Bayesian Model Comparison for Large Bayesian VARs after the COVID-19 Pandemic

Joshua C.C. Chan
Purdue University

Xuewen Yu
Fudan University

Wei Zhang
Purdue University

July 2024

Abstract

There is an increasing interest in applying variational Bayes techniques to estimating large Bayesian vector autoregressions (VARs) with stochastic volatility. However, less attention has been paid to the development of appropriate tools for comparing these high-dimensional models, especially among those designed to address COVID-19 outliers. This paper develops a marginal likelihood estimator that combines importance sampling and variational approximation for comparing large VARs with different time-varying volatility specifications and outlier adjustments. Through a Monte Carlo study, we show that the proposed approach is fast and able to identify the correct models. The effectiveness of the proposed method is further illustrated through an empirical application of comparing a variety of 180-variable VARs.

Keywords: Variational inference, large vector autoregression, marginal likelihood, Bayesian model comparison, stochastic volatility, outlier adjustment

1 Introduction

Since the seminal work of Bańbura et al. (2010), large Bayesian vector autoregressions (VARs) have become standard tools in empirical macroeconomics for forecasting and structural analysis. Prominent examples include Carriero et al. (2009), Koop (2013), Koop and Korobilis (2013), Bańbura et al. (2015), Korobilis and Pettenuzzo (2019) and Huber and Feldkircher (2019). More recently, there is a surge in interest in developing various stochastic volatility specifications for large Bayesian VARs (see, e.g., Carriero et al., 2016, 2019; Chan, 2020a; Tsionas et al., 2022), due to the increasing recognition of the importance of time-varying volatility in modeling macroeconomic and financial variables. Naturally, the unprecedented economic turbulence triggered by the COVID-19 pandemic hastens this upward trend.

However, an important bottleneck that impedes the routine application of large Bayesian VARs, particularly when flexible features such as stochastic volatility or outlier adjustments are included, is the computational burden of conventional Markov chain Monte Carlo (MCMC) methods. This motivates the use of variational Bayes approaches to approximate the posterior distributions in large VARs; recent papers include Koop and Korobilis (2018), Gefang et al. (2020, 2023), Chan and Yu (2022), and Bernardi et al. (2024). We contribute to this line of research by considering a related but unsolved problem of comparing these large Bayesian VARs with stochastic volatility and outlier adjustments. We tackle a key challenge for practitioners: multiple nonlinear, high-dimensional VARs are available for a particular dataset, but there are no adequate tools to compare or select among them.

We consider a variational importance sampling (VIS) method to estimate the marginal likelihoods of large VARs, by combining the variational Bayes and importance sampling techniques. More specifically, we first obtain the optimal density from the variational Bayes by minimizing the Kullback-Leibler divergence to the posterior distribution. This optimal density is then used as the importance sampling density to generate independent samples for the associated marginal likelihood estimator.[1] This distinguishes our

[1] There is a long tradition of using importance sampling methods to estimate the marginal likelihood or the posterior distribution. For example, Perrakis, Ntzoufras, and Tsionas (2014) propose using the product of marginal posteriors as an importance sampling density to estimate the marginal likelihood; Chan and Eisenstat (2015, 2018) use the cross-entropy method to obtain the optimal importance sampling density within a given parametric family of distributions. For approximating the posterior distribution,

approach from various marginal likelihood estimators that rely on MCMC methods (e.g., Gelfand and Dey, 1994; Newton and Raftery, 1994; Chib, 1995; Frühwirth-Schnatter and Wagner, 2008; Perrakis et al., 2014; Chan and Eisenstat, 2015, 2018). The proposed approach has the advantage of avoiding the use of MCMC draws, which are very costly to obtain in high-dimensional settings.

Our second contribution involves comparing alternative specifications of outlier adjustments in the context of large VARs with stochastic volatility. This is motivated by the new challenge for empirical macroeconomists caused by the extreme movements in many macroeconomic variables at the onset of the COVID-19 pandemic. For instance, in a dataset comprising 104 macroeconomic time-series constructed from the FRED-MD database, 32 variables reached unprecedented levels/rates in April 2020; four variables exceeded over ten times their previous record values. Such extreme variability can significantly impact parameter estimates and forecasts from standard VARs, as demonstrated by Schorfheide and Song (2021) and Bobeica and Hartwig (2023). Consequently, several recent papers, such as Lenza and Primiceri (2022) and Carriero et al. (2022b), have proposed different ways to address these COVID-19 outliers in the setting of Bayesian VARs. We demonstrate the usefulness of the proposed VIS method to evaluate these recently proposed outlier adjustments.

Our paper is closely related to two recent works. The first is Hajargasht and Woźniak (2020), who use the optimal density obtained from the variational Bayes method as a weighting density in the modified harmonic mean estimator of Geweke (1999). While they illustrate their method using a homoskedastic VAR of seven variables, we focus on large VARs with stochastic volatility. The second paper is Chan (2023), who proposes marginal likelihood estimators for large VARs with stochastic volatility. But since those estimators are constructed using MCMC draws, the computational burden becomes excessive when the dimension of the VAR is very large (e.g., over 50 variables). We circumvent this computational issue by using the variational Bayes approach instead of MCMC methods.

Using datasets of different sizes constructed from the FRED-QD database, we show that parameter estimates from the variational Bayes approach are as accurate as those produced by MCMC. In addition, through a series of Monte Carlo experiments, we demon-

Dellaportas and Tsionas (2019) consider using a product of univariate Student-t densities and a copula function, where the parameters are obtained using importance sampling.

strate that the variational Bayes approach can dramatically reduce the computational time. For instance, for a dataset consisting of 100 variables and 500 observations, MCMC takes around 20 hours, while the variational Bayes method takes only 3 minutes. More importantly, the Monte Carlo results show that both the VIS estimator and the variational lower bound can be used to correctly select the true models.

We illustrate the methodology via a Bayesian model comparison exercise using two datasets. The first dataset is the same as that in Carriero et al. (2022b), which consists of 16 monthly variables and covers the period from March 1959 to March 2021. The second dataset is constructed from the FRED-QD database that includes 180 variables and spans from September 1959 to December 2023. We find that VARs with stochastic volatility are decidedly favored over the standard homoskedastic VAR for both datasets. This result is consistent with the growing body of evidence that underscores the significance of stochastic volatility in modeling both medium and large macroeconomic datasets. Furthermore, the time-varying volatility model of Lenza and Primiceri (2022) is outperformed by other stochastic volatility VARs. Among the latter models, the medium dataset shows a slight preference for the outlier specification proposed in Carriero et al. (2022b), whereas the large dataset with 180 time-series prefers a standard VAR with stochastic volatility.

The rest of the paper is organized as follows. Section 2 describes a variety of VARs with different time-varying volatility specifications and outlier components. Section 3 outlines the basic theory on variational Bayes, particularly the mean-field approximation. Section 4 develops the marginal likelihood estimator that combines the variational Bayes method and importance sampling. We then illustrate the proposed approach with a simple linear regression and compare the estimates with alternative methods. In Section 5, we conduct a series of Monte Carlo experiments to evaluate the accuracy of the variational Bayes estimates and to assess whether the proposed marginal likelihood estimator can correctly identify the true models. We also investigate the appropriateness of using the variational lower bound as a criterion for model selection. Section 6 presents the empirical application in which we compare various VARs with different types of time-varying volatility and outlier adjustments. Lastly, Section 7 concludes.

2 Large VARs with Stochastic Volatility and Outlier Adjustments

We begin this section by presenting the baseline model—a reduced-form VAR with the Cholesky stochastic volatility developed by Cogley and Sargent (2005) that is especially suitable for modeling large datasets (Carriero et al., 2019). Next, we describe a few recently proposed specifications that can be added to this baseline model to account for COVID-19 outliers. Lastly, we outline a data-driven Minnesota prior that is particularly useful for high-dimensional VARs.

2.1 A reduced-form VAR with stochastic volatility

Let $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ be an $n \times 1$ vector of variables that is observed over the periods $t = 1, \ldots, T$. Consider the following reduced-form VAR with p lags:

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{1}$$

where $\mathbf{a}_0$ denotes an $n \times 1$ vector of intercepts, and $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $n \times n$ coefficient matrices. Following Cogley and Sargent (2005), the covariance matrix of the innovations is modeled using n stochastic volatility processes in order to account for the potential heteroskedasticity and time-varying covariances. In particular,

$$\boldsymbol{\Sigma}_t^{-1} = \mathbf{B}_0'\mathbf{D}_t^{-1}\mathbf{B}_0, \tag{2}$$

where $\mathbf{D}_t = \text{diag}(\mathrm{e}^{h_{1,t}}, \ldots, \mathrm{e}^{h_{n,t}})$, and $\mathbf{B}_0$ is an $n \times n$ lower triangular matrix with ones on the diagonal. Each element of $\mathbf{h}_t = (h_{1,t}, \ldots, h_{n,t})'$ follows a random walk process

$$h_{i,t} = h_{i,t-1} + u_{i,t}^h, \quad u_{i,t}^h \sim \mathcal{N}(0, \sigma_i^2)$$

for $t = 1, 2, \ldots, T$, and the initial condition $h_{i,0}$ is treated as an unknown parameter to estimate. We refer to this baseline stochastic volatility model as VAR-SV.

The VAR-SV model contains a different stochastic volatility process for each of its n variables, enhancing its flexibility. However, this feature demands extensive posterior computations, particularly when employing conventional MCMC algorithms. To mitigate these computational demands, we adopt the equation-by-equation approach—based on a triangularization of the VAR—developed by Carriero et al. (2019, 2022a). But instead of using MCMC methods, we employ the variational Bayes approach designed to approximate the posterior distribution efficiently.

2.2 Stochastic volatility with outlier adjustments

Next, we discuss three modeling strategies that have been used in the literature to account for COVID-19 outliers. The first strategy explicitly specifies an outlier component by using a discrete mixture of distributions. The second strategy characterizes the infrequent occurrences of outliers using the t distribution that has more mass at the tails than the Gaussian. The last strategy takes advantage of the known timing of the COVID-19 pandemic, and treats it as a deterministic break in the covariance matrix. It also allows for the potentially elevated volatility after the outbreak of the pandemic.

Specification 1: An explicit outlier component

The first specification introduces outlier indicators that have a discrete mixture representation that is proposed by Stock and Watson (2016) and is later adapted to VAR settings in Carriero et al. (2022b). More specifically, the outlier indicators enter the model in a diagonal matrix of scale factors, denoted $\mathbf{O}_t$, with diagonal elements $o_{i,t}$ that are mutually independent over all i and t. With $\mathbf{B}_0$ and $\mathbf{D}_t$ specified as before, the covariance matrix now takes the form:

$$\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{O}_t\mathbf{D}_t\mathbf{O}_t'(\mathbf{B}_0^{-1})'.$$

The outlier indicator $o_{i,t}$ is assumed to have a mixture distribution that distinguishes between regular observations $o_{i,t} = 1$ and outliers with $o_{i,t} \geqslant 2$. The probability that outliers in variable i occur is $p_{\mathbf{o}_i}$. We follow Carriero et al. (2022b) and assume that when the outliers occur, they follow a uniform distribution on $(2, 20)$, i.e., $o_{i,t} \sim \mathcal{U}(2, 20)$. The outlier probability $p_{\mathbf{o}_i}$ is assumed to have a beta prior $\mathcal{B}(a_{p_{\mathbf{o}_i}}, b_{p_{\mathbf{o}_i}})$.[2] We refer to this

[2]In practice, the hyperparameters $a_{p_{\mathbf{o}_i}}$ and $b_{p_{\mathbf{o}_i}}$ are calibrated so that the mean outlier frequency is once every 4 years in quarterly or monthly data; see, e.g., Carriero et al. (2022b) for details.

outlier model as the VAR-SVO model.

Specification 2: Student-t distributed innovations

The second specification extends the VAR-SV model by incorporating the latent variables $q_{i,t}$, for $i = 1, \ldots, n$, $t = 1, \ldots, T$. In particular, the squares of the latent variables are mutually $i.i.d.$ over all i and t and have an inverse-gamma distribution:

$$q_{i,t}^2 \sim \mathcal{IG}\left(\frac{l_i}{2}, \frac{l_i}{2}\right).$$

Let $\mathbf{Q}_t = \text{diat}(q_{1,t}, \ldots, q_{n,t})$. Then, the error covariance matrix of the VAR takes the form

$$\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{Q}_t\mathbf{D}_t\mathbf{Q}_t'(\mathbf{B}_0^{-1})'.$$

Under this set-up, the vector of innovations can be written as $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{Q}_t\mathbf{D}_t^{\frac{1}{2}}\mathbf{v}_t$, where $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. It is important to note that the product $q_{i,t}v_{i,t}$ (scaled by $\mathbf{B}_0^{-1}\mathbf{D}_t^{\frac{1}{2}}$) has a student-$t$ distribution with l_i degree of freedom, since $v_{i,t} \sim \mathcal{N}(0, 1)$ and $l_i/q_{i,t}^2 \sim \chi_{l_i}^2$. We therefore call this extension the VAR-SVt model.

Specification 3: Common volatility with a deterministic break date

The third modeling strategy, proposed by Lenza and Primiceri (2022), is tailored to the COVID-19 pandemic, where the break date t^* is known. Specifically, the error covariance matrix now takes the form:

$$\mathbf{\Sigma}_t = s_t^2\mathbf{\Sigma}, \tag{3}$$

where s_t, for $t = 1, \ldots, T$, are latent variables to be estimated. To model the extreme volatility at the onset of the COVID-19 pandemic, the standard deviations of the shocks in March 2020 are scaled by an unknown parameter $\bar{s}_0$; similarly for April and May 2020, with two additional parameters $\bar{s}_1$ and $\bar{s}_2$. Afterward, the volatility is assumed to decay at a constant rate ρ. To summarize, we have

$$s_{t^*} = \bar{s}_0, \quad s_{t^*+1} = \bar{s}_1, \quad s_{t^*+2} = \bar{s}_2, \quad s_{t^*+j} = 1 + (\bar{s}_2 - 1)\rho^{j-2}, \quad j = 3, \ldots, T.$$

This modeling approach is similar to the common stochastic volatility model introduced by Carriero et al. (2016), where the error covariance matrix is scaled by a common, time-varying factor representing the overall macroeconomic volatility. The main difference is that here s_t does not follow a stochastic process, but is a deterministic function of a few parameters.

Compared to other outlier adjustments, this specification is more restrictive due to the constant proportionality, as it effectively models the comovements in the error variances using a shared volatility factor. But it is more parsimonious, and might work particularly well for the COVID-19 outliers, where the timing of their occurrences is known. In addition, it allows for persistent changes in volatility after the onset of the pandemic. We refer to this model with a common volatility and a deterministic break VAR-CVD.

2.3 Data-driven Minnesota priors

We now provide an overview of the priors on the VAR coefficients. Details of priors on other parameters are available in Appendix A. In general, we assume the same priors on the common parameters across models. When it is not applicable, we opt for analogous priors, thereby ensuring comparability among models.

In high-dimensional settings such as large VARs, it is important to impose shrinkage priors to avoid overfitting. There is a vast literature on shrinkage priors for Bayesian VARs. Commonly-used priors include the Minnesota priors (Litterman, 1986; Doan et al., 1984; Giannone et al., 2015; Chan, 2021), the normal-gamma prior (Griffin and Brown, 2010; Huber and Feldkircher, 2019), the horseshoe prior (Follett and Yu, 2019) and the SSVS prior (George et al., 2008). These priors are useful for variable selection and improving forecasting performance in large VARs.

Among these shrinkage priors, the Minnesota priors stands out as the most prominent, primarily due to its ease of use and remarkable performance in forecasting applications.[3] We use a version that has two useful features. First, it incorporates cross-variable shrinkage, i.e., the prior belief that the coefficients on other variables' lags are on average smaller than those on own lags. This feature has been shown to improve forecasting per-

[3] For a more detailed discussion about the Minnesota priors, see, e.g., Koop et al. (2010), Karlsson (2013) and Chan (2020b).

formance; see, e.g., Carriero et al. (2015), Cross et al. (2020) and Chan (2021). Second, the hyperparameters that control the overall shrinkage strength are estimated from the data rather than being set at some subjective values. This adaptive feature has been consistently shown to yield better forecasting results, as demonstrated in a growing body of empirical works such as Giannone et al. (2015), Amir-Ahmadi et al. (2020) and Chan (2023).

Specifically, let $\boldsymbol{\alpha}_i = (a_{i,0}, \mathbf{A}_{i,1}, \dots, \mathbf{A}_{i,p})'$ denote the intercept and coefficients in the i-th equation, where $\mathbf{A}_{i,j}$ is the i-th row of $\mathbf{A}_j$, for $i = 1, \dots, n$. Consider the hierarchical normal prior of the form $\boldsymbol{\alpha}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{\boldsymbol{\alpha}_i})$. The prior covariance matrix $\mathbf{V}_{\boldsymbol{\alpha}_i}$ is assumed to be diagonal and it depends on two hyperparameters: κ_1 and κ_2. The former controls the shrinkage strength on coefficients associated with own lags, whereas the latter controls those on lags of other variables. See Appendix A for details. Similarly, for the free elements in the i-th row of the impact matrix $\mathbf{B}_0$, denoted as $\boldsymbol{\beta}_i$, $i = 2, \dots, n$, we assume that $\boldsymbol{\beta}_i$ has a hierarchical normal prior: $\boldsymbol{\beta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{\boldsymbol{\beta}_i})$, where $\mathbf{V}_{\boldsymbol{\beta}_i}$ is a diagonal matrix and it depends on a hyperparameter κ_3. The hyperparameters κ_1, κ_2, and κ_3 are treated as unknown parameters, each with a hierarchical gamma prior.

3 Overview of Variational Bayes

Variational inference has been gaining popularity as a practical approach for conducting Bayesian inference in situations where the computational demands of MCMC methods are excessive. It is therefore not surprising that many papers have employed variational inference in fitting high-dimensional models, such as large VARs (see e.g., Gefang et al., 2020, 2023; Chan and Yu, 2022; Bernardi et al., 2024), state space models (Loaiza-Maya et al., 2022; Quiroz et al., 2023), copulas (Loaiza-Maya and Smith, 2019; Smith et al., 2020; Deng et al., 2024), quantile regressions (Prüser and Huber, 2024) and multinomial probit models (Loaiza-Maya and Nibbering, 2023). In this section, we outline the basic theory of the variational Bayes approach; Blei et al. (2017) provides a recent review of variational inference.

3.1 The approximate inference and variational lower bound

The key idea of variational inference is to approximate the posterior distribution by a probability distribution with density $q(\boldsymbol{\theta})$ which belongs to some tractable family of distributions $\mathcal{Q}$, such as Gaussians. The best variational approximation $q^* \in \mathcal{Q}$ is found by minimizing a certain measure of how the approximating density q is different from the target $p(\boldsymbol{\theta}|\mathbf{y})$.

The most common type of variational inference is known as variational Bayes (VB) which uses the Kullback-Leibler divergence (KL-divergence) as the choice of dissimilarity function. This specific choice makes the minimization tractable. The KL-divergence of approximating density $q(\boldsymbol{\theta})$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ is defined as

$$\mathrm{KL}(q||p(\,\cdot\,|\mathbf{y})) = \int q(\boldsymbol{\theta}) \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{y})} \mathrm{d}\boldsymbol{\theta},$$

and the best VB approximation $q^* \in \mathcal{Q}$ can therefore be obtained by minimizing the KL-divergence

$$q^* = \underset{q\in\mathcal{Q}}{\operatorname{argmin}} \left\{\mathrm{KL}(q||p(\,\cdot\,|\mathbf{y}))\right\}. \tag{4}$$

It is easy to see that $\mathrm{KL}(q||p(\,\cdot\,|\mathbf{y})$ can be rewritten as

$$\begin{aligned}\mathrm{KL}(q||p(\,\cdot\,|\mathbf{y}) &= \mathbb{E}_q\left[\log q(\boldsymbol{\theta})\right] - \mathbb{E}_q\left[\log p(\boldsymbol{\theta}|\mathbf{y})\right] \\ &= \mathbb{E}_q\left[\log q(\boldsymbol{\theta})\right] - \mathbb{E}_q\left[\log p(\boldsymbol{\theta}, \mathbf{y})\right] + \log p(\mathbf{y}).\end{aligned} \tag{5}$$

From (5), it is not hard to see that minimizing $\mathrm{KL}(q||p(\,\cdot\,|\mathbf{y}))$ is equivalent to maximizing the following function

$$\mathrm{VLB}(q) = \mathbb{E}_q\left[\log p(\boldsymbol{\theta}, \mathbf{y})\right] - \mathbb{E}_q\left[\log q(\boldsymbol{\theta})\right], \tag{6}$$

which is called variational lower bound (VLB), also known as evidence lower bound. To see the intuition behind the VLB, first notice that

$$\mathrm{VLB}(q) = \mathbb{E}_q\left[\log p(\mathbf{y}|\boldsymbol{\theta})\right] - \mathrm{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta})) \tag{7}$$

$$= \log p(\mathbf{y}) - \mathrm{KL}(q(\boldsymbol{\theta}||p(\boldsymbol{\theta}|\mathbf{y})) \tag{8}$$

Equation (7) illustrates the principle that maximizing the VLB involves prioritizing densities that not only accurately capture the observed data but also remain closely aligned with the priors. Further, Equation (8) establishes the VLB as an actual lower bound on the log marginal density since the KL divergence is non-negative ($\mathrm{KL}(\,\cdot\,) \geqslant 0$). This intrinsic relationship between the VLB and the log marginal density renders the VLB a useful criterion for model selection.

3.2 Mean-field variational Bayes

Without any constraint on the density family $\mathcal{Q}$, the best approximating density q^* is nothing but the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. However, in order for this problem to be tractable, we need to impose some constraint(s) on the family $\mathcal{Q}$. The most commonly used constraint is assuming that all the parameters in the vector $\boldsymbol{\theta}$ are mutually independent and each governed by a distinct factor in the variational density

$$q(\boldsymbol{\theta}) = \Pi_{j=1}^{K} q_j(\theta_j). \tag{9}$$

Each density $q_j(\,\cdot\,)$, $j = 1, ..., K$, is then chosen to maximize the VLB of equation (6).

The variational family used in this context is referred to as the mean-field variational family, and the corresponding approach is known as mean-field variational Bayes. There is a growing body of literature on expanding the variational family, including structured variational inference that permits dependencies among the variables (see, e.g., Barber and Wiegerinck, 1998; Hoffman and Blei, 2015), adaptive variational inference that adaptively adjust the variational family as needed during the optimization process to better fit the posterior (see, e.g., Ranganath et al., 2016), normalizing flows that transform the base distribution into a more complicated distribution using a series of invertible transformations (see, e.g., Rezende and Mohamed, 2015), variational Rényi inference that extends the traditional variational inference by minimizing the Rényi divergence between the approximate and the posteriors rather than the KL-divergence (see, e.g., Li and Turner, 2016), to name a few. These methods can potentially improve the approximation, but they usually come with a more difficult-to-solve variational optimization problem. For this reason, we focus on mean-field variational Bayes. In addition, to handle the high-dimensional latent variables in the stochastic volatility models, we adapt the global approximation of the joint distribution of the latent states proposed in Chan and Yu (2022)—which is shown to be fast to obtain and more accurate than alternatives—to our reduced-form VARs. Appendix A contains the estimation details.

4 Model Selection using the Marginal Likelihood

When multiple models are available and they are high-dimensional and nonlinear, a major challenge for practitioners is the lack of adequate tools for comparing these models. In this section, we develop an effective approach to conduct model comparison in these complex settings. To that end, we first provide some background on the marginal likelihood and its significance in Bayesian model comparison. Subsequently, we demonstrate how one can obtain a marginal likelihood estimator through the fusion of variational Bayes and importance sampling. Lastly, we illustrate how the proposed approach works and conduct a comparison of the estimates generated by our method with those of two closely-related alternatives in the context of a linear regression with a closed-form marginal likelihood.

4.1 Overview of the marginal likelihood

One advantage of employing the Bayesian approach is the ability to compare models using the Bayes factor, which is defined as the ratio of the marginal likelihoods of two competing models. Suppose we want to compare K models $\{M_1, \ldots, M_K\}$, where each model M_k is defined by a likelihood function $p(\mathbf{y}|\boldsymbol{\theta}_k, M_k)$ and a prior on the model specific parameter vector $\boldsymbol{\theta}_k$ denoted by $p(\boldsymbol{\theta}_k|M_k)$. The Bayes factor in favor of M_i, against M_j is defined as

$$\mathrm{BF}_{i,j} = \frac{p(\mathbf{y}|M_i)}{p(\mathbf{y}|M_j)},$$

where $p(\mathbf{y}|M_k)$ is the marginal likelihood under model M_k, $k = i, j$, computed by

$$p(\mathbf{y}|M_k) = \int_{\boldsymbol{\theta}_k} p(\mathbf{y}|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k)\mathrm{d}\boldsymbol{\theta}_k. \tag{10}$$

In practice, if $\mathrm{BF}_{ij} = 100$, then model M_i is 100 times more likely than model M_j given the data. For a textbook treatment of the Bayes factor and its role in Bayesian model comparison, see Chan et al. (2019).

One advantage of using the marginal likelihood in a high-dimensional setting is that it contains a "penalty" for model complexity. This ensures that the marginal likelihood naturally prefers simpler models that adequately explain the data over more complex ones, unless the additional complexity significantly increases the model's explanatory power. This "penalty" comes into play in two main ways. First, for models with more parameters, the prior distribution $p(\boldsymbol{\theta}_k)$ tends to be spread over a larger parameter space. Unless there is strong prior information that tightly constrains the parameters, this spreading means that any specific set of parameter values is generally less likely a priori, reducing the marginal likelihood for complex models with more parameters, assuming the prior is properly normalized. Second, in a high-dimensional setting, a model may fit the training data better (higher likelihood), but this does not necessarily translate to a better marginal likelihood. The integral over all parameters averages the likelihood over all possible parameter values, not just the best-fitting ones. If adding more parameters only improves the fit by capturing noise rather than genuine data patterns, this improvement will not significantly enhance the marginal likelihood. Thus, the process inherently penalizes over-fitting.

While the Bayes factor is conceptually straightforward, its computation can be challenging, particularly when dealing with high-dimensional, non-nested models. This is because calculating the marginal likelihood in equation (10) involves integrating the likelihood function with respect to the prior distribution of the parameters. Therefore, the computational burden of computing the marginal likelihood scales with the dimension of the parameter space.

An extensive literature exists on estimating the marginal likelihood using MCMC methods. For instance, important advances include Gelfand and Dey (1994), Newton and Raftery (1994), Chib (1995), Chib and Jeliazkov (2001), Frühwirth-Schnatter and Wagner (2008), Friel and Pettitt (2008), Li et al. (2023), among many others. While these models are widely used in practice, they are computationally infeasible for computing the marginal likelihoods of large VARs with stochastic volatility due to the large number of VAR coefficients and latent variables.

As a computationally feasible alternative, we develop model comparison tools based on the variational approximation of the posterior distribution. Earlier works have investigated if the variational lower bound can be used as a model selection criterion; examples include McGrory and Titterington (2007) for mixture models, Bernardo et al. (2003) for models with incomplete data and Penny (2012) for general linear models and dynamic causal models. In addition, Hajargasht and Woźniak (2020) use the variational approximation in conjunction with the modified harmonic mean estimator of Geweke (1999) to compute the marginal likelihoods of homoskedastic VARs with different shrinkage priors. We further develop this line of research by focusing on large VARs with stochastic volatility. In addition, we consider an importance sampling estimator based on the variational approximation. This is motivated by the observation that the optimal density obtained by minimizing the Kullback-Leibler divergence from the posterior distribution can serve as a convenient choice for the importance sampling density.

4.2 Variational importance sampling

Let $\boldsymbol{\theta}$ denote the parameters of interest, $p(\mathbf{y}|\boldsymbol{\theta})$ denote the posterior distribution, and $q(\boldsymbol{\theta})$ denote the importance densities. Note that we can rewrite equation (10) as the expectation of $[p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})/q(\boldsymbol{\theta})]$ with respect to the importance sampling density, as shown in equation (11)

$$\begin{aligned} p(\mathbf{y}) &= \int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \\ &= \int \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}q(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}. \\ &= \mathbb{E}_q\left[\frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right]. \end{aligned} \tag{11}$$

The importance sampling estimator can therefore be obtained from

$$\hat{p}_{IS}(\mathbf{y}) = \frac{1}{M}\sum_{m=1}^{M}\frac{p\left(\mathbf{y}|\boldsymbol{\theta}^{(m)}\right)p\left(\boldsymbol{\theta}^{(m)}\right)}{q\left(\boldsymbol{\theta}^{(m)}\right)}, \tag{12}$$

where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$ are M independent draws obtained from the importance sampling density $g(\,\cdot\,)$ that dominates $p(\mathbf{y}|\,\cdot\,)p(\,\cdot\,)$, i.e., $q(\mathbf{x}) = 0 \Rightarrow p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = 0$.

The estimator in (12) is unbiased and simulation consistent for any density q that dominates $p(\mathbf{y}|\,\cdot\,)p(\,\cdot\,)$. However, in practice, the performance of this estimator depends heavily on the choice of the importance sampling density. Chan and Eisenstat (2015) used the cross-entropy method to obtain the "best" importance sampling density by choosing the optimal parameters in a parametric family that minimizes the Kullback-Leibler divergence between the posterior density and the importance sampling density. This was later used in Chan (2023) to compare different specifications of stochastic volatility in VARs. Both approaches, however, rely on MCMC draws, which are costly to obtained in very high-dimensional settings.

Instead, here we obtain the densities that minimize the Kullback-Leibler divergence using the variational Bayes approach, without relying on MCMC draws. We call this method variational importance sampling (VIS). The algorithm is summarized in Algorithm 1 in the Appendix B. It is noteworthy that in the adaptive importance sampling based on cross-entropy method (CEAIS) proposed in Chan and Eisenstat (2015), the Kullback-Leibler divergence is the divergence of the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ from the importance sampling density $q(\boldsymbol{\theta})$, whereas in the variational Bayes, it is the divergence of the approximating density $q(\boldsymbol{\theta})$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. These two quantities are not equal because the Kullback-Leibler divergence is not symmetric, i.e., $\mathrm{KL}(p(\boldsymbol{\theta}|\mathbf{y})||q(\boldsymbol{\theta})) \neq \mathrm{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{y}))$. Nevertheless, in many applications, such as the illustration in the following section, these two approaches give similar estimates, suggesting that both are accurate approximations of the posterior distribution.

One important concern of using the optimal density obtained from the variational Bayes approach as the importance sampling density is that it tends to under-represent the variance of the posterior density. This is a common effect in mean-field variational inference; see Blei et al. (2017) for details. This could be problematic for importance sampling, because for the estimator in (12) to work well, the variance of the importance sampling weights should be finite. Checking this requirement is often possible in simple problems, but it is difficult in high-dimensional settings. To ensure that this finite-variance condition holds, one strategy is to implement the so-called defensive importance sampling (DIS) proposed by Hesterberg (1995). Specifically, instead of directly using the original importance sampling density $q(\boldsymbol{\theta})$, one puts some weight $\gamma \in (0, 1)$ on the prior $p(\boldsymbol{\theta})$ and uses the mixture

$$q_\gamma(\boldsymbol{\theta}) = \gamma p(\boldsymbol{\theta}) + (1 - \gamma)q(\boldsymbol{\theta}),$$

as the importance sampling density. One can then show that the weight function $w(\boldsymbol{\theta}) =$

$p(\boldsymbol{\theta})/q_\gamma(\boldsymbol{\theta})$ is bounded by $1/\gamma$, and therefore the variance of the importance sampling weights is finite.

4.3 An illustration: a linear regression with a closed-form marginal likelihood

We consider a simple example of the linear regression to illustrate how the algorithm works in high-dimensional settings. In particular, consider the following linear regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n),$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$, $\mathbf{x}_j = (x_{1j}, \ldots, x_{nj})'$, $j = 1, \ldots, k$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$. We assume the following natural conjugate prior

$$(\boldsymbol{\beta}|\sigma^2) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Lambda_0^{-1}), \quad \sigma^2 \sim \mathcal{IG}(\nu, S).$$

The log marginal likelihood is then available in closed-form:

$$\begin{aligned}\log p(\mathbf{y}) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log\det(\Lambda_0) - \frac{1}{2}\log\det(\mathbf{X}'\mathbf{X} + \Lambda_0) \\ +\nu\log(S) - \log\Gamma(\nu) + \log\Gamma(\widetilde{\nu}) - \widetilde{\nu}\log(\widetilde{S}),\end{aligned}$$

where $\widetilde{\nu} = \frac{T}{2} + \nu$, $\widetilde{S} = \frac{1}{2}\left(\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \Lambda_0)^{-1}\mathbf{X}'\mathbf{y}\right)$.

In the following Monte Carlo experiments, the data are generated as follows. We set $\sigma^2 = 3$ and sample k iid draws for $\boldsymbol{\beta}$ from the normal distribution: $\mathcal{N}(0, 0.3^2)$. In order to compare the estimates under different dimensions of parameters, we generate 9 datasets with different sizes. In the smallest dataset, $n = 500, k = 10$, and in the largest dataset, $n = 10,000, k = 200$, as shown in Table 1. In terms of the hyperparameters in the priors, we set $S = 10$ and $\nu = 4$. When $n/k < 100$, we set $\Lambda_0 = 2.4\mathbf{I}_k$. Otherwise, we set $\Lambda_0 = 0.3\mathbf{I}_k$, in order to impose more shrinkage on $\boldsymbol{\beta}$ when the dimension of parameters is high.

Table 1 reports the log marginal likelihood estimates using three approaches: the cross-entropy approach of Chan and Eisenstat (2015) (CEAIS), the defensive importance sampling of Hesterberg (1995) (DIS) and the proposed variational importance sampling (VIS).

It also reports the true log marginal likelihood values (TRUE) computed using the closed-form formula and the variational lower bounds (VLB). The results show that the estimates from the three methods are virtually identical to the true values for all the datasets, but their standard errors vary. For all the cases, the VIS gives the smallest standard errors, especially in higher-dimensional settings. For example, when $k = 200$, the standard error for the VIS is around 0.002, about 10 times smaller than those of CEAIS and DIS, highlighting the efficiency of the VIS estimator.

Table 1: Log marginal likelihood estimates of the linear regression.

(n, k)	TRUE	CEAIS	DIS	VIS	VLB
(500, 10)	-1013	-1013	-1013	-1013	-1031
		(0.002)	(0.003)	(0.001)	
(500, 20)	-1024	-1024	-1024	-1024	-1061
		0.002	0.003	0.003	
(500, 50)	-1154	-1154	-1154	-1154	-1246
		(0.005)	(0.007)	(0.004)	
(1000, 10)	-2002	-2002	-2002	-2002	-2020
		(0.001)	(0.003)	(0.001)	
(1000, 20)	-2000	-2000	-2000	-2000	-2036
		(0.002)	(0.003)	(0.002)	
(1000, 50)	-2176	-2176	-2176	-2176	-2268
		(0.005)	(0.006)	(0.002)	
(10000, 50)	-19833	-19833	-19833	-19833	-19925
		(0.004)	(0.005)	(0.002)	
(10000, 100)	-20196	-20196	-20196	-20196	-20380
		(0.009)	(0.008)	(0.001)	
(10000, 200)	-20642	-20642	-20642	-20642	-21010
		(0.025)	(0.023)	(0.002)	

This table shows the log marginal likelihood estimates using the three approaches: the cross-entropy approach of Chan and Eisenstat (2015) (CEAIS), the defensive importance sampling of Hesterberg (1995) (DIS) and the proposed variational importance sampling (VIS). The standard errors are reported in parenthesis. The second column reports the true value of the log marginal likelihood (TRUE). The last column reports the variational lower bounds (VLB).

5 Performance of the VB Algorithm and the Log Marginal Likelihood Estimator

In this section, we conduct several Monte Carlo experiments to assess the performance of the variational Bayes (VB) algorithms and the log marginal likelihood estimator. Specifically, in the first subsection, we focus on the comparison of the accuracy of estimates and computational burden using VB and MCMC methods. The second subsection then assesses the capability of the variational importance sampling estimator in correctly identifying the true models. We also explore the appropriateness of using the variational lower bound as a model selection criterion.

5.1 Variational Bayes vs MCMC: Computational time and accuracy

We first evaluate the computational time to fit VAR-SV using VB and MCMC for small ($n = 5$), medium ($n = 20$) and large ($n = 50, 100$) datasets with $T = 300, 500, 1000$ observations. Table 2 reports the results. For small and medium datasets, employing MCMC remains practical. For instance, fitting a 5-variable VAR model with 300 observations via MCMC can be completed in approximately one minute, yielding 10,000 posterior draws. However, the computational demand escalates significantly with larger models; a 100-variable VAR model with a sample size of $T = 500$ necessitates around 20 hours for MCMC estimation. In stark contrast, the VB method requires merely about 3 minutes. This efficiency gains become particularly advantageous in applications such as macroeconomic forecasting that involves recursive estimation with an expanding window or model comparison that requires estimation for multiple models.

Table 2: The computational time (in seconds) to fit an n-variable VAR-SV with a sample size T using MCMC (to obtain 10,000 posterior draws) and VB (to converge). All VARs have $p = 4$ lags.

T	n	MCMC	VB
300	5	64.5	0.3
	20	280.9	1.1
	50	2735.5	12.9
	100	16692.3	89.1
500	5	73.2	0.5
	20	301.3	1.6
	50	2864.1	62.5
	100	74466.2	199.4
1000	5	142.3	1.3
	20	505.4	3.2
	50	5212.0	31.9
	100	94636.6	388.2

Next, we compare the estimates from our VB approach against the standard MCMC approach using the FRED-QD data. Specifically, we use the vintage of "2024-02", consisting of the data from September 1959 to December 2023. After transforming the raw data to stationary data and removing the columns with missing values, we obtain a data set with 180 variables ($n = 180$) and 258 observations ($T = 258$). Next, we randomly choose 5, 10 and 50 variables ($n = 5, 10, 50$) and compare the accuracy of the estimates of VAR coefficients $\mathbf{A}$, impact matrix coefficient $\mathbf{B}_0$, and stochastic volatility using VB and MCMC approaches.

Figure 1 reports a scatter plot comparing the estimates of the VAR coefficients ($\mathbf{A}$) obtained via VB and MCMC methods. Similarly, Figure 2 displays the comparison of the estimates of the impact matrix coefficients ($\mathbf{B}_0$). These figures demonstrate that the posterior means derived from the VB and MCMC methods are nearly indistinguishable. While there are instances of slight discrepancies, the estimates are similar.

In Figure 3, we compare the estimates of stochastic volatility using VB and MCMC. For better presentation, we used the results from a VAR that consists of 5 key macroeconomic variables: real GDP, personal consumption expenditures, real private fixed investment, unemployment rate, and CPI. Again, estimates of stochastic volatility from VB and MCMC are quite similar.

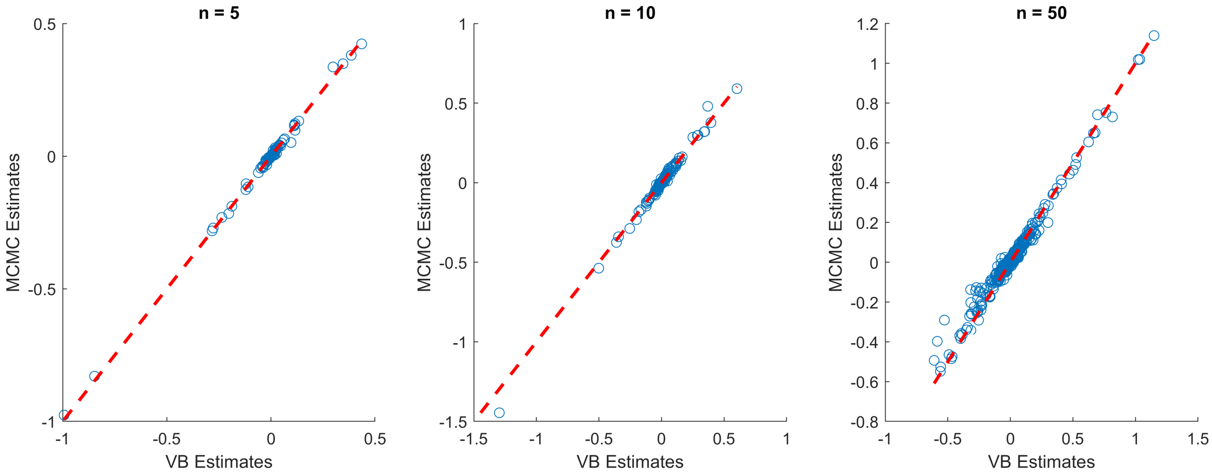

Figure 1: Scatter plots of the estimates of the VAR coefficients from VB and MCMC. The dashed red line is the 45-degree line.

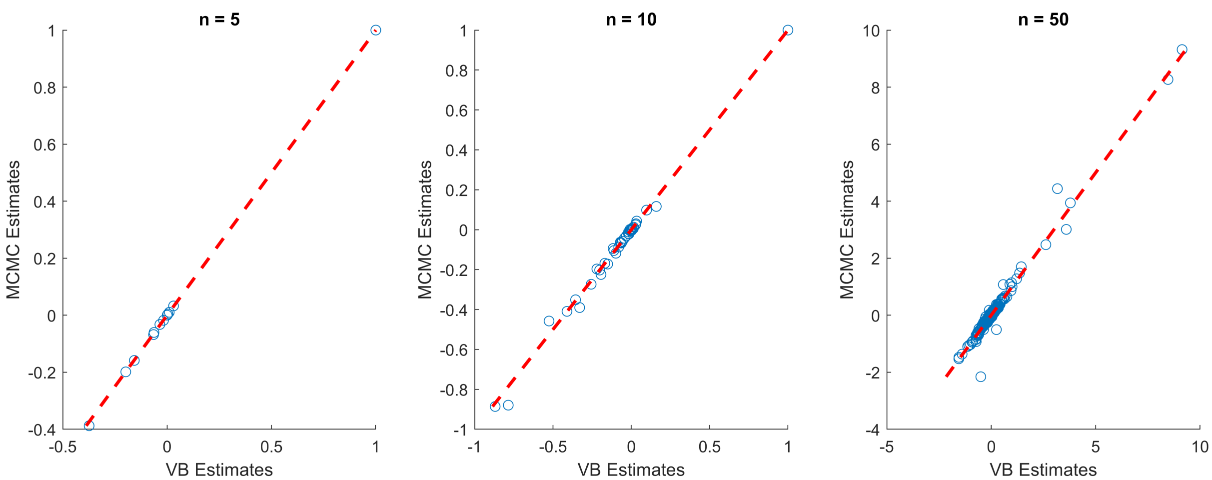

Figure 2: Scatter plots of the estimates of the impact matrix coefficients from VB and MCMC. The dashed red line is the 45-degree line.

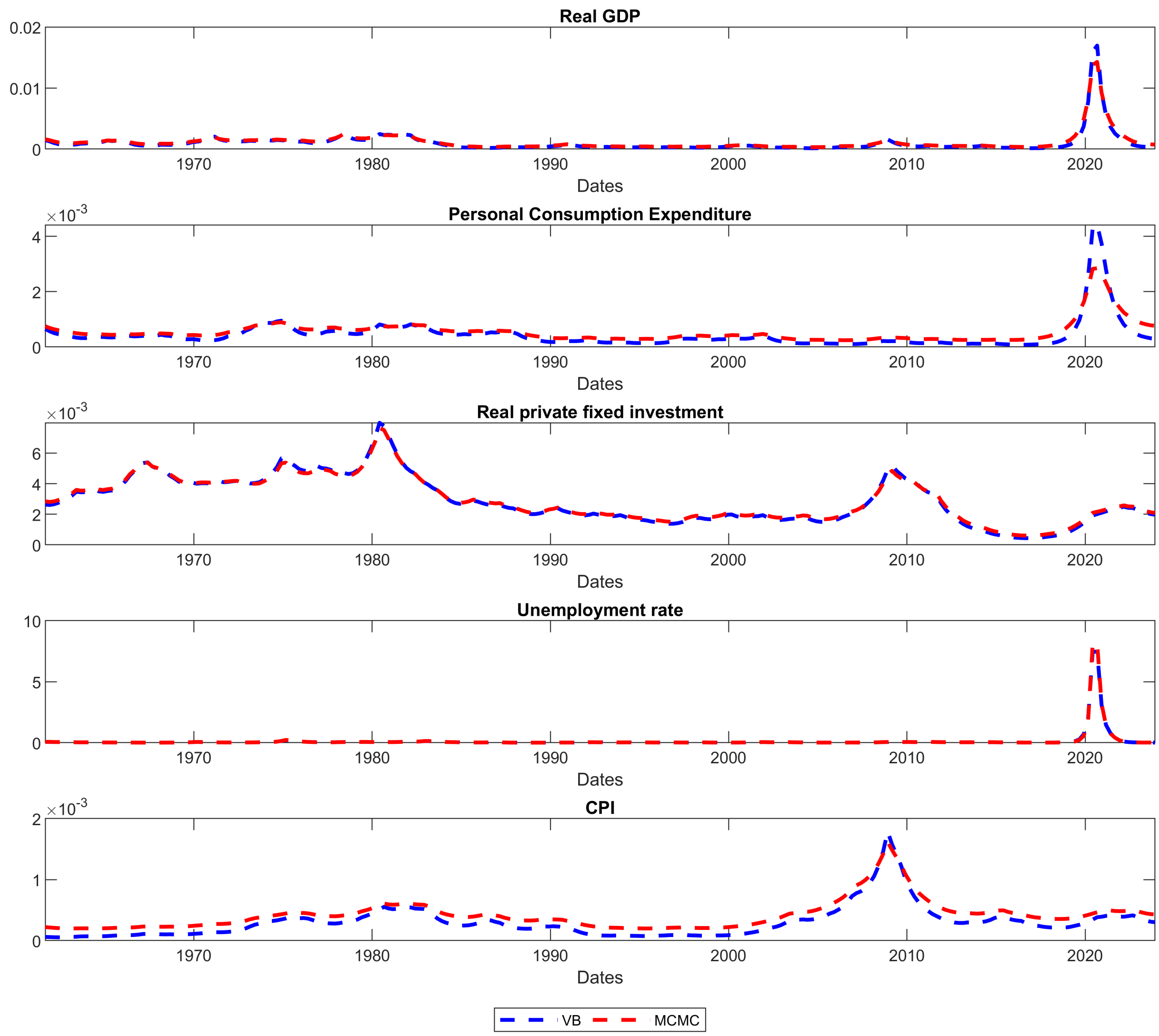

Figure 3: Estimates of the stochastic volatility from VB (blue line) and MCMC (red line).

5.2 Can the VIS estimator identify the correct models?

Next, we delve into the question of whether we can distinguish the four VARs with stochastic volatility, namely, VAR-SV, VAR-SVO, VAR-SVt, and VAR-CVD, using the variational importance sampling estimator of the marginal likelihood. To that end, we generate 100 datasets from each of the four models. Each dataset consists of 30 variables ($n = 30$), 300 observations ($T = 300$) and 4 lags ($p = 4$). We generate the intercepts from $\mathcal{U}(0, 1/3)$. The free elements in the impact matrix are *i.i.d.* $\mathcal{N}(0, 0.5^2)$. The diagonal elements of the first VAR coefficient matrix are *i.i.d.* $\mathcal{U}(-0.2, 0.4)$ and the off-diagonal elements are $\mathcal{U}(-0.2, 0.6)$; all elements of the j-th VAR coefficient matrix are *i.i.d.* $\mathcal{N}(0, (0.1/j)^2)$, $j = 1, \ldots, p$.

For VAR-SVO, the outlier parameter $o_{i,t}$ is assigned the value of 1 with a probability 15/16, and is randomly drawn from a discrete uniform distribution from 2 to 9 with probability 1/16. The parameter $q_{i,t}^2$ is drawn from inverse gamma distribution: $\mathcal{IG}(3, 20)$, for $i = 1, \ldots, n$, and $t = 1, \ldots, T$. The variance for evolution process of latent factor $\mathbf{h}_i$, σ_i is set to 0.1 for $i = 1, \ldots, n$ and $\mathbf{h}_i$ is drawn from a random walk process with variance σ_i and the initial factor is set to be 0.

For VAR-CVD, we specify $\rho = 0.8$, $\bar{s}_0 = 15$, $\bar{s}_1 = 70$, and $\bar{s}_2 = 20$. These values closely approximate the estimates provided in Lenza and Primiceri (2022). In addition, We set $t^* = 80$, and $\mathbf{\Sigma}$ is randomly drawn from an inverse-Wishart distribution with a mean of $5\mathbf{I}_n$, where $\mathbf{I}_n$ represents the identity matrix of size n.

In the first experiment, we generate 100 datasets from VAR-SV. For each dataset, we then compute the log marginal likelihoods of VAR-SVO, VAR-SVt and VAR-CVD in comparison to that of the true model VAR-SV. To be specific, we subtract the latter log marginal likelihood from the log marginal likelihoods of VAR-SVO, VAR-SVt and VAR-CVD. Given that a model is preferred by the data if it has larger log marginal likelihood value, a difference that is negative implies that the correct model is favored. Results in Figure 4 show that for all the datasets the correct model VAR-SV is more favored in comparison to the other two specifications.

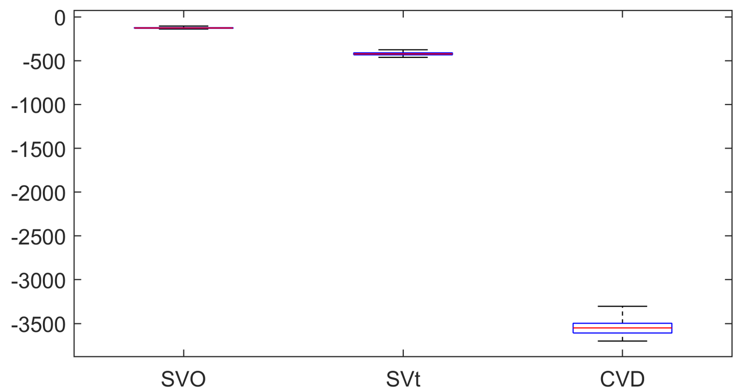

Figure 4: Boxplots of log marginal likelihoods under VAR-SVO (left), VAR-SVt (middle) and VAR-CVD (right) relative to the true model (VAR-SV). A negative value indicates that the correct model is favored.

Next, we generated 100 datasets from VAR-SVO. For each dataset, we then compute the log marginal likelihoods of VAR-SV, VAR-SVt and VAR-CVD, relative to that of the true model. The results are shown in Figure 5. The results show that for all datasets, the correct model VAR-SVO is favored compared to the other two specifications.

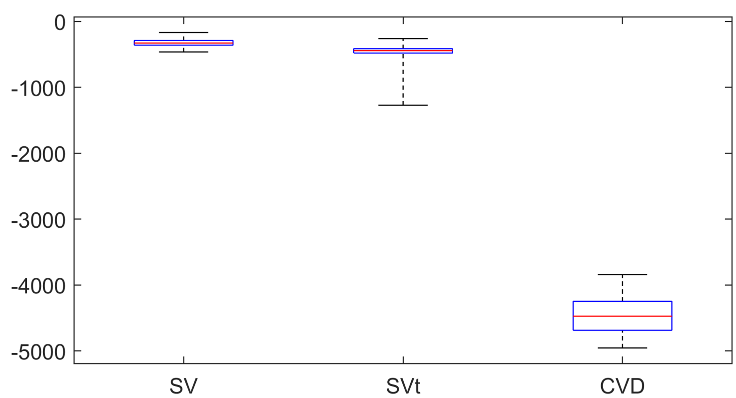

Figure 5: Boxplots of log marginal likelihoods under VAR-SV (left), VAR-SVt (middle) and VAR-CVD (right) relative to the true model (VAR-SVO). A negative value indicates that the correct model is favored.

In the third experiment, we generate 100 datasets from VAR-SVt. Again, we compute the log marginal likelihoods of VAR-SV, VAR-SVt and VAR-CVD, relative to that of the true model. The results are shown in Figure 6. It is clear that for all datasets, the correct model VAR-SVt is favored compared to the others.

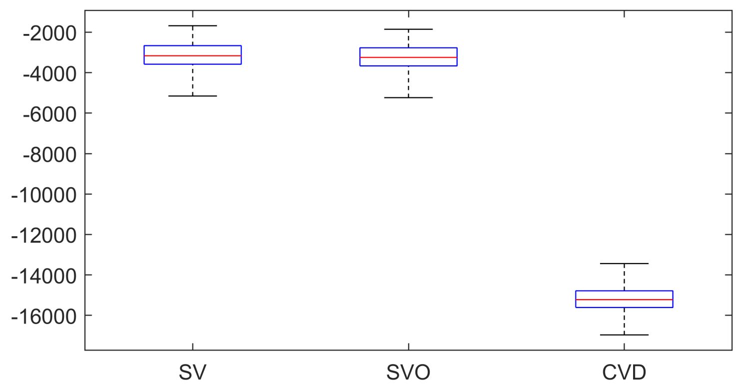

Figure 6: Boxplots of log marginal likelihoods under VAR-SV (left), VAR-SVO (middle) and VAR-CVD(right) relative to the true model (VAR-SVt). A negative value indicates that the correct model is favored.

At last, we generate 100 datasets from VAR-CVD. We compute the log marginal likelihoods of VAR-SV, VAR-SVO and VAR-CVD, relative to that of the true model. The results are shown in Figure 7. It is clear that for all datasets, the correct model VAR-CVD is favored over the others.

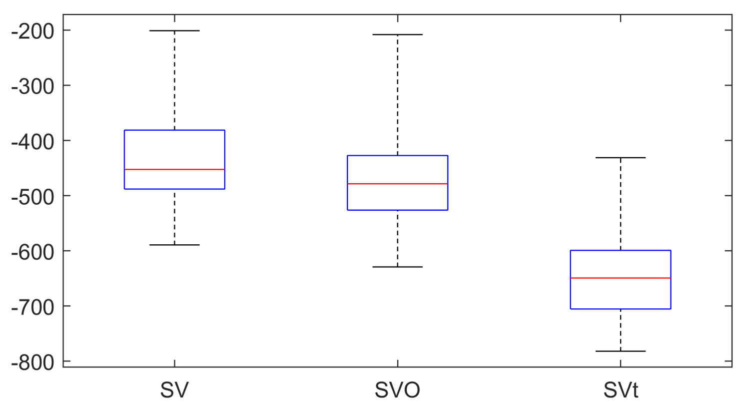

Figure 7: Boxplots of log marginal likelihoods under VAR-SV (left), VAR-SVO (middle) and VAR-SVt (right) relative to the true model (VAR-CVD). A negative value indicates that the correct model is favored.

A key observation from these results is that employing a common volatility model like VAR-CVD results in a markedly large disparity in the log marginal likelihood, when the true model is a more flexible stochastic volatility model such as VAR-SV, VAR-SVO, or VAR-SVt, as shown in Figures 4 - 6. For example, when the underlying model is VAR-SV, the mean differences in log marginal likelihood for VAR-SVO and VAR-SVt are approximately -127 and -422, respectively. In contrast, the difference with VAR-CVD is around $-3,550$. This observation serves as a valuable guide in deciding whether to opt for a more flexible stochastic volatility model or a common volatility model in practical applications.

5.3 Can the variational lower bound identify the correct models?

To assess the efficacy of variational lower bounds (VLBs) in aiding model selection, we examine the VLBs for the datasets generated in the preceding section. Subsequently, we calculate the difference between the VLBs of the true model and those of the competing models.

Figures 8-11 show the boxplots of the variational lower bounds under the true model of VAR-SV, VAR-SVO, VAR-SVt and VAR-CVD, respectively. Again, a negative value indicates that the correct model is favored.

In all the 100 generated datasets, the variational lower bounds consistently attain their highest values for the true model, thereby showcasing the discerning ability of the VLB in correctly identifying the model.

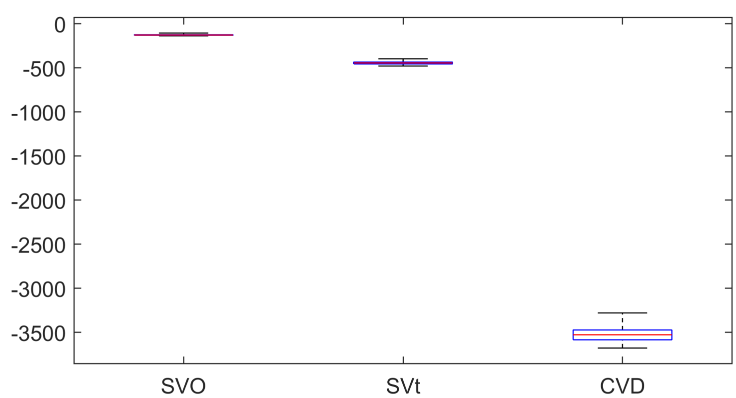

Figure 8: Histograms of variational lower bounds under VAR-SVO (left), VAR-SVt (middle) and VAR-CVD (right) relative to the true model (VAR-SV). A negative value indicates that the correct model is favored.

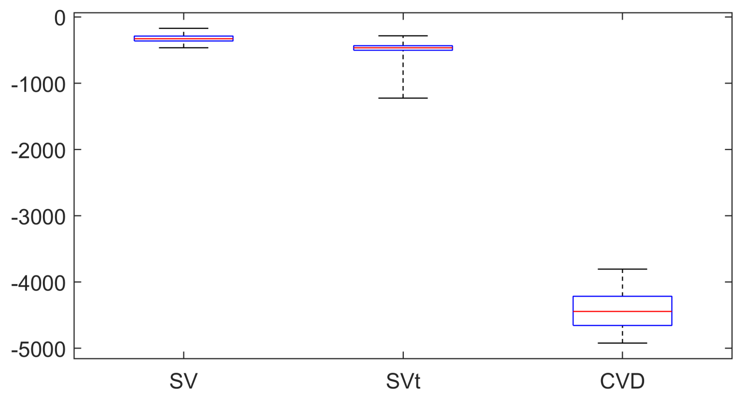

Figure 9: Histograms of variational lower bounds under VAR-SV (left), VAR-SVt (middle) and VAR-CVD (right) relative to the true model (VAR-SVO). A negative value indicates that the correct model is favored.

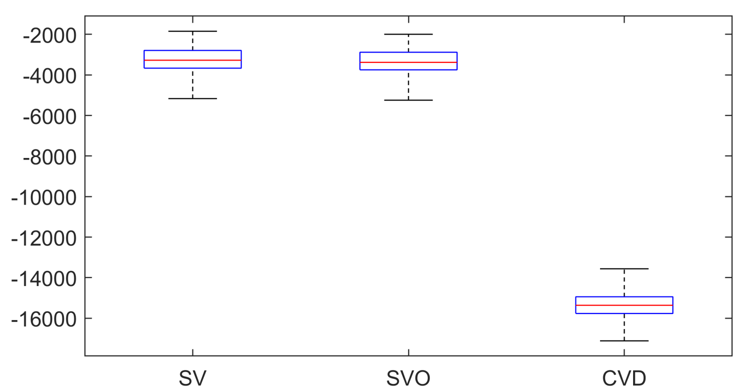

Figure 10: Histograms of variational lower bounds under VAR-SV (left), VAR-SVO (middle) and VAR-CVD (right) relative to the true model (VAR-SVt). A negative value indicates that the correct model is favored.

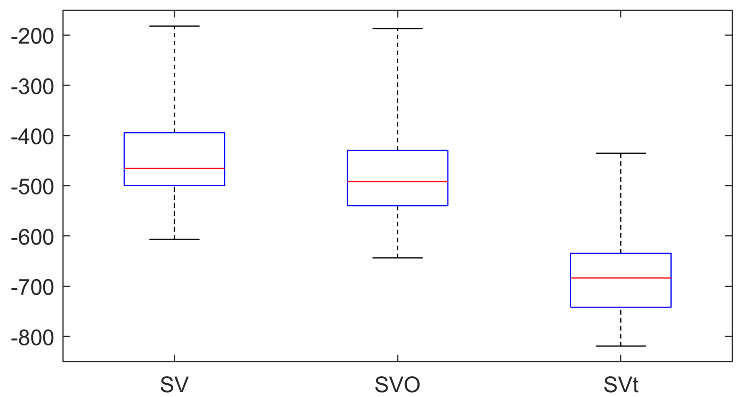

Figure 11: Histograms of variational lower bounds under VAR-SV (left), VAR-SVO (middle) and VAR-SVt (right) relative to the true model (VAR-CVD). A negative value indicates that the correct model is favored.

6 Empirical Application

We demonstrate the proposed methodology using an empirical application that compares different stochastic volatility specifications and outlier components in the context of Bayesian VARs.

6.1 Data

We employ two datasets of different sizes in this application. The first dataset is the same as that in Carriero et al. (2022b), which consists of 16 monthly variables, including real income, real consumption, industrial production, and inflation indexes. The list of the variables and their transformations are outlined in Appendix C. This dataset covers the period from March 1959 to March 2021. We include $p = 12$ lags in the VARs to fit this monthly dataset. Taking the first 12 observations as the initial values, we use the remaining 733 observations for estimation.

The second dataset is constructed from the FRED-QD database at the Federal Reserve Bank of St. Louis. We use the "2024-02" vintage, spanning from September 1959 to December 2023 with 258 observations. Only variables with complete data for the entire sample period are selected. The first 8 observations serve as initial values, culminating in a dataset of dimension 180×250. The raw data is transformed based on the code provided by McCracken and Ng (2020). The VARs for this quarterly dataset incorporate $p = 4$ lags. Further details of the dataset are provided in Appendix C.

6.2 Model comparison results

We compare a variety of VARs with different stochastic volatility and outlier specifications using the two datasets described above. More specifically, we consider a VAR with the stochastic volatility model (VAR-SV) of Cogley and Sargent (2005), the outlier-augmented version (VAR-SVO) developed in Carriero et al. (2022b), a variant with the Student-t innovations (VAR-SVt) and the common volatility model with a deterministic break date (VAR-CVD) proposed in Lenza and Primiceri (2022). As a benchmark, we also include a standard homoskedastic VAR.

Table 3 reports the log marginal likelihood estimates alongside the variational lower bounds of the five VARs across the two model dimensions. It is evident that the models incorporating any form of stochastic volatility are decidedly favored over the standard homoskedastic VAR for both datasets. For instance, the difference between the log marginal likelihoods of VAR-SV and the homoskedastic VAR is 2,922 for the 16-variable dataset, highlighting overwhelming preference for the stochastic volatility model. This finding is consistent with the growing body of evidence that underscores the importance of time-varying volatility in fitting both medium and large macroeconomic datasets.

Table 3: Log marginal likelihood estimates (numerical standard errors) and variational lower bounds of a standard homoskedastic VAR, VAR-SV, VAR-SVO, VAR-SVt, and VAR-CVD.

		VAR	VAR-SV	VAR-SVO	VAR-SVt	VAR-CVD
16 variables	Log-ML	-20,523	-17,601	**-17,594**	-17,697	-20,488
		(0.7)	(0.7)	(1.0)	(1.1)	(0.6)
	VLB	-20,541	-17,625	**-17,620**	-17,743	-20,499
180 variables	Log-ML	140,905	**199,943**	199,306	198,388	141,065
		(4.4)	(3.8)	(3.4)	(4.8)	(4.7)
	VLB	140,679	**199,699**	199,060	198,076	140,839

In addition, the common volatility model VAR-CVD is outperformed by the three stochastic volatility models for both datasets. For example, the log marginal likelihood difference between VAR-SV and VAR-CVD is 2,887 for the 16-variable dataset, and this difference increases to 58,878 for the 180-variable dataset. This suggests that the common volatility assumption might be too restrictive. Another possibility is that time-varying volatility is important throughout the sample, not only after the onset of the COVID-19 pandemic (recall that VAR-CVD assumes homoskedastic errors before the known volatility break).

Among the VARs featuring stochastic volatility, the 16-variable dataset shows a slight preference for VAR-SVO, suggesting that the outlier component enhances model-fit relative to the increase in model complexity. However, for the 180-variable dataset, VAR-SV has a higher log marginal likelihood. This could be attributed to several reasons. First, outlier adjustments may offer significant benefits in smaller models, yet their marginal contributions can diminish in larger models where the larger set of variables can better explain the outlier-induced variations. Second, with increasing model complexity, there is a greater risk of overfitting in which the model captures noise rather than the true underlying process. As discussed in Section 4, the marginal likelihood inherently penalizes model complexity; if the improvement in fit is marginal, the marginal likelihood value of the more complex model would be smaller.

Finally, the inferences drawn from the variational lower bounds align with those from the marginal likelihood, affirming the viability of using the variational lower bound as an alternative measure for model comparison.

7 Conclusion

This paper has tackled the problem of model selection in the context of large Bayesian VARs that account for time-varying volatility and outlier adjustments. We have considered variational approximations to the joint posterior distributions of these models, along with importance sampling estimators for the marginal likelihoods. Our Monte Carlo experiments affirmed that the proposed methodology significantly expedites the estimation process relative to traditional MCMC approaches, while also being able to identify the correct models. Moreover, our results suggested that the variational lower bound might be a viable alternative for model comparison.

The effectiveness and applicability of our approach were further validated through its application to medium and large VARs with stochastic volatility and outlier adjustments. The results showed that for both model sizes, incorporating stochastic volatility and adjustments for outliers generally improves the model-fit. Overall, whether the inclusion of the outlier component in addition to stochastic volatility is beneficial appears to depend on the size of the dataset. For the 16-variable dataset, the inclusion of outlier adjustments seems to provide benefits, whereas for the 180-variable dataset, the benefit is less clear.

References

Amir-Ahmadi, P., C. Matthes, and M.-C. Wang (2020): "Choosing prior hyperparameters: With applications to time-varying parameter models," *Journal of Business & Economic Statistics*, 38, 124–136.

Bańbura, M., D. Giannone, and M. Lenza (2015): "Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections," *International Journal of Forecasting*, 31, 739–756.

Bańbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92.

Barber, D. and W. Wiegerinck (1998): "Tractable variational structures for approximating graphical models," *Advances in Neural Information Processing Systems*, 11.

Bernardi, M., D. Bianchi, and N. Bianco (2024): "Variational inference for large Bayesian vector autoregressions," *Journal of Business & Economic Statistics*, 1–17.

Bernardo, J., M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West, et al. (2003): "The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures," *Bayesian statistics*, 7, 210.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017): "Variational inference: A review for statisticians," *Journal of the American statistical Association*, 112, 859–877.

Bobeica, E. and B. Hartwig (2023): "The COVID-19 shock and challenges for inflation modelling," *International Journal of Forecasting*, 39, 519–539.

Carriero, A., J. Chan, T. E. Clark, and M. Marcellino (2022a): "Corrigendum to "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors"[J. Econometrics 212 (1)(2019) 137–154]," *Journal of Econometrics*, 227, 506–512.

Carriero, A., T. E. Clark, and M. Marcellino (2015): "Bayesian VARs: specification choices and forecast accuracy," *Journal of Applied Econometrics*, 30, 46–73.

——— (2016): "Common drifting volatility in large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390.

——— (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2022b): "Addressing COVID-19 outliers in BVARs with stochastic volatility," *Review of Economics and Statistics*, 1–38.

Carriero, A., G. Kapetanios, and M. Marcellino (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25, 400–417.

Chan, J. C. C. (2020a): "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business and Economic Statistics*, 38, 68–79.

——— (2020b): "Large Bayesian Vector Autoregressions," *Macroeconomic Forecasting in the Era of Big Data*, 95–125.

——— (2021): "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37, 1212–1226.

——— (2023): "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*, 235, 1419–1446.

Chan, J. C. C. and E. Eisenstat (2015): "Marginal likelihood estimation with the Cross-Entropy method," *Econometric Reviews*, 34, 256–285.

——— (2018): "Bayesian model comparison for time-varying parameter VARs with stochastic volatility," *Journal of Applied Econometrics*, 33, 509–532.

Chan, J. C. C., G. Koop, D. J. Poirier, and J. L. Tobias (2019): *Bayesian Econometric Methods*, vol. 7, Cambridge University Press.

Chan, J. C. C. and X. Yu (2022): "Fast and accurate variational inference for large Bayesian VARs with stochastic volatility," *Journal of Economic Dynamics and Control*, 143, 104505.

Chib, S. (1995): "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90, 1313–1321.

CHIB, S. AND I. JELIAZKOV (2001): "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270–281.

COGLEY, T. AND T. J. SARGENT (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic dynamics*, 8, 262–302.

CROSS, J. L., C. HOU, AND A. POON (2020): "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity," *International Journal of Forecasting*, 36, 899–915.

CUINGNET, R. (2023): "On the Computation of the Logarithm of the Modified Bessel Function of the Second Kind," *arXiv preprint arXiv:2308.11964*.

DELLAPORTAS, P. AND M. G. TSIONAS (2019): "Importance sampling from posterior distributions using copula-like approximations," *Journal of Econometrics*, 210, 45–57.

DENG, L., M. S. SMITH, AND W. MANEESOONTHORN (2024): "Large Skew-t Copula Models and Asymmetric Dependence in Intraday Equity Returns," *Journal of Business & Economic Statistics*, 1–40.

DOAN, T., R. LITTERMAN, AND C. SIMS (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3, 1–100.

FOLLETT, L. AND C. YU (2019): "Achieving parsimony in Bayesian vector autoregressions with the horseshoe prior," *Econometrics and Statistics*, 11, 130–144.

FRIEL, N. AND A. N. PETTITT (2008): "Marginal likelihood estimation via power posteriors," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70, 589–607.

FRÜHWIRTH-SCHNATTER, S. AND H. WAGNER (2008): "Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling," *Computational Statistics & Data Analysis*, 52, 4608–4624.

GEFANG, D., G. KOOP, AND A. POON (2020): "Computationally efficient inference in large Bayesian mixed frequency VARs," *Economics Letters*, 191, 109120.

——— (2023): "Forecasting using variational Bayesian inference in large vector autoregressions with hierarchical shrinkage," *International Journal of Forecasting*, 39, 346–363.

Gelfand, A. E. and D. K. Dey (1994): “Bayesian model choice: asymptotics and exact calculations,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 56, 501–514.

George, E. I., D. Sun, and S. Ni (2008): “Bayesian stochastic search for VAR model restrictions,” *Journal of Econometrics*, 142, 553–580.

Geweke, J. (1999): “Using simulation methods for Bayesian econometric models: Inference, development, and communication,” *Econometric Reviews*, 18, 1–73.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): “Prior selection for vector autoregressions,” *Review of Economics and Statistics*, 97, 436–451.

Griffin, J. E. and P. J. Brown (2010): “Inference with normal-gamma prior distributions in regression problems,” *Bayesian Analysis*, 5, 171–188.

Hajargasht, G. and T. Woźniak (2020): “Accurate computation of marginal data densities using variational Bayes,” *arXiv preprint arXiv:1805.10036*.

Hesterberg, T. (1995): “Weighted average importance sampling and defensive mixture distributions,” *Technometrics*, 37, 185–194.

Hoffman, M. D. and D. M. Blei (2015): “Structured stochastic variational inference,” in *Artificial Intelligence and Statistics*, 361–369.

Huber, F. and M. Feldkircher (2019): “Adaptive shrinkage in Bayesian vector autoregressive models,” *Journal of Business & Economic Statistics*, 37, 27–39.

Karlsson, S. (2013): “Forecasting with Bayesian vector autoregression,” *Handbook of Economic Forecasting*, 2, 791–897.

Koop, G. and D. Korobilis (2013): “Large time-varying parameter VARs,” *Journal of Econometrics*, 177, 185–198.

——— (2018): “Variational Bayes inference in high-dimensional time-varying parameter models,” *arXiv preprint arXiv:1809.03031*.

Koop, G., D. Korobilis, et al. (2010): “Bayesian multivariate time series methods for empirical macroeconomics,” *Foundations and Trends® in Econometrics*, 3, 267–358.

Koop, G. M. (2013): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28, 177–203.

Korobilis, D. and D. Pettenuzzo (2019): "Adaptive hierarchical priors for high-dimensional vector autoregressions," *Journal of Econometrics*, 212, 241–271.

Lenza, M. and G. E. Primiceri (2022): "How to estimate a vector autoregression after March 2020," *Journal of Applied Econometrics*, 37, 688–699.

Li, Y. and R. E. Turner (2016): "Rényi divergence variational inference," *Advances in Neural Information Processing Systems*, 29.

Li, Y., N. Wang, and J. Yu (2023): "Improved marginal likelihood estimation via power posteriors and importance sampling," *Journal of Econometrics*, 234, 28–52.

Litterman, R. B. (1986): "Forecasting with Bayesian vector autoregressions—five years of experience," *Journal of Business & Economic Statistics*, 4, 25–38.

Loaiza-Maya, R. and D. Nibbering (2023): "Fast variational Bayes methods for multinomial probit models," *Journal of Business & Economic Statistics*, 41, 1352–1363.

Loaiza-Maya, R. and M. S. Smith (2019): "Variational Bayes estimation of discrete-margined copula models with application to time series," *Journal of Computational and Graphical Statistics*, 28, 523–539.

Loaiza-Maya, R., M. S. Smith, D. J. Nott, and P. J. Danaher (2022): "Fast and accurate variational inference for models with many latent variables," *Journal of Econometrics*, 230, 339–362.

McCracken, M. and S. Ng (2020): "FRED-QD: A quarterly database for macroeconomic research," Tech. rep., National Bureau of Economic Research.

McGrory, C. A. and D. Titterington (2007): "Variational approximations in Bayesian model selection for finite mixture distributions," *Computational Statistics & Data Analysis*, 51, 5352–5367.

Newton, M. A. and A. E. Raftery (1994): "Approximate Bayesian inference with the weighted likelihood bootstrap," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 56, 3–26.

Penny, W. D. (2012): "Comparing dynamic causal models using AIC, BIC and free energy," *Neuroimage*, 59, 319–330.

Perrakis, K., I. Ntzoufras, and E. G. Tsionas (2014): "On the use of marginal posteriors in marginal likelihood estimation via importance sampling," *Computational Statistics & Data Analysis*, 77, 54–69.

Prüser, J. and F. Huber (2024): "Nonlinearities in macroeconomic tail risk through the lens of big data quantile regressions," *Journal of Applied Econometrics*, 39, 269–291.

Quiroz, M., D. J. Nott, and R. Kohn (2023): "Gaussian variational approximations for high-dimensional state space models," *Bayesian Analysis*, 18, 989–1016.

Ranganath, R., D. Tran, and D. Blei (2016): "Hierarchical variational models," in *International Conference on Machine Learning*, PMLR, 324–333.

Rezende, D. and S. Mohamed (2015): "Variational inference with normalizing flows," in *International Conference on Machine Learning*, PMLR, 1530–1538.

Schorfheide, F. and D. Song (2021): "Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic," Tech. rep., National Bureau of Economic Research.

Smith, M. S., R. Loaiza-Maya, and D. J. Nott (2020): "High-dimensional copula variational approximation through transformation," *Journal of Computational and Graphical Statistics*, 29, 729–743.

Stock, J. H. and M. W. Watson (2016): "Core inflation and trend inflation," *Review of Economics and Statistics*, 98, 770–784.

Tsionas, M. G., M. Izzeldin, and L. Trapani (2022): "Estimation of large dimensional time varying VARs using copulas," *European Economic Review*, 141, 103952.

A Online Appendix: Estimation Details

In this appendix, we provide the details of the variational Bayes approximation of the posterior distribution for the reduced-form VARs with stochastic volatility and outlier component. As mentioned in Section 2, the hierarchical Minnesota prior is applied.

A.1 Reduced Form Large VARs with Stochastic Volatility

Recall the model

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{A}_1\mathbf{y}_{t-1} + ... + \mathbf{A}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t),$$

where a Cholesky stochastic volatility is incorporated, i.e., $\boldsymbol{\Sigma}_t^{-1} = \mathbf{B}_0'\mathbf{D}_t^{-1}\mathbf{B}_0$, $\mathbf{D}_t = \mathrm{diag}(\mathrm{e}^{h_{1,t}}, ..., \mathrm{e}^{h_{n,t}})$, and $\mathbf{B}_0$ is an $n \times n$ lower triangular matrix with ones on the diagonal. Each element of $\mathbf{h}_t = (h_{1,t}, ..., h_{n,t})'$ follows a random walk process

$$h_{i,t} = h_{i,t-1} + u_{i,t}^h, \quad u_{i,t}^h \sim \mathcal{N}(0, \sigma_i^2)$$

for $t = 1, 2, ..., T$, and the initial condition $h_{i,0}$ is treated as an unknown parameter to estimate. Let $\mathbf{h}_0 = (h_{1,0}, ..., h_{n,0})'$.

Let $\boldsymbol{\alpha}_i$ denotes the $k \times 1$ vector that consists of the intercept and VAR coefficients in the i-th equation, and $\boldsymbol{\beta}_i$ represents the $(i-1) \times 1$ vector of free elements in the i-th row of the impact matrix $\mathbf{B}_0$. Then, the parameters for the i-th equation are $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i$, $h_{i,0}$ and σ_i^2. We adopt a hierarchical Minnesota prior:

$$\begin{aligned}
&(\boldsymbol{\alpha}_i|\boldsymbol{\kappa}) \sim \mathcal{N}(\boldsymbol{\alpha}_{0,i}, \mathbf{V}_{\boldsymbol{\alpha}_i}), \quad (\boldsymbol{\beta}_i|\boldsymbol{\kappa}) \sim \mathcal{N}(\boldsymbol{\beta}_{0.i}, \mathbf{V}_{\boldsymbol{\beta}_i}) \\
&h_{i,0} \sim \mathcal{N}(0, V_{h_{i,0}}), \quad \sigma_i^2 \sim \mathcal{IG}(\nu_i, S_i),
\end{aligned}$$

where $\boldsymbol{\kappa}$ is a vector of hyperparameters that is described in more detail below. We set the prior mean for $\boldsymbol{\alpha}_{0,i}$ to 0, in order to shrink the VAR coefficients to zero. For $\mathbf{V}_{\boldsymbol{\alpha}_i}$, we assume it to be diagonal with the k-th diagonal element $V_{\boldsymbol{\alpha}_i,kk}$ set to be

$$V_{\boldsymbol{\alpha}_i,kk} = \begin{cases} \frac{\kappa_1}{l^2}, & \text{for the coefficient on the } l\text{-th lag of variable } i, \\ \frac{\kappa_2 s_i^2}{l^2 s_j^2} & \text{for the coefficient on the } l\text{-th lag of variable } j, j \neq i, \\ 100 s_i^2, & \text{for the intercept,} \end{cases}$$

where s_r^2 denotes the sample variance of the residuals from an AR(4) model for the variable r for $r = 1, ..., n$. In addition, we set the prior mean vector $\boldsymbol{\beta}_{0,i}$ to be zero to shrink the impact matrix to the identity matrix. The prior covariance matrix $\mathbf{V}_{\boldsymbol{\beta}_i}$ is assumed to be diagonal, where the j-th diagonal element is set to be $\kappa_3 s_i^2 / s_j^2$. Finally, we treat the shrinkage hyperparameters $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3)'$ as unknown parameters to be estimated with hierarchical gamma priors $\kappa_i \sim \mathcal{G}(c_{j,1}, c_{j,2})$, $j = 1, 2, 3$.

Let $\mathbf{y}_i = (y_{i,1}, ..., y_{i,T})'$ denote the vector of observed values for the i-th variable for $i = 1, ..., n$. Similarly we define $\mathbf{h}_i = (h_{i,1}, ..., h_{i,T})'$. Next, we stack $\mathbf{y} = (\mathbf{y}_1', ..., \mathbf{y}_n')'$, $\mathbf{h} = (\mathbf{h}_1', ..., \mathbf{h}_n')'$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', ..., \boldsymbol{\alpha}_n')'$. Similarly, we define $\boldsymbol{\beta}_i = (\beta_1, ..., \beta_{i-1})'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_n')'$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, ..., \sigma_n^2)'$.

In addition, in order to differentiate between the expectation of the inverse of a variable and the inverse of the expectation of a variable, we use $\overline{x^{-1}}$ to denote the former, and $\bar{x}^{-1}$ for the latter.

Now, we approximate $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa} | \mathbf{y})$ using the product of densities

$$\begin{aligned} q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}) &= q(\boldsymbol{\kappa}) \prod_{i=1}^n q(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \mathbf{h}_i, h_{i,0}, \sigma_i^2) \\ &= q(\boldsymbol{\kappa}) \prod_{i=1}^n q(\boldsymbol{\alpha}_i) q(\boldsymbol{\beta}_i), q(\mathbf{h}_i) q(h_{i,0}) q(\sigma_i^2) \end{aligned}$$

In what follows, we derive the explicit forms of each of these optimal marginal densities and their associated parameters.

The Optimal Density $q^*_{\boldsymbol{\alpha}_i}$

The optimal density $q^*_{\boldsymbol{\alpha}_i}$ has the form

$$q^*_{\boldsymbol{\alpha}_i} \propto \exp\left\{ \mathbb{E}_{-\boldsymbol{\alpha}_i} \left[\log p(\boldsymbol{\alpha}_i | \mathbf{y}, \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}) \right] \right\},$$

where the expectation is taken with respect to the marginal density $q_{-\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa})$. $\boldsymbol{\alpha}_{-i}$ denotes the intercepts and VAR coefficients except for those in the i-th equation.

First, define $\mathbf{A}_{i=0}$ to be a $k \times n$ matrix that has exactly the same elements as $\mathbf{A} = (\mathbf{a}_0, \mathbf{A}_1, ..., \mathbf{A}_p)'$ except for the i-th element, which is set to be zero, i.e., $\mathbf{A}_{i=0} = (\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_{i+1}, ..., \boldsymbol{\alpha}_n)$. Next, denote $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-p})'$, and $\mathbf{B}_{0,1:n,i}$ to be the i-th column of $\mathbf{B}_0$. Further, let $\mathbf{z}^i$ denote $\mathrm{vec}((\mathbf{Y} - \mathbf{X}\mathbf{A}_{i=0})\mathbf{B}'_0)$, $\mathbf{W}^i$ denote $\mathbf{B}_{0,1:n,i} \otimes \mathbf{X}$, $\mathbf{D}$ denote $\mathrm{diag}(\mathbf{D}_1, ..., \mathbf{D}_n)$, with $\mathbf{D}_i = \mathrm{diag}(\mathrm{e}^{h_{i,1}}, ..., \mathrm{e}^{h_{i,T}})$.

From Chan (2023), we have

$$(\boldsymbol{\alpha}_i | \mathbf{y}, \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}) \sim \mathcal{N}(\widehat{\boldsymbol{\alpha}}_i, \mathbf{K}_{\boldsymbol{\alpha}_i}^{-1}),$$

where $\boldsymbol{\alpha}_{-i} = (\boldsymbol{\alpha}'_1, ..., \boldsymbol{\alpha}'_{i-1}, \boldsymbol{\alpha}'_{i+1}, ..., \boldsymbol{\alpha}'_n)'$,

$$\mathbf{K}_{\boldsymbol{\alpha}_i} = \mathbf{V}_{\boldsymbol{\alpha}_i}^{-1} + \mathbf{W}^{i\prime}\mathbf{D}^{-1}\mathbf{W}^i, \quad \widehat{\boldsymbol{\alpha}}_i = \mathbf{K}_{\boldsymbol{\alpha}_i}^{-1}(\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}\boldsymbol{\alpha}_{0,i} + \mathbf{W}^{i\prime}\mathbf{D}^{-1}\mathbf{z}^i).$$

The log-density is therefore

$$\log p(\boldsymbol{\alpha}_i | \mathbf{y}, \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}) = c_{\boldsymbol{\alpha}_i} - \frac{1}{2}\boldsymbol{\alpha}'_i\mathbf{K}_{\boldsymbol{\alpha}_i}\boldsymbol{\alpha}_i + \boldsymbol{\alpha}'_i(\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}\boldsymbol{\alpha}_{0,i} + \mathbf{W}^{i\prime}\mathbf{D}^{-1}\mathbf{z}^i), \quad \text{(A.13)}$$

where $c_{\boldsymbol{\alpha}_i}$ is a term not dependent on $\boldsymbol{\alpha}_i$. After taking the expectation, we essentially get an approximating density $\mathcal{N}(\bar{\widehat{\boldsymbol{\alpha}}}_i, \widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1})$, where

$$\begin{aligned}
\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i} &= \mathbb{E}_{-\boldsymbol{\alpha}_i}[\mathbf{K}_{\boldsymbol{\alpha}_i}] = \mathbb{E}_{\boldsymbol{\alpha}_i}\left[\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1} + \mathbf{W}^{i\prime}\mathbf{D}^{-1}\mathbf{W}^i\right] \\
&= \mathbb{E}_{-\boldsymbol{\alpha}_i}\left[\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1} + (\mathbf{B}'_{0,1:n,i} \otimes \mathbf{X}')\mathbf{D}^{-1}(\mathbf{B}_{0,1:n,i} \otimes \mathbf{X})\right] \\
&= \overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}} + \sum_{j=1}^{n}\left(\bar{B}_{0,j,i}^2 + V_{\bar{B}_{0,j,i}}\mathbf{X}'\bar{\mathbf{D}}_j^{-1}\mathbf{X}\right) \\
&= \overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}} + (\bar{\mathbf{B}}'_{0,1:n,i} \otimes \mathbf{X}')\overline{\mathbf{D}^{-1}}(\bar{\mathbf{B}}_{0,1:n,i} \otimes \mathbf{X}) + (\boldsymbol{\sigma}'_{\bar{\mathbf{B}}_{0,1:n,i}} \otimes \mathbf{X}')\overline{\mathbf{D}^{-1}}(\boldsymbol{\sigma}_{\bar{\mathbf{B}}_{0,1:n,i}} \otimes \mathbf{X})),
\end{aligned}$$

in which $\boldsymbol{\sigma}'_{\bar{\mathbf{B}}_{0,1:n,i}}$ is a $n \times 1$ vector stacked by $V_{\bar{B}_{0,j,i}}^{1/2}$, $j = 1, ..., n$, $\bar{B}_{0,j,i}$ and $V_{\bar{B}_{0,j,i}}$ are the corresponding mean and variance of the approximating distribution,

$$\begin{aligned}\bar{\widehat{\boldsymbol{\alpha}}}_i &= \widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1}\mathbb{E}_{-\boldsymbol{\alpha}_i}\left[\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}\boldsymbol{\alpha}_{0,i} + \mathbf{W}^{i'}\mathbf{D}^{-1}\mathbf{z}^i\right]\\ &= \widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1}\mathbb{E}_{-\boldsymbol{\alpha}_i}\left[\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}\boldsymbol{\alpha}_{0,i} + (\mathbf{B}'_{0,1:n,i}\otimes\mathbf{X}')\mathbf{D}^{-1}\mathrm{vec}((\mathbf{Y}-\mathbf{X}\mathbf{A}_{i=0})\mathbf{B}'_0)\right]\\ &= \widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1}\left\{\overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}}\boldsymbol{\alpha}_{0,i} + \mathbb{E}_{-\boldsymbol{\alpha}_i}\left[(\mathbf{B}'_{0,1:n,i}\otimes\mathbf{X}')\mathbf{D}^{-1}\mathrm{vec}((\mathbf{Y}-\mathbf{X}\mathbf{A}_{i=0})\mathbf{B}'_0)\right]\right\},\end{aligned}$$

Note that

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\alpha}_i}\left[(\mathbf{B}'_{0,1:n,i}\otimes\mathbf{X}')\mathbf{D}^{-1}\mathrm{vec}((\mathbf{Y}-\mathbf{X}\mathbf{A}_{i=0})\mathbf{B}'_0)\right]\\ &\quad= (\bar{\mathbf{B}}'_{0,1:n,i}\otimes\mathbf{X}')\overline{\mathbf{D}^{-1}}\mathrm{vec}((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}}_{i=0})\bar{\mathbf{B}}'_0) + (\mathbb{1}'_n\otimes\mathbf{X}')\overline{\mathbf{D}^{-1}}\mathrm{vec}((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}}_{i=0})\boldsymbol{\Omega}^{i'}),\end{aligned}$$

where $\boldsymbol{\Omega}^i = (\boldsymbol{\Omega}_1^i, ..., \boldsymbol{\Omega}_n^i)'$, in which given m, $\boldsymbol{\Omega}_m^i$ is a $n\times 1$ vector stores the covariance between $B_{0,m,i}$ and $B_{0,m,j}$, $j = 1, ..., n$, i.e., $\boldsymbol{\Omega}_{m,j}^i = \mathrm{Cov}(B_{0,m,i}, B_{0,m,j})$.

In addition, we have

$$\bar{V}_{\boldsymbol{\alpha}_i,kk}^{-1} = \begin{cases} l^2\overline{\kappa_1^{-1}}, & \text{for the coefficient on the } l\text{-th lag of variable } i,\\ \frac{l^2 s_j^2\overline{\kappa_2^{-1}}}{s_i^2} & \text{for the coefficient on the } l\text{-th lag of variable } j, j\neq i,\\ \frac{1}{100 s_i^2}, & \text{for the intercept,}\end{cases}$$

where $\overline{\kappa_1^{-1}} = \mathbb{E}_{q^*(\kappa_1)}[\kappa_1^{-1}]$, and $\overline{\kappa_2^{-1}} = \mathbb{E}_{q^*(\kappa_2)}[\kappa_2^{-1}]$.

The Optimal Density $q^*_{\boldsymbol{\beta}_i}$

The optimal density $q^*_{\boldsymbol{\beta}_i}$ has the form

$$q^*_{\boldsymbol{\beta}_i} \propto \exp\left\{\mathbb{E}_{-\boldsymbol{\beta}_i}\left[\log p(\boldsymbol{\beta}_i|\mathbf{y}, \boldsymbol{\alpha}, \mathbf{h}_i, \boldsymbol{\kappa})\right]\right\}$$

where the expectation is taken with respect to the marginal density $q_{-\boldsymbol{\beta}_i}(\boldsymbol{\alpha}, \mathbf{h}_i, \boldsymbol{\kappa})$. From Chan (2023), we know

$$(\boldsymbol{\beta}_i|\mathbf{y}, \boldsymbol{\alpha}, \mathbf{h}_i, \boldsymbol{\kappa}) \sim \mathcal{N}\left(\widehat{\boldsymbol{\beta}}_i, \mathbf{K}_{\boldsymbol{\beta}_i}^{-1}\right),$$

where

$$\mathbf{K}_{\boldsymbol{\beta}_i} = \mathbf{V}_{\boldsymbol{\beta}_i}^{-1} + \mathbf{E}'_i\mathbf{D}_i^{-1}\mathbf{E}_i, \quad \widehat{\boldsymbol{\beta}}_i = \mathbf{K}_{\boldsymbol{\beta}_i}^{-1}\left(\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}\boldsymbol{\beta}_{0,i} + \mathbf{E}'_i\mathbf{D}_i^{-1}\boldsymbol{\varepsilon}_i\right), \tag{A.14}$$

in which $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})'$, and $\mathbf{E}_i = (\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_{i-1})$, $\boldsymbol{\varepsilon}_i = \mathbf{E}_i\boldsymbol{\beta}_i + \boldsymbol{\eta}_i$, and $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_i)$.

We can get an approximating density $\mathcal{N}\left(\bar{\bar{\boldsymbol{\beta}}}_i, \widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}^{-1}\right)$, where

$$\begin{aligned}\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i} &= \mathbb{E}_{-\boldsymbol{\beta}_i}\left[\mathbf{V}_{\boldsymbol{\beta}_i}^{-1} + \mathbf{E}_i'\mathbf{D}_i^{-1}\mathbf{E}_i\right] \\ &= \overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}} + \mathbb{E}_{-\boldsymbol{\beta}_i}[\mathbf{E}_i'\mathbf{D}_i^{-1}\mathbf{E}_i],\end{aligned}$$

where

$$\mathbb{E}_{-\boldsymbol{\beta}_i}\left[\mathbf{E}_i'\mathbf{D}_i^{-1}\mathbf{E}_i\right] = \mathbb{E}_{-\boldsymbol{\beta}_i}\left[(\mathbf{S}_{i-1}'\mathbf{Y}' - \mathbf{S}_{i-1}'\mathbf{A}'\mathbf{X}')\mathbf{D}_i^{-1}(\mathbf{Y}\mathbf{S}_{i-1} - \mathbf{X}\mathbf{A}\mathbf{S}_{i-1})\right],$$

in which

$$\mathbb{E}_{-\boldsymbol{\beta}_i}[\mathbf{S}_{i-1}'\mathbf{A}'\mathbf{X}'\mathbf{D}_i^{-1}\mathbf{X}\mathbf{A}\mathbf{S}_{i-1}] = \mathbf{S}_{i-1}'\bar{\mathbf{A}}'\mathbf{X}'\bar{\mathbf{D}}_{i,\cdot}^{-1}\mathbf{X}\bar{\mathbf{A}}\mathbf{S}_{i-1} + \mathbf{S}_{i-1}'\mathbf{G}_i\mathbf{S}_{i-1}.$$

Note that $\mathbf{S}_{i=1} = [\mathbf{I}_{i-1}, \mathbf{O}_{(i-1)\times(n-i+1)}]'$, which is a selection matrix of dimension $n \times (i-1)$. $\mathbf{G}_i = \text{diag}\left(\text{tr}\left(\mathbf{X}'\bar{\mathbf{D}}_{i,\cdot}^{-1}\mathbf{X}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_1}^{-1}\right), ...\text{tr}\left(\mathbf{X}'\bar{\mathbf{D}}_{i,\cdot}^{-1}\mathbf{X}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_n}^{-1}\right)\right)$ is a block-diagonal matrix. In addition, $\bar{\mathbf{V}}_{\boldsymbol{\beta}_{i,j}}^{-1} = s_j^2\overline{\kappa_3^{-1}}/s_i^2$, where $\overline{\kappa_3^{-1}} = \mathbb{E}_{q^*(\kappa_3)}[\kappa_3^{-1}]$. Therefore, we have

$$\begin{aligned}\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i} &= \overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}} + \bar{\mathbf{E}}_i'\overline{\mathbf{D}_i^{-1}}\bar{\mathbf{E}}_i + \mathbf{S}_{i-1}'\mathbf{G}_i\mathbf{S}_{i-1} \\ \bar{\bar{\boldsymbol{\beta}}}_i &= \widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}^{-1}\left(\overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}}\boldsymbol{\beta}_{0,i} + \bar{\mathbf{E}}_i\overline{\mathbf{D}_i^{-1}}\bar{\boldsymbol{\varepsilon}}_i\right),\end{aligned}$$

where $\bar{\boldsymbol{\varepsilon}}_i = \left(\mathbf{Y} - \mathbf{X}\bar{\mathbf{A}}\right)\mathbf{e}_i$, and $\mathbf{e}_i$ is a selection matrix of dimension $n \times 1$, which is a unit vector with its i-th element being 1.

The Optimal Density $q_{\boldsymbol{\kappa}}^*$

The optimal density $q_{\boldsymbol{\kappa}}^* = q_{\kappa_1}^* q_{\kappa_2}^* q_{\kappa_3}^*$ has the form

$$\begin{aligned}q_{\kappa_2}^* &\propto \exp\left\{\mathbb{E}_{-\kappa_1}\left[\log p(\kappa_1|\boldsymbol{\alpha})\right]\right\}, \\ q_{\kappa_2}^* &\propto \exp\left\{\mathbb{E}_{-\kappa_1}\left[\log p(\kappa_2|\boldsymbol{\alpha})\right]\right\}, \\ q_{\kappa_3}^* &\propto \exp\left\{\mathbb{E}_{-\kappa_3}\left[\log p(\kappa_3|\boldsymbol{\beta})\right]\right\}.\end{aligned}$$

Define the index set S_{κ_1} that collects all the indexes (i, j) such that $\alpha_{i,j}$ is a coefficient associated with an own lag. That is, $S_{\kappa_1} = \{(i, j) : \alpha_{i,j} \text{ is a coefficient associated with an own lag}\}$. Similarly, define S_{κ_2} as the set that collects all the indexes (i, j) such that $\alpha_{i,j}$ is a coefficient associated with a lag of other variables. Lastly, define $S_{\kappa_3} = \{(i,j) : i = 2, ..., n, j = 1, ..., i-1\}$. From Chan (2023), we know

$$\begin{aligned}
(\kappa_1|\boldsymbol{\alpha}) &\sim \mathcal{GIG}\left(c_{1,1} - \frac{np}{2}, 2c_{1,2}, \sum_{(i,j)\in S_{\kappa_1}} \frac{(\alpha_{i,j} - \alpha_{0,i,j})^2}{C_{i,j}}\right), \\
(\kappa_2|\boldsymbol{\alpha}) &\sim \mathcal{GIG}\left(c_{2,1} - \frac{(n-1)np}{2}, 2c_{2,2}, \sum_{(i,j)\in S_{\kappa_2}} \frac{(\alpha_{i,j} - \alpha_{0,i,j})^2}{C_{i,j}}\right), \\
(\kappa_3|\boldsymbol{\beta}) &\sim \mathcal{GIG}\left(c_{3,1} - \frac{n(n-1)}{4}, 2c_{3,2}, \sum_{(i,j)\in S_{\kappa_3}} \frac{(\beta_{i,j} - \beta_{0,i,j})^2}{\widetilde{C}_{i,j}}\right).
\end{aligned}$$

In particular, we have

$$p(\kappa_1|\boldsymbol{\alpha}) \propto \kappa_1^{c_{1,1}-\frac{np}{2}-1} \exp\left\{-\frac{1}{2}\left[2c_{1,2}\kappa_1 + \kappa_1^{-1} \sum_{(i,j)\in S_{\kappa_1}} \frac{(\alpha_{i,j} - \alpha_{0,i,j})^2}{C_{i,j}}\right]\right\}$$

So that

$$\log p(\kappa_1|\boldsymbol{\alpha}) = c_{\kappa_1} + \Big(c_{1,1} - \frac{np}{2} - 1\Big)\log(\kappa_1) - \frac{1}{2}\left[2c_{1,2}\kappa_1 + \kappa_1^{-1} \sum_{(i,j)\in S_{\kappa_1}} \frac{(\alpha_{i,j} - \alpha_{0,i,j})^2}{C_{i,j}}\right],$$

where c_{κ_1} is the part that is independent of κ_1.

Taking the expectation regarding the parameters other than κ_1, and then taking the exponential form, we obtain $q^*(\kappa_1)$

$$\begin{aligned}
q^*(\kappa_1) &= \exp\{\mathbb{E}_{-\kappa_1}[\log p(\kappa_1|\boldsymbol{\alpha})]\} \\
&\propto \kappa_1^{c_{1,1}-\frac{np}{2}-1} \exp\left\{-\frac{1}{2}\left[2c_{1,2}\kappa_1 + \kappa_1^{-1} \sum_{(i,j)\in S_{\kappa_1}} \frac{\bar{\widehat{\alpha}}_{i,j}^2 + \bar{\sigma}_{\alpha_{i,j}}^2 - 2\alpha_{0.i,j}\bar{\widehat{\alpha}}_{i,j} + \alpha_{0,i,j}^2}{C_{i,j}}\right]\right\},
\end{aligned}$$

which is the kernal of a generlized-inverse-Gaussian distribution $\mathcal{GIG}(v_{\kappa_1}, a_{\kappa_1}, b_{\kappa_1})$, where

$$v_{\kappa_1} = c_{1,1} - \frac{np}{2}, \quad a_{\kappa_1} = 2c_{1,2}, \quad b_{\kappa_1} = \sum_{(i,j)\in S_{\kappa_1}} \frac{\bar{\widehat{\alpha}}_{i,j}^2 + \bar{\sigma}_{\alpha_{i,j}}^2 - 2\alpha_{0.i,j}\bar{\widehat{\alpha}}_{i,j} + \alpha_{0,i,j}^2}{C_{i,j}}.$$

Similarly, we can get the approximating densities for κ_2 and κ_3: $\mathcal{GIG}(v_{\kappa_r}, a_{\kappa_r}, b_{\kappa_r})$, $r = 2, 3$, where

$$v_{\kappa_2} = c_{2,1} - \frac{(n-1)np}{2}, \quad a_{\kappa_2} = 2c_{2,2}, \quad b_{\kappa_2} = \sum_{(i,j)\in S_{\kappa_2}} \frac{\bar{\widehat{\alpha}}_{i,j}^2 + \bar{\sigma}_{\alpha_{i,j}}^2 - 2\alpha_{0.i,j}\bar{\widehat{\alpha}}_{i,j} + \alpha_{0,i,j}^2}{C_{i,j}},$$

$$v_{\kappa_3} = c_{3,1} - \frac{n(n-1)}{4}, \quad a_{\kappa_3} = 2c_{3,2}, \quad b_{\kappa_3} = \sum_{(i,j)\in S_{\kappa_3}} \frac{\bar{\widehat{\beta}}_{i,j}^2 + \bar{\sigma}_{\beta_{i,j}}^2 - 2\beta_{0.i,j}\bar{\widehat{\beta}}_{i,j} + \beta_{0,i,j}^2}{\widetilde{C}_{i,j}},$$

where $\bar{\sigma}_{\alpha_{i,j}}^2$ and $\bar{\sigma}_{\beta_{i,j}}^2$ are the corresponding element in $\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1}$ and $\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}^{-1}$.

It is useful to know that if a random variable x follows a $\mathcal{GIG}$ distribution, it has the following properties

$$\begin{aligned}
\mathbb{E}[x] &= \sqrt{\frac{b}{a}} \frac{K_{v+1}(\sqrt{ab})}{K_v(\sqrt{ab})} \\
\mathbb{E}[1/x] &= \sqrt{\frac{a}{b}} \frac{K_{v+1}(\sqrt{ab})}{K_v(\sqrt{ab})} - \frac{2v}{b} \\
\mathbb{E}[\log x] &= \log\left(\sqrt{\frac{b}{a}}\right) + \frac{\partial \log K_v(\sqrt{ab})}{\partial v},
\end{aligned}$$

where K_v is a modified Bessel function of the second kind. Note that there is no analytical solution for $\frac{\partial \log K_v(\sqrt{ab})}{\partial v}$. There are several ways to obtain an approximation for $\mathbb{E}[\log x]$. For example, one can compute it using numerical integration. A pitfall of using numerical integration is that a proper sequence of support for x needs to be specified in the beginning. This sequence usually is an arithmetic sequence and a common difference should also be set. In practice, mis-specifying either the sequence length and the common difference is prone to induce computational instability. We found that a less costly and also relatively accurate way is randomly drawing a large sample of x from the $\mathcal{GIG}$ distribution, and taking the average.

Another computational issue is that when ν or $\sqrt{ab}$ are extreme values, for example, $n = 180$, $\nu = -64439$ and $\sqrt{ab} = 354.4$ in our application, functions from software such as MATLAB would give infinity as an answer when we are trying to calculate $\log K_\nu(\sqrt{ab})$. To solve this problem, we use the forward recursion algorithm proposed by Cuingnet (2023) (in Equation (23) and (24)) to compute the logarithm of the modified Bessel function of the second kind.

The Optimal Density $q^*_{h_{i,0}}$

Next, we derive the optimal density $q^*_{h_{i,0}}$, which takes the form

$$q^*_{h_{i,0}} \propto \exp\{\mathbb{E}_{h_{i,0}}[\log p(h_{i,0}|\mathbf{h}_i, \sigma_i^2]\},$$

where the expectation is taken with respect to the marginal density $q_{-h_{i,0}}(\sigma_i^2, \mathbf{h}_i) = q_{\sigma_i^2}(\sigma_i^2)q_{\mathbf{h}_i}(\mathbf{h}_i)$.

We have

$$\log p(h_{i,0}|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}_i, \sigma_i^2) = \log p(h_{i,0}|\mathbf{h}_i, \sigma_i^2) = c_{h_{i,0}} - \frac{1}{2\sigma_i^2}(h_{i,1} - h_{i,0})^2 - \frac{1}{2V_{h_{i,0}}}h_{i,0}^2,$$

where $c_{h_{i,0}}$ is a constant independent of $h_{i,0}$. Taking the expectation with respect to the marginal density $q_{-h_{i,0}}$, we have

$$\mathbb{E}_{-h_{i,0}}[\log p(h_{i,0}|\mathbf{h}_i, \sigma_i^2)] = c_{h_{i,0}} - \frac{1}{2}\mathbb{E}_{\sigma_i^2}\left[\frac{1}{\sigma_i^2}\right]\left[(\widehat{h}_{i,1} - h_{i,0})^2 + \widehat{d}_{i,1}\right] - \frac{1}{2V_{h_{i,0}}}h_{i,0}^2,$$

where $\widehat{d}_{i,1}$ is the first diagonal element of $\widehat{\mathbf{K}}_{\mathbf{h}_i}^{-1}$ and the expectation $\mathbb{E}_{\sigma_i^2}$ is taken with respect to the density $q_{\sigma_i^2}(\sigma_i^2)$ - this expectation can be computed analytically as shown in the next subsection. Finally, using standard linear regression results, one can show that $q^*_{h_{i,0}}$ is a normal distribution: $\mathcal{N}(\widehat{h}_{h_{i,0}}, \widehat{K}^{-1}_{h_{i,0}})$, where

$$\widehat{K}^{-1}_{h_{i,0}} = V^{-1}_{h_{i,0}} + \mathbb{E}_{\sigma_i^2}\left[\frac{1}{\sigma_i^2}\right], \quad \widehat{h}_{i,0} = \widehat{K}^{-1}_{h_{i,0}}\mathbb{E}_{\sigma_i^2}\left[\frac{1}{\sigma_i^2}\right]\widehat{h}_{i,1}.$$

The Optimal Density $q^*_{\sigma_i^2}$

The kernel of the optimal density $q^*_{\sigma_i^2}$ is given by

$$q^*_{\sigma_i^2} \propto \exp\left\{\mathbb{E}_{-\sigma_i^2}[\log p(\sigma_i^2|\mathbf{h}_i, h_{i,0})]\right\},$$

where the expectation is taken with respect to the marginal density $q_{-\sigma_i^2}(\mathbf{h}_i, h_{i,0}) = q_{h_{i,0}}(h_{i,0})q_{\mathbf{h}_i}(\mathbf{h}_i)$. To derive an explicit expression for $q^*_{\sigma_i^2}$, first note that

$$\log p(\sigma_i^2|\mathbf{h}_i, h_{i,0}) = c_{\sigma_i^2} - \frac{T}{2}\log\sigma_i^2 - \frac{1}{2\sigma_i^2}(\mathbf{h}_i - h_{i,0}\mathbb{1}_T)'\mathbf{H}'\mathbf{H}(\mathbf{h}_i - h_{i,0}\mathbb{1}_T) - \nu_i\log\sigma_i^2 - \frac{S_i}{\sigma_i^2}$$

where $c_{\sigma_i^2}$ is a constant not dependent on σ_i^2. Taking expectation with respect to the marginal density $q_{-\sigma_i^2}$ gives

$$\mathbb{E}_{-\sigma_i^2}[\log p(\sigma_i^2|\mathbf{h}_i, h_{i,0})] = c_{\sigma_i^2} - \left(\nu_i + \frac{T}{2}\right)\log \sigma_i^2 - \frac{S_i}{\sigma_i^2}$$
$$- \frac{1}{2\sigma_i^2}\left[(\widehat{\mathbf{h}}_i - \widehat{h}_{i,0})'\mathbf{H}'\mathbf{H}(\widehat{\mathbf{h}}_i - \widehat{h}_{i,0}) + \mathrm{tr}(\mathbf{H}'\mathbf{H}\widehat{\mathbf{K}}_{\mathbf{h}_i}^{-1}) + \widehat{K}_{h_{i,0}}^{-1}\right].$$

It is clear that on the right-hand side, it is the kernel of an inverse-gamma distribution: $\mathcal{IG}(\widehat{\nu}_i, \widehat{S}_i)$, where

$$\widehat{\nu}_i = \nu_i + \frac{T}{2}, \quad \widehat{S}_i = S_i + \frac{1}{2}\left[(\widehat{\mathbf{h}}_i - \widehat{h}_{i,0})'\mathbf{H}'\mathbf{H}(\widehat{\mathbf{h}}_i - \widehat{h}_{i,0}) + \mathrm{tr}(\mathbf{H}'\mathbf{H}\widehat{\mathbf{K}}_{\mathbf{h}_i}^{-1}) + \widehat{K}_{h_{i,0}}^{-1}\right].$$

It is also important to know that the expectation of $1/\sigma_i^2$ can be obtained analytically as

$$\mathbb{E}_{\sigma_i^2}\left[\frac{1}{\sigma_i^2}\right] = \frac{\widehat{\nu}_i}{\widehat{S}_i}.$$

The Optimal Density $q^*_{\mathbf{h}_i}$

The log of the conditional distribution of $\mathbf{h}_i$ is as follows

$$\log p(\mathbf{h}_i|\mathbf{y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, h_{i,0}, \sigma_i^2) = c_{\mathbf{h}_i} - \frac{1}{2}\sum_{t=1}^{T} h_{i,t} - \frac{1}{2}\sum_{t=1}^{T} \mathrm{e}^{-h_{i,t}}\widetilde{\varepsilon}_{i,t}^2 - \frac{1}{2\sigma_i^2}\sum_{t=1}^{T}(h_{i,t} - h_{i,t-1})^2,$$

where $c_{\mathbf{h}_i}$ is a constant not dependent on $\mathbf{h}_i$. Taking the expectation with respect to the marginal density $q_{-\mathbf{h}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta}, h_{i,0}, \sigma_i^2)$ gives

$$\mathbb{E}_{-\mathbf{h}_i}[\log p(\mathbf{h}_i|\mathbf{y}_i, \boldsymbol{\alpha}, \beta, h_{i,0}, \sigma_i^2)] = c_{\mathbf{h}_i} - \frac{1}{2}\sum_{t=1}^{T} h_{i,t} - \frac{1}{2}\sum_{t=1}^{T} \mathrm{e}^{-h_{i,t}}\widehat{s}_t^2$$
$$- \frac{1}{2}\mathbb{E}_{\sigma_i^2}\left[\frac{1}{\sigma_i^2}\right]\left(\sum_{t=2}^{T}(h_{i,t} - h_{i,t-1})^2 + (h_{i,1} - \widehat{h}_{i,0})^2 + \widehat{K}_{h_{i,0}}^{-1}\right),$$

where $\widehat{s}_t^2 = \mathbb{E}_{-\mathbf{h}_i}[\widetilde{\varepsilon}_{i,t}] = \mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\beta}}[(\mathbf{e}_t'(\mathbf{Y} - \mathbf{X}\mathbf{A})\mathbf{B}_{0,i,1:n})^2]$, $\mathbf{e}_t$ is a vector with its t-th element being 1. In addition, we have

$$\widehat{s}_t^2 = (\mathbf{e}_t'(\mathbf{Y} - \mathbf{X}\bar{\mathbf{A}})\bar{\mathbf{B}}_{0,i,1:n})^2 + \bar{\mathbf{B}}_{0,i,1:n}'\widetilde{\mathbf{G}}_t\bar{\mathbf{B}}_{0,i,1:n} + \mathrm{tr}\left((\mathbf{Y} - \mathbf{X}\bar{\mathbf{A}})'\mathbf{e}_t\mathbf{e}_t'(\mathbf{Y} - \mathbf{X}\bar{\mathbf{A}})\widehat{\mathbf{K}}_{\mathbf{B}_{0,i,1:n}^{-1}}\right),$$

where $\widetilde{\mathbf{G}}_t = \text{diag}\left(\text{tr}(\mathbf{X}'\mathbf{e}_t\mathbf{e}_t'\mathbf{X}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_1}^{-1}, ..., \text{tr}(\mathbf{X}'\mathbf{e}_t\mathbf{e}_t'\mathbf{X}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_n}^{-1})\right)$, $\widehat{\mathbf{K}}_{\mathbf{B}_{0,i,1:n}}$ is an $n \times n$ matrix with its first $(i-1) \times (i-1)$ elements being $\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}^{-1}$ and all other elements being zero. Therefore, the log kernel of $\widetilde{q}_{\mathbf{h}_i}^*$ has the following expression:

$$\log \widetilde{q}_{\mathbf{h}_i}^* = c_{\mathbf{h}_i} - \frac{1}{2}\sum_{t=1}^{T} h_{i,t} - \frac{1}{2}\sum_{t=1}^{T} \mathrm{e}^{-h_{i,t}}\widehat{s}_t^2 - \frac{1}{2}\mathbb{E}_{\sigma_i^2}\left[\frac{1}{\sigma_i^2}\right] \times \left(\sum_{t=2}^{T}(h_{i,t} - h_{i,t-1})^2 + (h_{i,1} - \widehat{h}_{i,0})^2 + \widehat{K}_{h_{i,0}}^{-1}\right).$$

Similar to the approach used in Chan and Yu (2022) for their VARSV model, we locate the optimal Gaussian density $q_{\mathbf{h}_i}^*$ by finding the mode of $\log \widetilde{q}_{\mathbf{h}_i}^*$ and employ it as the mean, and use the inverse negative Hessian of $\log \widetilde{q}_{\mathbf{h}_i}^*$ evaluated at the mode as the variance.

The Variational Lower Bound

Next, we derive the variational lower bound $\underline{p}(\mathbf{y}; q)$. To that end, we first compute the log ratio of the joint posterior density and the variational approximation:

$$
\begin{aligned}
&\log\left[\frac{p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},\mathbf{h}_0,\boldsymbol{\sigma}^2,\boldsymbol{\kappa})}{q(\boldsymbol{\kappa})\prod_{i=1}^n q(\boldsymbol{\alpha}_i)q(\boldsymbol{\beta}_i)q(\mathbf{h}_i)q(\mathbf{h}_{0,i})q(\sigma_i^2)}\right]\\
=&c_{\boldsymbol{\kappa}}+\sum_{i=1}^n\Bigg[c_i-\frac{1}{2}\mathbb{1}_T'\mathbf{h}_i-\frac{1}{2}((\mathbf{Y}-\mathbf{XA})\mathbf{B}_{0,i,1:n})'\mathbf{D}_i^{-1}((\mathbf{Y}-\mathbf{XA})\mathbf{B}_{0,i,1:n}))\\
&-\frac{T}{2}\log\sigma_i^2-\frac{1}{2\sigma_i^2}(\mathbf{h}_i-h_{i,0}\mathbb{1}_T)'\mathbf{H}'\mathbf{H}(\mathbf{h}_i-h_{i,0}\mathbb{1}_T)\\
&-\frac{1}{2}\log|\mathbf{V}_{\boldsymbol{\beta}_i}|-\frac{1}{2}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{0,i})'\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{0,i})-\frac{1}{2V_{h_{i,0}}}h_{i,0}^2-(\nu_i+1)\log\sigma_i^2-\frac{S_i}{\sigma_i^2}\\
&-\frac{1}{2}\log|\mathbf{V}_{\boldsymbol{\alpha}_i}|-\frac{1}{2}(\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_{0,i})'\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}(\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_{0,i})\Bigg]+\sum_{r=1}^3\Big[(c_{r,1}-1)\log\kappa_r-c_{r,2}\kappa_r\Big]\\
&+\sum_{i=1}^n\Bigg[\frac{1}{2}(\mathbf{h}_i-\widehat{\mathbf{h}}_i)'\widehat{\mathbf{K}}_{\mathbf{h}_i}(\mathbf{h}_i-\widehat{\mathbf{h}}_i)+\frac{1}{2}(\boldsymbol{\alpha}_i-\bar{\widehat{\boldsymbol{\alpha}}}_i)'\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_i-\bar{\widehat{\boldsymbol{\alpha}}}_i)\\
&+\frac{1}{2}(\boldsymbol{\beta}_i-\bar{\widehat{\boldsymbol{\beta}}}_i)'\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}(\boldsymbol{\beta}_i-\bar{\widehat{\boldsymbol{\beta}}}_i)+\frac{\widehat{K}_{h_{i,0}}}{2}(h_{i,0}-\widehat{h}_{i,0})^2+(\widehat{\nu}_i+1)\log\sigma_i^2+\frac{\widehat{S}_i}{\sigma_i^2}\Bigg]\\
&-\sum_{r=1}^3\left[(v_r-1)\log\kappa_r-\frac{a_r\kappa_r+b_r\kappa_r^{-1}}{2}\right],
\end{aligned}
$$

where $c_i = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log V_{h_{i,0}} - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}| + \nu_i\log S_i - \log\Gamma(\nu_i) - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\mathbf{h}_i}| - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}| - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}| - \frac{1}{2}\log\widehat{K}_{h_{i,0}} - \widehat{\nu}_i\log\widehat{S}_i + \log\Gamma(\widehat{\nu}_i)$, and $c_{\boldsymbol{\kappa}} = \sum_{r=1}^3\left[c_{r,1}\log c_{r,2}-\log\Gamma(c_{r,1})\right]-\sum_{r=1}^3\left[\frac{v_r}{2}(\log a_r-\log b_r)-\log(2K_{v_r}\sqrt{a_rb_r})\right]$, $K_{v_r}(\cdot)$ is a modified Bessel function of the second kind. Taking expectation of the above log ratio with respect to q, we obtain the following variational lower bound:

$$
\begin{aligned}
\underline{p}(\mathbf{y};q) &= \mathbb{E}_q\left\{\log\left[\frac{p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},\mathbf{h}_0,\boldsymbol{\sigma}^2,\boldsymbol{\kappa})}{q(\boldsymbol{\kappa})\prod_{i=1}^n q(\boldsymbol{\alpha}_i)q(\boldsymbol{\beta}_i)q(\mathbf{h}_i)q(\mathbf{h}_{0,i})q(\sigma_i^2)}\right]\right\} \\
&= c_{\boldsymbol{\kappa}} + \sum_{i=1}^n \Bigg[c_i - \frac{1}{2}\mathbb{1}_T'\widehat{\mathbf{h}}_i - \frac{1}{2}\left((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\bar{\mathbf{B}}_{0,i,1:n}\right)'\overline{\mathbf{D}_i^{-1}}\left((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\bar{\mathbf{B}}_{0,i,1:n}\right) \\
&\quad - \frac{1}{2}\bar{\mathbf{B}}_{0,i,1:n}'\mathbf{G}_i\bar{\mathbf{B}}_{0,i,1:n} - \frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})'\overline{\mathbf{D}_i^{-1}}(\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\widehat{\mathbf{K}}_{\mathbf{B}_{0,i,1:n}}^{-1}\right) \\
&\quad - \frac{1}{2}\frac{\hat{\nu}_i}{\hat{S}_i}\left[(\widehat{\mathbf{h}}_i - \widehat{h}_{i,0}\mathbb{1}_T)'\mathbf{H}'\mathbf{H}(\widehat{\mathbf{h}}_i - \widehat{h}_{i,0}\mathbb{1}_T) + \mathrm{tr}(\mathbf{H}'\mathbf{H}\widehat{\mathbf{K}}_{\mathbf{h}_i}^{-1}) + \widehat{K}_{h_{i,0}}^{-1}\right] \\
&\quad - \frac{1}{2V_{h_{i,0}}}(\widehat{h}_{i,0} + \widehat{K}_{h_{i,0}}^{-1}) - \frac{S_i\hat{\nu}_i}{\hat{S}_i} + \hat{\nu}_i - \frac{1}{2}\overline{\log|\mathbf{V}_{\boldsymbol{\beta}_i}|} - \frac{1}{2}\overline{\log|\mathbf{V}_{\boldsymbol{\alpha}_i}|} \\
&\quad - \frac{1}{2}(\bar{\bar{\boldsymbol{\beta}}}_i - \boldsymbol{\beta}_{i,0})'\overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}}(\bar{\bar{\boldsymbol{\beta}}}_i - \boldsymbol{\beta}_{i,0}) - \frac{1}{2}\mathrm{tr}\left(\overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}}\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}^{-1}\right) \\
&\quad - \frac{1}{2}(\bar{\bar{\boldsymbol{\alpha}}}_i - \boldsymbol{\alpha}_{i,0})'\overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}}(\bar{\bar{\boldsymbol{\alpha}}}_i - \boldsymbol{\alpha}_{i,0}) - \frac{1}{2}\mathrm{tr}\left(\overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1}\right) + \frac{1}{2}(T + k_i + 1)\Bigg] \\
&\quad + \sum_{r=1}^3\left[(c_{r,1} - v_r)\overline{\log\kappa_r} - \left(c_{r,2} - \frac{1}{2}a_r\right)\bar{\kappa}_r + \frac{1}{2}b_r\overline{\kappa_r^{-1}}\right],
\end{aligned}
\tag{A.15}
$$

where we use numerical methods to compute the mean $\overline{\log\kappa_r}$.

A.2 Reduced-Form Large VARs with Stochastic Volatility and Outlier Component

On the basis of the VARSV model, we discuss three modeling strategies that have been used in literature for outlier adjustment. In the first model, we assign greater importance to outliers that correspond to extreme but rare events. The second strategy characterizes frequent occurrences of small outliers as samples from an inverse-gamma distribution, effectively transforming the Gaussian innovations in the VAR-SV model into t-distributed shocks. The last strategy takes advantage of the known timing of COVID-19 and treats it as a deterministic break in the covariance matrix.

Specification 1: An explicit outlier component (VAR-SVO)
In SVO model, we introduce an outlier parameter that has a discrete mixture representation proposed in Stock and Watson (2016). In specific, the outliers enter the model in a diagonal matrix of scale factors, denoted $\mathbf{O}_t$, with diagonal elements $o_{i,t}$ that are mutually $i.i.d$ over all i and t.With $\mathbf{B}_0$ and $\mathbf{D}_t$ specified as before, the covariance matrix now takes the form:

$$\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{O}_t\mathbf{D}_t\mathbf{O}_t'(\mathbf{B}_0^{-1})'.$$

The outlier parameters $o_{i,t}$ is assumed to have a mixture distribution that distinguishes between regular observations $o_{i,t} = 1$ and outliers with $o_{i,t} \geqslant 2$. The probability that outliers in variable i occur is $p_{\mathbf{o}_i}$. We assume that when the outliers occur, they follow a uniform distribution on $(2, 20)$, i.e., $o_{i,t} \sim \mathcal{U}(2, 20)$. The outlier probability $p_{\mathbf{o}_i}$ is assumed to have a beta prior $\mathcal{B}(a_{p_{\mathbf{o}_i}}, b_{p_{\mathbf{o}_i}})$, where in practice the hyperparameters $a_{p_{\mathbf{o}_i}}$ and $b_{p_{\mathbf{o}_i}}$ are calibrated so that the mean outlier frequency is once every 4 years in quarterly data.

Specification 2: Student-t distributed innovations (VAR-SVt)
The SV-t model expands upon the SV model by incorporating an additional parameter $q_{i,t}$, for $i = 1, ..., n$, $t = 1, ..., T$. In specific, we let the squares of the new parameter have inverse-gamma distribution:

$$q_{i,t}^2 \sim \mathcal{IG}\left(\frac{l_i}{2}, \frac{l_i}{2}\right).$$

Let $\mathbf{Q}_t$ denote the new state matrix, in which the diagonal elements $q_{i,t}$ are mutually $i.i.d.$ over all i and t. With this specification, the covariance matrix of the VAR takes the form

$$\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{Q}_t\mathbf{D}_t\mathbf{Q}_t'(\mathbf{B}_0^{-1})'.$$

Define $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{D}_t^{\frac{1}{2}}\mathbf{Q}_t\mathbf{v}_t$, where $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. It is important to note that the i-th residual $q_{i,t} \cdot v_{i,t}$ (adjusted by $\mathbf{B}_0^{-1}$ and scaling by $\mathbf{D}^{\frac{1}{2}}$) has a student-t distribution with l_i degrees of freedom because $v_{i,t} \sim \mathcal{N}(0,1)$ and $l_i/q_{i,t}^2 \sim \chi_{l_i}^2$.

Specification 3. Common volatility with a deterministic break date (VAR-CVD)

In the VAR-CVD model, the exact timing of the change in volatility during the COVID-19 pandemic is regarded as deterministic, denoted as t^*. The covariance matrix takes the form:

$$\mathbf{\Sigma}_t = s_t^2\mathbf{\Sigma}, \tag{A.16}$$

where s_t, for $t = 1, ..., T$, is latent and to be estimated. When we work with a monthly VAR, since March 2020 was the first month of abnormal data variation, the standard deviation of the March shocks is scaled by an unknown parameter $\bar{s}_0$, and the same goes for April and May, with two other unknown parameters $\bar{s}_1$ and $\bar{s}_2$. Then a persistent process is assumed for s_t after May 2020. Specifically, the residual variance after May decays at a constant monthly rate, ρ, which is also treated as an unknown parameter to be estimated. To put these assumptions in equations, we have

$$s_{t^*} = \bar{s}_0, \quad s_{t^*+1} = \bar{s}_1, \quad s_{t^*+2} = \bar{s}_2, \quad s_{t^*+j} = 1 + (\bar{s}_2 - 1)\rho^{j-2}, \quad j = 3, ..., T.$$

Variational Inference for the Reduced-Form VAR

We define $\mathbf{o}_i = (o_{i,1}, ..., o_{i,T})'$, and $\mathbf{q}_i = (q_{i,1}, ..., q_{i,T})'$. For the VAR-SVO model, we approximate $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}, \mathbf{o}|\mathbf{y})$ using the product of densities

$$\begin{aligned} q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}, \mathbf{o}) &= q(\boldsymbol{\kappa})\prod_{i=1}^{n} q(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \mathbf{h}_i, h_{i,0}, \sigma_i^2, \mathbf{o}_i, \mathbf{q}_i) \\ &= q(\boldsymbol{\kappa})\prod_{i=1}^{n} q(\boldsymbol{\alpha}_i)q(\boldsymbol{\beta}_i), q(\mathbf{h}_i)q(h_{i,0})q(\sigma_i^2)q(\mathbf{o}_i). \end{aligned}$$

For VAR-SV-t, we approximate $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}, \mathbf{q}|\mathbf{y})$ using the product of densities

$$\begin{aligned} q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{h}_0, \boldsymbol{\sigma}^2, \boldsymbol{\kappa}, \mathbf{q}) &= q(\boldsymbol{\kappa}) \prod_{i=1}^{n} q(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \mathbf{h}_i, h_{i,0}, \sigma_i^2, \mathbf{q}_i) \\ &= q(\boldsymbol{\kappa}) \prod_{i=1}^{n} q(\boldsymbol{\alpha}_i) q(\boldsymbol{\beta}_i), q(\mathbf{h}_i) q(h_{i,0}) q(\sigma_i^2) q(\mathbf{q}_i). \end{aligned}$$

For VAR-CVD, we denote the unknown parameters $\bar{s}_2$ and ρ by a vector $\boldsymbol{\theta} = (\bar{s}_2, \rho)'$. Then we approximate $p(\boldsymbol{\alpha}, \bar{s}_0, \bar{s}_1, \boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{\Sigma}|\mathbf{y})$ using the product of densities

$$q(\boldsymbol{\alpha}, \bar{s}_0, \bar{s}_1, \boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{\Sigma}) = q(\boldsymbol{\alpha}) q(\bar{s}_0) q(\bar{s}_1) q(\boldsymbol{\theta}) q(\boldsymbol{\kappa}) q(\boldsymbol{\Sigma}).$$

In what follows, we derive the explicit forms of each of these optimal marginal densities and their associated parameters. For VAR-SVO and VAR-SVt, we omit the details for obtaining $q^*_{\boldsymbol{\alpha}_i}$, $q^*_{\boldsymbol{\beta}_i}$, $q^*(\boldsymbol{\kappa})$, $q^*(h_{i,0})$, and $q^*(\boldsymbol{\sigma}^2)$ in this section because it is similar to that in VARSV without the outlier component.

1. VAR-SVO

The Optimal Density $q^*_{\mathbf{o}_i}$

The optimal density $q^*_{\mathbf{o}_i}$ takes the form

$$q^*_{\mathbf{o}_i} \propto \exp\{\mathbb{E}_{-\mathbf{o}_i}[\log p(\mathbf{o}_i|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}_i, p_{\mathbf{o}_i})]\}.$$

The conditional distribution of $\mathbf{o}_i$ is as follows

$$p(\mathbf{o}_i|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}_i, p_{\mathbf{o}_i}) \propto \prod_{t=1}^{T} (o_{i,t}^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} o_{i,t}^{-2} e^{-h_{i,t}} \widetilde{\varepsilon}_{i,t}^2\right\} (1 - p_{\mathbf{o}_i})^{\mathrm{I}(o_{i,t}=1)} \left(\frac{p_{\mathbf{o}_i}}{19}\right)^{\mathrm{I}(o_{i,t} \geqslant 2)}.$$

The log of the optimal density $\log(q^*_{\mathbf{o}_i})$ therefore takes the form

$$\log(q^*_{\mathbf{o}_i}) = C_{\mathbf{o}_i} - \frac{1}{2}\sum_{t=1}^{T}\left[\log o_{i,t}^2 - \frac{1}{2}o_{i,t}^{-2}e^{-\bar{h}_{i,t}+\frac{1}{2}\widehat{d}_{i,t}}\widehat{s}_t^2\right] + T_1\mathbb{E}_{p_{\mathbf{o}_i}}[\log(1-p_{\mathbf{o}_i})] \\ + (T-T_1)\mathbb{E}_{p_{\mathbf{o}_i}}[\log(p_{\mathbf{o}_i}/19)],$$

where $T_1 = \sum_{t=1}^{T}\mathrm{I}(o_{i,t}=1)$, i.e., the number of elements in $\mathbf{o}_i$ that is equal to 1.

We follow Stock and Watson (2016) to discretize the support of $\mathbf{o}_{i,t}$ to simplify estimation. In specific, we use a grid with grid points at $1, 2, ..., 20$. The prior of $o_{i,t}$ then becomes a discrete distribution that has probability $1-p_{\mathbf{o}_i}$ at 1 and probability $p_{\mathbf{o}_i}/19$ at $j = 2, ..., 20$. The likelihood can also be easily evaluated at these grid points. Finally we compute the expectation and variance based on the likelihood at the corresponding points.

It is important to note that in this process we are able to compute $C_{\mathbf{o}_i}$ as well. In specific, since $\sum_{j=1}^{20} q_{o_{i,t}}(o_{i,t}=j) = 1$, we have

$$\exp(C_{o_{i,t}}) = \frac{1}{M_{o_{i,t}}},$$

where

$$M_{o_{i,t}} = \sum_{o_{i,t}=1}^{20}(o_{i,t}^2)^{-\frac{1}{2}}\exp\Bigg\{-\frac{1}{2}o_{i,t}^{-2}\mathrm{e}^{-\bar{h}_{i,t}+\frac{1}{2}\widehat{d}_{i,t}}\widehat{s}_t^2 \\ + \mathrm{I}(o_{i,t}=1)\mathbb{E}_{p_{\mathbf{o}_i}}[\log(1-p_{\mathbf{o}_i})] + \mathrm{I}(o_{i,t}\geqslant 2)\mathbb{E}_{p_{\mathbf{o}_i}}[\log(p_{\mathbf{o}_i}/19)]\Bigg\}.$$

This will be useful for us to compute the variational lower bound later on.

The Optimal Density $q^*_{p_{\mathbf{o}_i}}$

The optimal density $q^*_{p_{\mathbf{o}_i}}$ takes the form

$$q^*_{p_{\mathbf{o}_i}} \propto \exp\{\mathbb{E}_{-p_{\mathbf{o}_i}}[\log p(p_{\mathbf{o}_i}|\mathbf{o}_i)]\},$$

The conditional distribution of $p_{\mathbf{o}_i}$ is as follows

$$p(p_{\mathbf{o}_i}|\mathbf{o}_i) \propto p_{\mathbf{o}_i}^{a_{p_{\mathbf{o}_i}}+(T-T_1)}(1-p_{\mathbf{o}_i})^{b_{p_{\mathbf{o}_i}}+T_1}.$$

The log of the optimal density $\log(p^*_{\mathbf{o}_i})$ therefore takes the form

$$\log(p^*_{\mathbf{o}_i}) = C_{p_{\mathbf{o}_i}} + (a_{p_{\mathbf{o}_i}} + (T-T_1))\log p_{\mathbf{o}_i} + (b_{p_{\mathbf{o}_i}} + T_1)\log(1-p_{\mathbf{o}_i})$$

So that

$$\mathbb{E}_{-p_{\mathbf{o}_i}}[\log(p^*_{\mathbf{o}_i})] = C_{p_{\mathbf{o}_i}} + (a_{p_{\mathbf{o}_i}} + (T-\bar{T}_1))\log p_{\mathbf{o}_i} + (b_{p_{\mathbf{o}_i}} + \bar{T}_1)\log(1-p_{\mathbf{o}_i}),$$

where $\bar{T}_1 \equiv \mathbb{E}_{\mathbf{o}_i}[T_1] = \mathbb{E}_{\mathbf{o}_i}[\sum_{t=1}^T \mathrm{I}(o_{i,t}=1)] = \sum_{t=1}^T q^*_{o_{i,t}}(o_{i,t}=1)$. Therefore, the approximating density is a beta distribution: $\mathcal{B}(a_{p_{\mathbf{o}_i}} + (T-\bar{T}_1)), b_{p_{\mathbf{o}_i}} + \bar{T}_1)$.

The Variational Lower Bound

Next, we derive the variational lower bound $\underline{p}(\mathbf{y};q)$. To that end, we first compute the log ratio of the joint posterior density and the variational approximation:

$$
\begin{aligned}
&\log\left[\frac{p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},\mathbf{h}_0,\boldsymbol{\sigma}^2,\mathbf{o},\boldsymbol{\kappa},\mathbf{p_o})}{q(\boldsymbol{\kappa})\prod_{i=1}^n q(\boldsymbol{\alpha}_i)q(\boldsymbol{\beta}_i)q(\mathbf{h}_i)q(\mathbf{h}_{0,i})q(\sigma_i^2)q(\mathbf{o}_i)q(p_{\mathbf{o}_i})}\right] \\
=& c_{\boldsymbol{\kappa}} + \sum_{i=1}^n \Bigg[c_i - \frac{1}{2}\mathbb{1}_T'(\mathbf{h}_i + \log \mathbf{o}_i^2) - \frac{1}{2}((\mathbf{Y}-\mathbf{XA})\mathbf{B}_{0,i,1:n})'\mathbf{M}_i^{-1}((\mathbf{Y}-\mathbf{XA})\mathbf{B}_{0,i,1:n})) \\
& - \frac{T}{2}\log\sigma_i^2 - \frac{1}{2\sigma_i^2}(\mathbf{h}_i - h_{i,0}\mathbb{1}_T)'\mathbf{H}'\mathbf{H}(\mathbf{h}_i - h_{i,0}\mathbb{1}_T) \\
& - \frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i})'\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}) - \frac{1}{2V_{h_{i,0}}}h_{i,0}^2 - (\nu_i+1)\log\sigma_i^2 - \frac{S_i}{\sigma_i^2} \\
& - \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_{0,i})'\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_{0,i}) + T_1\log(1-p_{\mathbf{o}_i}) \\
& + (T-T_1)\log\left(\frac{p_{\mathbf{o}_i}}{19}\right) + (a_{p_{\mathbf{o}_i}}-1)\log(p_{\mathbf{o}_i}) + (b_{p_{\mathbf{o}_i}}-1)\log(1-p_{\mathbf{o}_i}) \Bigg] \\
& + \sum_{r=1}^3\Big[(c_{r,1}-1)\log\kappa_r - c_{r,2}\kappa_r\Big] - \Big[\sum_{r=1}^3(v_r-1)\log\kappa_r - \frac{a_r\kappa_r + b_r\kappa_r^{-1}}{2}\Big] \\
& + \sum_{i=1}^n \Bigg\{ \frac{1}{2}(\mathbf{h}_i - \widehat{\mathbf{h}}_i)'\widehat{\mathbf{K}}_{\mathbf{h}_i}(\mathbf{h}_i - \widehat{\mathbf{h}}_i) + \frac{1}{2}(\boldsymbol{\alpha}_i - \bar{\widehat{\boldsymbol{\alpha}}}_i)'\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_i - \bar{\widehat{\boldsymbol{\alpha}}}_i) \\
& + \frac{1}{2}(\boldsymbol{\beta}_i - \bar{\widehat{\boldsymbol{\beta}}}_i)'\mathbf{K}_{\boldsymbol{\beta}_i}(\boldsymbol{\beta}_i - \bar{\widehat{\boldsymbol{\beta}}}_i) + \frac{\bar{K}_{h_{i,0}}}{2}(h_{i,0} - \widehat{h}_{i,0})^2 + (\widehat{\nu}_i + 1)\log\sigma_i^2 + \frac{\widehat{S}_i}{\sigma_i^2} \\
& - \Bigg[C_{\mathbf{o}_i} + \sum_{t=1}^T \Big\{ \log(o_{i,t}^{-1}) - \frac{1}{2}o_{i,t}^{-2}\mathrm{e}^{-\bar{h}_{i,t}+\frac{1}{2}\widehat{d}_{i,t}}\widehat{s}_t^2 \Big\} + T_1\mathbb{E}_{p_{\mathbf{o}_i}}[\log(1-p_{\mathbf{o}_i})] \\
& + (T-T_1)\mathbb{E}_{p_{\mathbf{o}_i}}[\log(p_{\mathbf{o}_i}/19)] + (a_{p_{\mathbf{o}_i}} + T - \bar{T}_1 - 1)\log(p_{\mathbf{o}_i}) \\
& + (b_{p_{\mathbf{o}_i}} + \bar{T}_1 - 1)\log(1-p_{\mathbf{o}_i}) - \log\Gamma(a_{p_{\mathbf{o}_i}} + T - \bar{T}_1) - \log\Gamma(b_{p_{\mathbf{o}_i}} + \bar{T}_1) \Bigg] \Bigg\} \qquad ,
\end{aligned}
$$

where $\mathbf{M} = \mathbf{O}^2 \cdot \mathbf{D}$, $c_i = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log V_{h_{i,0}} - \frac{1}{2}\log|\mathbf{V}_{\boldsymbol{\beta}_i}| - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}| + \nu_i\log S_i - \log\Gamma(\nu_i) - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\mathbf{h}_i}| - \frac{1}{2}\log|\mathbf{V}_{\boldsymbol{\alpha}_i}| - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}| - \frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}| - \frac{1}{2}\log\widehat{K}_{h_{i,0}} - \widehat{\nu}_i\log\widehat{S}_i + \log\Gamma(\widehat{\nu}_i) + \log\Gamma(a_{p_{\mathbf{o}_i}} + b_{p_{\mathbf{o}_i}}) - \log\Gamma(a_{p_{\mathbf{o}_i}}) - \log\Gamma(b_{p_{\mathbf{o}_i}}) - \log\Gamma(a_{p_{\mathbf{o}_i}} + b_{p_{\mathbf{o}_i}} + T)$, and $c_{\boldsymbol{\kappa}} = \sum_{r=1}^3[c_{r,1}\log c_{r,2} - \log\Gamma(c_{r,1})] - \sum_{r=1}^3\left[\frac{v_r}{2}(\log a_r - \log b_r) - \log(2K_{v_r}\sqrt{a_r b_r})\right]$, $K_{v_r}(\cdot)$ is a modified Bessel function of the second kind. Taking expectation of the above log ratio with respect to q, we obtain the following variational lower bound:

$$\begin{aligned}
\underline{p}(\mathbf{y};q) &= \mathbb{E}_q\left\{\log\left[\frac{p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},\mathbf{h}_0,\boldsymbol{\sigma}^2,\mathbf{o},\mathbf{p_o})}{\prod_{i=1}^n q(\boldsymbol{\alpha}_i)q(\boldsymbol{\beta}_i)q(\mathbf{h}_i)q(\mathbf{h}_{0,i})q(\sigma_i^2)q(\mathbf{o}_i)q(p_{\mathbf{o}_i})}\right]\right\} \\
&= \sum_{i=1}^n\left[c_i - \frac{1}{2}\mathbb{1}_T'\widehat{\mathbf{h}}_i - \frac{1}{2}\left((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\bar{\mathbf{B}}_{0,i,1:n}\right)'\overline{\mathbf{M}_i^{-1}}\left((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\bar{\mathbf{B}}_{0,i,1:n}\right)\right. \\
&\quad - \frac{1}{2}\bar{\mathbf{B}}_{0,i,1:n}'\mathbf{G}_i\bar{\mathbf{B}}_{0,i,1:n} - \frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})'\overline{\mathbf{M}_i^{-1}}(\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\widehat{\mathbf{K}}_{\mathbf{B}_{0,i,1:n}}^{-1}\right) \\
&\quad - \frac{1}{2V_{h_{i,0}}}(\widehat{h}_{i,0}+\bar{K}_{h_{i,0}}^{-1}) - \frac{1}{2}(\bar{\widehat{\boldsymbol{\beta}}}_i-\boldsymbol{\beta}_{i,0})'\overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}}(\bar{\widehat{\boldsymbol{\beta}}}_i-\boldsymbol{\beta}_{i,0}) - \frac{1}{2}\mathrm{tr}\left(\overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}}\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}^{-1}\right) \\
&\quad - \frac{1}{2}(\bar{\widehat{\boldsymbol{\alpha}}}_i-\boldsymbol{\alpha}_{i,0})'\overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}}(\bar{\widehat{\boldsymbol{\alpha}}}_i-\boldsymbol{\alpha}_{i,0}) - \frac{1}{2}\mathrm{tr}\left(\overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1}\right) + \frac{1}{2}(T+k_i+1) \\
&\quad - C_{\mathbf{o}_i} + \frac{1}{2}\sum_{t=1}^T\overline{o_{i,t}^{-2}}\mathrm{e}^{-\bar{h}_{i,t}+\frac{1}{2}\widehat{d}_{i,t}}\widehat{s}_t^2 - (T-\bar{T}_1)\overline{\log p_{\mathbf{o}_i}} - \bar{T}_1\overline{\log(1-p_{\mathbf{o}_i})} \\
&\quad \left. + \log\Gamma(a_{p_{\mathbf{o}_i}}+T-\bar{T}_1) + \log\Gamma(b_{p_{\mathbf{o}_i}}+\bar{T}_1)\right] \\
&\quad + \sum_{r=1}^3\left[(c_{r,1}-v_r)\overline{\log\kappa_r} - \left(c_{r,2}-\frac{1}{2}a_r\right)\bar{\kappa}_r + \frac{1}{2}b_r\overline{\kappa_r^{-1}}\right],
\end{aligned} \tag{A.17}$$

where $\overline{\log p_{\mathbf{o}_i}} = \mathbb{E}_{p_{\mathbf{o}_i}}[\log p_{\mathbf{o}_i}] = \psi(a_{p_{\mathbf{o}_i}}+T-\bar{T}_1) - \psi(a_{p_{\mathbf{o}_i}}+b_{p_{\mathbf{o}_i}}+T)$, and $\overline{\log(1-p_{\mathbf{o}_i})} = \mathbb{E}_{p_{\mathbf{o}_i}}[\log(1-p_{\mathbf{o}_i})] = \psi(b_{p_{\mathbf{o}_i}}+\bar{T}_1) - \psi(a_{p_{\mathbf{o}_i}}+b_{p_{\mathbf{o}_i}}+T)$, $\psi(\cdot)$ is the digamma function, and $\overline{o_{i,t}^{-2}} = \mathbb{E}_{o_{i,t}}[o_{i,t}^{-2}]$.

2. VAR-SVt

We omit the details for obtaining the optimal densities $q^*_{\boldsymbol{\alpha}_i}, q^*_{\boldsymbol{\beta}_i}, q^*_{\mathbf{h}}, q^*_{\mathbf{h}_0}, q^*_{\boldsymbol{\sigma}^2}$ in this section and focus on $q^*_{\mathbf{q}_i}$, because they are similar to those in the model VAR-SVO.

The Optimal Density $q^*_{\mathbf{q}_i}$

The optimal density $q^*_{\mathbf{q}_i}$ takes the form

$$q^*_{\mathbf{q}_i} \propto \exp\{\mathbb{E}_{-\mathbf{q}_i}[\log p(\mathbf{q}_i|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}_i)]\},$$

The conditional distribution of $\mathbf{q}_i$ is as follows

$$\begin{aligned} p(\mathbf{q}_i|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{h}_i) &\propto \prod_{t=1}^{T} (q^2_{i,t})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} q^{-2}_{i,t} \mathrm{e}^{-h_{i,t}} \widehat{\varepsilon}^2_{i,t}\right\} (q^2_{i,t})^{-\frac{l_i}{2}-1} \exp\left\{-\frac{\frac{l_i}{2}}{q^2_{i,t}}\right\} \\ &= \prod_{t=1}^{T} (q^2_{i,t})^{-\frac{l_i}{2}-\frac{1}{2}-1} \exp\left\{-\frac{1}{q^2_{i,t}}\left(\frac{l_i}{2} + \frac{1}{2}\mathrm{e}^{-h_{i,t}} \widehat{\varepsilon}^2_{i,t}\right)\right\}. \end{aligned}$$

The expectation of the log of the optimal density $\mathbb{E}_{-\mathbf{q}_i}[\log p(\mathbf{q}_i| \cdot)]$ gives

$$\begin{aligned} \mathbb{E}_{-\mathbf{q}_i}[\log p(\mathbf{q}_i| \cdot)] = C_{\mathbf{q}_i} + \sum_{t=1}^{T} \Bigg[&\left(-\frac{l_i}{2} - \frac{1}{2} - 1\right) \log(q^2_{i,t}) \\ &- \frac{1}{q^2_{i,t}} \left(\frac{l_i}{2} + \frac{1}{2}\mathbb{E}_{h_{i,t}}[\mathrm{e}^{-h_{i,t}}] \widehat{s}^2_t\right) \Bigg], \end{aligned}$$

where $C_{\mathbf{q}_i}$ is a constant independent of $\mathbf{q}_i$. It is clear that the optimal density $q^*_{\mathbf{q}_{i,t}}$ is an inverse-gamma distribution: $\mathcal{IG}(\widehat{\nu}_{q_i}, \widehat{S}_{q_{i,t}})$, where

$$\widehat{\nu}_{q_i} = \frac{l_i}{2} + \frac{1}{2}, \quad \widehat{S}_{q_{i,t}} = \frac{l_i}{2} + \frac{1}{2}\mathbb{E}_{h_{i,t}}[\mathrm{e}^{-h_{i,t}}] \widehat{s}^2_t.$$

The optimal densities for other parameters are quite similar to SV model, so we are omitting the details here.

The Variational Lower Bound

Next, we derive the variational lower bound $\underline{p}(\mathbf{y}; q)$. To that end, we first compute the log ratio of the joint posterior density and the variational approximation:

$$
\begin{aligned}
&\log\left[\frac{p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},\mathbf{h}_0,\boldsymbol{\sigma}^2,\mathbf{q})}{\prod_{i=1}^n q(\boldsymbol{\alpha}_i)q(\boldsymbol{\beta}_i)q(\mathbf{h}_i)q(\mathbf{h}_{0,i})q(\sigma_i^2)q(\mathbf{q}_i)}\right]\\
&=\sum_{i=1}^n\Bigg[c_i-\frac{1}{2}\mathbb{1}_T'(\mathbf{h}_i+\log\mathbf{q}_i^2)-\frac{1}{2}((\mathbf{Y}-\mathbf{X}\mathbf{A})\mathbf{B}_{0,i,1:n})'\mathbf{F}_i^{-1}((\mathbf{Y}-\mathbf{X}\mathbf{A})\mathbf{B}_{0,i,1:n}))\\
&\quad-\frac{T}{2}\log\sigma_i^2-\frac{1}{2\sigma_i^2}(\mathbf{h}_i-h_{i,0}\mathbb{1}_T)'\mathbf{H}'\mathbf{H}(\mathbf{h}_i-h_{i,0}\mathbb{1}_T)-\frac{1}{2}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{0,i})'\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_{0,i})\\
&\quad-\frac{1}{2V_{h_{i,0}}}h_{i,0}^2-(\nu_i+1)\log\sigma_i^2-\frac{S_i}{\sigma_i^2}-\frac{1}{2}(\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_{0,i})'\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}(\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_{0,i})\\
&\quad-\left(\frac{l_i}{2}+1\right)\mathbb{1}_T'\log\mathbf{q}_i^2-\frac{l_i}{2}\mathbb{1}_T'\mathbf{q}_i^{-2}\Bigg]+\sum_{r=1}^3\Big[(c_{r,1}-1)\log\kappa_r-c_{r,2}\kappa_r\Big]\\
&\quad+\sum_{i=1}^n\Bigg\{\frac{1}{2}(\mathbf{h}_i-\widehat{\mathbf{h}}_i)'\widehat{\mathbf{K}}_{\mathbf{h}_i}(\mathbf{h}_i-\widehat{\mathbf{h}}_i)+\frac{1}{2}(\boldsymbol{\alpha}_i-\bar{\widehat{\boldsymbol{\alpha}}}_i)'\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_i-\bar{\widehat{\boldsymbol{\alpha}}}_i)\\
&\quad+\frac{1}{2}(\boldsymbol{\beta}_i-\bar{\widehat{\boldsymbol{\beta}}}_i)'\mathbf{K}_{\boldsymbol{\beta}_i}(\boldsymbol{\beta}_i-\bar{\widehat{\boldsymbol{\beta}}}_i)+\frac{\bar{K}_{h_{i,0}}}{2}(h_{i,0}-\widehat{h}_{i,0})^2+(\widehat{\nu}_i+1)\log\sigma_i^2+\frac{\widehat{S}_i}{\sigma_i^2}\\
&\quad-\Bigg[\sum_{t=1}^T\left(\widehat{\nu}_{q_i}\log\widehat{S}_{q_{i,t}}\right)-\left(\frac{l_i}{2}+\frac{1}{2}+1\right)\mathbb{1}_T'\log\mathbf{q}_i^2-\sum_{t=1}^T\left(\frac{\widehat{S}_{q_{i,t}}}{q_{i,t}^2}\right)\\
&\quad+\sum_{r=1}^3(v_r-1)\log\kappa_r-\frac{a_r\kappa_r+b_r\kappa_r^{-1}}{2}\Bigg]\Bigg\},
\end{aligned}
$$

where $c_i=-\frac{T}{2}\log(2\pi)-\frac{1}{2}\log V_{h_{i,0}}-\frac{1}{2}\log|\mathbf{V}_{\boldsymbol{\beta}_i}|-\frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}|+\nu_i\log S_i-\log\Gamma(\nu_i)-\frac{1}{2}\log|\widehat{\mathbf{K}}_{\mathbf{h}_i}|-\frac{1}{2}\log|\mathbf{V}_{\boldsymbol{\alpha}_i}|-\frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}|-\frac{1}{2}\log|\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}|-\frac{1}{2}\log\widehat{K}_{h_{i,0}}-\widehat{\nu}_i\log\widehat{S}_i+\log\Gamma(\widehat{\nu}_i)+\frac{Tl_i}{2}(\log l_i-\log 2)-T\log\Gamma\left(\frac{l_i}{2}\right)+T\log\Gamma(\widehat{\nu}_{q_i})$, $\mathbf{F}_i=\mathbf{Q}_i^2\cdot\mathbf{D}_i$, and $c_{\boldsymbol{\kappa}}=\sum_{r=1}^3\left[c_{r,1}\log c_{r,2}-\log\Gamma(c_{r,1})\right]-\sum_{r=1}^3\left[\frac{v_r}{2}(\log a_r-\log b_r)-\log(2K_{v_r}\sqrt{a_rb_r})\right]$, $K_{v_r}(\cdot)$ is a modified Bessel function of the second kind. Let $\mathbf{Q}_i$ denote the diagonal matrix of which the diagonal elements are $\mathbf{q}_i=(q_{i,1},...q_{i,T})'$, i.e., $\mathbf{Q}_i=\mathrm{diag}(q_{i,1},...q_{i,T})$. Taking expectation of the above log ratio with respect to q, we obtain the following variational lower bound:

$$
\begin{aligned}
\underline{p}(\mathbf{y};q) &= \mathbb{E}_q\left\{\log\left[\frac{p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{h},\mathbf{h}_0,\boldsymbol{\sigma}^2,\mathbf{o},\mathbf{p_o})}{\prod_{i=1}^n q(\boldsymbol{\alpha}_i)q(\boldsymbol{\beta}_i)q(\mathbf{h}_i)q(\mathbf{h}_{0,i})q(\sigma_i^2)q(\mathbf{o}_i)q(p_{\mathbf{o}_i})}\right]\right\} \\
&= \sum_{i=1}^n\Bigg[c_i - \frac{1}{2}\mathbb{1}_T'\widehat{\mathbf{h}}_i - \frac{1}{2}\big((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\bar{\mathbf{B}}_{0,i,1:n}\big)'\,\overline{\mathbf{F}_i^{-1}}\,\big((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\bar{\mathbf{B}}_{0,i,1:n}\big) \\
&\quad - \frac{1}{2}\bar{\mathbf{B}}_{0,i,1:n}'\mathbf{G}_i\bar{\mathbf{B}}_{0,i,1:n} - \frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})'\overline{\mathbf{F}_i^{-1}}(\mathbf{Y}-\mathbf{X}\bar{\mathbf{A}})\widehat{\mathbf{K}}_{\mathbf{B}_{0,i,1:n}}^{-1}\right) \\
&\quad - \frac{1}{2V_{h_{i,0}}}(\widehat{h}_{i,0}+\bar{K}_{h_{i,0}}^{-1}) - \frac{1}{2}(\bar{\widehat{\boldsymbol{\beta}}}_i-\boldsymbol{\beta}_{i,0})'\overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}}(\bar{\widehat{\boldsymbol{\beta}}}_i-\boldsymbol{\beta}_{i,0}) - \frac{1}{2}\mathrm{tr}\left(\overline{\mathbf{V}_{\boldsymbol{\beta}_i}^{-1}}\widehat{\mathbf{K}}_{\boldsymbol{\beta}_i}^{-1}\right) \\
&\quad - \frac{1}{2}(\bar{\widehat{\boldsymbol{\alpha}}}_i-\boldsymbol{\alpha}_{i,0})'\overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}}(\bar{\widehat{\boldsymbol{\alpha}}}_i-\boldsymbol{\alpha}_{i,0}) - \frac{1}{2}\mathrm{tr}\left(\overline{\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}_i}^{-1}\right) + \frac{1}{2}(T+k_i+1) \\
&\quad + \frac{1}{2}\sum_{t=1}^T \mathrm{e}^{-\bar{h}_{i,t}+\frac{1}{2}\widehat{d}_{i,t}}\overline{q_{i,t}^{-2}}\widehat{s}_t^2 - \sum_{t=1}^T\widehat{\nu}_{q_i}\log\widehat{S}_{q_{i,t}}\Bigg] \\
&\quad + c_{\boldsymbol{\kappa}} + \sum_{r=1}^3\left[(c_{r,1}-v_r)\overline{\log\kappa_r} - \left(c_{r,2}-\frac{1}{2}a_r\right)\bar{\kappa}_r + \frac{1}{2}b_r\overline{\kappa_r^{-1}}\right],
\end{aligned} \tag{A.18}
$$

where $\overline{q_{i,t}^{-2}} = \mathbb{E}_{q_{i,t}^2}[q_{i,t}^{-2}] = \frac{\widehat{\nu}_{q_i}}{\widehat{S}_{q_{i,t}}}$.

3. VAR-CVD

The optimal density $q^*_{\boldsymbol{\alpha}}$

The optimal density $q^*_{\boldsymbol{\alpha}}$ has the form

$$q^*_{\boldsymbol{\alpha}} \propto \exp\left\{\mathbb{E}_{-\boldsymbol{\alpha}}\left[\log p(\boldsymbol{\alpha}|\mathbf{y}, \mathbf{s}, \boldsymbol{\kappa}, \boldsymbol{\Sigma})\right]\right\},$$

where the expectation is taken with respect to the marginal density $q_{-\boldsymbol{\alpha}}(\mathbf{s}, \boldsymbol{\kappa}, \boldsymbol{\Sigma})$.

From Chan (2020b), we have

$$(\boldsymbol{\alpha}|\mathbf{y}, \mathbf{s}, \boldsymbol{\kappa}, \boldsymbol{\Sigma}) \sim \mathcal{N}(\widehat{\boldsymbol{\alpha}}, \mathbf{K}_{\boldsymbol{\alpha}}^{-1}),$$

where

$$\mathbf{K}_{\boldsymbol{\alpha}} = \mathbf{V}_{\boldsymbol{\alpha}}^{-1} + \mathbf{X}'(\mathbf{S}^{-2} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{X}, \quad \widehat{\boldsymbol{\alpha}} = \mathbf{K}_{\boldsymbol{\alpha}}^{-1}\left(\mathbf{V}_{\alpha}^{-1}\boldsymbol{\alpha}_0 + \mathbf{X}'(\mathbf{S}^{-2} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{y})\right).$$

The log-density is therefore

$$\log p(\boldsymbol{\alpha}|\cdot) = c_{\boldsymbol{\alpha}} - \frac{1}{2}\boldsymbol{\alpha}'\mathbf{K}_{\boldsymbol{\alpha}}\boldsymbol{\alpha} + \boldsymbol{\alpha}'\left(\mathbf{V}_{\alpha}^{-1}\boldsymbol{\alpha}_0 + \mathbf{X}'(\mathbf{S}^{-2} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{y})\right),$$

where $c_{\boldsymbol{\alpha}}$ is a term not dependent on $\boldsymbol{\alpha}$. After taking the expectation, we essentially get an approximating density $\mathcal{N}(\bar{\widehat{\boldsymbol{\alpha}}}, \widehat{\mathbf{K}}_{\alpha}^{-1})$, where

$$\begin{aligned}
\widehat{\mathbf{K}}_{\boldsymbol{\alpha}} &= \mathbb{E}_{-\boldsymbol{\alpha}_i}\left[\mathbf{V}_{\boldsymbol{\alpha}}^{-1} + \mathbf{X}'(\mathbf{S}^{-2} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{X}\right] \\
&= \overline{\mathbf{V}}_{\boldsymbol{\alpha}_i}^{-1} + \mathbf{X}'(\overline{\mathbf{S}}^{-2} \otimes \overline{\boldsymbol{\Sigma}}^{-1})\mathbf{X},
\end{aligned}$$

$$\begin{aligned}
\bar{\widehat{\boldsymbol{\alpha}}} &= \mathbb{E}_{-\boldsymbol{\alpha}_i}\left[\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\left(\mathbf{V}_{\alpha}^{-1}\boldsymbol{\alpha}_0 + \mathbf{X}'(\mathbf{S}^{-2} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{y}\right)\right] \\
&= \widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\left[\overline{\mathbf{V}}_{\boldsymbol{\alpha}_i}^{-1}\boldsymbol{\alpha}_0 + \mathbf{X}'(\overline{\mathbf{S}}^{-2} \otimes \overline{\boldsymbol{\Sigma}}^{-1})\mathbf{y}\right].
\end{aligned}$$

The Optimal Density $q^*_{\boldsymbol{\Sigma}}$

The optimal density $q^*_{\boldsymbol{\Sigma}}$ has the form

$$q^*_{\boldsymbol{\Sigma}} \propto \exp\left\{\mathbb{E}_{-\boldsymbol{\Sigma}}\left[\log p(\boldsymbol{\Sigma}|\mathbf{y}, \mathbf{s}, \boldsymbol{\kappa}, \boldsymbol{\alpha})\right]\right\},$$

where the expectation is taken with respect to the marginal density $q_{-\boldsymbol{\Sigma}}(\mathbf{s}, \boldsymbol{\kappa}, \boldsymbol{\alpha})$. From

Chan (2020b), we know

$$(\mathbf{\Sigma}|\cdot) \sim \mathcal{IW}\left(\nu + T, \mathbf{\Psi} + \sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\alpha})s_t^{-2}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\alpha})'\right).$$

After taking the expectation, we obtain the approximating density $\mathcal{IW}(\hat{\nu}, \hat{\mathbf{\Psi}})$, where

$$\hat{\nu} = \nu + T, \quad \hat{\mathbf{\Psi}} = \mathbf{\Psi} + \sum_{t=1}^{T} \bar{s}_t^{-2}\left(\mathbf{x}_t\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{x}_t' + \left(\mathbf{y}_t - \mathbf{x}_t\bar{\widehat{\boldsymbol{\alpha}}}\right)\left(\mathbf{y}_t - \mathbf{x}_t\bar{\widehat{\boldsymbol{\alpha}}}\right)'\right).$$

The expectation of $\mathbf{\Sigma}$ is therefore $\hat{\mathbf{\Psi}}/(\hat{\nu} - n - 1)$. It is also useful to know that $\mathbf{\Sigma}$ follows an inverse Wishart distribution $\mathcal{IW}_n(\hat{\nu}, \hat{\mathbf{\Psi}})$ if $\mathbf{\Sigma}^{-1}$ follows a Wishart distribution $\mathcal{W}_n(\hat{\nu} + n - 1, \mathbf{\Psi}^{-1})$. Therefore, the expectation of $\mathbf{\Sigma}^{-1}$ is $(\hat{\nu} + n - 1)\mathbf{\Psi}^{-1}$.

Optimal Density $q^*_{\bar{s}_0}$ and $q^*_{\bar{s}_1}$

The optimal density $q^*_{\bar{s}_0}$ has the form

$$q^*_{\bar{s}_0} \propto \exp\left\{\mathbb{E}_{-\bar{s}_0}\left[\log p(\bar{s}_0|\mathbf{y}_{t^*}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \mathbf{\Sigma}, \rho)\right]\right\},$$

where the expectation is taken with respect to the marginal density $q_{-\bar{s}_0}(\boldsymbol{\alpha}, \boldsymbol{\kappa}, \mathbf{\Sigma}, \rho)$, t^* denotes the time period of the pandemic (March 2020).

The log-density is

$$\log p(\bar{s}_0) = C_{\bar{s}_0} - \frac{n}{2}\log \bar{s}_0^2 - \frac{(\mathbf{y}_{t^*} - \mathbf{X}_{t^*}\boldsymbol{\alpha})'\,\mathbf{\Sigma}^{-1}\,(\mathbf{y}_{t^*} - \mathbf{X}_{t^*}\boldsymbol{\alpha})}{2\bar{s}_0^2} - \log \bar{s}_0^2.$$

After taking the expectation, we obtain the approximating density for $\bar{s}_0^2$: $\mathcal{IG}(\nu_{\bar{s}_0}, \phi_{\bar{s}_0})$, where

$$\nu_{\bar{s}_0} = \frac{n+1}{2}, \quad \phi_{\bar{s}_0} = \frac{1}{2}\left[\left(\mathbf{y}_{t^*} - \mathbf{X}_{t^*}\bar{\widehat{\boldsymbol{\alpha}}}\right)'\,\overline{\mathbf{\Sigma}}^{-1}\left(\mathbf{y}_{t^*} - \mathbf{X}_{t^*}\bar{\widehat{\boldsymbol{\alpha}}}\right) + \mathrm{tr}\left(\overline{\mathbf{\Sigma}}^{-1}\mathbf{X}_{t^*}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{X}_{t^*}'\right)\right].$$

Similarly, we have the approximating density for $\bar{s}_1^2$: $\mathcal{IG}(\nu_{\bar{s}_1}, \phi_{\bar{s}_1})$, where

$$\nu_{\bar{s}_1} = \frac{n+1}{2}, \quad \phi_{\bar{s}_1} = \frac{1}{2}\left[\left(\mathbf{y}_{t^*+1} - \mathbf{X}_{t^*+1}\bar{\widehat{\boldsymbol{\alpha}}}\right)'\,\overline{\mathbf{\Sigma}}^{-1}\left(\mathbf{y}_{t^*+1} - \mathbf{X}_{t^*+1}\bar{\widehat{\boldsymbol{\alpha}}}\right) + \mathrm{tr}\left(\overline{\mathbf{\Sigma}}^{-1}\mathbf{X}_{t^*+1}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{X}_{t^*+1}'\right)\right].$$

Optimal Density $q^*_{\boldsymbol{\theta}}$

The optimal density $q^*_{\boldsymbol{\theta}}$ has the form

$$q^*_{\boldsymbol{\theta}} \propto \exp\left\{\mathbb{E}_{-\boldsymbol{\theta}}\left[\log p(\boldsymbol{\theta}|\mathbf{y}_{t^*+2:T}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\Sigma})\right]\right\},$$

where the expectation is taken with respect to the marginal density $q_{-\boldsymbol{\theta}}(\boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\Sigma})$, $t^*+2 : T$ denotes the time periods from the period $t^* + 2$ and onwards.

The log-density is

$$\begin{aligned}\log p(\boldsymbol{\theta}|\cdot) =& C_{\boldsymbol{\theta}} - \left(\frac{n}{2}+1\right)\log \bar{s}_2^2 - n\sum_{t=t^*+3}^{T}\log s_t - \frac{1}{2}\sum_{t=t^*+3}^{T}\frac{(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\alpha})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\alpha})}{s_t^2} \\ &- \frac{(\mathbf{y}_{t^*+2} - \mathbf{X}_{t^*+2}\boldsymbol{\alpha})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_{t^*+2} - \mathbf{X}_{t^*+2}\boldsymbol{\alpha})}{2\bar{s}_2^2} + (a-1)\log\rho + (b-1)\log(1-\rho),\end{aligned}$$

where $s_t = 1 + (\bar{s}_2 - 1)\rho^{t-t^*-2}$.

After taking the expectation, we have

$$\begin{aligned}\mathbb{E}_{-\boldsymbol{\theta}}(\log p(\boldsymbol{\theta}|\cdot)) =& C_{\boldsymbol{\theta}} - \left(\frac{n}{2}+1\right)\log \bar{s}_2^2 - n\sum_{t=t^*+3}^{T}\log s_t \\ &- \frac{1}{2}\sum_{t=t^*+3}^{T}\frac{\left[(\mathbf{y}_t - \mathbf{X}_t\bar{\widehat{\boldsymbol{\alpha}}})'\bar{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_t - \mathbf{X}_t\bar{\widehat{\boldsymbol{\alpha}}}) + \mathrm{tr}(\bar{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_t\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{X}_t')\right]}{s_t^2} \\ &- \frac{1}{2\bar{s}_2^2}\left[(\mathbf{y}_{t^*+2} - \mathbf{X}_{t^*+2}\bar{\widehat{\boldsymbol{\alpha}}})'\bar{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_{t^*+2} - \mathbf{X}_{t^*+2}\bar{\widehat{\boldsymbol{\alpha}}}) + \mathrm{tr}(\bar{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{t^*+2}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{X}_{t^*+2}')\right] \\ &+ (a-1)\log\rho + (b-1)\log(1-\rho).\end{aligned}$$

Clearly this is not a standard density function. In this paper, we use grid approximation for this density. In specific, we define a two-dimensional grid for $\boldsymbol{\theta}$, and evaluate the log-density on each point of the grid. Finally, we obtain the approximation of $q^*(\boldsymbol{\theta})$, $\mathbb{E}(\boldsymbol{\theta})$ and $\mathbb{E}(s_t)$, for $t = t^* + 2, ..., T$.

We omit the details for obtaining the optimal density $q^*_{\boldsymbol{\kappa}}$ since they are the same as in VAR-SV.

Variational Lower Bound

Next, we derive the variational lower bound $\underline{p}(\mathbf{y}; q)$. To that end, we first compute the log ratio of the joint posterior density and the variational approximation:

$$
\begin{aligned}
&\log\left[\frac{p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\Sigma},\boldsymbol{\kappa},s_0^2,s_1^2,\boldsymbol{\theta})}{q(\boldsymbol{\kappa})q(\boldsymbol{\alpha})q(\boldsymbol{\Sigma})q(\boldsymbol{\kappa})q(s_0^2)q(s_1^2)q(\boldsymbol{\theta})}\right]\\
&=-\frac{1}{2}\log\det\left(2\pi\mathbf{S}^2\otimes\boldsymbol{\Sigma}\right)-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'(\mathbf{S}^2\otimes\boldsymbol{\Sigma})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})\\
&\quad-\frac{1}{2}\log\det(2\pi\mathbf{V}_{\boldsymbol{\alpha}})-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)'\mathbf{V}_{\boldsymbol{\alpha}}^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\\
&\quad+\frac{\nu}{2}\log\det(\boldsymbol{\Psi})-\frac{n\nu}{2}\log(2)-\log\Gamma_n\left(\frac{\nu}{2}\right)-\frac{\nu+n+1}{2}\log\det(\boldsymbol{\Sigma})-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1})\\
&\quad+2\log\left(\frac{1}{2}\right)-\frac{3}{2}\log\bar{s}_0^2-\frac{3}{2}\log\bar{s}_1^2-\log\bar{s}_2^2\\
&\quad+(a-1)\log\rho+(b-1)\log(1-\rho)-\log\mathcal{B}(a,b)\\
&\quad-\Bigg[-\frac{1}{2}\log\det(2\pi\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1})-\frac{1}{2}(\boldsymbol{\alpha}-\widehat{\bar{\boldsymbol{\alpha}}})'\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}-\widehat{\bar{\boldsymbol{\alpha}}})\\
&\quad+\frac{\nu+T}{2}\log\det(\widehat{\boldsymbol{\Psi}})-\frac{(\nu+T)n}{2}\log 2-\log\Gamma_n\left(\frac{\nu+T}{2}\right)-\frac{\nu+T+n+1}{2}\log\det(\boldsymbol{\Sigma})-\frac{1}{2}\mathrm{tr}(\widehat{\boldsymbol{\Psi}}\boldsymbol{\Sigma}^{-1})\\
&\quad+\nu_{s_0}\log\phi_{s_0}-\log\Gamma(\nu_{s_0})-(\nu_{s_0}+1)\log s_0^2-\frac{\phi_{s_0}}{s_0^2}\\
&\quad+\nu_{s_1}\log\phi_{s_1}-\log\Gamma(\nu_{s_1})-(\nu_{s_1}+1)\log s_1^2-\frac{\phi_{s_1}}{s_1^2}\\
&\quad+C_{\boldsymbol{\theta}}-\left(\frac{n}{2}+1\right)\log\bar{s}_2^2-n\sum_{t=t^*+3}^{T}\log(1+(\bar{s}_2-1)\rho^{t-t^*-2})\\
&\quad-\frac{1}{2}\sum_{t=t^*+3}^{T}\frac{\left[(\mathbf{y}_t-\mathbf{X}_t\widehat{\bar{\boldsymbol{\alpha}}})'\bar{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_t-\mathbf{X}_t\widehat{\bar{\boldsymbol{\alpha}}})+\mathrm{tr}(\bar{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_t\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{X}_t')\right]}{(1+(\bar{s}_2-1)\rho^{t-t^*-2})^2}\\
&\quad-\frac{1}{2\bar{s}_2^2}\left[(\mathbf{y}_{t^*+2}-\mathbf{X}_{t^*+2}\widehat{\bar{\boldsymbol{\alpha}}})'\bar{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_{t^*+2}-\mathbf{X}_{t^*+2}\widehat{\bar{\boldsymbol{\alpha}}})+\mathrm{tr}(\bar{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{t^*+2}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{X}_{t^*+2}')\right]\\
&\quad+(a-1)\log\rho+(b-1)\log(1-\rho)\Bigg]\\
&\quad+c_{\boldsymbol{\kappa}}+\sum_{r=1}^{2}\left[(c_{r,1}-1)\log\kappa_r-c_{r,2}\kappa_r\right]-\sum_{r=1}^{2}\left[(\upsilon_r-1)\log\kappa_r-\frac{a_r\kappa_r+b_r\kappa_r^{-1}}{2}\right],
\end{aligned}
$$

where $c_{\boldsymbol{\kappa}} = \sum_{r=1}^{2} \left[c_{r,1} \log c_{r,2} - \log \Gamma(c_{r,1})\right] - \sum_{r=1}^{2} \left[\frac{v_r}{2}(\log a_r - \log b_r) - \log(2K_{v_r}\sqrt{a_r b_r})\right]$, $K_{v_r}(\cdot)$ is a modified Bessel function of the second kind. Taking expectation of the above log ratio with respect to q, we obtain the following variational lower bound:

$$
\begin{aligned}
\underline{p}(\mathbf{y}; q) &= \mathbb{E}_q \left\{ \log \left[\frac{p(\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}, \boldsymbol{\kappa}, s_0^2, s_1^2, s_2, \rho)}{q(\boldsymbol{\kappa})q(\boldsymbol{\alpha})q(\boldsymbol{\Sigma})q(\boldsymbol{\kappa})q(s_0^2)q(s_1^2)q(s_2)q(\rho)} \right] \right\} \\
&= -\frac{1}{2}\mathbb{E}[\log |\mathbf{V}_{\boldsymbol{\alpha}}|] - \frac{1}{2}(\bar{\widehat{\boldsymbol{\alpha}}} - \boldsymbol{\alpha}_0)'\overline{\mathbf{V}_{\boldsymbol{\alpha}^{-1}}}(\bar{\widehat{\boldsymbol{\alpha}}} - \boldsymbol{\alpha}_0) - \frac{1}{2}\mathrm{tr}(\overline{\mathbf{V}_{\boldsymbol{\alpha}}^{-1}}\widehat{\mathbf{K}}_{\alpha}^{-1}) \\
&\quad + \frac{1}{2}\log |\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}| + \frac{1}{2}(n^2 p + n) \\
&\quad - \frac{\nu + T}{2}\log |\widehat{\boldsymbol{\Psi}}| + \frac{1}{2}\sum_{t=t^*}^{T} \left(\mathbb{E}[s_t^{-2}] \left[\mathrm{tr}(\bar{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\widehat{\mathbf{K}}_{\boldsymbol{\alpha}}^{-1}\mathbf{X}_t') + (\mathbf{y} - \mathbf{X}_t\bar{\widehat{\boldsymbol{\alpha}}})'\bar{\boldsymbol{\Sigma}}^{-1}(\mathbf{y} - \mathbf{X}_t\bar{\widehat{\boldsymbol{\alpha}}}) \right] \right) \\
&\quad - \nu_{s_0}\log \phi_{s_0} + \log \Gamma(\nu_{s_0}) - \nu_{s_1}\log \phi_{s_1} + \log \Gamma(\nu_{s_1}) \\
&\quad - C_{\boldsymbol{\theta}} + c_{\boldsymbol{\kappa}} + \sum_{r=1}^{2} \left[(c_{r,1} - v_r)\overline{\log \kappa_r} - \left(c_{r,2} - \frac{1}{2}a_r \right) \bar{\kappa}_r + \frac{1}{2}b_r\overline{\kappa_r^{-1}} \right] + C,
\end{aligned}
$$

where $C = -\frac{nT}{2}\log 2\pi + \frac{\nu}{2}\log |\boldsymbol{\Psi}| - \frac{n\nu}{2}\log 2 - \log \Gamma_n\left(\frac{\nu}{2}\right) - 2\log(2) + \frac{(\nu+T)n}{2}\log 2 + \log \Gamma_n\left(\frac{\nu+T}{2}\right)$.

B Online Appendix: Variational Importance Sampling

Algorithm 1 Variational Importance Sampling

Input: The optimal density $q^*(\boldsymbol{\theta})$, prior density $p(\boldsymbol{\theta})$, dataset $\mathbf{y}$, sample size M

Output: Logarithm of the estimate of marginal likelihood for the data $\mathbf{y}$: $\log \widehat{p}_{IS}$

for i=1 to M **do**

Draw $\widetilde{\boldsymbol{\theta}}^{(i)} \sim q^*(\boldsymbol{\theta})$

Compute $\log p\left(\mathbf{y}|\widetilde{\boldsymbol{\theta}}^{(i)}\right)$, $\log p\left(\widetilde{\boldsymbol{\theta}}^{(i)}\right)$ and $\log q^*\left(\widetilde{\boldsymbol{\theta}}^{(i)}\right)$

Compute $\log \widehat{p}_{IS}^{(i)} = \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\theta}}^{(i)}\right) + \log p\left(\widetilde{\boldsymbol{\theta}}^{(i)}\right) - \log q^*\left(\widetilde{\boldsymbol{\theta}}^{(i)}\right)$

end for

Compute $\log \widehat{p}_{IS} = \log\left(1/M \sum_{i=1}^{M} \exp\left(\log \widehat{p}_{IS}^{(i)}\right)\right)$

return $\log \widehat{p}_{IS}$

C Online Appendix: Data

Two datasets are used in our application. The first dataset is presented in Carriero et al. (2022b). This dataset consists of 16 monthly variables, including real income, real consumption, industrial production, inflation indexes, etc. The specifics of these variables are outlined in Table 1 of Carriero et al. (2022b). For better reference, we reprint their variables and transformation in Table C.4.

Table C.4: List of Variables used in Carriero et al. (2022b)

Variable	FRED-MD code	Transformation
Real Income	RPI	$\Delta\log(x_t) \times 1200$
Real Consumption	DPCERA3M086SBEA	$\Delta\log(x_t) \times 1200$
IP	INDPRO	$\Delta\log(x_t) \times 1200$
Capacity Utilization	CUMFNS	
Unemployment Rate	UNRATE	
Nonfarm Payrolls	PAYEMS	$\Delta\log(x_t) \times 1200$
Hours	CES0600000007	
Hourly Earnings	CES0600000008	$\Delta\log(x_t) \times 1200$
PPI (Fin. Goods)	WPSFD49207	$\Delta\log(x_t) \times 1200$
PCE Prices	PCEPI	$\Delta\log(x_t) \times 1200$
Housing Starts	HOUST	$\log(x_t)$
S&P 500	SP500	$\Delta\log(x_t) \times 1200$
USD / GBP FX Rate	EXUSUKx	$\Delta\log(x_t) \times 1200$
5-Year Yield	GS5	
10-Year Yield	GS10	
Baa Spread	BAAFFM	

Note: This table is reprinted based on Table 1 in Carriero et al. (2022b). This data set is published in their Github website: `https://github.com/elmarmertens/CCMMoutlierVAR-code/blob/master/README.md`. This data set is obtained from the "2021-04" vintage of FRED-MD database, spanning from 03/01/1959 to 03/01/2021.

The second data set used in our application are from the FRED-QD database. We first transformed the raw data using the "tcode" provided by McCracken and Ng (2020). Then we conducted ADF test for each series, and found that four series, including capacity utilization: manufacturing, average weekly hours of production and nonsupervisory employees: manufacturing, help-wanted index, and Moody's Seasoned Baa Corporate

Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity, cannot reject the null hypothesis of the ADF test. We plot these four series in Figure C.12. Since we assume that the prior means of the Minnesota prior are zeros for all series, we transform these series into stationary series before estimation by taking first differences of their logarithms.

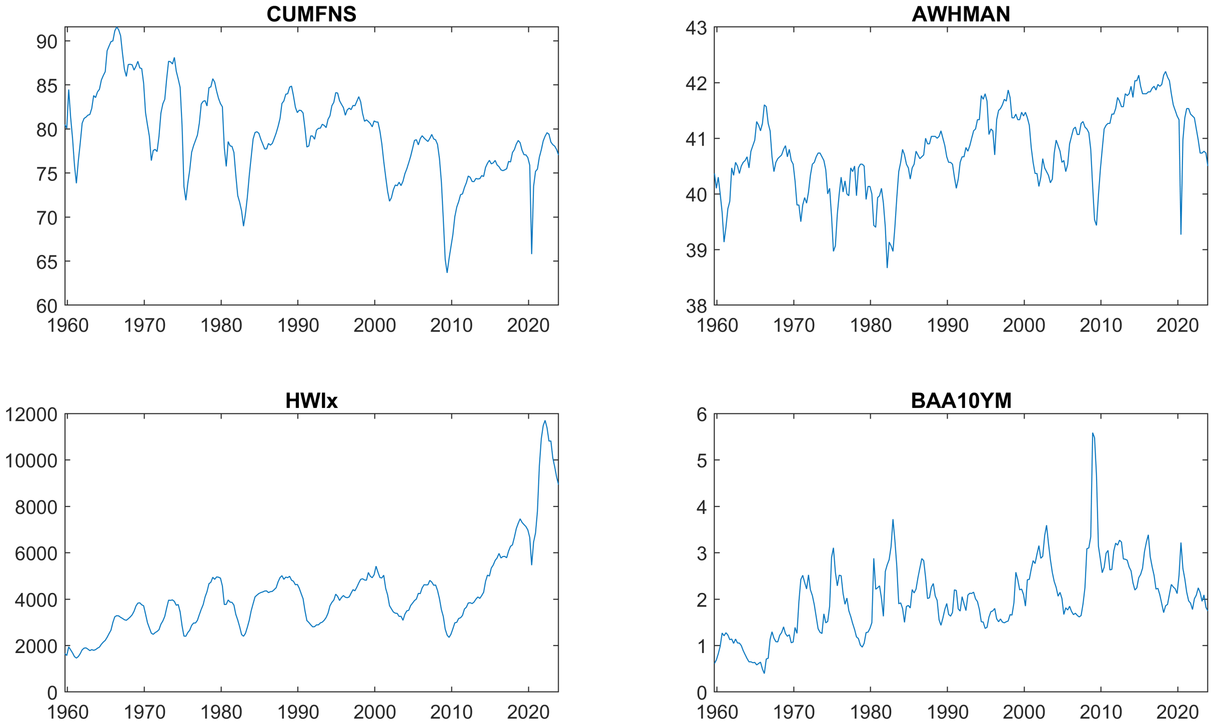

Figure C.12: Four non-stationary series in FRED-QD data set - capacity utilization: manufacturing, average weekly hours of production and nonsupervisory employees: manufacturing, help-wanted index, and Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (from left to right, up to bottom). The p-values for the Augmented Dickey-Fuller test for the four series are 0.54, 0.64, 0.96, and 0.26, respectively.

For better presentation, we summarize the list of the variables and their transformations in Tables C.5 - C.8.

Table C.5: List of Variables

Variable	FRED-QD code	Transformation
Real GDP	GDPC1	$\Delta \log(x_t)$
Real Personal Consumption Expenditures	PCECC96	$\Delta \log(x_t)$
Real personal consumption expenditures: Durable goods	PCDGx	$\Delta \log(x_t)$
Real Personal Consumption Expenditures: Services	PCESVx	$\Delta \log(x_t)$
Real Personal Consumption Expenditures: Nondurable Goods	PCNDx	$\Delta \log(x_t)$
Real Gross Private Domestic Investment	GPDIC1	$\Delta \log(x_t)$
Real private fixed investment	FPIx	$\Delta \log(x_t)$
Real Gross Private Domestic Investment: Fixed Investment: Nonresidential: Equipment	Y033RC1Q027SBEAx	$\Delta \log(x_t)$
Real private fixed investment: Nonresidential	PNFIx	$\Delta \log(x_t)$
Real private fixed investment: Residential	PRFIx	$\Delta \log(x_t)$
Shares of gross domestic product: Gross private domestic investment: Change in private inventories	A014RE1Q156NBEA	
Real Government Consumption Expenditures & Gross Investment	GCEC1	$\Delta \log(x_t)$
Real Government Consumption Expenditures and Gross Investment: Federal	A823RL1Q225SBEA	
Real government state and local consumption expenditures	SLCEx	$\Delta \log(x_t)$
Real Exports of Goods & Services	EXPGSC1	$\Delta \log(x_t)$
Real Imports of Goods & Services	IMPGSC1	$\Delta \log(x_t)$
Real Disposable Personal Income	DPIC96	$\Delta \log(x_t)$
Nonfarm Business Sector: Real Output	OUTNFB	$\Delta \log(x_t)$
Business Sector: Real Output	OUTBS	$\Delta \log(x_t)$
Industrial Production Index	INDPRO	$\Delta \log(x_t)$
Industrial Production: Final Products	IPFINAL	$\Delta \log(x_t)$
Industrial Production: Consumer Goods	IPCONGD	$\Delta \log(x_t)$
Industrial Production: Materials	IPMAT	$\Delta \log(x_t)$
Industrial Production: Durable Materials	IPDMAT	$\Delta \log(x_t)$
Industrial Production: Nondurable Materials	IPNMAT	$\Delta \log(x_t)$
Industrial Production: Durable Consumer Goods	IPDCONGD	$\Delta \log(x_t)$
Industrial Production: Durable Goods: Automotive products	IPB51110SQ	$\Delta \log(x_t)$
Industrial Production: Nondurable Consumer Goods	IPNCONGD	$\Delta \log(x_t)$
Industrial Production: Business Equipment	IPBUSEQ	$\Delta \log(x_t)$
Industrial Production: Consumer energy products	IPB51220SQ	$\Delta \log(x_t)$
Capacity Utilization: Manufacturing	CUMFNS	
All Employees: Total nonfarm	PAYEMS	$\Delta \log(x_t)$
All Employees: Total Private Industries	USPRIV	$\Delta \log(x_t)$
All Employees: Manufacturing	MANEMP	$\Delta \log(x_t)$
All Employees: Service-Providing Industries	SRVPRD	$\Delta \log(x_t)$
All Employees: Goods-Producing Industries	USGOOD	$\Delta \log(x_t)$
All Employees: Durable goods	DMANEMP	$\Delta \log(x_t)$
All Employees: Nondurable goods	NDMANEMP	$\Delta \log(x_t)$
All Employees: Construction	USCONS	$\Delta \log(x_t)$
All Employees: Education & Health Services	USEHS	$\Delta \log(x_t)$
All Employees: Financial Activities	USFIRE	$\Delta \log(x_t)$
All Employees: Information Services	USINFO	$\Delta \log(x_t)$
All Employees: Professional & Business Services	USPBS	$\Delta \log(x_t)$
All Employees: Leisure & Hospitality	USLAH	$\Delta \log(x_t)$
All Employees: Other Services	USSERV	$\Delta \log(x_t)$
All Employees: Mining and logging	USMINE	$\Delta \log(x_t)$
All Employees: Trade, Transportation & Utilities	USTPU	$\Delta \log(x_t)$
All Employees: Government	USGOVT	$\Delta \log(x_t)$
All Employees: Retail Trade	USTRADE	$\Delta \log(x_t)$

Table C.6: List of Variables Continued

Variable	FRED-QD code	Transformation
All Employees: Wholesale Trade	USWTRADE	$\Delta\log(x_t)$
All Employees: Government: Federal	CES9091000001	$\Delta\log(x_t)$
All Employees: Government: State	CES9092000001	$\Delta\log(x_t)$
All Employees: Government: Local	CES9093000001	$\Delta\log(x_t)$
Civilian Employment	CE16OV	$\Delta\log(x_t)$
Civilian Labor Force	CIVPART	Δx_t
Civilian Unemployment Rate	UNRATE	Δx_t
Unemployment Rate less than 27 weeks	UNRATESTx	Δx_t
Unemployment Rate for more than 27 weeks	UNRATELTx	Δx_t
Unemployment Rate - 16 to 19 years	LNS14000012	Δx_t
Unemployment Rate - 20 years and over, Men	LNS14000025	Δx_t
Unemployment Rate - 20 years and over, Women	LNS14000026	Δx_t
Number of Civilians Unemployed - Less Than 5 Weeks	UEMPLT5	$\Delta\log(x_t)$
Number of Civilians Unemployed for 5 to 14 Weeks (Thousands	UEMP5TO14	$\Delta\log(x_t)$
Number of Civilians Unemployed for 15 to 26 Weeks	UEMP15T26	$\Delta\log(x_t)$
Number of Civilians Unemployed for 27 Weeks and Over	UEMP27OV	$\Delta\log(x_t)$
Employment Level - Part-Time for Economic Reasons, All Industries	LNS12032194	$\Delta\log(x_t)$
Business Sector: Hours of All Persons	HOABS	$\Delta\log(x_t)$
Nonfarm Business Sector: Hours of All Persons	HOANBS	$\Delta\log(x_t)$
Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	AWHMAN	
Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)	AWOTMAN	Δx_t
Help-Wanted Index	HWIx	$\Delta\log(x_t)$
Housing Starts: Total: New Privately Owned Housing Units Started	HOUST	$\Delta\log(x_t)$
Privately Owned Housing Starts: 5-Unit Structures or More	HOUST5F	$\Delta\log(x_t)$
Housing Starts in Midwest Census Region	HOUSTMW	$\Delta\log(x_t)$
Housing Starts in Northeast Census Region	HOUSTNE	$\Delta\log(x_t)$
Housing Starts in South Census Region	HOUSTS	$\Delta\log(x_t)$
Housing Starts in West Census Region	HOUSTW	$\Delta\log(x_t)$
Real Manufacturers' New Orders: Durable Goods	AMDMNOx	$\Delta\log(x_t)$
Real Value of Manufacturers' Unlled Orders for Durable Goods Industries	AMDMUOx	$\Delta\log(x_t)$
Personal Consumption Expenditures: Chain-type Price Index	PCECTPI	$\Delta^2\log(x_t)$
Personal Consumption Expenditures Excluding Food and Energy	PCEPILFE	$\Delta^2\log(x_t)$
Gross Domestic Product: Chain-type Price Index	GDPCTPI	$\Delta^2\log(x_t)$
Gross Private Domestic Investment: Chain-type Price Index	GPDICTPI	$\Delta^2\log(x_t)$
Business Sector: Implicit Price Deflator	IPDBS	$\Delta^2\log(x_t)$
Personal consumption expenditures: Goods	DGDSRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Durable goods	DDURRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Services	DSERRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Nondurable goods	DNDGRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Services: Household consumption expenditures	DHCERG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Durable goods: Motor vehicles and parts	DMOTRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Durable goods: Furnishings and durable household equipment	DFDHRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Durable goods: Recreational goods and vehicles	DREQRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Durable goods: Other durable goods	DODGRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Nondurable goods: Food and beverages purchased for off-premises consumption	DFXARG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Nondurable goods: Clothing and footwear	DCLORG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Nondurable goods: Gasoline and other energy goods	DGOERG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Nondurable goods: Other nondurable goods	DONGRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Services: Housing and utilities	DHUTRG3Q086SBEA	$\Delta^2\log(x_t)$
Personal consumption expenditures: Services: Health care	DHLCRG3Q086SBEA	$\Delta^2\log(x_t)$

Table C.7: List of Variables Continued

Variable	FRED-QD code	Transformation
Personal consumption expenditures: Transportation services	DTRSRG3Q086SBEA	$\Delta^2 \log(x_t)$
Personal consumption expenditures: Recreation services	DRCARG3Q086SBEA	$\Delta^2 \log(x_t)$
Personal consumption expenditures: Services: Food services and accommodations	DFSARG3Q086SBEA	$\Delta^2 \log(x_t)$
Personal consumption expenditures: Financial services and insurance	DIFSRG3Q086SBEA	$\Delta^2 \log(x_t)$
Personal consumption expenditures: Other services	DOTSRG3Q086SBEA	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: All Items	CPIAUCSL	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	CPILFESL	$\Delta^2 \log(x_t)$
Producer Price Index by Commodity for Final Demand: Finished Goods	WPSFD49207	$\Delta^2 \log(x_t)$
Producer Price Index for All Commodities	PPIACO	$\Delta^2 \log(x_t)$
Producer Price Index by Commodity for Final Demand: Personal Consumption Goods (Finished Consumer Goods)	WPSFD49502	$\Delta^2 \log(x_t)$
Producer Price Index by Commodity for Finished Consumer Foods	WPSFD4111	$\Delta^2 \log(x_t)$
Producer Price Index by Commodity Industrial Commodities	PPIIDC	$\Delta^2 \log(x_t)$
Producer Price Index by Commodity Intermediate Materials: Supplies & Components	WPSID61	$\Delta^2 \log(x_t)$
Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum	WPU0561	$\Delta \log(x_t)$
Real Average Hourly Earnings of Production and Nonsupervisory Employees: Construction	CES2000000008x	$\Delta \log(x_t)$
Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing	CES3000000008x	$\Delta \log(x_t)$
Nonfarm Business Sector: Real Compensation Per Hour	COMPRNFB	$\Delta \log(x_t)$
Business Sector: Real Compensation Per Hour	RCPHBS	$\Delta \log(x_t)$
Nonfarm Business Sector: Real Output Per Hour of All Persons	OPHNFB	$\Delta \log(x_t)$
Business Sector: Real Output Per Hour of All Persons	OPHPBS	$\Delta \log(x_t)$
Business Sector: Unit Labor Cost	ULCBS	$\Delta \log(x_t)$
Nonfarm Business Sector: Unit Labor Cost	ULCNFB	$\Delta \log(x_t)$
Nonfarm Business Sector: Unit Nonlabor Payments	UNLPNBS	$\Delta \log(x_t)$
Effective Federal Funds Rate	FEDFUNDS	Δx_t
3-Month Treasury Bill: Secondary Market Rate	TB3MS	Δx_t
6-Month Treasury Bill: Secondary Market Rate	TB6MS	Δx_t
1-Year Treasury Constant Maturity Rate	GS1	Δx_t
10-Year Treasury Constant Maturity Rate	GS10	Δx_t
Moody's Seasoned Aaa Corporate Bond Yield	AAA	Δx_t
Moody's Seasoned Baa Corporate Bond Yield	BAA	Δx_t
Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity	BAA10YM	
6-Month Treasury Bill Minus 3-Month Treasury Bill, secondary market	TB6M3Mx	
1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market	GS1TB3Mx	
10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market	GS10TB3Mx	
Monetary Base	BOGMBASEREALx	$\Delta \log(x_t)$
Real M1 Money Stock	M1REAL	$\Delta \log(x_t)$
Real M2 Money Stock	M2REAL	$\Delta \log(x_t)$
Real Commercial and Industrial Loans, All Commercial Banks	BUSLOANSx	$\Delta \log(x_t)$
Real Consumer Loans at All Commercial Banks	CONSUMERx	$\Delta \log(x_t)$
Total Real Nonrevolving Credit Owned and Securitized, Outstanding	NONREVSLx	$\Delta \log(x_t)$
Real Real Estate Loans, All Commercial Banks	REALLNx	$\Delta \log(x_t)$
Total Consumer Credit Outstanding	TOTALSLx	$\Delta \log(x_t)$
Switzerland / U.S. Foreign Exchange Rate	EXSZUSx	$\Delta \log(x_t)$
Japan / U.S. Foreign Exchange Rate	EXJPUSx	$\Delta \log(x_t)$
U.S. / U.K. Foreign Exchange Rate	EXUSUKx	$\Delta \log(x_t)$
Canada / U.S. Foreign Exchange Rate	EXCAUSx	$\Delta \log(x_t)$
Shares of gross domestic product: Exports of goods and services	B020RE1Q156NBEA	Δx_t
Shares of gross domestic product: Imports of goods and services	B021RE1Q156NBEA	Δx_t
Industrial Production: Manufacturing (SIC)	IPMANSICS	$\Delta \log(x_t)$
Industrial Production: Residential Utilities	IPB51222S	$\Delta \log(x_t)$

Table C.8: List of Variables Continued

Variable	FRED-QD code	Transformation
Industrial Production: Fuels	IPFUELS	$\Delta \log(x_t)$
Average (Mean) Duration of Unemployment	UEMPMEAN	Δx_t
Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	CES0600000007	Δx_t
Total Reserves of Depository Institutions	TOTRESNS	$\Delta^2 \log(x_t)$
Reserves Of Depository Institutions, Nonborrowed	NONBORRES	$\Delta(x_t/x_{t-1} - 1.0)$
5-Year Treasury Constant Maturity Rate	GS5	Δx_t
3-Month Treasury Constant Maturity Minus Federal Funds Rate	TB3SMFFM	
5-Year Treasury Constant Maturity Minus Federal Funds Rate	T5YFFM	
Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	AAAFFM	
Producer Price Index: Crude Materials for Further Processing	WPSID62	$\Delta^2 \log(x_t)$
Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals	PPICMM	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: Apparel	CPIAPPSL	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: Transportation	CPITRNSL	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: Medical Care	CPIMEDSL	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: Commodities	CUSR0000SAC	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: Durables	CUSR0000SAD	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: Services	CUSR0000SAS	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: All Items Less Food	CPIULFSL	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: All items less shelter	CUSR0000SA0L2	$\Delta^2 \log(x_t)$
Consumer Price Index for All Urban Consumers: All items less medical	CUSR0000SA0L5	$\Delta^2 \log(x_t)$
Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing	CES0600000008	$\Delta^2 \log(x_t)$
Consumer Motor Vehicle Loans Outstanding Owned by Finance Companies	DTCOLNVHFNM	$\Delta^2 \log(x_t)$
Total Consumer Loans and Leases Outstanding Owned and Securitized by Finance Companies	DTCTHFNM	$\Delta^2 \log(x_t)$
Securities in Bank Credit at All Commercial Banks	INVEST	$\Delta^2 \log(x_t)$
Ratio of Help Wanted/No. Unemployed	HWIURATIOx	Δx_t
Total Business Inventories	BUSINVx	$\Delta \log(x_t)$
Total Business: Inventories to Sales Ratio	ISRATIOx	Δx_t
Nonrevolving consumer credit to Personal Income	CONSPIx	Δx_t
Nikkei Stock Average	NIKKEI225	$\Delta \log(x_t)$
S&P's Common Stock Price Index: Composite	S&P 500	$\Delta \log(x_t)$
S&P's Common Stock Price Index: Industrials	S&P: indust	$\Delta \log(x_t)$

D Online Appendix: Estimation Results

Figure D.13 shows the outlier estimates across time for the 16 macroeconomic series. Figure D.14 shows the outlier-adjusted volatility for the 16 macroeconomic series. Similarly, Figure D.15 shows the estimates for $\mathbf{Q}^2$ across time for the 16 macroeconomic series. Figure D.16 shows the adjusted volatility for the 16 macroeconomic series.

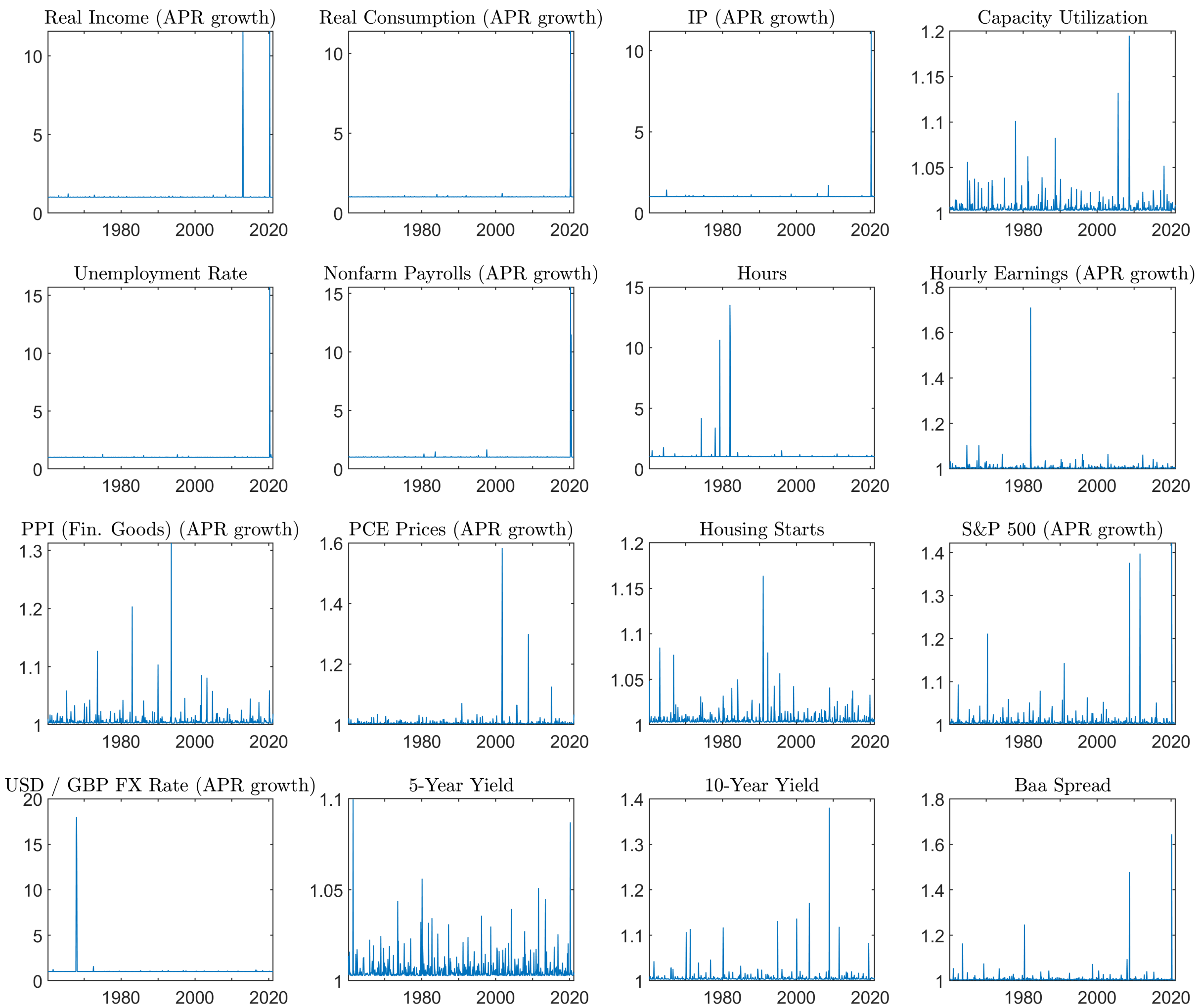

Figure D.13: Outlier estimates for the 16 macroeconomic series from March 1960 to March 2021.

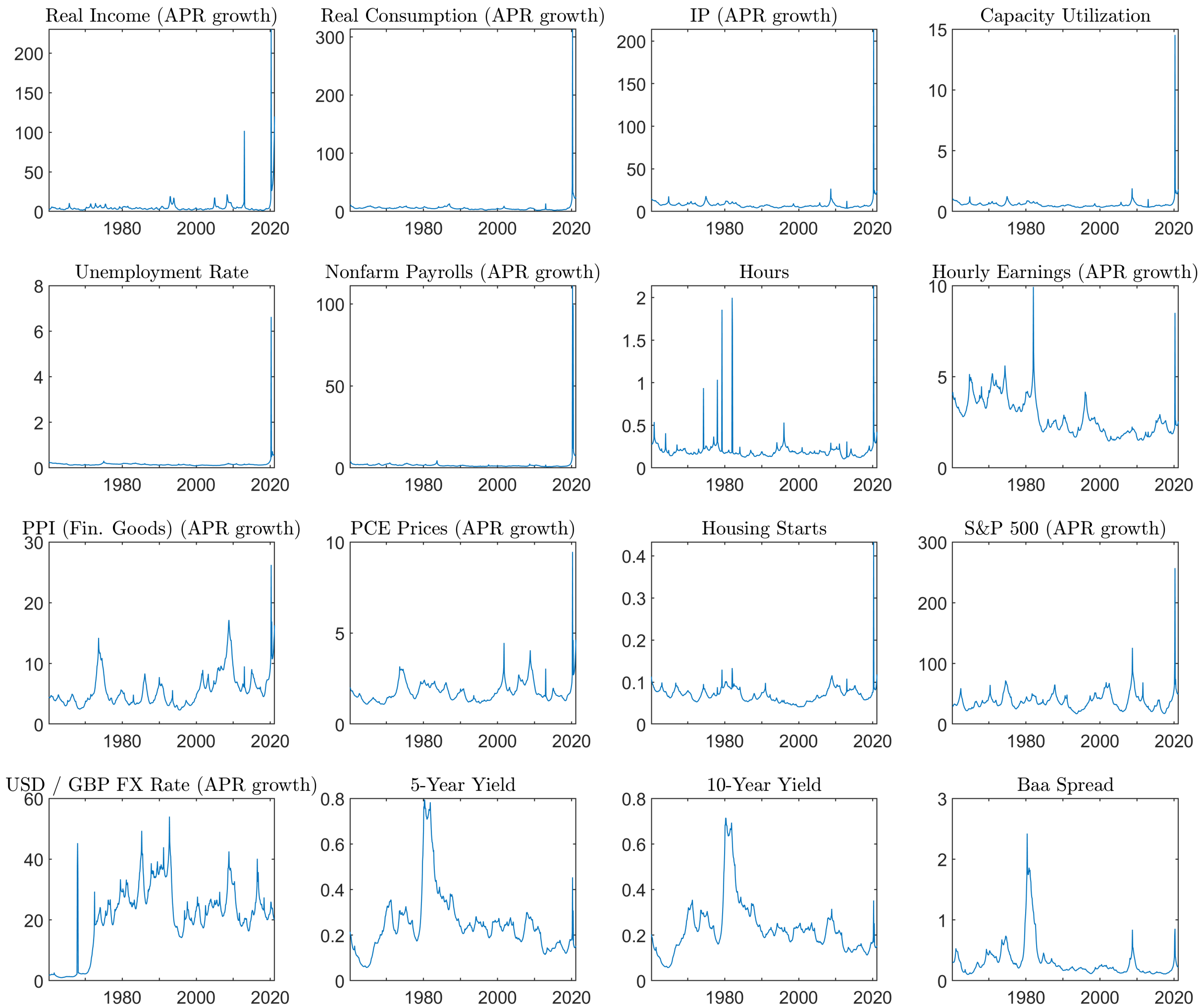

Figure D.14: Volatility estimates for the 16 macroeconomic series using VAR-SVO. Specifically, it is given by the square root of diagonal element of $\widehat{\boldsymbol{\Sigma}}_t = \widehat{\mathbf{B}}_0^{-1}\widehat{\mathbf{O}}_t\widehat{\mathbf{D}}_t\widehat{\mathbf{O}}_t'(\widehat{\mathbf{B}}_0^{-1})'$

.

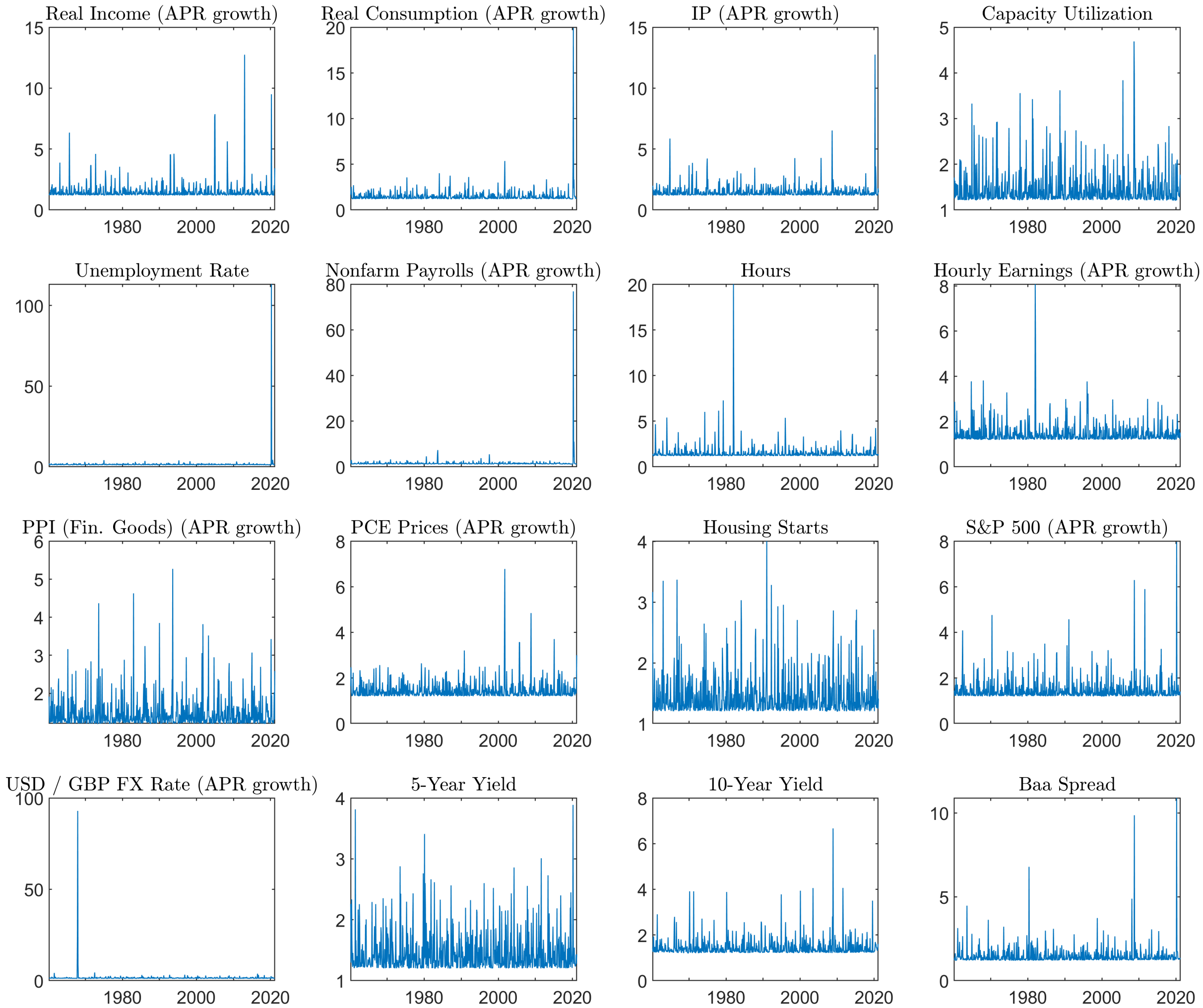

Figure D.15: Estimates for $\mathbf{Q}^2$ for the 16 macroeconomic series from March 1960 to March 2021.

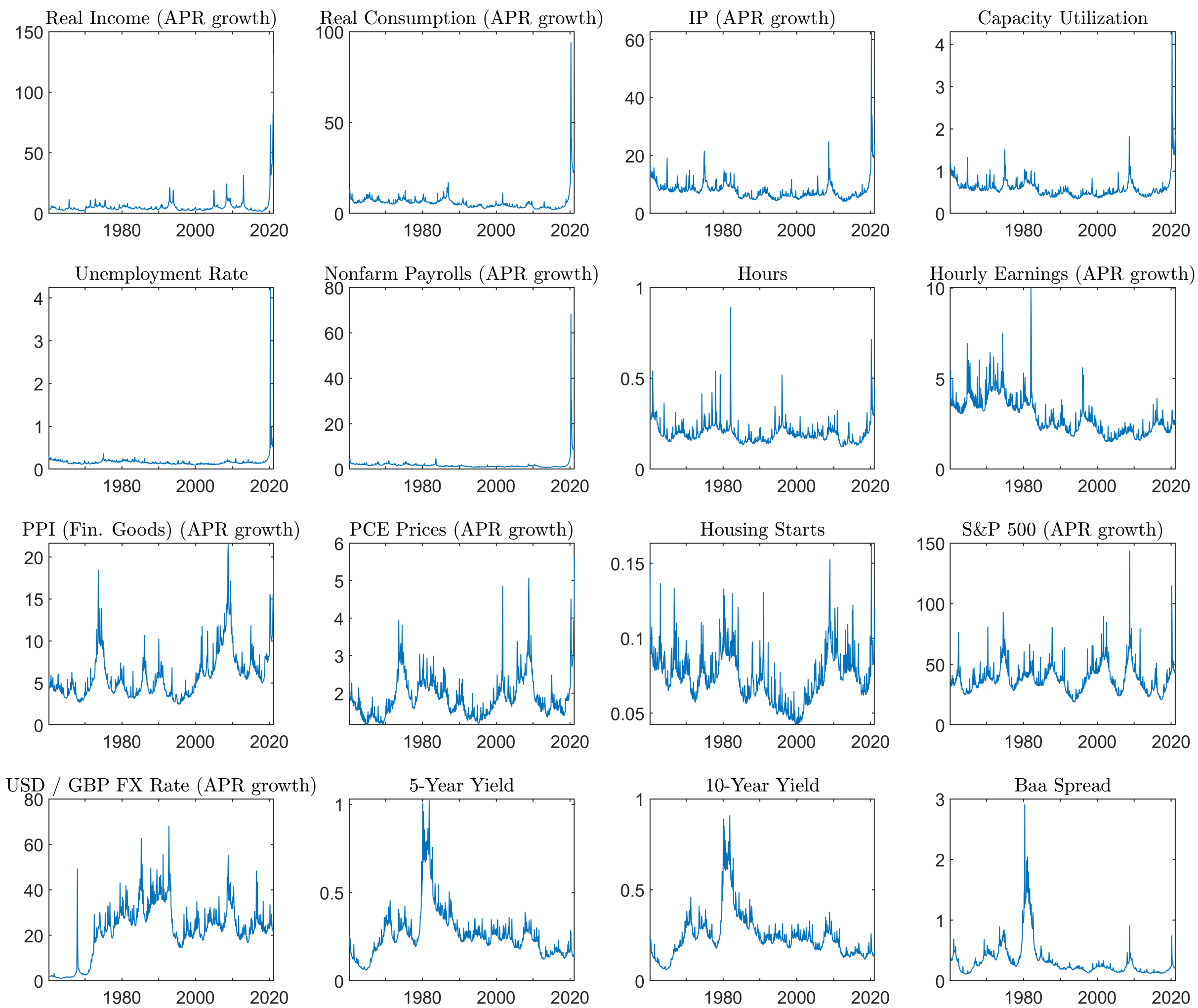

Figure D.16: Volatility estimates for the 16 macroeconomic series using VAR-SVt. Specifically, it is given by the square root of diagonal element of $\widehat{\boldsymbol{\Sigma}}_t = \widehat{\mathbf{B}}_0^{-1}\widehat{\mathbf{Q}}_t\widehat{\mathbf{D}}_t\widehat{\mathbf{Q}}_t'(\widehat{\mathbf{B}}_0^{-1})'$

.