Y}_u}{\big(1 + \mathsf{Y}_u' \Lambda_{0,s_u} \mathsf{Y}_u\big) z_u' \Sigma_{s_u} z_u} & \text{if} \quad s_u \in \delta \end{cases} \tag{141}$$

for $u = t+1, \dots, T$. Therefore, an importance sampling method that estimates the conditional probabilities $\mathbb{P}(X_u > x_u|\mathcal{F}_t)$ for $u = t+1, \dots, T$ as follows: for $\ell = 1, \dots, \mathcal{L}$,

- generate $\big(\bar{y}_t^c(\ell), \pi_{\bar{s}_t^c(\ell)}(\ell), \Sigma_{\bar{s}_t^c(\ell)}(\ell), \bar{s}_t^c(\ell)\big)$ using the general simulation method,
- for $u = t+1, \dots, T$,
  - calculate $\theta_{s_u(\ell)}^*(x_u, \ell)$ using equation (141),
  - if $s_u(\ell) \in \gamma(\ell)$, generate $\pi_{s_u(\ell)}^*(\ell)$ from
$$\mathcal{N}\Big(A_{s_u(\ell)}(\ell) b_{s_u(\ell)}(\ell) + \theta_{s_u(\ell)}^*(x_u, \ell)(\mathsf{Y}_u(\ell) \otimes I_n) z_u, A_{s_u}(\ell)\Big), \tag{142}$$
  - if $s_u(\ell) \in \delta(\ell)$, generate $\pi_{s_u(\ell)}^*(\ell)$ from
$$\mathcal{N}\Big(\pi_{0,s_u(\ell)} + \theta_{s_u(\ell)}^*(x_u, \ell)(\Lambda_{0,s_u(\ell)} \mathsf{Y}_u(\ell) \otimes \Sigma_{s_u(\ell)}(\ell)) z_u, \Lambda_{0,s_u(\ell)} \otimes \Sigma_{s_u(\ell)}(\ell)\Big), \tag{143}$$
  - generate $y_u^*(\ell)$ from
$$\mathcal{N}\Big(\Pi_{s_u}^* \mathsf{Y}_u(\ell) + \theta_{s_u(\ell)}^*(x_u, \ell) \Sigma_{s_u(\ell)} z_u, \Sigma_{s_u(\ell)}(\ell)\Big), \tag{144}$$
    where $\pi_{s_u(\ell)}^*(\ell) = \text{vec}(\Pi_{s_u(\ell)}^*(\ell))$,
  - calculate $L_u^*(\ell)$ using equation (137), where $X_u^*(\ell) = z_u' y_u^*(\ell)$,
- and for $u = t+1, \dots, T$, estimate the probabilities $\mathbb{P}(X_u > x_u|\mathcal{F}_t)$ by
$$\hat{\mathbb{P}}(X_u > x_u|\mathcal{F}_t) = \frac{1}{\mathcal{L}} \sum_{\ell=1}^{\mathcal{L}} 1_{\{X_u^*(\ell) > x_u\}} L_u^*(\ell), \tag{145}$$

where calculations of the matrix $A_{s_u(\ell)}(\ell)$ and vectors $b_{s_u(\ell)}(\ell)$ and $\mathsf{Y}_u(\ell)$ are based on the vectors $\bar{y}_t$ and $\bar{y}_t^c(\ell)$.

4 Numerical Results

Now we consider numerical results. For means of illustration, we have chosen three companies, listed in the S&P 500 index from different sectors. In order to increase the number of price and dividend observation points, we take quarterly data instead of yearly data. Our data covers a period from Q1, 1985 to Q4, 2023. That leads to $t = 156$ observations for Johnson & Johnson (J&J), PepsiCo, and Walmart. All quarterly price and dividend data have been collected from Yahoo Finance.

4.1 Dividend Discount Model

Before we move to the numerical results, we provide a brief review of the dividend discount model (DDM), which is useful for the numerical results. Let us assume that there are m companies, and the companies will not default in the future. For DDM with default risk, we refer to Battulga et al. (2022). As mentioned before, the basic idea of all DDMs is that the market price of a stock equals the sum of the stock's next period price and dividend discounted at the required rate of return. Therefore, for successive prices of i–th company, the following relation holds

$$P_{i,t} = (1 + k_{i,t})P_{i,t-1} - d_{i,t}, \quad i = 1, \dots, m, \ t = 1, 2, \dots, \tag{146}$$

where $k_{i,t}$ is the required rate of return on stock, $P_{i,t}$ is the stock price, and $d_{i,t}$ is the dividend, respectively, at time t of i–th company. In vector form, the above equation is written by

$$P_t = (i_m + k_t) \odot P_{t-1} - d_t, \quad t = 1, 2, \dots, \tag{147}$$

where $k_{s_t} := (k_{1,t}, \dots, k_{m,t})'$ is an $(m \times 1)$ vector of the required rate of returns on stocks at time t, $P_t := (P_{1,t}, \dots, P_{m,t})'$ is an $(m \times 1)$ price vector at time t, and $d_t := (d_{1,t}, \dots, d_{m,t})'$ is an $(m \times 1)$ dividend vector at time t of the companies. On the other hand, we assume that a dividend of i–th company is proportional to stock price of the company. To price dividend paying option, Merton (1974) used the assumption. Consequently, dividends of the companies are modeled by

$$d_t = \alpha_t \odot P_{t-1}, \quad t = 1, 2, \dots, \tag{148}$$

where $\alpha_t = (\alpha_{1,t}, \dots, \alpha_{m,t})'$ is an $(m \times 1)$ vector of dividend–to–price ratios at time t of the companies. Consequently, by equation (147) and (148), the price vector at time t is written by

$$P_t = (i_m + k_t - \alpha_t)P_{t-1}. \tag{149}$$

To the vector of dividend–to–price ratios take positive values, instead of equation (148), we work a vector of log dividend–to–price ratios, which is given by the following equation

$$\tilde{d}_t = \tilde{\alpha}_t - \tilde{P}_{t-1}, \tag{150}$$

where $\tilde{d}_t := \ln(d_t)$ is an $(m \times 1)$ log dividend vector at time t, $\tilde{P}_{t-1} := \ln(P_{t-1})$ is an $(m \times 1)$ log price vector at time $t-1$, $\tilde{\alpha}_t$ is the $(m \times 1)$ vector of log dividend–to–price ratios at time t. As a result, equation (149) becomes

$$P_t = \big(i_m + k_t - \exp\{\tilde{\alpha}_t\}\big)P_{t-1}. \tag{151}$$

It follows from equations (147) and (150) that for $t = 1, 2, \dots$, vectors of the required rate of returns and log dividend–to–price ratios of the companies are obtained by

$$k_t = (P_t + d_t) \oslash P_{t-1} - i_m \tag{152}$$

and

$$\tilde{\alpha}_t = \ln(d_t \oslash P_{t-1}), \tag{153}$$

respectively. By repeating equation (151), for $t = 0, 1, \ldots$ and $r = 0, 1, \ldots$, we obtain the price vectors at time $t + r$ of the companies

$$P_{t+r} = \prod_{j=1}^{r} \big(i_m + k_{t+j} - \exp\{\tilde{\alpha}_{t+j}\}\big) \odot P_t, \tag{154}$$

where for $q = 1, 2, \ldots$ and generic $(m \times 1)$ vectors $o_1, \ldots, o_q$, $\prod_{j=1}^{q} o_j = o_1 \odot \cdots \odot o_q$ is an element–wise product of the vectors $o_1, \ldots, o_q$ and with convention $\prod_{j=q}^{q-1} o_j = i_m$.

4.2 Maximum Likelihood Estimation

It is not difficult to show that the required rate of returns of the selected companies are AR(0) processes, see Figure 1. Thus, for each firm, we model the required rate of returns by the following AR(0) process

$$k_{i,t} = c_{i,s_t} + \xi_{i,t}, \quad i = 1, 2, 3, \tag{155}$$

where the regime–switching process s_t takes one of the values $\{1, 2, 3\}$ and c_{i,s_t} is a constant, corresponding to i–th company and the regime s_t. We suppose that a variance of the residual process $\xi_{i,t}$ in the regime s_t of the i–th company equals σ^2_{i,s_t}.

We present maximum likelihood estimations of the parameters for the selected companies in Table 1. Maximum likelihood estimation method can be found in Hamilton (1990, 1994), see also Battulga (2022, 2023a). The 2–9th rows of Table 1 correspond to that the required rate of returns of the companies are modeled by the regime–switching process with three regimes and the 10–13th rows of the same Table correspond to that the required rate of returns of the companies take constant values (the regime–switching process takes one regime).

Table 1: Maximum Likelihood Estimations of Parameters of Selected Companies

Row	Parameters	Johnson & Johnson			Pepsi			Walmart		
2	$c_{i,j}$	15.64%	4.26%	–6.41%	6.29%	2.73%	–13.90%	15.09%	1.74%	–5.60%
3	$\sigma_{i,j}$	0.047	0.052	0.054	0.098	0.055	0.022	0.096	0.066	0.083
4	$\hat{\mathsf{P}}_i$	0.222	0.384	0.395	0.908	0.092	0.000	0.460	0.000	0.540
5		0.099	0.645	0.256	0.000	0.933	0.067	0.000	0.932	0.068
6		0.232	0.768	0.000	0.890	0.000	0.110	0.683	0.164	0.153
7	$\tau_{i,j}$	1.285	2.814	1.000	10.842	14.865	1.124	1.851	14.789	1.181
8	$\pi_{i,j}$	0.146	0.634	0.220	0.404	0.554	0.042	0.269	0.517	0.213
9	$k_{i,\infty}$	3.58%			3.47%			3.77%		
10	c_i	3.58%			3.54%			4.02%		
11	$k_{i,L}$	2.24%			2.18%			2.25%		
12	$k_{i,U}$	4.92%			4.90%			5.80%		
13	σ	0.084			0.086			0.112		

In order to obtain estimations of the parameters, which correspond to the 2–9th rows of Table 1 we assume that the regime–switching process s_t follows a Markov chain with three regimes, namely, up regime (regime 1), normal regime (regime 2), and down regime (regime 3). For the normal regime, most of the time of the regime–switching process s_t lasts in regime 2. Since explanations are comparable for the other companies, we give explanations only for J&J. In the 2nd row of Table 1, for each company $i = 1, 2, 3$, we provide estimations of the parameters $c_{i,1}, c_{i,2}, c_{i,3}$. For J&J, in regimes 1, 2, and 3, estimations of the required rate of return are 15.64%, 4.26%, and –6.41%, respectively. For example, in the normal regime, the required rate of return of J&J could be 4.26% on average.

In the 3rd row of Table 1, we present parameter estimations of standard deviations of the error random variables $\xi_{i,t}$ for the selected companies. For J&J, the parameter estimations of the standard deviations equal 0.047 (regime 1), 0.052 (regime 2), and 0.054 (regime 3). The 13th row of Table 1 corresponds to the parameter estimations of standard deviations, in which the required rate of returns of the companies are modeled by regime–switching process with one regime. For J&J, the parameter estimation equals 0.084. As we compare the 9th row and 13th row of the Table, we can see that the estimations that correspond to the regime–switching process with three regimes are significantly lower than the ones that correspond to the regime–switching process with one regime, except the up regime of PepsiCo.

The 4–6th rows of Table 1 correspond to the transition probability matrix $\hat{\mathsf{P}}$. For the selected companies, their transition probability matrices $\hat{\mathsf{P}}$s are ergodic, where ergodic means that one of the eigenvalues of $\hat{\mathsf{P}}$ is unity and that all other eigenvalues of $\hat{\mathsf{P}}$ are inside the unit circle, see Hamilton (1994). From the 4th row of Table 1, one can deduce that if the required rate of return of J&J is in the up regime, then in the next period, it will switch to the up regime with a probability of 0.22, the normal regime with a probability of 0.384, or the down regime with a probability of 0.395. If the required rate of return of J&J in the normal regime, corresponding to row 5 of the Table, then in the next period, it will switch to the up regime with a probability of 0.099, the normal regime with a probability of 0.645, or the down regime with a probability of 0.256. Finally, if the required rate of return of J&J is in the down regime, then in the next period, it will switch to the up regime with a probability of 0.232 or the normal regime with a probability of 0.768 because of the down regime's zero probability, see 6th row of the same Table.

We provide the average persistence times of the regimes in the 7th row of Table 1. The average persistence time of the regime s_t is defined by $\tau_{s_t} := 1/(1 - p_{s_t s_t})$ for $s_t \in \{1, 2, 3\}$. From Table 1, one can conclude that up, normal, and down regimes of J&J's required rate of return will persist on average for 1.285, 2.814, and 1.000 quarters, respectively.

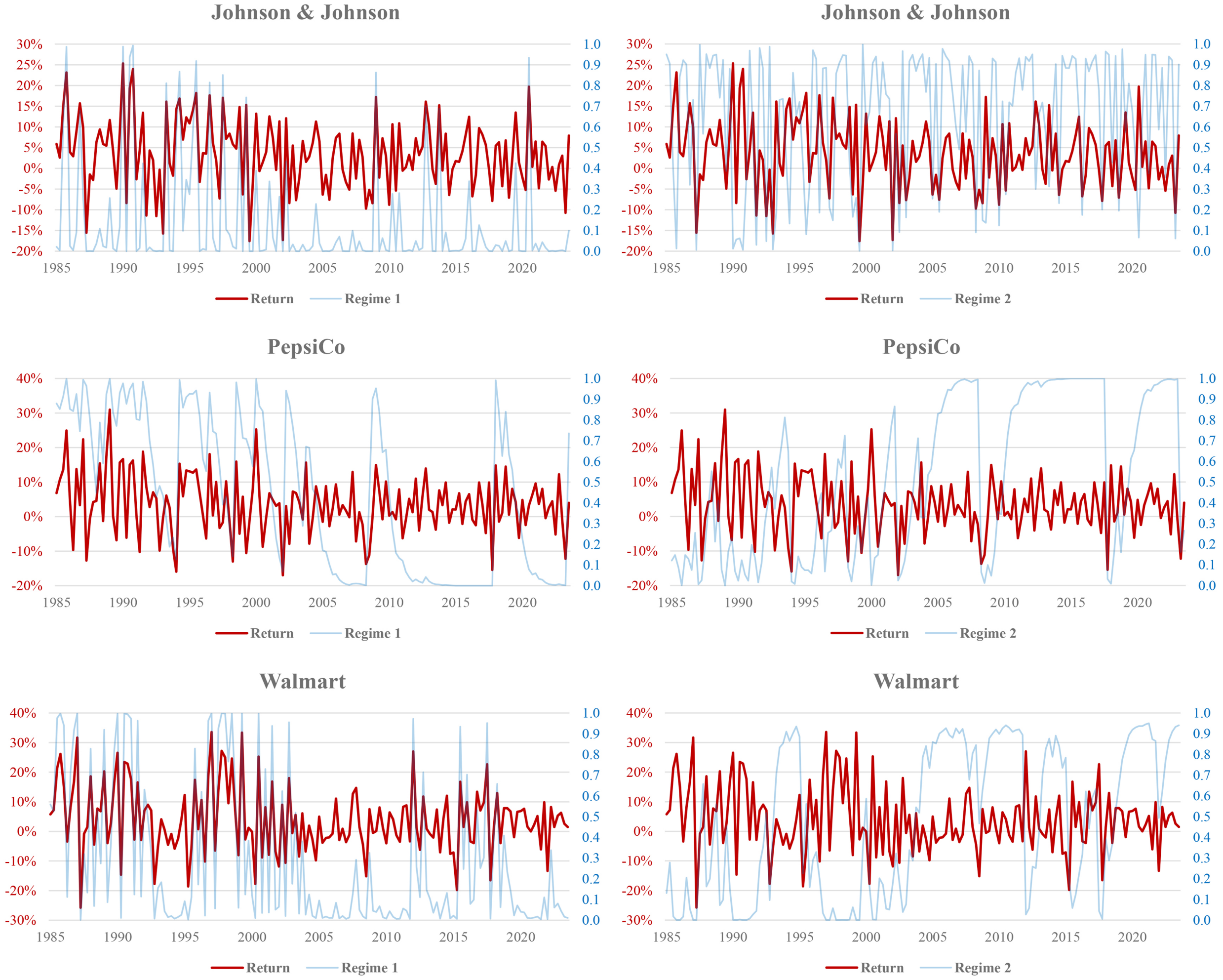
In the 8th row of Table 1, we give ergodic probabilities π of the selected companies. Ergodic probability vector π of an ergodic Markov chain is obtained from an equation $\hat{\mathsf{P}}\pi = \pi$. The ergodic probability vector represents long–run probabilities, which do not depend on the initial probability vector $z_{1|0}$. After sufficiently long periods, the required rate of return of J&J will be in the up regime with a probability of 0.146, the normal regime with a probability of 0.634, or the down regime with a probability of 0.220, which are irrelevant to initial regimes.

The 9th row of Table 1 is devoted to long–run expectations of the required rate of returns of the selected companies. The long–run expectation of the required rate of return is defined by $k_\infty := \lim_{t\to\infty} \mathbb{E}(k(s_t))$. For J&J, it equals 3.58%. So that, after long periods, the average required rate of return of J&J converges to 3.58%.

Finally, for the one regime case, parameter estimations of the required rate of returns at time Q4, 2023 of the firms are presented in row 10 of the Table, while the corresponding 95% confidence intervals are included in rows 11 and 12 below. Here we use formulas in Battulga (2023a) to calculate the parameter estimation and confidence bands. Note that since the required rate of return estimation expresses the average quarterly return of the companies, we can convert them yearly using a formula $(1 + k)^4 - 1$. The Table further illustrates average returns (3.58% for J&J) and return variability, as the return is supposed to lie within the (2.24%, 4.92%) interval with a 95% probability.

For the selected firms, plotting the smoothed probabilistic inferences with a return series will be interesting. For each period $u = 1, \dots, t$, and each firm, the smoothed inferences are calculated by equation (123), and the return series are calculated by equation (152). In Figure 1, we plotted the resulting series as a function of period u. In Figure 1, the left axis corresponds to the return series, while the right axis corresponds to the smoothed inference series for each company. From the Figure and the 9th and 13th rows of Table 1, we can deduce that the regime–switching processes with three regimes are more suited to explain the required rate of return series as compared to the regime–switching processes with one regime.

Figure 1: Returns VS Regime Probabilities of Selected Companies



4.3 Bayesian Estimation

Now, we aim to obtain future theoretical prices of the selected companies using equation (154). To get the future prices of the firms, we assume that the vectors of the required rate of returns and the log dividend–to–price ratios of the firms are placed on the Bayesian MS–VAR process y_t. We also assume that the process y_t is modeled by the Bayesian MS–VAR process of order 4, which is given by the following equation

$$y_t = a_{0,s_t} + A_{1,s_t} y_{t-1} + \dots + A_{4,s_t} y_{t-4} + \xi_t, \ \ t = 1, \dots, T, \tag{156}$$

where $y_t = (k_t', \tilde{\alpha}_t')'$ is a (6×1) vector of endogenous variables with $k_t := (k_{1,t}, k_{2,t}, k_{3,t})'$ and $\tilde{\alpha}_t := (\tilde{\alpha}_{1,t}, \tilde{\alpha}_{2,t}, \tilde{\alpha}_{3,t})'$, $\xi_t = (\xi_{1,t}, \dots, \xi_{6,t})'$ is a (6×1) residual process, a_{0,s_t} is a (6×1) random coefficient vector at regime s_t, and for $i = 1, \dots, 4$, A_{i,s_t} are (6×6) random coefficient matrices at regime s_t.

For the vector of log dividend–to–price ratios $\tilde{\alpha}_t$ of the selected companies, it can be shown that the ratios follow the unit root process. On the other hand, the vector of the required rate of returns k_t of the selected companies is a stationary process, see Figure 1. Consequently, conditional expectations of diagonal elements of the random matrix A_{1,s_t} are chosen by $\phi := (0, 0, 0, 1, 1, 1)'$.

We suppose that the scale matrix V_{0,s_t} is a diagonal matrix and its diagonal elements, corresponding to the process of log dividend–to–price ratios are estimated by sample variances of the processes, modeled by univariate AR(4). For other diagonal elements, corresponding to the required rate of returns, we use the maximum likelihood estimations of parameters, which are given in Table 1. Therefore, for each regime $s_t = 1, 2, 3$, we choose the scale matrix V_{0,s_t} by

$$\begin{aligned} V_{0,1} &:= \operatorname{diag}\big\{0.047^2, 0.098^2, 0.096^2, 0.092^2, 0.094^2, 0.133^2\big\} \\ V_{0,2} &:= \operatorname{diag}\big\{0.052^2, 0.055^2, 0.066^2, 0.092^2, 0.094^2, 0.133^2\big\} \\ V_{0,3} &:= \operatorname{diag}\big\{0.054^2, 0.022^2, 0.083^2, 0.092^2, 0.094^2, 0.133^2\big\}. \end{aligned} \tag{157}$$

Also, based on the same Table, for each regime, we choose conditional expectation of the random vector a_{0,s_t} by $c_1 := \mathbb{E}\big[a_{0,s_t}\big|\Sigma_{s_t}, s_t = 1, \mathcal{F}_0\big] = (0.15, 0.06, 0.15, 0, 0, 0)'$, $c_2 := \mathbb{E}\big[a_{0,s_t}\big|\Sigma_{s_t}, s_t = 2, \mathcal{F}_0\big] = (0.04, 0.03, 0.02, 0, 0, 0)'$, and $c_3 := \mathbb{E}\big[a_{0,s_t}\big|\Sigma_{s_t}, s_t = 3, \mathcal{F}_0\big] = (-0.07, -0.14, -0.05, 0, 0, 0)'$. For the other hyperparameters, we choose $\varepsilon_j = \lambda_{1,j} = 20$ and $\lambda_{2,i} = \tau_{i,j} = 1$ for each regime $j = 1, 2, 3$ and each endogenous variable $i = 1, \dots, 6$. As a result, the matrices of the dummy variables $\hat{y}^\circ_{s_t}$ and $\hat{\mathsf{Y}}^\circ_{s_t}$, which are given in equations (113) and (114) are represented by

$$\hat{y}^\circ_{s_t} := 20\big[c_{s_t} : \operatorname{diag}\{0, 0, 0, 1, 1, 1\} : 0_{[6\times 20]}\big] \tag{158}$$

and

$$\hat{\mathsf{Y}}^\circ_{s_t} := 20 \begin{bmatrix} 1 & 0_{[1\times 24]} \\ 0_{[24\times 1]} & \big(\operatorname{diag}\{1, 2, 3, 4\} \otimes \operatorname{diag}\{1, 1, 1, 1, 1, 1\}\big) \end{bmatrix}, \tag{159}$$

where dimensions of the matrices of the dummy variables $\hat{y}^\circ_{s_t}$ and $\hat{\mathsf{Y}}^\circ_{s_t}$ are (6×25) and (25×25), respectively.

In order to obtain future eight quarters' theoretical prices and their 95% confidence bands of the selected companies, we use equation (154) and make a simulation of $\mathcal{L} = 10,000$. The following Table presents the results of the theoretical prices and confidence bands.

Table 2: Future Theoretical Prices and 95% Confidence Bands of Selected Companies

Companies		P_t	P_{t+1}	P_{t+2}	P_{t+3}	P_{t+4}	P_{t+5}	P_{t+6}	P_{t+7}	P_{t+8}
J&J	T/Price	158.90	163.44	167.00	170.74	174.29	178.03	181.79	185.65	189.84
	L/Bound	158.90	143.04	137.33	133.97	131.30	129.46	127.21	125.87	124.01
	U/Bound	158.90	184.10	199.14	213.59	225.62	238.55	250.06	261.65	273.48
PepsiCo	T/Price	168.53	171.08	173.25	175.43	177.33	179.30	181.38	183.39	185.45
	L/Bound	168.53	150.54	144.19	140.09	136.47	133.36	130.78	128.37	125.59
	U/Bound	168.53	191.43	203.46	213.40	222.56	231.38	239.09	248.31	255.65
Walmart	T/Price	55.08	57.00	58.75	60.70	62.48	64.47	66.40	68.40	67.38
	L/Bound	55.08	47.94	45.42	44.31	43.28	42.58	41.76	41.05	40.62
	U/Bound	55.08	66.19	72.91	79.20	84.91	91.24	97.23	102.37	109.35

For Table 2, T/Price, L/Bound, and U/Bound mean theoretical price, lower bound, and upper bound, respectively. The column P_t represents real prices at time Q4, 2023 of the selected companies. For each $u = 1, \dots, 8$, the column P_{t+u} expresses future prices at time $t + u$ (after u quarters from Q4, 2023) of the companies. In particular, the column P_{t+4} represents future prices in Q4, 2024. The same column further explains that the theoretical price in Q4, 2024 of J&J equals 174.29, and the price in Q4, 2024 of the firm will lie within the (131.30,225.62) interval with a 95% probability. The same explanations hold for other columns and other companies. From the Table, one can deduce that histograms of the future prices are right skewed, and the skewness increases as time increases.

5 Conclusion

In this paper, for the general Bayesian MS–VAR process, we obtain some useful density functions for Monte–Carlo simulations. The density functions tell us that conditional on the regime vectors and initial information, the vector of endogenous variables is independent of model's some random components. Thus, one only needs the prior distributions to calculate the density functions, and the results have yet to be explored before.

In a particular case of the general Bayesian MS–VAR process, we also get closed–form density functions of the random components of the model. In particular, we find that joint distributions of future values of the random coefficient matrix are matrix variate student distribution, see equation (48). Hence, one can analyze impulse response by directly generating the coefficient matrix from the density. Also, we provide a new density function, which has yet to be introduced before of future values of the endogenous variables; see equations (44) and (77). Thus, future studies may concentrate on marginal density functions and direct simulation methods for the density function. Further, we obtain a characteristic function of the random coefficient matrix, which can be used to calculate the forecast of the endogenous variables.

In the paper, we developed Monte–Carlo simulation algorithms. The simulation method's novelty is that it removes the regime vector duplication. As a result, our proposed Monte–Carlo simulation method departs from the previous simulation methods with regime switching. We also provide importance sampling method to estimate probability of a rare event, corresponding to the future endogenous variables. Since the method can be used to calculate quantiles, in this case, the quantiles of the future endogenous variables become more reliable than navy simulation methods.

References

Albert, J. H., & Chib, S. (1993). Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *Journal of Business & Economic Statistics*, *11*(1), 1–15. doi: 10.2307/1391303

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian Vector Autoregressions. *Journal of Applied Econometrics*, *25*(1), 71–92.

Battulga, G. (2022). Stochastic ddm with regime–switching process. *Numerical Algebra, Control and Optimization*, 1–27. doi: 10.3934/naco.2022031

Battulga, G. (2023a). Parameter Estimation Methods of Required Rate of Return on Stock. *International Journal of Theoretical and Applied Finance*, *26*(8), 2450005.

Battulga, G. (2023b). Rainbow Options with Bayesian MS-VAR Process. *Mongolian Mathematical Journal*, *26*(24), 1-16.

Battulga, G. (2024a). Equity–Linked Life Insurances on Maximum of Several Assets. *to appear in Numerical Algebra, Control & Optimization*. Available at: `https://arxiv.org/abs/2112.10447`.

Battulga, G. (2024b). Options Pricing under Bayesian MS–VAR Process. *to appear in Numerical Algebra, Control & Optimization*. Available at: `https://arxiv.org/abs/2109.05998`.

Battulga, G., Jacob, K., Altangerel, L., & Horsch, A. (2022). Dividends and Compound Poisson–Process: A new Stochastic Stock Price Model. *International Journal of Theoretical and Applied Finance*, *25*(3), 2250014.

Butler, R. W. (1998). Generalized inverse gaussian distributions and their wishart connections. *Scandinavian journal of statistics*, *25*(1), 69–75.

D'Amico, G., & De Blasis, R. (2020). A Review of the Dividend Discount Model: from Deterministic to Stochastic Models. *Statistical Topics and Stochastic Models for Dependent Data with Applications*, 47–67.

Glasserman, P., Heidelberger, P., & Shahabuddin, P. (2000). Variance reduction techniques for estimating value-at-risk. *Management Science*, *46*(10), 1349–1364.

Glasserman, P., & Li, J. (2005). Importance sampling for portfolio credit risk. *Management science*, *51*(11), 1643–1656.

Goldfeld, S. M., & Quandt, R. E. (1973). A markov model for switching regressions. *Journal of Econometrics*, *1*(1), 3–15.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hamilton, J. D. (1990). Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, *45*(1-2), 39–70.

Hamilton, J. D. (1994). *Time Series Econometrics*. Princeton University Press, Princeton.

Karlsson, S. (2013). Forecasting with bayesian vector autoregression. *Handbook of economic forecasting*, *2*, 791–897.

Kim, C.-J. (1994). Dynamic Linear Models with Markov–Switching. *Journal of Econometrics*, *60*(1–2), 1–22.

Krolzig, H.-M. (1997). *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis* (Vol. 454). Springer Science & Business Media.

Litterman, R. (1979). Techniques of forecasting using vector autoregressions. *Federal Reserve of Minneapolis Working Paper 115*.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis* (2nd ed.). Springer Berlin Heidelberg.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, *29*(2), 449–470.

Miranda-Agrippino, S., & Ricco, G. (2018). Bayesian vector autoregressions. *Bank of England working paper*.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the american statistical association*, *53*(284), 873–880.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48.

Tiao, G. C., & Box, G. E. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, *76*(376), 802–816.

Tong, H. (1983). *Threshold models in non-linear time series analysis* (Vol. 21). Springer Science & Business Media.

Williams, J. B. (1938). *The Theory of Investment Value*. Harvard University Press.

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R* (2nd ed.). CRC press.