

Dynamic Correlation Multivariate Stochastic Volatility Black-Litterman With Latent Factors

Mihnea S. Andrei¹, Sujit K. Ghosh² & Jian Zou¹

¹ Department of Mathematical Sciences, Worcester Polytechnic Institute, USA

² Department of Statistics, North Carolina State University, USA

Correspondence: Mihnea S. Andrei, Department of Mathematical Sciences, Worcester Polytechnic Institute, USA

Received: December 6, 2020 Accepted: January 6, 2021 Online Published: January 11, 2021

doi:10.5539/ijsp.v10n2p1 URL: <https://doi.org/10.5539/ijsp.v10n2p1>

Abstract

In finance, it is often of interest to study market volatility for portfolios that may consist of a large number of assets using multivariate stochastic volatility models. However, such models, though useful, do not usually incorporate investor views that might be available. In this paper we introduce a novel hierarchical Bayesian methodology of modeling volatility for a large portfolio of assets that incorporates investor's personal views of the market via the Black-Litterman (BL) model. We extend the scope and use of BL models by using it within a multivariate stochastic volatility model based on latent factors for dimensionality reduction but allows for time varying correlations. Detailed derivations of MCMC algorithm are provided with an illustration with *S* & *P*500 asset returns. Moreover, sensitivity analysis for the confidence levels that the investor has in their personal views is also explored. Numerical results show that the proposed method provides flexible interpretation based on the investor's uncertainty in personal beliefs, and converges to the empirical sample estimate when their confidence level of the market becomes weak.

Keywords: Bayesian statistics, factor models, MCMC, portfolio allocation

1. Introduction

In finance, portfolio optimization is one of the quintessential practices and often the goal is to dynamically adjust the portfolio allocation scheme depending on the evolving market volatility. A majority of the stochastic models for asset allocation either use historical data to account for market volatility using a static model (e.g., using multivariate normal distributions after suitable transformations of the returns) or use dynamic models based on various versions of so-called stochastic volatility models that allow for the market volatility to vary over time. However, neither of these models take into account the investor's opinions or experience for future asset allocation along with the historical data. Almost three decades ago, (Black & Litterman, 1992) introduced a novel idea that allows to incorporate investor's views with inputted degrees of uncertainty into standard portfolio theory based on Markowitz models. One of the primary goals of this paper is to embed the Black-Litterman (BL henceforth) model into a dynamic correlation multivariate stochastic volatility (DCMSV henceforth) model (See e.g., (Ku et al., 2014)).

The Black-Litterman model is a portfolio allocation model developed by Fischer Black and Robert Litterman in the early 1990's while they were working for Goldman Sachs. The article published by (He & Litterman, 2002) provided sufficient mathematical details of the BL model that paved the way for other researchers. Previous research on the Black and Litterman model was centered around various optimization methods (Idzorek, n.d.; Bertsimas et al., 2012). The main idea is to solve the optimality condition for the Markowitz mean variance portfolio allocation based on investors personal beliefs. It may not be clear how to generalize this approach to other constrained asset allocation problems. In the past decades more and more researchers realized that the model could be embedded into a Bayesian framework because, as we will see in the next section, equations (1) and (2) are actually Bayesian priors on the means of the returns. One of the authors has 2 papers using precisely this Bayesian approach to Black-Litterman (please see (Andrei & Hsu, 2020) and (Andrei & Hsu, 2021)). Once this observation is made, there is a myriad of other different approaches: some from purely statistical points of view and others coming from finance. Very few incorporate time, while combining ideas from finance and Bayesian statistics (Avramov & Zhou, 2010; Polson & Tew, 2000). In their papers, the time dynamics are embedded through a factor multivariate regression model.

In Section 2, we provide a more detailed background of the BL model before we introduce our proposed version that attempts to combine the BL model with the DCMSV model.

As the BL model is predicated on using investor's opinions it is prudent to consider a full Bayesian hierarchical framework

that allows for the use of formal prior distributions on the historical mean values of the returns to express (*apriori*) views that investors have about future asset performance. In this paper, we propose a novel BL-DCMSV model based on latent factors that incorporates investors views as a prior on the mean of the asset returns. Our full Bayesian approach will also allow the covariance matrix of the returns to be stochastic and evolve over time. Our method has several advantages over the existing method. First, we can directly incorporate the investors preferences and market insights into the model. This is one of the main reasons that popularized the BL model in the finance industry. Second, by using a hierarchical Bayesian framework, we avoid running into issues where the estimate is ill-conditioned for high dimensional applications. This is a well-known problem that arises in estimating large covariance matrices with the sample covariance matrix. Regularization techniques are often used to solve this issue (Ledoit & Wolf, 2004; Wang & Zou, 2010; Zou et al., 2018), which introduces other tuning parameter selection problems. Third, our Bayesian framework naturally incorporates uncertainty of investor’s personal beliefs, and converges to the empirical sample estimate when their confidence level of the market becomes weak, as illustrated in the numerical studies from Section 5. This feature allows an easy stress test under different expert input scenarios.

The paper is organized as follows. In Section 3, we present an extension of the DCMSV model based on latent factors that incorporates investors views as a prior on the mean parameter of the returns. In Section 4, we present the details of the Markov Chain Monte Carlo (MCMC) methods that we employ to obtain the estimates of the parameters. Section 5 includes numerical illustrations of our algorithm using daily returns from the whole *S&P500* and finally we conclude with some final remarks on the possible alternatives and extensions of our models in Section 6.

2. Black-Litterman Asset Allocation Model

Consider a portfolio consisting of p assets and the goal is to obtain an asset allocation model based on the return values. Let the return value for the i -th asset be denoted by $r_i = \frac{P_{1i} - P_{0i}}{P_{0i}}$, where P_{1i} denotes the current price of the i -th stock and P_{0i} the corresponding initial price of that stock. Based on standard practice used in finance, let us assume that the vector of asset returns, denoted by $\mathbf{r} = (r_1, r_2, \dots, r_p)^T$, follows a multivariate normal distribution:

$$\mathbf{r} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1}$$

Black and Litterman proposed the following Capital Asset Pricing Model (CAPM henceforth) prior for the mean of the return:

$$\boldsymbol{\mu} \sim N_p(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma}) \tag{2}$$

In finance, $\boldsymbol{\pi}$ is called the market equilibrium returns and they can be found using the formula:

$$\boldsymbol{\pi} = \delta\boldsymbol{\Sigma}\mathbf{w}_{eq}$$

where the parameter δ is the investor’s risk aversion parameter, \mathbf{w}_{eq} is the vector of p equilibrium weights, and the parameter τ in equation (2) indicates the uncertainty in the CAPM prior. It has been shown in (Cheng, 2008) that one can derive the above equation using the assumptions of the traditional Capital Asset Pricing Model with the extra assumption represented by equations (1) and (2).

In addition to the CAPM prior, Black and Litterman also take investor’s views into consideration. Suppose that the investor has v views. His or her views can be expressed in the equation:

$$\mathbf{P}\boldsymbol{\mu} \sim N_v(\mathbf{q}_0, \boldsymbol{\Omega}_0) \tag{3}$$

where the matrix \mathbf{P} is a $v \times p$ matrix, \mathbf{q}_0 is a $v \times 1$ vector, and $\boldsymbol{\Omega}_0$ is a $v \times v$ matrix, usually diagonal. Each row in \mathbf{P} and \mathbf{q}_0 represent a personal view. By combining equations (2) and (3) it follows that when $\boldsymbol{\Omega}_0$ has very small diagonal elements, one may view $\boldsymbol{\mu}$ as having a multivariate normal distribution approximately supported on the plane $\mathbf{P}\boldsymbol{\mu} \approx \mathbf{q}_0$. To illustrate all terms in equation (3), we consider an example with $p = 4$ assets: Apple Inc (AAPL), Facebook Inc. (FB), Alphabet Inc Class C (GOOG) and Microsoft Corporation (MSFT). Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)^T$ represent the vector of mean returns for AAPL, FB, GOOG, and MSFT, respectively. Suppose that the investor has $v = 2$ views and believes that FB will outperform AAPL by 2% and that GOOG will have returns that amount to 5%. In this case,

$$\mathbf{P} = \begin{matrix} & \dots & \text{AAPL} & \dots & \text{FB} & \dots & \text{GOOG} & \dots & \text{MSFT} & \dots \\ \text{view1} & \mathbf{0}^T & -1 & \mathbf{0}^T & 1 & \mathbf{0}^T & 0 & \mathbf{0}^T & 0 & \mathbf{0}^T \\ \text{view2} & \mathbf{0}^T & 0 & \mathbf{0}^T & 0 & \mathbf{0}^T & 1 & \mathbf{0}^T & 0 & \mathbf{0}^T \end{matrix}, \mathbf{q}_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$$

We remember here that the matrix of views \mathbf{P} is of size $v \times p$ (please see equation (3)). In the simulations we did for the whole *S&P500*, the number of stocks actively traded on a daily basis is 488. Therefore, in practice, the matrix \mathbf{P} can have 2 rows (corresponding to $v = 2$ views) and 488 columns (a column for each stock). Just as presented above, the matrix \mathbf{P}

can have very few entries that are nonzero (here, $\mathbf{0}$ denotes a vector full of zeros). The covariance matrix $\mathbf{\Omega}_0$ is usually a diagonal matrix. The diagonal elements represent the variance of each view. A small value reflects a high confidence in the view and a high value reflects a small confidence in the view.

In (He & Litterman, 2002), the authors discussed the model in detail. They reported that, by combining prior information in equations (2) and (3), the mean of the return follows a normal distribution

$$\boldsymbol{\mu} \sim N(\bar{\boldsymbol{\mu}}, \bar{\mathbf{M}}^{-1}) \tag{4}$$

where

$$\bar{\boldsymbol{\mu}} = [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}_0^{-1}\mathbf{P}]^{-1}[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^T\boldsymbol{\Omega}_0^{-1}\mathbf{q}_0]$$

and

$$\bar{\mathbf{M}}^{-1} = [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}_0^{-1}\mathbf{P}]^{-1}$$

Please notice that the combined prior mean $\bar{\boldsymbol{\mu}}$ is closer to the CAPM prior mean $\boldsymbol{\pi}$ when the uncertainty parameter τ is small, or equivalently, when we are more certain about the CAPM prior.

The unconditional (marginal) distribution of \mathbf{r} is therefore given by

$$\mathbf{r} \sim N(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}), \tag{5}$$

which is obtained by combining equations (1) and (4) and then integrating the vector $\boldsymbol{\mu}$ out, where $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \bar{\mathbf{M}}^{-1}$. Based on the unconditional mean $\bar{\boldsymbol{\mu}}$ and on the covariance matrix $\bar{\boldsymbol{\Sigma}}$ in equation (5), the optimal portfolio can be determined by using the standard mean-variance optimization method to maximize $\mathbf{w}^T\bar{\boldsymbol{\mu}} - \frac{\delta}{2}\mathbf{w}^T\bar{\boldsymbol{\Sigma}}\mathbf{w}$ with respect to the weights \mathbf{w} for an investor with the risk aversion parameter δ . (He & Litterman, 2002) reported that the optimal portfolio weights \mathbf{w}^* can be expressed as

$$\mathbf{w}^* = \frac{1}{1 + \tau}(\mathbf{w}_{eq} + \mathbf{P}^T\mathbf{w}) \tag{6}$$

where

$$\mathbf{w} = \frac{\tau}{\delta}\boldsymbol{\Omega}_0^{-1}\mathbf{q}_0 - \frac{1}{1 + \tau}\mathbf{A}^{-1}\mathbf{P}\boldsymbol{\Sigma}\mathbf{w}_{eq} - \frac{1}{\delta(1 + \tau)}\mathbf{A}^{-1}\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T\boldsymbol{\Omega}_0^{-1}\mathbf{q}_0$$

and

$$\mathbf{A} = \frac{1}{\tau}\boldsymbol{\Omega}_0 + \frac{1}{1 + \tau}\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T.$$

The optimal portfolio weights \mathbf{w}^* can be obtained by plugging in the CAPM prior mean $\boldsymbol{\pi}$, the uncertainty parameter τ , personal views parameters \mathbf{P} , \mathbf{q}_0 , $\boldsymbol{\Omega}_0$, and the covariance matrix $\boldsymbol{\Sigma}$, estimated from historical data. Moreover, from a statistical point of view, the investors views prior (equation (3)) and the CAPM prior (equation (2)) are inconsistent.

Instead, in this paper, we will propose a statistical approach, indeed, a hierarchical Bayesian statistical approach for a large number of stocks. A full Bayesian factor model with stochastic volatility models on each common factor is introduced. The purpose of the factor model is to reduce the dimension of the problem. Otherwise, we would encounter computational and memory allocation problems, as we will see later in the implementation and results sections. The purpose of the stochastic volatilities on the common factors is to model the time varying covariance matrix of the returns. We will keep the investor’s views prior since we still want investors to be able to input their personal views, but we will drop the inconsistent CAPM prior, which contained only historical information. In Section 5, we will see in our simulations using daily returns of the whole S&P500 from January 2nd 2014 to December 29th 2017 that our Bayesian version still uses historical data, especially when the investor is not confident at all in their views (which works according to our intuition).

3. Dynamic Correlation Multivariate Stochastic Volatility Black-Litterman With Latent Factors

Let \mathbf{y}_t for $t = 1, 2, \dots$, denote the vector of returns at the time t relative to the previous time point $t - 1$. More specifically, the i -th component $y_{i,t}$ of \mathbf{y}_t is obtained as $(P_{i,t} - P_{i,t-1})/P_{i,t-1}$, where $P_{i,t}$ denotes the stock price of the i -th asset at time t . One of the most limiting aspects of the Black-Litterman model, as introduced in the previous section, is that it assumes that $\mathbf{y}_t \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e., the covariances among the returns are fixed over time and $\boldsymbol{\Sigma}$ is typically estimated by sample covariance matrix based on say T time points. However when the number of assets is large (e.g., $p = 488$ for the stocks in S&P500 that are actively traded) and we use only a year worth of data ($T = 250$ trading days in a year), then the estimation of the $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ using the sample covariance matrix is poor. We however take a different approach and generalize the equation (1) in the original Black-Litterman model by allowing $\boldsymbol{\Sigma}$ to vary with time by utilizing the framework presented in (Ku et al., 2014) but introducing an additional mean parameter that incorporates investor’s views, represented by equation (3). Moreover, by using a Bayesian inferential framework that allows for (low

dimensional) latent factors, we show that the even when $p > T$, the inference about the asset allocation can be carried out within such a generalized model that not only allows for incorporating the investor’s views but also allows for the correlations to vary dynamically with time t .

3.1 The DCMSV-BL Model

The main idea used by (Ku et al., 2014) is that even when we consider a large portfolio consisting of p assets, there might only be a few, $q \ll p$ latent market factors that may be driving the prices of the p assets. As illustrated by (Ku et al., 2014) often $q = 2$ or 3 suffices even when $p \geq 50$ in many practical scenarios. Of course, this is a strong assumption and it is generally not known how many q latent factors are necessary to adequately capture the cross correlations among the assets, but the model allows for trying different values and then a predictive criteria is used to determine the number of latent factors. For simplicity, we describe the model for a fixed $q \ll p$ latent factors denoted by \mathbf{f}_t , a $q \times 1$ vector and with a $p \times q$ loading matrix \mathbf{B} , the proposed model can be expressed as follows:

$$\begin{cases} \mathbf{y}_t = \boldsymbol{\mu} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t \\ \boldsymbol{\epsilon}_t \stackrel{iid.}{\sim} \mathcal{N}_p(0, \boldsymbol{\Omega}) \\ \boldsymbol{\Omega} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) \end{cases} \Rightarrow \boxed{\mathbf{y}_t | \mathbf{B}, \mathbf{f}_t, \boldsymbol{\Omega}, \boldsymbol{\mu} \sim \mathcal{N}_p(\boldsymbol{\mu} + \mathbf{B}\mathbf{f}_t, \boldsymbol{\Omega})} \quad (7)$$

On this mean $\boldsymbol{\mu}$, we introduce the Black-Litterman prior represented previously by equation (3):

$$\boxed{\mathbf{P}\boldsymbol{\mu} \sim \mathcal{N}_p(\mathbf{q}_0, \boldsymbol{\Omega}_0)} \quad (8)$$

Now, we introduce stochastic volatility models as priors on the common factors \mathbf{f}_t :

$$\begin{cases} \mathbf{f}_t = \mathbf{V}_t^{\frac{1}{2}} \boldsymbol{\epsilon}_t, \mathbf{V}_t = \text{diag}(e^{h_{t1}}, \dots, e^{h_{tq}}) \\ \boldsymbol{\epsilon}_t | \mathbf{Q}_t \sim \mathcal{N}_q(0, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t} = \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}}) \end{cases} \Rightarrow \boxed{\mathbf{f}_t | \mathbf{h}_t, \mathbf{Q}_t \sim \mathcal{N}_q\left(0, \mathbf{V}_t^{\frac{1}{2}} \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{V}_t^{\frac{1}{2}} = \mathbf{V}_t^{\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t} \mathbf{V}_t^{\frac{1}{2}}\right)} \quad (9)$$

The time varying covariance parameters \mathbf{Q}_t ’s are modeled using the so-called inverse Wishart autoregressive models:

- (i) an Autoregressive Inverse Wishart process on \mathbf{Q}_t , for all $t = \{0, 1, \dots, T - 1\}$, with starting point $\mathbf{Q}_0 = \mathbf{I}_q$:

$$\boxed{\mathbf{Q}_{t+1} | k, \mathbf{A}, \mathbf{Q}_t \sim \mathcal{W}_q^{-1}\left(k, k\mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}}\right)} \quad (10)$$

- (ii) an AR(1) process on the volatilities $\mathbf{h}_t = [h_{t1} \ \dots \ h_{tq}]^T$, for all $t = \{1, 2, \dots, T\}$:

$$\begin{cases} \mathbf{h}_{t+1} = \boldsymbol{\mu}_h + \boldsymbol{\Phi} \circ (\mathbf{h}_t - \boldsymbol{\mu}_h) + \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_t \sim \mathcal{N}_q(0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \text{diag}(\sigma_{\eta_1}^2, \dots, \sigma_{\eta_q}^2)) \end{cases} \Rightarrow \boxed{\mathbf{h}_{t+1} \sim \mathcal{N}_q(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_{\boldsymbol{\eta}})} \quad (11)$$

The above specifications provide the conditional sampling distributions of the return vectors and the corresponding prior distributions. The joint posterior distribution of the dynamic and fixed parameters is proportional to the product of the probability density functions of the boxed equations (7), (8), (9), (10) and (11). Clearly, the exact posterior distributions of the parameters can not be obtained in closed form and, therefore, we employ the well known Markov Chain Monte Carlo (MCMC) methods for making inference. In the next few subsections, we derive the full conditionals and then derive the full MCMC sampling scheme.

3.2 Full Conditional Distribution for \mathbf{f}_t

In order to obtain the posterior for \mathbf{f}_t , we collect the terms that depend on it, which are represented by equations (7) and (9):

$$\pi(\mathbf{f}_t | \cdot) \propto \exp\left\{-\frac{1}{2} ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B}\mathbf{f}_t)^T \boldsymbol{\Omega}^{-1} ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B}\mathbf{f}_t)\right\} \exp\left\{-\frac{1}{2} \mathbf{f}_t^T \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \mathbf{f}_t\right\}$$

We combine the 2 exponentials, expand the quadratic term in the first one and we obtain that in the exponentials we have the following term multiplied by $-\frac{1}{2}$:

$$(\mathbf{y}_t - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}) - 2(\mathbf{y}_t - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} \mathbf{B} \mathbf{f}_t + \mathbf{f}_t^T \mathbf{B}^T \boldsymbol{\Omega}^{-1} \mathbf{B} \mathbf{f}_t + \mathbf{f}_t^T \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \mathbf{f}_t = \mathbf{f}_t^T \left(\mathbf{B}^T \boldsymbol{\Omega}^{-1} \mathbf{B} + \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \right) \mathbf{f}_t - 2 \left(\mathbf{B}^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}) \right)^T \mathbf{f}_t + (\mathbf{y}_t - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu})$$

Since we will make use of the following identity repeatedly, which can be proven by expanding the right hand side, let us introduce it here:

Lemma 1. For any n -dimensional vectors \mathbf{x}, \mathbf{b} and for any $n \times n$ invertible matrix \mathbf{M} , the following holds:

$$\mathbf{x}^T \mathbf{M} \mathbf{x} - 2\mathbf{b}^T \mathbf{x} = (\mathbf{x} - \mathbf{M}^{-1} \mathbf{b})^T \mathbf{M} (\mathbf{x} - \mathbf{M}^{-1} \mathbf{b}) - \mathbf{b}^T \mathbf{M}^{-1} \mathbf{b}$$

Coming back to our equation, the term $(\mathbf{y}_t - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu})$ doesn't depend on \mathbf{f}_t , hence the posterior for \mathbf{f}_t doesn't depend on it. Using **Lemma 1** with $\mathbf{x} = \mathbf{f}_t, \mathbf{M} = \mathbf{B}^T \boldsymbol{\Omega}^{-1} \mathbf{B} + \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}}, \mathbf{b} = \mathbf{B}^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu})$, we obtain:

$$\left(\mathbf{f}_t - \left(\mathbf{B}^T \boldsymbol{\Omega}^{-1} \mathbf{B} + \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \right)^{-1} \mathbf{B}^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}) \right)^T \left(\mathbf{B}^T \boldsymbol{\Omega}^{-1} \mathbf{B} + \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \right) \cdot \left(\mathbf{f}_t - \left(\mathbf{B}^T \boldsymbol{\Omega}^{-1} \mathbf{B} + \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \right)^{-1} \mathbf{B}^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}) \right)$$

Please remember that this whole term was multiplied by $-\frac{1}{2}$ and it was plugged inside an exponential function ($\exp \{ \}$). This is exactly the kernel of the following multivariate normal distribution:

$$\mathbf{f}_t | \cdot \stackrel{indep.}{\sim} \mathcal{N}_q \left(\boldsymbol{\Sigma}_{\mathbf{f}_t} \mathbf{B}^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}), \boldsymbol{\Sigma}_{\mathbf{f}_t} = \left(\mathbf{B}^T \boldsymbol{\Omega}^{-1} \mathbf{B} + \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_t}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \right)^{-1} \right) \tag{12}$$

3.3 Full Conditional Distribution for \mathbf{B}

Again, the only distributions in the likelihood that depend on \mathbf{B} are equation (7) and the prior that we introduce on each entry in the matrix:

(i) Factors: $\mathbf{y}_t | \boldsymbol{\mu}, \mathbf{B}, \mathbf{f}_t, \boldsymbol{\Omega} \sim \mathcal{N}_p(\boldsymbol{\mu} + \mathbf{B} \mathbf{f}_t, \boldsymbol{\Omega})$

(ii) Prior: $\mathbf{B}_{ji} \sim \mathcal{N}(0, c_0^2), \forall j = \overline{1, p}, i = \overline{1, q}$

The kernel of the posterior for \mathbf{B} only depends on:

$$\pi(\mathbf{B} | \cdot) \propto \exp \left\{ -\frac{1}{2} \left[\sum_{t=1}^T ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B} \mathbf{f}_t)^T \boldsymbol{\Omega}^{-1} ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B} \mathbf{f}_t) + \sum_{i=1}^q \sum_{j=1}^p \frac{\mathbf{B}_{ji}^2}{c_0^2} \right] \right\}. \tag{13}$$

The problem with combining the two sums is that, as they are written, the first one goes over time, while the other two go over what we can call "space" (i.e. in this case the p variables and the q factors). In order to be able to combine the sums into a single one, we need them to be indexed in the same way.

The first step is to expand the quadratic term from the exponential over the rows (the p entries in \mathbf{y}_t) and to show that the following identity holds:

$$\sum_{t=1}^T ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B} \mathbf{f}_t)^T \boldsymbol{\Omega}^{-1} ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B} \mathbf{f}_t) = \sum_{t=1}^T \sum_{r=1}^p ((y_{tr} - \mu_r) - \mathbf{B}_r^T \mathbf{f}_t)^2 \cdot \frac{1}{\sigma_r^2}. \tag{14}$$

Here, the index going over rows is r and \mathbf{B}_r denotes the r^{th} row in the matrix \mathbf{B} written as a column vector. This identity holds because if we look at one of the terms inside the sum on the left hand side, we obtain that:

$$\begin{aligned} ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B}\mathbf{f}_t)^T \boldsymbol{\Omega}^{-1} ((\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{B}\mathbf{f}_t) &= \begin{bmatrix} \mathbf{y}_{1t} - \boldsymbol{\mu}_1 - \mathbf{B}_1^T \mathbf{f}_t \\ \vdots \\ \mathbf{y}_{pt} - \boldsymbol{\mu}_p - \mathbf{B}_p^T \mathbf{f}_t \end{bmatrix}^T \begin{bmatrix} \frac{1}{\sigma_1^2} & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \frac{1}{\sigma_p^2} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1t} - \boldsymbol{\mu}_1 - \mathbf{B}_1^T \mathbf{f}_t \\ \vdots \\ \mathbf{y}_{pt} - \boldsymbol{\mu}_p - \mathbf{B}_p^T \mathbf{f}_t \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{y}_{1t} - \boldsymbol{\mu}_1 - \mathbf{B}_1^T \mathbf{f}_t & \dots & \mathbf{y}_{pt} - \boldsymbol{\mu}_p - \mathbf{B}_p^T \mathbf{f}_t \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \frac{1}{\sigma_p^2} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1t} - \boldsymbol{\mu}_1 - \mathbf{B}_1^T \mathbf{f}_t \\ \vdots \\ \mathbf{y}_{pt} - \boldsymbol{\mu}_p - \mathbf{B}_p^T \mathbf{f}_t \end{bmatrix} \\ &= \left[\frac{1}{\sigma_1^2} (\mathbf{y}_{1t} - \boldsymbol{\mu}_1 - \mathbf{B}_1^T \mathbf{f}_t) \quad \dots \quad \frac{1}{\sigma_p^2} (\mathbf{y}_{pt} - \boldsymbol{\mu}_p - \mathbf{B}_p^T \mathbf{f}_t) \right] \begin{bmatrix} \mathbf{y}_{1t} - \boldsymbol{\mu}_1 - \mathbf{B}_1^T \mathbf{f}_t \\ \vdots \\ \mathbf{y}_{pt} - \boldsymbol{\mu}_p - \mathbf{B}_p^T \mathbf{f}_t \end{bmatrix} = \sum_{r=1}^p (\mathbf{y}_{rt} - \boldsymbol{\mu}_r - \mathbf{B}_r^T \mathbf{f}_t)^2 \cdot \frac{1}{\sigma_r^2} \end{aligned}$$

Hence, the identity in equation (14) holds. Similarly, by starting from the right hand side, we can also show that:

$$\sum_{t=1}^T \sum_{r=1}^p (\mathbf{y}_{rt} - \boldsymbol{\mu}_r - \mathbf{B}_r^T \mathbf{f}_t)^2 \cdot \frac{1}{\sigma_r^2} = \sum_{r=1}^p (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1} - \mathbf{F}^T \mathbf{B}_r)^T \frac{1}{\sigma_r^2} (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1} - \mathbf{F}^T \mathbf{B}_r) \tag{15}$$

For simplicity, we use the following notations:

- (i) $\mathbf{y} = [\mathbf{y}_1 \quad \dots \quad \mathbf{y}_T]$ is now a matrix that has as columns time and as rows the variables.
- (ii) $\mathbf{F} = [\mathbf{f}_1 \quad \dots \quad \mathbf{f}_T]$ is a matrix that has as columns the common factors.
- (iii) \mathbf{B}_r is the r^{th} row in the matrix \mathbf{B} written as a column vector.

Hence, using the identities represented by equations (14) and (15), we actually showed that:

$$\sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{B}\mathbf{f}_t)^T \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{B}\mathbf{f}_t) = \sum_{r=1}^p (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1} - \mathbf{F}^T \mathbf{B}_r)^T \frac{1}{\sigma_r^2} (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1} - \mathbf{F}^T \mathbf{B}_r)$$

So, we managed to re-write the sum in the first term in the exponential from equation (13). We are left with re-writing the second sum:

$$\sum_{i=1}^q \sum_{j=1}^p \frac{\mathbf{B}_{ji}^2}{c_0^2} = \frac{1}{c_0^2} \sum_{r=1}^p \mathbf{B}_r^T \mathbf{B}_r$$

Now that finally both sums are indexed in the same way, we can re-write the exponential in the posterior for \mathbf{B} represented by equation (13):

$$\sum_{r=1}^p \left[(\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1} - \mathbf{F}^T \mathbf{B}_r)^T \frac{1}{\sigma_r^2} (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1} - \mathbf{F}^T \mathbf{B}_r) + \mathbf{B}_r^T \mathbf{B}_r \cdot \frac{1}{c_0^2} \right]$$

Let us focus on expanding the first quadratic term:

$$(\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1})^T \frac{1}{\sigma_r^2} (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1}) - 2 (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1})^T \frac{1}{\sigma_r^2} \mathbf{F}^T \mathbf{B}_r + \mathbf{B}_r^T \mathbf{F} \frac{1}{\sigma_r^2} \mathbf{F}^T \mathbf{B}_r + \frac{1}{c_0^2} \mathbf{B}_r^T \mathbf{B}_r$$

The kernel of the posterior for \mathbf{B} depends only on the terms that contain \mathbf{B} . Hence, the first quadratic term is a constant for the posterior for \mathbf{B} :

$$-2 \left(\mathbf{F} \frac{1}{\sigma_r^2} (\mathbf{y}_r - \boldsymbol{\mu}_r \mathbf{1}) \right)^T \mathbf{B}_r + \mathbf{B}_r^T \left(\frac{1}{\sigma_r^2} \mathbf{F} \mathbf{F}^T + \frac{1}{c_0^2} \mathbf{I}_q \right) \mathbf{B}_r$$

New, we can apply **Lemma 1** with $\mathbf{x} = \mathbf{B}_r$, $\mathbf{M} = \frac{1}{\sigma_r^2} \mathbf{F} \mathbf{F}^T + \frac{1}{c_0^2} \mathbf{I}_q$, $\mathbf{b} = \mathbf{F} \frac{1}{\sigma_r^2} (\mathbf{y}_r - \mu_r \mathbf{1})$ and we obtain that it is actually equal to:

$$\left(\mathbf{B}_r - \left(\frac{1}{\sigma_r^2} \mathbf{F} \mathbf{F}^T + \frac{1}{c_0^2} \mathbf{I}_q \right)^{-1} \mathbf{F} \frac{1}{\sigma_r^2} (\mathbf{y}_r - \mu_r \mathbf{1}) \right)^T \left(\frac{1}{\sigma_r^2} \mathbf{F} \mathbf{F}^T + \frac{1}{c_0^2} \mathbf{I}_q \right) \left(\mathbf{B}_r - \left(\frac{1}{\sigma_r^2} \mathbf{F} \mathbf{F}^T + \frac{1}{c_0^2} \mathbf{I}_q \right)^{-1} \mathbf{F} \frac{1}{\sigma_r^2} (\mathbf{y}_r - \mu_r \mathbf{1}) \right)$$

This implies that the posterior distribution of the rows in the matrix of loadings \mathbf{B}_r is:

$$\mathbf{B}_r | \cdot \stackrel{indep.}{\sim} \mathcal{N}_q \left(\frac{1}{\sigma_r^2} \Sigma_{\mathbf{B}_r} \mathbf{F} (\mathbf{y}_r - \mu_r \mathbf{1}), \Sigma_{\mathbf{B}_r} = \left(\frac{1}{\sigma_r^2} \mathbf{F} \mathbf{F}^T + \frac{1}{c_0^2} \mathbf{I}_q \right)^{-1} \right) \tag{16}$$

3.4 Full Conditional Distribution for μ

Similarly to what we did in the previous section, the only distributions in the likelihood that depend on μ are

(i) Factors: $\mathbf{y}_t | \mu, \mathbf{B}, \mathbf{f}_t, \Omega \sim \mathcal{N}_p(\mu + \mathbf{B} \mathbf{f}_t, \Omega)$

(ii) Views prior: $\mathbf{P} \mu \sim \mathcal{N}_v(\mathbf{q}_0, \Omega_0)$

Hence, by a similar argument as before, we focus on the kernel of μ from the likelihood. The posterior for μ is proportional to the following function:

$$\pi(\mu | \cdot) \propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \mu - \mathbf{B} \mathbf{f}_t)^T \Omega^{-1} (\mathbf{y}_t - \mu - \mathbf{B} \mathbf{f}_t) \right\} \exp \left\{ -\frac{1}{2} (\mathbf{P} \mu - \mathbf{q}_0)^T \Omega_0^{-1} (\mathbf{P} \mu - \mathbf{q}_0) \right\}$$

Let us simplify the formulas that we will be working with by introducing the following notation $\mathbf{y}_t^* = \mathbf{y}_t - \mathbf{B} \mathbf{f}_t$. With this, let us focus only on the terms in the exponential, without the multiplication factor $-\frac{1}{2}$:

$$\sum_{t=1}^T (\mathbf{y}_t^* - \mu)^T \Omega^{-1} (\mathbf{y}_t^* - \mu) + (\mathbf{P} \mu - \mathbf{q}_0)^T \Omega_0^{-1} (\mathbf{P} \mu - \mathbf{q}_0) \tag{17}$$

Now, we notice that the first sum (which was multiplied by $-\frac{1}{2}$ at an exponential) is the likelihood of a sample $\mathbf{y}_1^*, \dots, \mathbf{y}_T^* \stackrel{iid.}{\sim} \mathcal{N}_p(\mu, \Omega)$. Hence, we can use the following well known identity for this first sum:

Lemma 2. *The following equality holds, where $\bar{\mathbf{y}}^* = \frac{\sum_{t=1}^T \mathbf{y}_t^*}{T}$:*

$$\sum_{t=1}^T (\mathbf{y}_t^* - \mu)^T \Omega^{-1} (\mathbf{y}_t^* - \mu) = \sum_{t=1}^T (\mathbf{y}_t^* - \bar{\mathbf{y}}^*)^T \Omega^{-1} (\mathbf{y}_t^* - \bar{\mathbf{y}}^*) + T(\bar{\mathbf{y}}^* - \mu)^T \Omega^{-1} (\bar{\mathbf{y}}^* - \mu)$$

Using this identity, we can go back into equation (17) and we obtain that:

$$\begin{aligned} & \sum_{t=1}^T (\mathbf{y}_t^* - \mu)^T \Omega^{-1} (\mathbf{y}_t^* - \mu) + (\mathbf{P} \mu - \mathbf{q}_0)^T \Omega_0^{-1} (\mathbf{P} \mu - \mathbf{q}_0) \\ &= \sum_{t=1}^T (\mathbf{y}_t^* - \bar{\mathbf{y}}^*)^T \Omega^{-1} (\mathbf{y}_t^* - \bar{\mathbf{y}}^*) + T(\bar{\mathbf{y}}^* - \mu)^T \Omega^{-1} (\bar{\mathbf{y}}^* - \mu) + (\mathbf{P} \mu - \mathbf{q}_0)^T \Omega_0^{-1} (\mathbf{P} \mu - \mathbf{q}_0) \end{aligned}$$

The first sum does not depend on μ and, just like before, we will focus only on its kernel. After expanding the quadratic terms we obtain:

$$T \mu^T \Omega^{-1} \mu - 2 T \bar{\mathbf{y}}^{*T} \Omega^{-1} \mu + \mu^T \mathbf{P}^T \Omega_0^{-1} \mathbf{P} \mu - 2 \mathbf{q}_0^T \Omega_0^{-1} \mathbf{P} \mu = \mu^T (T \Omega^{-1} + \mathbf{P}^T \Omega_0^{-1} \mathbf{P}) \mu - 2 (T \bar{\mathbf{y}}^{*T} \Omega^{-1} + \mathbf{q}_0^T \Omega_0^{-1} \mathbf{P}) \mu$$

We apply again **Lemma 1** with $\mathbf{x} = \mu$, $\mathbf{M} = T \Omega^{-1} + \mathbf{P}^T \Omega_0^{-1} \mathbf{P}$, $\mathbf{b} = T \Omega^{-1} \bar{\mathbf{y}}^* + \mathbf{P}^T \Omega_0^{-1} \mathbf{q}_0$ and we obtain that the posterior for μ is proportional to (again, multiplied by $-\frac{1}{2}$ at an exponential):

$$\left(\mu - (T \Omega^{-1} + \mathbf{P}^T \Omega_0^{-1} \mathbf{P})^{-1} (T \Omega^{-1} \bar{\mathbf{y}}^* + \mathbf{P}^T \Omega_0^{-1} \mathbf{q}_0) \right)^T (T \Omega^{-1} + \mathbf{P}^T \Omega_0^{-1} \mathbf{P}) \left(\mu - (T \Omega^{-1} + \mathbf{P}^T \Omega_0^{-1} \mathbf{P})^{-1} (T \Omega^{-1} \bar{\mathbf{y}}^* + \mathbf{P}^T \Omega_0^{-1} \mathbf{q}_0) \right)$$

This is the kernel of a normal distribution:

$$\mu | \cdot \sim \mathcal{N} \left(\Sigma \mu (T \Omega^{-1} \bar{\mathbf{y}}^* + \mathbf{P}^T \Omega_0^{-1} \mathbf{q}_0), \Sigma \mu = (T \Omega^{-1} + \mathbf{P}^T \Omega_0^{-1} \mathbf{P})^{-1} \right)$$

3.5 Full Conditional Distribution for \mathbf{Q}_t

Again, the only distributions that depend on \mathbf{Q}_t , for $t = \{1, \dots, T - 1\}$ are:

- (i) $\mathbf{Q}_t|k, \mathbf{A}, \mathbf{Q}_{t-1} \sim \mathcal{W}_q^{-1} \left(k, k\mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_{t-1}^{\frac{d}{2}} \right)$
- (ii) $\mathbf{Q}_{t+1}|k, \mathbf{A}, \mathbf{Q}_t \sim \mathcal{W}_q^{-1} \left(k, k\mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}} \right)$
- (iii) $\mathbf{f}_t|\mathbf{h}_t, \mathbf{Q}_t \sim \mathcal{N}_q \left(0, \mathbf{V}_t^{-\frac{1}{2}} \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{V}_t^{-\frac{1}{2}} = \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{\epsilon}_t \mathbf{V}_t^{-\frac{1}{2}} \right)$

Hence, the kernel of the posterior for \mathbf{Q}_t , if $t = \{1, 2, \dots, T - 1\}$ is:

$$\begin{aligned} \pi(\mathbf{Q}_t|\cdot) \propto & |\mathbf{Q}_t|^{-\frac{k+q+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left(k\mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{Q}_t^{-1} \right) \right\} \left| k\mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}} \right|^{\frac{k}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left(k\mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}} \mathbf{Q}_{t+1}^{-1} \right) \right\} \\ & \cdot \left| \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{\epsilon}_t \mathbf{V}_t^{-\frac{1}{2}} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{f}_t^T \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \mathbf{V}_t^{-\frac{1}{2}} \mathbf{f}_t \right\} \end{aligned}$$

But from (9), we know that $\mathbf{f}_t = \mathbf{V}_t^{-\frac{1}{2}} \boldsymbol{\epsilon}_t \Rightarrow \boldsymbol{\epsilon}_t = \mathbf{V}_t^{-\frac{1}{2}} \mathbf{f}_t$. Therefore, by substituting this into the last exponential, by multiplying and dividing $\exp \left\{ \frac{1}{2} \boldsymbol{\epsilon}_t^T \mathbf{Q}_t^{-1} \boldsymbol{\epsilon}_t \right\}$ and by remembering that we introduced the notation $\boldsymbol{\Sigma} \boldsymbol{\epsilon}_t = \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}}$, we obtain:

$$\begin{aligned} \pi(\mathbf{Q}_t|\cdot) \propto & |\mathbf{Q}_t|^{\frac{kd-k-q-1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left(k\mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{Q}_t^{-1} \right) - \frac{1}{2} \boldsymbol{\epsilon}_t^T \mathbf{Q}_t^{-1} \boldsymbol{\epsilon}_t \right\} \\ & \cdot \exp \left\{ \frac{1}{2} \boldsymbol{\epsilon}_t^T \mathbf{Q}_t^{-1} \boldsymbol{\epsilon}_t \right\} \left| \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \right|^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \left(k\mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}} \mathbf{Q}_{t+1}^{-1} \right) \right\} \\ & \cdot \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}_t^T \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \boldsymbol{\epsilon}_t \right\} \end{aligned}$$

Remark 1. Please notice that some of the quadratic terms are scalars:

(i) since $\frac{1}{2} \boldsymbol{\epsilon}_t^T \mathbf{Q}_t^{-1} \boldsymbol{\epsilon}_t \in \mathbb{R}$ and by also using the fact that $\text{Tr}(\cdot)$ is cyclically commutative, we obtain that:

$$\frac{1}{2} \boldsymbol{\epsilon}_t^T \mathbf{Q}_t^{-1} \boldsymbol{\epsilon}_t = \frac{1}{2} \text{Tr} \left(\boldsymbol{\epsilon}_t^T \mathbf{Q}_t^{-1} \boldsymbol{\epsilon}_t \right) = \frac{1}{2} \text{Tr} \left(\mathbf{Q}_t^{-1} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right) = \frac{1}{2} \text{Tr} \left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \mathbf{Q}_t^{-1} \right)$$

(ii) since $\frac{1}{2} \boldsymbol{\epsilon}_t^T \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \boldsymbol{\epsilon}_t \in \mathbb{R}$ and again by using the fact that $\text{Tr}(\cdot)$ is cyclically commutative, we obtain that:

$$\frac{1}{2} \boldsymbol{\epsilon}_t^T \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \boldsymbol{\epsilon}_t = \frac{1}{2} \text{Tr} \left(\boldsymbol{\epsilon}_t^T \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \boldsymbol{\epsilon}_t \right) = \frac{1}{2} \text{Tr} \left(\text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right)$$

Using those identities in our previous equation, we obtain:

$$\begin{aligned} \pi(\mathbf{Q}_t|\cdot) \propto & |\mathbf{Q}_t|^{-\frac{k+1+q+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left(\left(k\mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_{t-1}^{\frac{d}{2}} + \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right) \mathbf{Q}_t^{-1} \right) \right\} \\ & \cdot \left| \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \right|^{-\frac{1}{2}} |\mathbf{Q}_t|^{\frac{kd+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left(k\mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}} \mathbf{Q}_{t+1}^{-1} \right) \right\} \\ & \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \left(\left(\text{diag}(\mathbf{Q}_t)^{\frac{1}{2}} \mathbf{Q}_t^{-1} \text{diag}(\mathbf{Q}_t)^{\frac{1}{2}} - \mathbf{Q}_t^{-1} \right) \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right) \right\} \\ & = \mathcal{W}^{-1} \left(k + 1, k\mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_{t-1}^{\frac{d}{2}} + \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right) \left| \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-\frac{1}{2}} \right|^{-\frac{1}{2}} |\mathbf{Q}_t|^{\frac{kd+1}{2}} \\ & \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \left(k\mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}} \mathbf{Q}_{t+1}^{-1} \right) \right\} \exp \left\{ -\frac{1}{2} \text{Tr} \left(\left(\text{diag}(\mathbf{Q}_t)^{\frac{1}{2}} \mathbf{Q}_t^{-1} \text{diag}(\mathbf{Q}_t)^{\frac{1}{2}} - \mathbf{Q}_t^{-1} \right) \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right) \right\} \end{aligned}$$

That is, the posterior for \mathbf{Q}_t for $t = \{1, \dots, T - 1\}$ is distributed as an Inverse Wishart times another term. This is exactly what was reported in (Asai & McAleer, 2009). As noticed in the paper and as we will see in here in the implementation

section, a Metropolis-Hastings algorithm is deployed to sample from this distribution and only the extra terms matter in the ratio and not $\mathcal{W}^{-1}\left(k+1, k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}} + \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\right)$, which denotes the kernel of the Inverse Wishart, with the specified parameters, computed at \mathbf{Q}_t .

When it comes to the distributions that depend on \mathbf{Q}_T , we only have one Inverse Wishart, instead of 2, as we have seen for \mathbf{Q}_t when $t = \{1, \dots, T - 1\}$:

- (i) $\mathbf{Q}_T|k, \mathbf{A}, \mathbf{Q}_{T-1} \sim \mathcal{W}_p^{-1}\left(k, k\mathbf{Q}_{T-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{T-1}^{\frac{d}{2}}\right)$
- (ii) $\mathbf{f}_T|\mathbf{h}_T, \mathbf{Q}_T \sim \mathcal{N}_q\left(0, \mathbf{V}_T^{\frac{1}{2}}\text{diag}(\mathbf{Q}_T)^{-\frac{1}{2}}\mathbf{Q}_T\text{diag}(\mathbf{Q}_T)^{-\frac{1}{2}}\mathbf{V}_T^{\frac{1}{2}} = \mathbf{V}_T^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\epsilon}_T\mathbf{V}_T^{\frac{1}{2}}\right)$

The steps are the exact same as before. As (Asai & McAleer, 2009) report, we obtain that the posterior for \mathbf{Q}_T is:

$$\pi(\mathbf{Q}_T|\cdot) \propto \mathcal{W}^{-1}\left(k+1, k\mathbf{Q}_{T-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{T-1}^{\frac{d}{2}} + \boldsymbol{\epsilon}_T\boldsymbol{\epsilon}_T^T\right) \cdot \exp\left\{-\frac{1}{2}\text{Tr}\left(\left(\text{diag}(\mathbf{Q}_T)^{\frac{1}{2}}\mathbf{Q}_T^{-1}\text{diag}(\mathbf{Q}_T)^{\frac{1}{2}} - \mathbf{Q}_T^{-1}\right)\boldsymbol{\epsilon}_T\boldsymbol{\epsilon}_T^T\right)\right\}$$

3.6 Full Conditional Distribution for \mathbf{A}

The only distributions that depend on \mathbf{A} are:

- (i) $\mathbf{Q}_t|k, \mathbf{A}, \mathbf{Q}_{t-1} \sim \mathcal{W}_q^{-1}\left(k, k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right), \forall t = \{1, \dots, T\}$
- (ii) $\mathbf{A} \sim \mathcal{W}_q^{-1}\left(\nu_0, \mathbf{C}_0^{-1}\right)$

Hence, by multiplying all of them and focusing only on the kernel of \mathbf{A} , we obtain:

$$\begin{aligned} \pi(\mathbf{A}|\cdot) &\propto \prod_{t=1}^T \left|k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right|^{\frac{k}{2}} \exp\left\{-\frac{1}{2}\text{Tr}\left(k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\right)\right\} |\mathbf{C}_0^{-1}|^{\frac{\nu_0}{2}} |\mathbf{A}|^{-\frac{\nu_0+q+1}{2}} \exp\left\{-\frac{1}{2}\text{Tr}\left(\mathbf{C}_0^{-1}\mathbf{A}^{-1}\right)\right\} \\ &\propto |\mathbf{A}|^{-\frac{\nu_0+Tk+q+1}{2}} \exp\left\{-\frac{1}{2}\text{Tr}\left(\left(k\sum_{t=1}^T\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}} + \mathbf{C}_0^{-1}\right)\mathbf{A}^{-1}\right)\right\} \end{aligned}$$

Here we used the property that $\text{Tr}(\cdot)$ is cyclically commutative and additive, just as we did in the previous section when we computed the posterior for \mathbf{Q}_t . We obtain that the posterior for \mathbf{A} is the same as the one reported by (Asai & McAleer, 2009):

$$\mathbf{A}|\cdot \sim \mathcal{W}_q^{-1}\left(\nu_0 + Tk, k\sum_{t=1}^T\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}} + \mathbf{C}_0^{-1}\right)$$

3.7 Full Conditional Distribution for d

The only distributions that depend on d are:

- (i) $\mathbf{Q}_t|k, \mathbf{A}, \mathbf{Q}_{t-1} \sim \mathcal{W}_q^{-1}\left(k, k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right), \forall t = \{1, \dots, T\}$
- (ii) $d \sim \mathcal{U}(-1, 1)$

Hence, by multiplying all of them and focusing on the kernel of d , we obtain:

$$\pi(d|\cdot) \propto \prod_{t=1}^T \left|k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right|^{\frac{k}{2}} \exp\left\{-\frac{1}{2}\text{Tr}\left(k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\right)\right\} 1_{(-1,1)}(d)$$

By using the fact that $\text{Tr}(\cdot)$ is cyclically commutative, we know that:

$$\text{Tr}\left(\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\right) = \text{Tr}\left(\mathbf{A}^{-1}\left(\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right)\right)$$

When we will write this probability density function (pdf), we will think about the implementation also. It turns out that it is not a pdf of any commonly known distribution, and in order to sample from it we will need an algorithm such as Adaptive Rejection Metropolis-Hastings. There already exists a pre-programmed function in **R** called *arms()* which takes as input the $\log(\cdot)$ of the pdf. Therefore, by also preparing the pdf for implementation, we obtain that the posterior for d is:

$$\begin{aligned} \pi(d|\cdot) &\propto \prod_{t=1}^T \exp\left\{\frac{dk}{2} \ln |\mathbf{Q}_{t-1}|\right\} \cdot \exp\left\{-\frac{1}{2}Tr\left(\mathbf{A}^{-1}\left(k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right)\right)\right\} \\ &\Rightarrow \pi(d|\cdot) \propto \exp\left\{\frac{dk}{2} \sum_{t=1}^T \ln |\mathbf{Q}_{t-1}| - \frac{1}{2}Tr\left(\mathbf{A}^{-1}\left(k \sum_{t=1}^T \mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right)\right)\right\} \end{aligned}$$

3.8 Full Conditional Distribution for k

Similarly to what we did before, the only distributions that depend on k are:

- (i) $\mathbf{Q}_t|k, \mathbf{A}, \mathbf{Q}_{t-1} \sim \mathcal{W}_q^{-1}\left(k, k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right), \forall t = \{1, \dots, T\}$
- (ii) $k \sim \exp(\lambda_0)$

By multiplying them, focusing on the kernel of k , and using the cyclical commutativity and the additivity of $Tr(\cdot)$, we obtain:

$$\begin{aligned} \pi(k|\cdot) &\propto \prod_{t=1}^T \left(\frac{1}{2^{\frac{kq}{2}}\Gamma_q\left(\frac{k}{2}\right)} \left|k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right|^{\frac{k}{2}} |\mathbf{Q}_t|^{-\frac{q+k+1}{2}}\right) \prod_{t=1}^T \left(\exp\left\{-\frac{1}{2}Tr\left(k\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{A}^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\right) - \lambda_0 k\right\}\right) \\ &\propto k^{\frac{qTk}{2}} \left(\prod_{t=1}^T |\mathbf{Q}_{t-1}|^{\frac{dk}{2}}\right) |\mathbf{A}|^{-\frac{Tk}{2}} \left(\prod_{t=1}^T |\mathbf{Q}_t|^{-\frac{k}{2}}\right) \left(\frac{1}{2^{\frac{kq}{2}}\Gamma_q\left(\frac{k}{2}\right)}\right)^T \exp\left\{-\frac{1}{2}Tr\left(k\left(\sum_{t=1}^T \mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right)\mathbf{A}^{-1}\right) - \lambda_0 k\right\} \\ &= \exp\left\{\frac{Tkq}{2} \ln k + \frac{dk}{2} \sum_{t=1}^T \ln |\mathbf{Q}_{t-1}| - \frac{k}{2} \sum_{t=1}^T \ln |\mathbf{Q}_t| - \frac{Tk}{2} \ln |\mathbf{A}|\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}Tr\left(k\left(\sum_{t=1}^T \mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right)\mathbf{A}^{-1}\right) - \lambda_0 k - \frac{kTq}{2} \ln 2\right\} \left(\frac{1}{\Gamma_q\left(\frac{k}{2}\right)}\right)^T \end{aligned}$$

Lemma 3. The q^{th} -variate gamma distribution is connected to the univariate one by the following identity:

$$\Gamma_q(a) = \pi^{\frac{q(q-1)}{4}} \prod_{j=1}^q \Gamma\left(a + \frac{1-j}{2}\right)$$

Hence, if we apply the above lemma for $a = \frac{k}{2}$ and we focus only on the kernel of k , we obtain that:

$$\Gamma_q\left(\frac{k}{2}\right) \propto \prod_{j=1}^q \Gamma\left(\frac{k+1-j}{2}\right)$$

Again, this posterior is not from a common distribution and in the implementation, we will use the same Adaptive Rejection Metropolis-Hastings that we used at sampling from the posterior for d . For this, we need the $\ln(\cdot)$ of the probability density function:

$$\begin{aligned} \pi(k|\cdot) &\propto \exp\left\{\frac{Tkq}{2} \ln \frac{k}{2} + \frac{dk}{2} \sum_{t=1}^T \ln |\mathbf{Q}_{t-1}| - \frac{k}{2} \sum_{t=1}^T \ln |\mathbf{Q}_t| - \frac{Tk}{2} \ln |\mathbf{A}|\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}Tr\left(k\left(\sum_{t=1}^T \mathbf{Q}_{t-1}^{\frac{d}{2}}\mathbf{Q}_t^{-1}\mathbf{Q}_{t-1}^{\frac{d}{2}}\right)\mathbf{A}^{-1}\right) - \lambda_0 k - T \sum_{j=1}^q \ln \Gamma\left(\frac{k+1-j}{2}\right)\right\} \end{aligned}$$

3.9 Full Conditional Distribution for σ_i^2

The distributions that depend on σ_i^2 for $i = \{1, \dots, p\}$ are:

$$(i) \mathbf{y}_t = \boldsymbol{\mu} + \mathbf{B}\mathbf{f}_t + \mathbf{e}_t \Leftrightarrow \mathbf{y}_t | \mathbf{B}, \mathbf{f}_t, \boldsymbol{\Omega}, \boldsymbol{\mu} \sim \mathcal{N}_p \left(\boldsymbol{\mu} + \mathbf{B}\mathbf{f}_t, \boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \sigma_p^2 \end{bmatrix} \right)$$

$$(ii) \sigma_i^2 \stackrel{indep.}{\sim} \text{InvGamma}(\alpha_0, \beta_0)$$

Since $\boldsymbol{\Omega}$ is diagonal, from the first assumption we obtain that:

$$\mathbf{e}_{it} = \mathbf{y}_{it} - \boldsymbol{\mu}_i - \mathbf{B}_i^T \mathbf{f}_t \stackrel{indep.}{\sim} \mathcal{N}(0, \sigma_i^2)$$

Together with the fact that $\sigma_i^2 \stackrel{indep.}{\sim} \text{InvGamma}(\alpha_0, \beta_0)$, we obtain that the posterior for σ_i^2 is:

$$\pi(\sigma_i^2 | \cdot) \propto (\sigma_i^2)^{\alpha_0 - 1} \exp \left\{ -\frac{\beta_0}{\sigma_i^2} \right\} \left(\exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{t=1}^T \mathbf{e}_{it}^2 \right\} \prod_{t=1}^T \frac{1}{\sigma_i} \right) (\sigma_i^2)^{-(\alpha_0 + \frac{T}{2}) - 1} \exp \left\{ -\frac{\beta_0 + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_{it}^2}{\sigma_i^2} \right\}$$

That is the posterior for σ_i^2 is an Inverse Gamma:

$$\sigma_i^2 \stackrel{indep.}{\sim} \text{InvGamma} \left(\alpha_0 + \frac{T}{2}, \beta_0 + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_{it}^2 \right), \forall i = \{1, \dots, p\}$$

4. The Full MCMC Sampling Algorithm

Now that we have derived our full conditional posteriors, we can combine these full conditionals to build the full MCMC sampler. At each step j in the MCMC sampler, a Metropolis-Hastings algorithm will be needed for sampling \mathbf{Q}_t for all $t = \{1, \dots, T\}$. We showed in Section 3.5 that the exact distribution for $\mathbf{Q}_t, t = \{1, \dots, T - 1\}$ is

$$\begin{aligned} \pi(\mathbf{Q}_t | \cdot) \propto & \mathcal{W}^{-1} \left(k + 1, k \mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_{t-1}^{\frac{d}{2}} + \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right) \left| \text{diag}(\mathbf{Q}_t) \right|^{-\frac{1}{2}} \left| \mathbf{Q}_t \text{diag}(\mathbf{Q}_t) \right|^{-\frac{1}{2}} \left| \mathbf{Q}_t \right|^{\frac{kd+1}{2}} \\ & \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \left(k \mathbf{Q}_t^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_t^{\frac{d}{2}} \mathbf{Q}_{t+1}^{-1} \right) \right\} \exp \left\{ -\frac{1}{2} \text{Tr} \left(\left(\text{diag}(\mathbf{Q}_t) \right)^{\frac{1}{2}} \mathbf{Q}_t^{-1} \text{diag}(\mathbf{Q}_t) \right)^{\frac{1}{2}} - \mathbf{Q}_t^{-1} \right\} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \end{aligned}$$

Let us consider the following trivial approximation to the above probability density function:

$$\pi^*(\mathbf{Q}_t | \cdot) \propto \mathcal{W}^{-1} \left(k + 1, k \mathbf{Q}_{t-1}^{\frac{d}{2}} \mathbf{A}^{-1} \mathbf{Q}_{t-1}^{\frac{d}{2}} + \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \right)$$

With this new notation, we obtain that $\pi(\mathbf{Q}_t | \cdot) \propto \pi^*(\mathbf{Q}_t | \cdot) \cdot f(\mathbf{Q}_t)$, where $f(\mathbf{Q}_t)$ is the extra part outside of the Inverse Wishart kernel. Then, at the $j + 1^{\text{st}}$ step in the MCMC sampler, we sample a candidate $\tilde{\mathbf{Q}}_t \sim \pi^*(\mathbf{Q}_t | \cdot)$ and accept it (i.e. $\mathbf{Q}_t^{(j+1)} = \tilde{\mathbf{Q}}_t$) with probability $\min(\rho_t, 1)$, where

$$\rho_t = \frac{\pi(\tilde{\mathbf{Q}}_t | \cdot)}{\pi(\mathbf{Q}_t^{(j)} | \cdot)} \cdot \frac{\pi^*(\mathbf{Q}_t^{(j)} | \cdot)}{\pi^*(\tilde{\mathbf{Q}}_t | \cdot)} = \frac{\pi^*(\tilde{\mathbf{Q}}_t | \cdot) \cdot f(\tilde{\mathbf{Q}}_t)}{\pi^*(\mathbf{Q}_t^{(j)} | \cdot) \cdot f(\mathbf{Q}_t^{(j)})} \cdot \frac{\pi^*(\mathbf{Q}_t^{(j)} | \cdot)}{\pi^*(\tilde{\mathbf{Q}}_t | \cdot)} = \frac{f(\tilde{\mathbf{Q}}_t)}{f(\mathbf{Q}_t^{(j)})}$$

Similarly, when $t = T$, the Inverse-Wishart probability density functions cancel out and we are left with $\rho_T = \frac{f(\tilde{\mathbf{Q}}_T)}{f(\mathbf{Q}_T^{(j)})}$.

Using the Metropolis-Hastings step that we just discussed, we arrive at the following MCMC Sampler:

MCMC Sampler for BL-DCMSV model

- 1: $\mathbf{f}_t^{(j+1)} | \cdot \stackrel{indep.}{\sim} \mathcal{N}_q \left(\Sigma_{\mathbf{f}_t}^{(j)} \mathbf{B}^{(j)T} \Omega^{(j)-1} (\mathbf{y}_t - \boldsymbol{\mu}^{(j)}), \Sigma_{\mathbf{f}_t}^{(j)} = \left(\mathbf{B}^{(j)T} \Omega^{(j)-1} \mathbf{B}^{(j)} + \mathbf{V}_t^{(j)-\frac{1}{2}} \Sigma_{\boldsymbol{\epsilon}_t}^{(j)-1} \mathbf{V}_t^{(j)-\frac{1}{2}} \right)^{-1} \right)$
- 2: $\mathbf{B}_r^{(j+1)} | \cdot \stackrel{indep.}{\sim} \mathcal{N}_q \left(\frac{1}{\sigma_r^{(j)2}} \Sigma_{\mathbf{B}_r}^{(j)} \mathbf{F}^{(j+1)} (\mathbf{y}_r - \boldsymbol{\mu}_r^{(j)} \mathbf{1}), \Sigma_{\mathbf{B}_r}^{(j)} = \left(\frac{1}{\sigma_r^{(j)2}} \mathbf{F}^{(j+1)} \mathbf{F}^{(j+1)T} + \frac{1}{c_0^2} \mathbf{I}_q \right)^{-1} \right)$
- 3: $\boldsymbol{\mu}^{(j+1)} | \cdot \sim \mathcal{N} \left(\Sigma_{\boldsymbol{\mu}}^{(j)} \left(T \Omega^{(j)-1} \bar{\mathbf{y}}^s + \mathbf{P}^T \Omega_0^{-1} \mathbf{q}_0 \right), \Sigma_{\boldsymbol{\mu}}^{(j)} = \left(T \Omega^{(j)-1} + \mathbf{P}^T \Omega_0^{-1} \mathbf{P} \right)^{-1} \right)$
- 4: $\sigma_i^{(j+1)2} \stackrel{indep.}{\sim} InvGamma \left(\alpha_0 + \frac{T}{2}, \beta_0 + \frac{1}{2} \sum_{t=1}^T \mathbf{e}_i^{(j+1)2} \right)$, where $\mathbf{e}_i^{(j+1)} = \mathbf{y}_i - \boldsymbol{\mu}_i^{(j+1)} - \mathbf{B}_i^{(j+1)T} \mathbf{f}_i^{(j+1)}$
- 5: $\mathbf{h}_t^{(j+1)} = stochvol(\boldsymbol{\epsilon}_t^{(j+1)}) \Rightarrow \mathbf{V}_t^{(j+1)} = diag \left(\mathbf{e}_t^{(j+1)} \right)$, $\boldsymbol{\epsilon}_t^{(j+1)} = \mathbf{V}_t^{(j+1)-\frac{1}{2}} \mathbf{f}_t^{(j+1)}$
- 6: In this step we need to implement the Metropolis-Hastings step discussed:
 - If $t = \{1, \dots, T - 1\}$ we sample from:

$$\mathbf{Q}_t^{(j+1)} \sim \mathcal{W}^{-1} \left(k^{(j)} + 1, k^{(j)} \mathbf{Q}_{t-1}^{(j+1) \frac{d(j)}{2}} \mathbf{A}^{(j)-1} \mathbf{Q}_{t-1}^{(j+1) \frac{d(j)}{2}} + \boldsymbol{\epsilon}_t^{(j+1)} \boldsymbol{\epsilon}_t^{(j+1)T} \right) \left| \Sigma_{\boldsymbol{\epsilon}_t}^{(j)} \right|^{-\frac{1}{2}} \left| \mathbf{Q}_t^{(j)} \right|^{\frac{k^{(j)} d(j) + 1}{2}}$$

$$\exp \left\{ -\frac{1}{2} Tr \left(k^{(j)} \mathbf{Q}_t^{(j+1) \frac{d(j)}{2}} \mathbf{A}^{(j)-1} \mathbf{Q}_t^{(j+1) \frac{d(j)}{2}} \mathbf{Q}_{t+1}^{(j)-1} \right) \right\} \exp \left\{ -\frac{1}{2} Tr \left(\left(\Sigma_{\boldsymbol{\epsilon}_t}^{(j)-1} - \mathbf{Q}_t^{(j)-1} \right) \boldsymbol{\epsilon}_t^{(j+1)} \boldsymbol{\epsilon}_t^{(j+1)T} \right) \right\}$$

- If $t = T$ we sample from:

$$\mathbf{Q}_T^{(j+1)} \sim \mathcal{W}^{-1} \left(k^{(j)} + 1, k^{(j)} \mathbf{Q}_{T-1}^{(j+1) \frac{d(j)}{2}} \mathbf{A}^{(j+1)-1} \mathbf{Q}_{T-1}^{(j+1) \frac{d(j)}{2}} + \boldsymbol{\epsilon}_T^{(j+1)} \boldsymbol{\epsilon}_T^{(j+1)T} \right) \cdot \exp \left\{ -\frac{1}{2} Tr \left(\left(\Sigma_{\boldsymbol{\epsilon}_T}^{(j)-1} - \mathbf{Q}_T^{(j)-1} \right) \boldsymbol{\epsilon}_T^{(j+1)} \boldsymbol{\epsilon}_T^{(j+1)T} \right) \right\}$$

$$7: \mathbf{A}^{(j+1)} | \cdot \sim \mathcal{W}_q^{-1} \left(\nu_0 + T k^{(j)}, k^{(j)} \sum_{t=1}^T \mathbf{Q}_{t-1}^{(j+1) \frac{d(j)}{2}} \mathbf{Q}_t^{(j+1)-1} \mathbf{Q}_{t-1}^{(j+1) \frac{d(j)}{2}} + \mathbf{C}_0^{-1} \right)$$

MCMC Sampler for BL-DCMSV model

- 8: Use the *arms()* function in *R* to sample:

$$d^{(j+1)} | \cdot \sim \exp \left\{ \frac{d^{(j)} k^{(j)}}{2} \sum_{t=1}^T \ln |\mathbf{Q}_{t-1}^{(j+1)}| - \frac{1}{2} Tr \left(\mathbf{A}^{(j+1)-1} \left(k^{(j)} \sum_{t=1}^T \mathbf{Q}_{t-1}^{(j+1) \frac{d(j)}{2}} \mathbf{Q}_t^{(j+1)-1} \mathbf{Q}_{t-1}^{(j+1) \frac{d(j)}{2}} \right) \right) \right\}$$

- 9: Use the *arms()* function in *R* to sample:

$$k^{(j+1)} | \cdot \sim \exp \left\{ \frac{T k^{(j)} q}{2} \ln \frac{k^{(j)}}{2} + \frac{d^{(j+1)} k^{(j)}}{2} \sum_{t=1}^T \ln |\mathbf{Q}_{t-1}^{(j+1)}| - \frac{k^{(j)}}{2} \sum_{t=1}^T \ln |\mathbf{Q}_t^{(j+1)}| - \frac{T k^{(j)}}{2} \ln |\mathbf{A}^{(j+1)}| \right\}$$

$$\cdot \exp \left\{ -\frac{1}{2} Tr \left(k^{(j)} \left(\sum_{t=1}^T \mathbf{Q}_{t-1}^{(j+1) \frac{d(j+1)}{2}} \mathbf{Q}_t^{(j+1)-1} \mathbf{Q}_{t-1}^{(j+1) \frac{d(j+1)}{2}} \right) \mathbf{A}^{(j+1)-1} \right) \right\} \cdot \exp \left\{ -\lambda_0 k^{(j)} - T \sum_{j=1}^q \ln \Gamma \left(\frac{k^{(j)} + 1 - j}{2} \right) \right\}$$

By running this MCMC Sampler, we obtain samples from the posterior mean of the returns $\boldsymbol{\mu}^{(j+1)}$. In practice, the investor is interested in a portfolio of stocks, with optimal vector of weights \mathbf{w}^* such that $\mathbf{w}^T \mathbf{1} = 1$ and it maximizes the mean of the posterior portfolio (i.e. $\mathbf{w}^T \boldsymbol{\mu}$) while minimizing the risk (or variance $\mathbf{w}^T \boldsymbol{\Sigma}^{(j+1)} \mathbf{w}$). Please notice that the theoretical covariance matrix of the returns $\boldsymbol{\Sigma}$ under the assumptions of our model represented by equations (7), (8), (9), (10) and (11) is:

$$Var(\mathbf{y}_t) = E \left[(\mathbf{y}_t - \boldsymbol{\mu}) (\mathbf{y}_t - \boldsymbol{\mu})^T \right] \stackrel{(7)}{=} E \left[(\mathbf{B} \mathbf{f}_t + \mathbf{e}_t) (\mathbf{B} \mathbf{f}_t + \mathbf{e}_t)^T \right]$$

$$= E \left[\mathbf{B} \mathbf{f}_t \mathbf{f}_t^T \mathbf{B}^T \right] + 2E \left[\mathbf{B} \mathbf{f}_t \mathbf{e}_t^T \right] + E \left[\mathbf{e}_t \mathbf{e}_t^T \right] \stackrel{(7)}{=} \mathbf{B} E \left[\mathbf{f}_t \mathbf{f}_t^T \right] \mathbf{B}^T + 2\mathbf{B} E \left[\mathbf{f}_t \mathbf{e}_t^T \right] + \boldsymbol{\Omega}$$

In the last equation we used the fact that the common factors and errors are independent, which in this case it is equivalent to uncorrelated since they are both multivariate normal. Moreover, using assumption (9) and Tower Property, we know that $E[\mathbf{f}_t] = 0$ and using assumption (7) we also know that $E[\mathbf{e}_t] = 0$. Hence, we obtain that indeed $0 = Cov(\mathbf{f}_t, \mathbf{e}_t) = E[\mathbf{f}_t^T \mathbf{e}_t] - 0$. Therefore, we find that the covariance matrix of the returns is:

$$Var(\mathbf{y}_t) = \mathbf{B} Var(\mathbf{f}_t) \mathbf{B} + \boldsymbol{\Omega} \stackrel{(9)}{=} \mathbf{B} \mathbf{V}_t^{\frac{1}{2}} diag(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t diag(\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{V}_t^{\frac{1}{2}} \mathbf{B}^T + \boldsymbol{\Omega}$$

Therefore, those ideas tell us that the investor can even obtain a posterior sample of optimal weights by solving the

following quadratic programming problem at each step inside the MCMC Sampler:

$$\min_{\mathbf{w}^T \mathbf{1} = 1} \left\{ \mathbf{w}^T \boldsymbol{\mu}^{(j+1)} - \frac{1}{2} \mathbf{w}^T \boldsymbol{\Sigma}^{(j+1)} \mathbf{w} \right\}, \text{ where}$$

$$\boldsymbol{\Sigma}^{(j+1)} = \mathbf{B}^{(j+1)} \mathbf{V}_t^{(j+1)^{\frac{1}{2}}} \text{diag}(\mathbf{Q}_t^{(j+1)})^{-\frac{1}{2}} \mathbf{Q}_t^{(j+1)} \text{diag}(\mathbf{Q}_t^{(j+1)})^{-\frac{1}{2}} \mathbf{V}_t^{(j+1)^{\frac{1}{2}}} \mathbf{B}^{(j+1)T} + \boldsymbol{\Omega}^{(j+1)} \quad (18)$$

5. Numerical Illustrations

Just as we did in the example from Section 2, we choose the same 4 stocks (AAPL, FB, GOOG, MSFT), with daily returns from January 2nd 2014 to December 29th 2017. We will use the following inputs, where the columns are in order AAPL, FB, GOOG, MSFT and the rows represent the views:

$$\mathbf{P} = \begin{matrix} & \dots & \text{AAPL} & \dots & \text{FB} & \dots & \text{GOOG} & \dots & \text{MSFT} & \dots \\ \text{view1} & \mathbf{0}^T & -1 & \mathbf{0}^T & 1 & \mathbf{0}^T & 0 & \mathbf{0}^T & 0 & \mathbf{0}^T \\ \text{view2} & \mathbf{0}^T & 0 & \mathbf{0}^T & 0 & \mathbf{0}^T & 1 & \mathbf{0}^T & -1 & \mathbf{0}^T \end{matrix}, \mathbf{q}_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$$

We ran the MCMC Sampler with 3 parallel chains each for 2000 iterations and carried out the standard convergence diagnostics, as can be seen in the following Gelman, trace and density plots of the posterior mean $\boldsymbol{\mu}$:

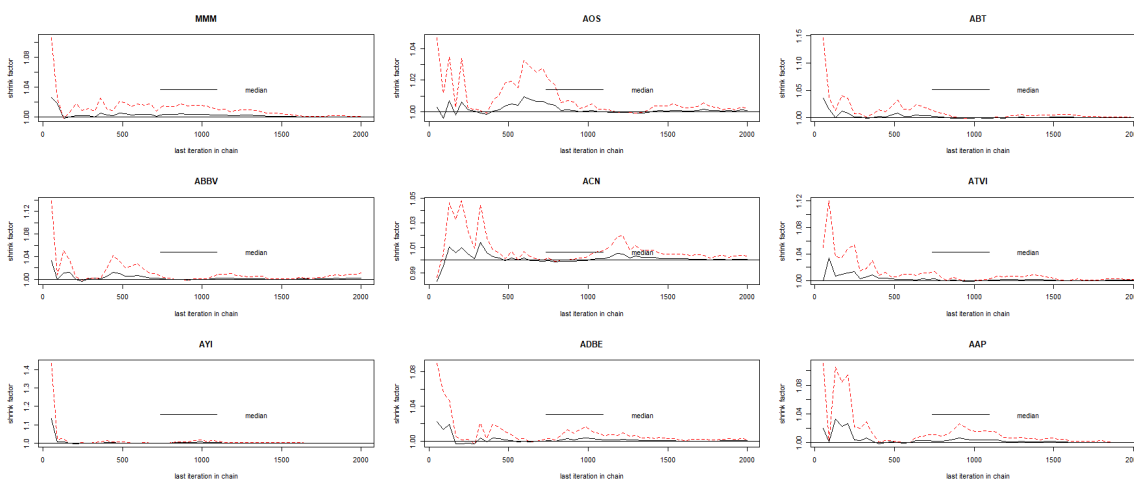


Figure 1. Gelman plots for samples from the posterior for the mean returns

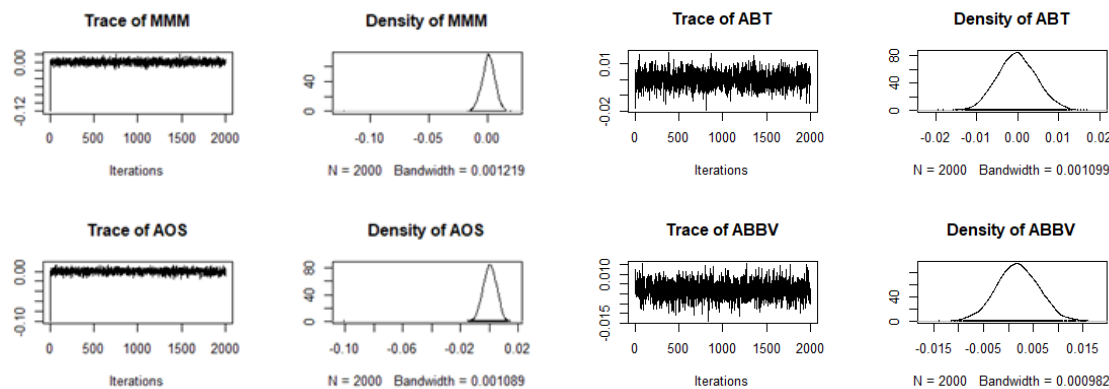


Figure 2. Trace and density plots for samples from the posterior for the mean returns

The MCMC Sampler takes approximately 27 hours to run for desired level of empirical conergence. Clearly, this approach is computationally expensive, therefore, the sensitivity analysis was ran in parallel on 144 cores at once (each core running the MCMC Sampler for 1 pair of confidence levels (ω_1, ω_2)).

But how will we conduct sensitivity analysis for the confidence that the investor has in his/her views (the diagonal elements of the Ω_0 matrix denoted by ω_i here)? Intuitively, we would expect the model to behave in the following way:

- (i) The more confident the investor is in the inputted views, the closer the model should follow them
- (ii) The less confident the investor is in the inputted views, the closer the model should follow history

This also suggests the way in which we will conduct sensitivity analysis. An exhaustive method was implemented that would compute for each pair of diagonal entries in Ω_0 (i.e. investor confidence levels) a posterior mean $\tilde{\mu}$. Using this, we can compute the Euclidean distance $\|\mathbf{P}\tilde{\mu} - \mathbf{q}_0\|_2$. Figure 3 has as 2 of the axes the 2 diagonal entries in Ω_0 and the third axis represents the distance $\|\mathbf{P}\tilde{\mu} - \mathbf{q}_0\|_2$. Figure 4 depicts the same distance obtained by using the model from (Andrei & Hsu, 2021) on the same data-set, with the same inputs \mathbf{P} and \mathbf{q}_0 :

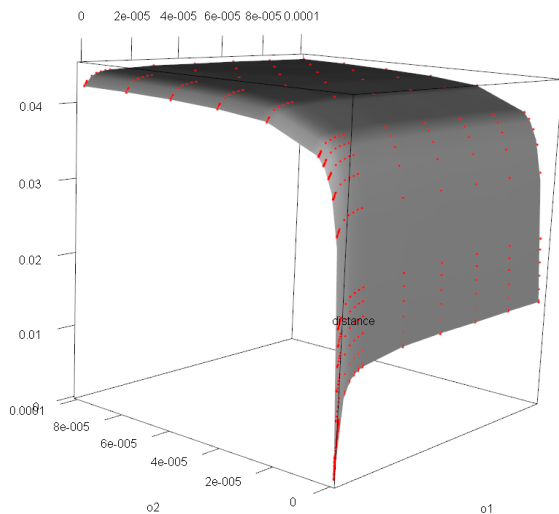


Figure 3. $\|\mathbf{P}\tilde{\mu} - \mathbf{q}_0\|_2$ for different confidence levels

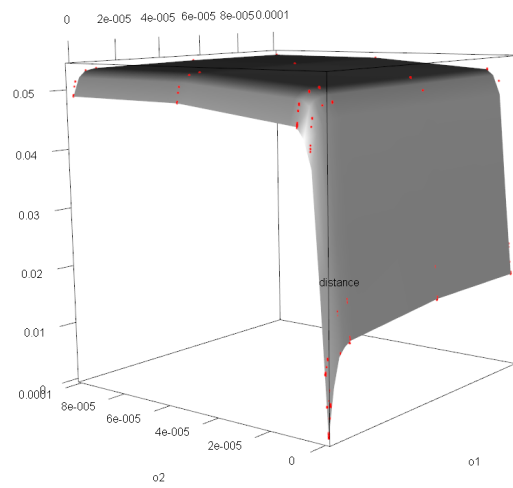


Figure 4. Distances from (Andrei & Hsu, 2021)

Both 3D plots confirm the fact that the models work according to our intuition:

- (i) the smaller ω_1/ω_2 from the plots (or ω_1/ω_2 from the theory), the more confident the investor is in the view since it is actually a variance. We can see in both 3D plots that the smaller ω_1/ω_2 , the smaller the distance $\|\mathbf{P}\tilde{\mu} - \mathbf{q}_0\|_2$ becomes.
- (ii) the larger ω_1/ω_2 (or ω_1/ω_2), the less confident the investor is in the view since, again, it is actually a variance. We can see in both 3D plots that the larger ω_1/ω_2 , the larger the distance $\|\mathbf{P}\tilde{\mu} - \mathbf{q}_0\|_2$ becomes and it plateaus to a certain value (the history $\|\mathbf{P}\bar{\mathbf{y}} - \mathbf{q}_0\|_2 = 0.0539$).

The only difference between the 2 plots is that the distance in the model presented in this paper converges to history slower than the distances from (Andrei & Hsu, 2021). This slower convergence to history is preferred in practice because the investor would like a more gradual mix between the historical data and the personal opinions, as the levels of confidence in those opinions change.

Let us also consider a testing data-set consisting of the month immediately following our historical data, which, in this case, is January 2018 and an initial capital of \$100,000. Over this month, we had the following returns $r_{FB} = 3.141879\%$, $r_{AAPL} = -3.070936\%$, $r_{GOOG} = 9.266661\%$ and $r_{MSFT} = 7.899943\%$. Hence, in reality, we had that the first view was $r_{FB} - r_{AAPL} = 6.212815\%$ (we inputted 2%) and the second view was $r_{GOOG} - r_{MSFT} = 1.366718\%$ (we inputted 5%). For each combination of confidence levels in those 2 views ω_1 and ω_2 , we compute the optimal portfolio weights using the procedure described in equation (18). Using those weights, we can easily compute the profits/losses for each combination of investor confidence levels. The mean across all pairs (ω_1, ω_2) is \$105,294,405, the minimum is \$70,232,381 and the maximum is \$154,101,151. Please keep in mind that those estimates do not include any sort of commissions, capital requirement for short positions (you could even short 500% of a certain stock), etc.

Similar conclusions as the ones we derived from Figure 3 can be drawn if we plot the prior density of the distance $\|\mathbf{P}\boldsymbol{\mu} - \mathbf{q}_0\|_2$ together with the posterior density of the distance $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for different confidence levels. Sampling from the prior is easily done if we make the following observation:

$$\begin{aligned}
 \mathbf{P}\boldsymbol{\mu} &\sim N_v(\mathbf{q}_0, \boldsymbol{\Omega}_0) \\
 \Rightarrow \mathbf{P}\boldsymbol{\mu} &\stackrel{d}{=} \mathbf{q}_0 + \boldsymbol{\Omega}_0^{1/2}\mathbf{Z}, \text{ where } \mathbf{Z} \sim N_v(\mathbf{0}, \mathbf{I}) \\
 \text{Thus, } \|\mathbf{P}\boldsymbol{\mu} - \mathbf{q}_0\|_2 &\stackrel{d}{=} \sqrt{\mathbf{Z}'\boldsymbol{\Omega}_0\mathbf{Z}} \\
 &\stackrel{d}{=} \sqrt{\sum_{k=1}^v \omega_{0k}\mathbf{Z}_k^2} \text{ where } \mathbf{Z}_k^2 \stackrel{iid}{\sim} \chi_1^2 \text{ and } \omega'_{0k} \text{ s are the eigenvalues of } \boldsymbol{\Omega}_0
 \end{aligned}$$

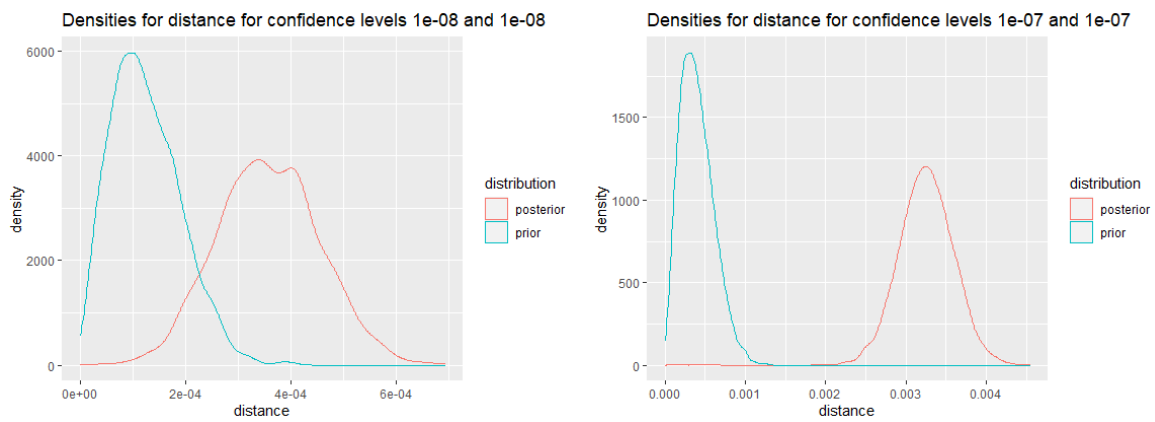


Figure 5. Density of $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for 10^{-8} confidence in Figure 6. Density of $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for 10^{-7} confidence in both views

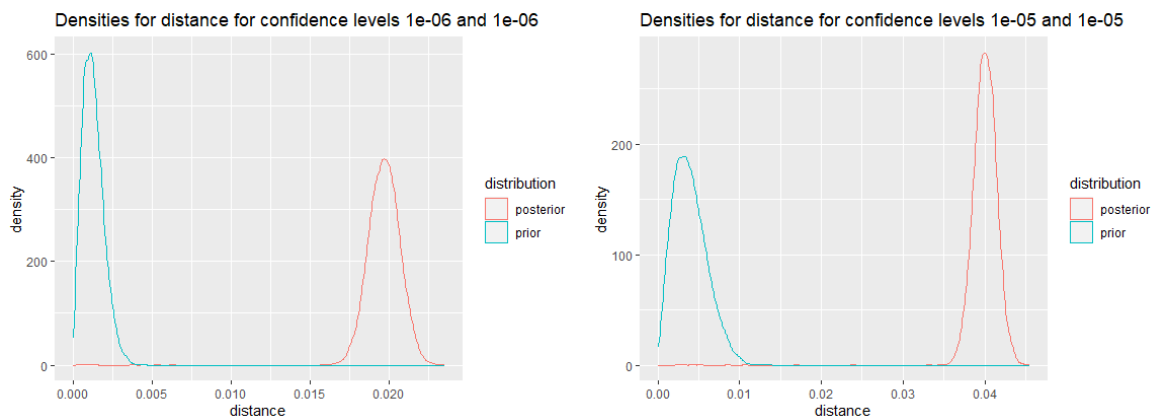


Figure 7. Density of $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for 10^{-6} confidence in Figure 8. Density of $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for 10^{-5} confidence in both views

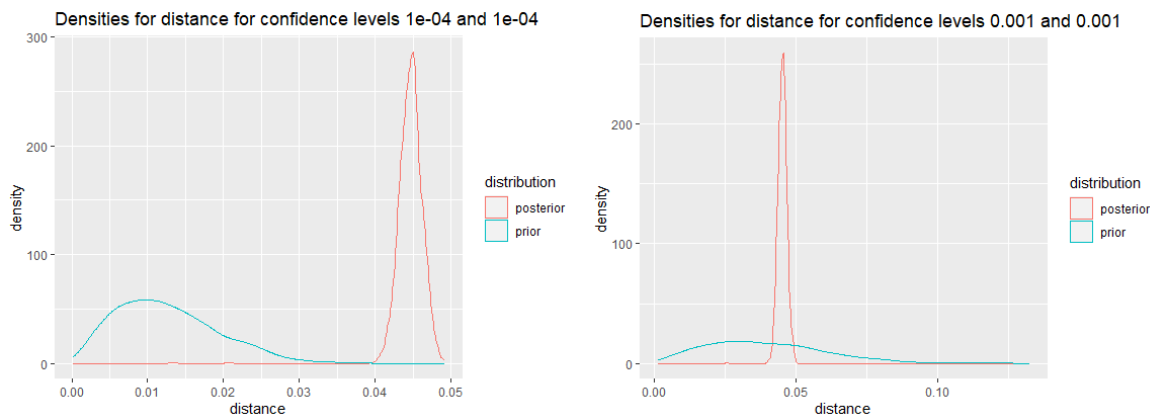


Figure 9. Density of $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for 10^{-4} confidence in both views
 Figure 10. Density of $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for 10^{-3} confidence in both views

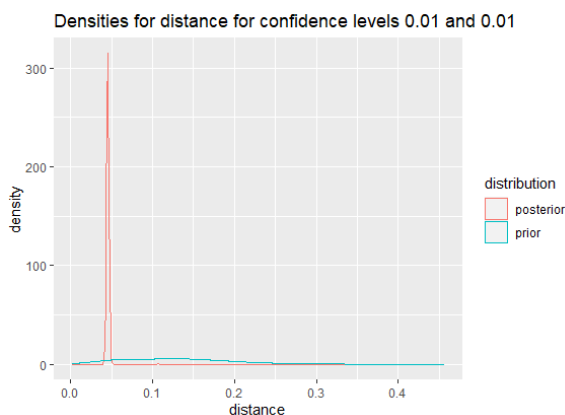


Figure 11. Density of $\|\mathbf{P}\tilde{\boldsymbol{\mu}} - \mathbf{q}_0\|_2$ for 10^{-2} confidence in both views

As we start from Figure 5 and we move to Figure 11, we sequentially go from the very confident investor to the least confident investor. We notice again that the model indeed works according to our intuition:

- (i) as the level of confidence that the investor has decreases, the 2 densities become further and further apart
- (ii) as the level of confidence the investor has decreases, we notice that the standard deviation of the posterior density becomes smaller and smaller. This is because the model starts putting more and more importance on history, as we have seen in the 3D plot represented by Figure 3.

6. Conclusions

In this paper, we propose a novel hierarchical Bayesian model based on latent factors that incorporates investor’s views as a prior on the asset returns, while allowing the covariance structure of the asset returns to be stochastic and dynamic. The advantages of our method are three folds. First, we can directly incorporate the investor’s preferences and market insights into the model via the BL model widely used in the financial industry. Second, we avoid ill-conditioned estimates for high dimensional applications by using a hierarchical Bayesian approach. Third, our Bayesian framework naturally incorporates uncertainty of investor’s personal beliefs, and converge to the empirical sample estimate when their confidence level of the market becomes weak. This feature facilitates an easy stress test under different expert input scenarios. In the future, we plan to investigate sensitivity analysis for profits/losses when holding a portfolio optimized using this alternative of BL, including margin requirements for short positions.

References

- Andrei, M. S., & Hsu, J. S. J. (2020, may). A bayesian approach for asset allocation. *International Journal of Statistics and Probability*, 9(4), 1.
- Andrei, M. S., & Hsu, J. S. J. (2021). A bayesian approach for large asset allocation. *International Journal of Statistics and Probability*.
- Asai, M., & McAleer, M. (2009, jun). The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, 150(2), 182–192.
- Avramov, D., & Zhou, G. (2010, dec). Bayesian portfolio analysis. *Annual Review of Financial Economics*, 2(1), 25–47.
- Bertsimas, D., Gupta, V., & Paschalidis, I. C. (2012). Inverse optimization: A new perspective on the black-litterman model. *Operations research*.
- Black, F., & Litterman, R. (1992, sep). Global portfolio optimization. *Financial Analysts Journal*, 48(5), 28–43.
- Cheng, I.-H. (2008, nov). Capm and black-litterman. *Princeton University*.
- He, G., & Litterman, R. (2002). The intuition behind black-litterman model portfolios. *SSRN Electronic Journal*.
- Idzorek, T. (n.d.).
In *Forecasting expected returns in the financial markets*.
- Ku, Y.-C., Bloomfield, P., & Ghosh, S. K. (2014, feb). A flexible observed factor model with separate dynamics for the factor volatilities and their correlation matrix. *Statistical Modelling: An International Journal*, 14(1), 1–20.
- Ledoit, O., & Wolf, M. (2004, feb). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411.
- Polson, N. G., & Tew, B. V. (2000, apr). Bayesian portfolio selection: An empirical analysis of the s&p 500 index 1970–1996. *Journal of Business & Economic Statistics*, 18(2), 164–173.
- Wang, Y., & Zou, J. (2010, apr). Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics*, 38(2), 943–978.
- Zou, J., Wang, F., & Wu, Y. (2018). Large portfolio allocation using high-frequency financial data. *Statistics and Its Interface*.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).