
ERASMUS SCHOOL OF ECONOMICS

# Bayesian Estimation of a Gaussian Macrofinance State Space Model

Thesis MSc Quantitative Finance

*Author:* Themis RALLIS *Student Number:* 477783

*Supervisor:* Assistant Prof. Rutger-Jan LANGE

*Co-reader:* Assistant Prof. Annika M. SCHNÜCKER

**Abstract**

This paper proposes the BLAME algorithm, an analytical Bayesian extension to the renowned EM algorithm. A simulation study is conducted to understand the strengths and weaknesses of the proposed modification and how these may affect forecasting performance. The simulation study finds that the classical EM method is powerful for latent coefficient estimation, whereas the BLAME method excels in observed coefficient estimation. When considering one-month-ahead US macroeconomic forecasts, the EM algorithm performs very well, being comparable to VARIMA and VEC models, but does not beat a random walk. The Bayesian algorithm performs worse than expected, producing respectable forecasts only for inflation, the output gap, and cyclical unemployment. Overall, this research remains a strong advocate for the theoretical soundness of the applied shrinkage techniques exemplified by the general BLAME algorithm.

*Date:* August 5, 2019

# Table of Contents

## Prologue

*"One should remember that estimating the true parameters of the system is possible only in simulated scenarios and in real applications the models used will be more or less wrong anyway. On the other hand, we should be careful not to ruin already probably inaccurate models with bad approximations of filters and smoothers."* – Simo Särkkä

I dedicate this work to my beloved mother, to whom I attribute my every strength of character and will. Love you always, mom.

I acknowledge that the works produced in this research are my own, and to the best of my knowledge solely reflect my own workings unless stated otherwise. I thank assistant professors Annika Schnücker and Rutger-Jan Lange for their useful comments. I would like to thank assistant professor Rutger-Jan Lange for his guidance and patience with the writing of this paper.

# 1 Introduction

The impact of macroeconomic policy on everyday life is indisputable. Changes in monetary policy are the government's way of preemptively reacting to future (unexpected) shifts in the country's economy. In many cases, these changes are made in an attempt to curb or control certain unwanted outcomes, such as hyperinflation. The ability to gain accurate future insight into an economy's movements both in the long and short run is thus an invaluable asset to any government/central bank. It is therefore of high priority to these central bodies to create a time series model that is as flexible as the economy it attempts to emulate. This does not necessarily mean that the model need be as complex as the economy, but rather that the model be as adaptable as possible when supplied with new incoming information. As we perceive the economy to be in different states through time, it is reasonable to assume that the parameters governing the system can be time-varying or, more specifically, non-deterministic random variables. The natural way to estimate such a time series model is to use Bayesian learning methods. This type of statistical inference allows us to continually *learn* about an economy's underlying dynamics, and shift parameters according to how the economy shifts in a natrual manner.

The primary aim of this research is to create such a Bayesian learning time series model that will challenge ordinary frequentist models such as VARs and VECMs (as in e.g. Ang and Piazzesi (2003), Stock and Watson (2009) and Stock and Watson (2011)) when it comes to in-sample fit and out-of-sample forecasting power. This is achieved by proposing a closed-form Bayesian extension to the well-established EM algorithm, named the *BLAME (Bayesian LineAr Minimisation of Energy* algorithm. In general, the application of the EM algorithm itself to multidimensional macroeconomic factor models has not been thoroughly explored in the literature, and so this research provides a new perspective on inferential methods for multidimensional factor models that do not rely on built-in optimisation routines.

When it comes to forecasting macroeconomic variables, a lot of effort has been put into dynamic factor models (DFMs), as in Stock and Watson (2002) who study statistically constructed diffusion indices as latent drivers of the observed variables. They conclude that a handful of condensed indices are strong sources of information for the general macroeconomy, especially at long forecasting horizons. A problem that typically exists in such models is deciding whether or not certain macroeconomic relations, such as the Phillips curve, hold at every time period, as

is studied by Stock and Watson (2009). In this research, this problem is dealt with by having a rolling estimation window. Should a relation not hold at a specific window, this will be reflected within the estimated parameters within this window. As the parameters are allowed to be time-varying, so will their influence on the model.

While it is appealing to construct the underlying macroeconomic drivers using statistical means (e.g. Principal Component Analysis), another string of literature shifts its focus to iteratively filtering out the latent drivers by means of state space estimation, which is also the focus of this research. The most straightforward application of state space filtering uses the celebrated Kalman (1960) filter, which assumes that the observation and state equations are linear and have Gaussian errors. Key advances in this area of research are attributed to the works of de Jong (2000), Dai and Singleton (2000) and Diebold et al. (2006), the latter of which include macroeconomic information in the observation equation. A key conclusion of Diebold et al. (2006) is that the macroeconomy has more explanatory power over the term structure of interest rates, rather than the reverse. This aligns with the natural process of US policy-making, which attempts to diagnose the state of the economy and adjust the Federal rate accordingly to achieve sustainable rates of inflation and unemployment. The attraction associated with the Kalman filter is its computational simplicity; no simulation is needed, as the integrals can be solved in closed form. Additionally, the Kalman estimator is also the best linear unbiased estimator (BLUE) within the linear Gaussian framework, see e.g. Hamilton (1994).

A recent simulation study by Christensen et al. (2014) carefully orchestrated several arbitrage-free Nelson-Siegel (AFNS) state space models of the term structure with and without stochastic volatility. A key finding is that there is a significant upward bias in the estimation of the mean-reversion parameter of the NS level factor, owing to its near unit-root dynamics. This upward bias decreases in magnitude for the respective slope and curvature factors, due to their lower persistence. Estimates of parameters for $\mathbb{Q}$-dynamic observations experience no bias whatsoever, so this problem is an inherent issue with the nature of the level factor behaving similarly to a random walk. They also find that by decreasing the frequency of the data from monthly to weekly observations this finite-sample upward bias to its latent state estimates is worsened. This occurs because at lower frequencies, the parameter standard deviations arising from the optimised likelihood are too low, a feature not heavily affected by the inclusion of stochastic volatility.

Consequently, more elaborate state space models have been investigated. One subset of models investigates nonlinearities in either the observation or state equation, or both. When nonlinearities are present, the standard Kalman filter fails to give accurate estimates, and thus other methods such as the extended Kalman filter (EKF) or the unscented Kalman filter (UKF) are used to capture the nonlinear nature of the model. A notable paper that employs these methods to (simulated) ATSMs is that of Christoffersen et al. (2014). They find that the UKF performs better than the EKF when the observation equation is nonlinear, and may perform just as well as a Bayesian filter at a fraction of the computational cost. An older study by Lund (1997) applies an iterative, extended Kalman filter (IEKF) to exponentially-ATSMs and estimated the parameters using quasi-maximum likelihood (QML). The author finds that the inconsistency associated with the theoretical QML estimator is economically insignificant, however this conclusion is based again on simulated in-sample fitting.

The Kalman filter is readily accompanied by the Rauch-Tung-Striebel (RTS) smoother (Rauch et al. (1965)), which performs a backward-propagating smoothing of the Kalman filtered states conditioned on the full data set. These "smoothed" state estimates are key to one of the most widely applied analytical parameter estimation techniques: the Expectation Maximisation (EM) algorithm. The algorithm is theoretically underpinned by the assumption that, given a set of observed data, researchers can estimate parameters that maximise the *expected* likelihood of the observed data. Creal and Wu (2015) apply the classical EM algorithm to a 3-factor term structure of interest rates, while also including and fine-tuning stochastic volatility. They find that the EM estimated parameters are identical to the estimated parameters using numerical optimisation routines. This highlights the prowess of the classical EM algorithm, even for a system as large as a 10-dimensional observed interest rate vector being driven by a 3-dimensional latent state vector. In a more recent study on sparse factor models, Maanan et al. (2018) develop a modification to the EM algorithm based on a maximum entropy principle for the purpose of model selection for multivariate time series data. In many cases, matrix sparsity, especially in the covariance matrices, can cause problems throughout the EM loop. If singular covariance matrices occur during the optimisation process, the likelihood diverges and the algorithm fails. When the number of parameters that need to be estimated become substantial, the possibility of this divergence increases and the algorithm becomes increasingly sensitive to the "initial guess" supplied by the researcher. This research proposes a modification to the EM algorithm that

detracts from a single initialisation of the estimation process by specifying prior distributions for the parameters in a Bayesian setting. This injection of extra information can be based on the researcher's prior belief, be data driven, or be based on macroeconomic theory. In the limit of diffuse priors, one then recovers the classical EM algorithm. Thus, the classical EM loop provides a starting point on which the proposed Bayesian algorithm builds, allowing for a straightforward extension to existing EM codes.

The classic frequentist method of parameter estimation and point forecasts becomes worrying when considering the possibility of extreme events. It is more informative to, say, policy-makers to describe a probability distribution of possible outcomes rather than a single value. This is where Bayesian inferential methods become prominent. Contrary to frequentist methods, Bayesians do not assume fixed unknown parameters. Rather, they assume the parameter values to be random variables that have uncertainty inherent to them. Although this method of parameter estimation is more natural, attempts to implement this more than 30 years ago were limited due to the lack of required computational power. At first, most attempts to apply Bayesian estimation methods to time series models were focused on deviations from the classic Kalman filter, specifically, nonlinear and/or non-Gaussian state space models, e.g. Dorfman and Havenner (1992), Carlin et al. (1992), Tanizaki and Mariano (1998) and Durbin and Koopman (2002). These papers investigate the effectiveness of particle filters, which rely on Monte Carlo simulations, on a variety of state space models including e.g. ones with a discrete observed variable or Student-$t$ errors. For a complete and self-contained overview on the various types of Bayesian models that can be applied to state space models, in particular those that include Markov chain Monte Carlo (MCMC) simulations, the reader is encouraged to read Andrieu et al. (2010).

Ever since the literature became aware of the strengths and flexibility of Bayesian modelling, such methods have become prominent in the financial econometrics literature, see e.g. Geweke (1989), Chib and Greenberg (1996), Lopes and Tsay (2010). Most applications have been geared towards estimation of parameters driving stochastic volatility models, such as in Chib et al. (2002), Jacquier et al. (2004), and Sakaria and Griffin (2017). However, and perhaps more relevant to this study, there has been a growing number of papers that apply Bayesian estimation methods to the term structure of interest rates. Even though this area of research is still in its infancy, there have been some notable advances over the last two decades. A study by Thompson (2008) uses

Bayesian updating to identify volatility parameters of the term structure, in particular for the short end. The author finds that the identification of volatility parameters relevant to the yield curve is highly sensitive to the estimation technique, but does not test the economic significance of the discrepancies in the obtained values. A much earlier study by Mikkelsen (2001) (in fact, the very first) successfully applies MCMC directly to the term structure of interest rates having assumed a nonlinear mapping of the latent variables to the observed yield curve dynamics. The author concludes that MCMC techniques are extremely well suited for parameter estimation problems when nonlinear mappings arise. Other notable works are those of Sanford and Martin (2005) and Batista and Laurini (2016). The former study applies a hybrid Gibbs/MCMC methodology to "augmented" yield curve data, while the latter apply a Hamiltonian Monte Carlo method, advocating the strengths of this variation of MCMC simulation.

What (almost) all previous works on this subject have lacked is their study of out-of-sample performance. Bayesian models are very rarely tested on this due to the massive computational strain their estimation demands. This research bridges that gap. The proposed Bayesian modification to the classical EM algorithm allows for a full Bayesian treatment of a macrofinance model entwining the term structure and the macroeconomy, while still remaining analytically tractable in terms of filtering and parameter estimation. The only simulation required is the sampling from parameter prior distributions. This is followed by solving closed-form equations for each set of prior values yielding a correct analytical posterior distribution of parameters. Because of this, a proper study of the Bayesian macrofinance model's out-of-sample forecasting performance becomes a reality and is directly comparable to classical models.

This research fits into the literature under the context of Bayesian state space models applied to financial time series, and continues a growing string of literature, namely those mentioned above. Where this research mostly differs is in the contents of the observation equation, as well as proposing a new modification to classical estimation techniques of multidimensional macroeconomic factor models. Almost all previous efforts have focused on modelling the entire (Gaussian) yield curve, which can cause problems due to its high persistence as Sanford and Martin (2005) point out. This research solely focuses on an affine model of the short rate for three primary reasons: (i) working under the ATSM framework defines all long-yield bonds as affine extensions of the short rate, hence, once the short rate is obtained, the rest of the yield curve is implicitly obtained; (ii) nonlinear mappings from states to observations is unavoidable

when considering the full term structure, thus, dealing with only the short rate circumvents this complication; (iii) adding lags of macroeconomic information as exogenous variables that drive the observation equation does not cause severe computational strain and completes a self-contained macrofinance model. As for the proposed modification, aptly named the *BLAME (Bayesian LineAr Minimisation of Energy)* algorithm, this applies as a direct extension to the classical EM algorithm and allows for a full Bayesian treatment of linear Gaussian state space models. This proposition solves several of the well-documented problems associated with the classical EM algorithm, namely the degeneracy of covariance matrices arising because systems become too large and too sparse. In addition to this, hard constraints are also implemented and studied in both the classical and Bayesian algorithms. In order to understand the power of the BLAME algorithm, a simulation study is conducted, and then the estimation techniques are poised against each other on real US macroeconomic data. The simulation study shows that the Bayesian algorithms' in-sample fit is just as good as its classical counterparts'. On the one hand, the (constrained) EM algorithm is stronger at identifying the coefficients of the latent dynamical coefficients, whereas (constrained) BLAME is better at identifying the observation-driven coefficients. When applied to US macroeconomic data on monthly short rate, inflation, output gap, and cyclical unemployment, the Bayesian algorithms perform significantly worse than their classical counterparts. They tend to have difficulties identifying the system parameters when shifting from window to window, which causes their one-step-ahead forecasts to be as much as 60 times worse than a simple random walk. More specifically, for inflation, output gap and unemployment, BLAME performs around 5 times worse than a random walk, and is over 20 times worse for short rate forecasting. The problem does not seem to be in the theoretical soundness of the algorithm itself, but rather in the naive assumption made on the number of iterations required for convergence, as well as the number of prior samples being too low.

The remainder of the paper is structured as follows. Section 2 describes the macrofinance model in detail. Section 3 presents the state space formulation of the macrofinance model, and Section 4 is devoted to the BLAME algorithm. Section 5 applies all estimation techniques to a simulated data set to understand the strengths and weaknesses of each method, while these techniques are then applied to a real macroeconomic data set in Section 6. Finally, the results from the macroeconomic data are discussed in Section 7, with the concluding remarks of the study in Section 8.

## 2 Macrofinance Model

Here I allocate an entire section to introduce the work of Rudebusch and Wu (2008) (henceforth RW), whose macrofinance model this research builds upon. I then outline the different choices this research takes, mostly deviating in the estimation methods. The results of the RW study and those of this one are not directly comparable, as the RW estimated macroeconomic coefficients are estimated in a state space-like structure using six interest rates rather than the short rate alone. However, the interpretations and theoretical underpinnings of the latent drivers and their relations to the observed variables are essentially the same, and these will later be investigated and directly compared.

RW design an ATSM that combines the no-arbitrage specification of the yield curve with well-established macroeconomic dynamics of inflation and the output gap. They first build a "yields-only" model, which defines the short rate as an affine function of a constant and two normalised latent factors:

$$i_t = \delta_0 + \delta_1' f_t = \delta_0 + L_t + S_t, \tag{2.1}$$

where the latent factors are estimated to be the statistical level and slope factors (first two principal components), and $\delta_1 = [1\ 1]'$. They also consider a second short rate model, driven by the inflation and output gap of the economy:

$$i_t = r + \pi_t + g_\pi(\pi_t - \pi_t^*) + g_y y_t + u_t,$$

where $r$ is the equilibrium short term real interest rate (assumed constant), $\pi_t$ is the cumulative inflation over the past year, $\pi_t^*$ is the target inflation, and $y_t$ is the output gap. The coefficients, of course, are understood as the level of influence changes in the inflation and output gap have on the short rate. This specification of the short rate is based on the famous Taylor (1993) rule, which links the decisions made on monetary policy to the movements of the short rate. RW investigate the two models and find similarities in the dynamics driving the short rate. Specifically, they conclude that the level factor $L_t$ closely follows the movement of the nominal interest rate $r + \pi_t$, and can thus be taken as a close approximation. As a result, the remainder of the Taylor (1993) specification, $g_\pi(\pi_t - \pi_t^*) + g_y y_t$, is approximated by the slope factor $S_t$, representing the underlying business cycle of the economy.

For their macrofinance specification, RW take the short rate to be the affine sum of latent factors, as in the above yields-only model. The key aspect of this specification is that the underlying latent factors are not only autoregressive, but are also explained by the observed set of macroeconomic variables as well:

$$L_t = \rho_L L_{t-1} + (1-\rho_L)\pi_t + \varepsilon_{L,t}, \tag{2.2}$$

$$S_t = \rho_S S_{t-1} + (1-\rho_S)(g_y y_t + g_\pi(\pi_t - L_t)) + u_{S,t}, \tag{2.3}$$

$$u_{S,t} = \rho_u u_{S,t-1} + \varepsilon_{S,t}, \tag{2.4}$$

where they specify that the latent variable $S_t$ has a serially correlated error, $u_{S,t}$, with its own dynamics. RW complete their macrofinance model by specifying autoregressive dynamics for the inflation and output gap with both forward- and backward-looking elements:

$$\pi_t = \mu_\pi \mathbb{E}_t[\pi_{t+1}] + (1-\mu_\pi)\pi_{t-1} + \alpha_y y_t + \varepsilon_{\pi,t}, \tag{2.5}$$

$$y_t = \mu_y \mathbb{E}_t[y_{t+1}] + (1-\mu_y)y_{t-1} - \beta_r(i_t - \mathbb{E}_t[\pi_{t+1}]) + \varepsilon_{y,t}. \tag{2.6}$$

In the context of macrofinance models, unemployment tends to be overlooked as a significant source of information. With the aim of capturing as much macroeconomic information as possible, I extend the RW specification by including what is known as the cyclical unemployment rate in the form of the difference between the real unemployment rate, $U_t$ and the natural unemployment rate, $U_t^*$:

$$\nu_t = U_t - U_t^*.$$

Cyclical unemployment is incorporated into the RW macrofinance model given as:

$$\pi_t = \mu_\pi L_t + (1-\mu_\pi)[\alpha_1 \pi_{t-1} + \alpha_2 \pi_{t-2}] + \alpha_y y_{t-1} - \alpha_\nu \nu_{t-1} + \varepsilon_{\pi,t}, \tag{2.7}$$

$$y_t = \mu_y \mathbb{E}_t[y_{t+1}] + (1-\mu_y)[\beta_1 y_{t-1} + \beta_2 y_{t-2}] - \beta_r(i_{t-1} - L_{t-1}) - \beta_\nu \nu_{t-1} + \varepsilon_{y,t}, \tag{2.8}$$

$$\nu_t = \mu_\nu \mathbb{E}_t[\nu_{t+1}] + (1-\mu_\nu)[\gamma_1 \nu_{t-1} + \gamma_2 \nu_{t-2}] + \varepsilon_{\nu,t}. \tag{2.9}$$

This specification of cyclic unemployment allows the model to capture (i) the Phillips curve for inflation and (ii) Okun's law for the output gap. This model is based on the macrofinance model developed by Garutti et al. (2019) in their unpublished paper.

# 3 Classical Parameter Estimation

In this section, I present the EM alongside its constrained counterpart. First, the state space macrofinance model to which all estimation techniques are applied is laid out. Deviations from the original RW set up will be pinpointed and justified.

## 3.1 General State Space Macrofinance Framework

Let the latent state $q$-vector be $\xi_t$. Let the observation $r$-vector be $x_t$, and thus the $g$-vector of explanatory macroeconomic variables be $z_t \subseteq x_t$, with $g \leq r$. Deviating from the RW macrofinance structure from Equations (2.2) and (2.3), the state transition equation will therefore be defined as:

$$\xi_{t+1} = \underset{[q\times q]}{\mathbf{F}}\xi_t + v_{t+1}, \qquad v_{t+1} \sim \mathcal{N}(0, \mathbf{Q}) \tag{3.1}$$

where the state vector is defined as $\xi_t = [L_t \; S_t]'$, and the error term is normally distributed with zero mean and $q \times q$ covariance matrix $\mathbf{Q}$ for $q = 2$. In the RW set up, they state that the latent variables are dependent on the observed inflation and output gap. As the latent variables in a state space model are precisely that, I separate them from the observed variables and do not specify a cyclic dependence in the model. In the filtering process, the dependence of the observed variables on the latent states is clear and one-sided; due to the iterative nature of the EM algorithm, an unidentified system may arise should a feedback effect of the observations be present in the state transition dynamics. Another deviation from the RW set up removes the serially correlated error term of $S_t$ and is absorbed into $v_t = [\varepsilon_{L,t} \; \varepsilon_{S,t}]'$. The correspondence between (the adjusted) Equations (2.2)-(2.3) and Equation (3.1) can be explicitly stated with the coefficients in the state system matrix written as

$$\mathbf{F} = \begin{bmatrix} \rho_L & 0 \\ \rho_S & (\rho_S - 1)g_\pi \end{bmatrix}.$$

Now, let the observation vector be defined as $x_t = i_t \cup z_t = [i_t \; \pi_t \; y_t \; v_t]'$. In this paper, as opposed to RW, I consider only the short rate and not the full term structure. Assume that the short rate's observations have an associated measurement error:

$$i_{t+1} = \delta_1'\xi_{t+1} + \varepsilon_{i,t+1},$$

where the latent state coefficients are not restricted to be normalised to 1, that is, $\delta_1 = [\delta_L \ \delta_S]'$, and a constant is not estimated. Therefore, the full observation equation driven by the state equation in (3.1) is given as:

$$x_{t+1} = \underset{[r\times q]}{\mathbf{H}}\, \xi_{t+1} + \underset{[r\times g]}{\mathbf{K}_1}\, z_t + \underset{[r\times g]}{\mathbf{K}_2}\, z_{t-1} + w_{t+1}, \qquad w_{t+1} \sim \mathcal{N}\big(0, \mathbf{R}\big) \tag{3.2}$$

where the error term, $w_t = [\varepsilon_{i,t} \ \varepsilon_{\pi,t} \ \varepsilon_{y,t} \ \varepsilon_{\nu,t}]$, has $r \times r$ covariance matrix $\mathbf{R}$, thus $r = 4$ and $g = 3$. The modified RW macrofinance specification from Equations (2.7)-(2.9) is preserved by setting:

$$\mathbf{H} = \begin{bmatrix} \delta_L & \delta_S \\ \mu_\pi & 0 \\ 0 & -\beta_r \\ \mu_{\nu,L} & \mu_{\nu,S} \end{bmatrix}, \ \mathbf{K}_1 = \begin{bmatrix} 0 & 0 & 0 \\ (1-\mu_\pi)\alpha_1 & \alpha_y & -\alpha_\nu \\ 0 & (1-\mu_y)\beta_1 & -\beta_\nu \\ 0 & 0 & (1-\mu_{\nu,L}-\mu_{\nu,S})\gamma_1 \end{bmatrix},$$

$$\mathbf{K}_2 = \begin{bmatrix} 0 & 0 & 0 \\ (1-\mu_\pi)\alpha_2 & 0 & 0 \\ 0 & (1-\mu_y)\beta_2 & 0 \\ 0 & 0 & (1-\mu_{\nu,L}-\mu_{\nu,S})\gamma_2 \end{bmatrix}.$$

There are a couple of important adjustments to the original RW set up: (i) the expectational quantity $\mathbb{E}_t[\pi_{t+1}]$ is set to be equivalent to the underlying state variable $L_t$; (ii) $\mathbb{E}_t[y_{t+1}]$ is omitted from the estimation specification since the direct influence of $S_t$ is felt through the value of $-\beta_r(i_t - \mathbb{E}_t[\pi_{t+1}])$; and (iii) the expectational quantity for cyclical unemployment $\mathbb{E}_t[\nu_{t+1}]$ is split into an affine sum of the latent factors with corresponding coefficients $\mu_{\nu,L}$ and $\mu_{\nu,S}$, so as to emulate an "underlying" Phillips curve and Okun's law.

The classical EM algorithm, originally developed by Dempster et al. (1977), is an iterative updating process that takes a set of initial parameters and approximates a lower bound for the marginal likelihood of the observations based on the full set of filtered, and thus smoothed, states. The expectation ('E') step uses a predefined set of parameters $\theta^{(k)}$ to calculate the expected value of the marginal likelihood across all states $\mathcal{Q}$, and then the maximisation ('M') step finds new parameters $\theta^{(k+1)}$ that maximise this expected value. This new set of parameters is used to calculate a new expected likelihood, and so on until convergence. In the context of this research, the relevant maximising quantities are shown in Appendix C.

### 3.2 Constrained System Matrices

The classical EM algorithm is unconstrained, that is, the resulting parameter values are unrestricted. When the dimensionality of the problem becomes substantial, which is mostly the case in dynamic multifactor term structure/DSGE models, reliable parameter estimates become harder to obtain. It is more desirable to impose certain macroeconomic restrictions on the system matrices, in this case, $\{\mathbf{F}, \mathbf{H}, \mathbf{K}_1, \mathbf{K}_2\}$. In principle, it is also possible to constrain the covariance matrices $\{\mathbf{Q}, \mathbf{R}\}$, but this is not explored in this research. The macrofinance model as given by RW has system matrix restrictions in the form of zeros in some elements. The analysis of the $\mathcal{Q}$-maximising quantities changes in this case, where one requires the maximisation of the constrained version of $\mathcal{Q}$. The full derivation of the $\mathcal{Q}_{\text{con}}$-maximising quantity for the $\mathbf{F}$ system matrix is given in Appendix D, and the key result is quoted below:

$$\text{vec}\left(\mathbf{F}_{\text{con}}\right) = \text{vec}\left(\mathbf{F}_{\text{unc}}\right) + \left(\mathbf{\Omega}^{-1}\mathbf{B}'\left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\right)\left(b - \mathbf{B}\text{vec}\left(\mathbf{F}_{\text{unc}}\right)\right), \tag{3.3}$$

where $\mathbf{\Omega}^{-1} = \left(\sum_{t=1}^{T}\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T}\right)^{-1} \otimes \mathbf{Q}$; it is given that the system matrix has linear constraints given by $\mathbf{B}\text{vec}\left(\mathbf{F}\right) = b$; and the vectorised form of the unconstrained solution is given as:

$$\text{vec}\left(\mathbf{F}_{\text{unc}}\right) = \sum_{t=1}^{T}\left[\left(\sum_{t=1}^{T}\hat{\xi}_{t-1|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T}\right)^{-1} \otimes \mathbf{I}_q\right]\left[\hat{\xi}_{t-1|T} \otimes \hat{\xi}_{t|T} + \text{vec}\left(\mathbf{P}_{t,t-1|T}\right)\right].$$

From the above expression, it is evident that the constrained version of the EM solution is equal to the unconstrained version plus a "penalty" that one incurs having imposed the necessary restrictions. Clearly, imposing $\mathbf{B} = b = 0$ recovers the classical unconstrained solution. All of the remaining system matrices have the same form as Equation (3.3), thus, I quote only the unconstrained vectorised solutions, as well as the form of their respective $\mathbf{\Omega}$'s:

$$\text{vec}\left(\mathbf{H}_{\text{unc}}\right) = \sum_{t=3}^{T}\left[\left(\sum_{t=3}^{T}\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T}\right)^{-1}\hat{\xi}_{t|T}\right] \otimes \left(x_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\right); \tag{3.4}$$

$$\underset{[rq\times rq]}{\mathbf{\Omega}^{-1}} = \left(\sum_{t=3}^{T}\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T}\right)^{-1} \otimes \mathbf{R},$$

$$\text{vec}\left(\mathbf{K}_{1,\text{unc}}\right) = \sum_{t=3}^{T}\left[\left(\sum_{t=3}^{T} z_{t-1}z_{t-1}'\right)^{-1} z_{t-1}\right] \otimes \left(x_t - \mathbf{H}_1\hat{\xi}_{t|T} - \mathbf{K}_2 z_{t-2}\right); \tag{3.5}$$

$$\underset{[rg\times rg]}{\mathbf{\Omega}^{-1}} = \left(\sum_{t=3}^{T} z_{t-1}z_{t-1}'\right)^{-1} \otimes \mathbf{R},$$

$$\text{vec}\left(\mathbf{K}_{2,\text{unc}}\right) = \sum_{t=3}^{T}\left[\left(\sum_{t=3}^{T} z_{t-2}z_{t-2}'\right)^{-1} z_{t-2}\right] \otimes \left(x_t - \mathbf{H}_1\hat{\xi}_{t|T} - \mathbf{K}_1 z_{t-1}\right); \tag{3.6}$$

$$\underset{[rg\times rg]}{\mathbf{\Omega}^{-1}} = \left(\sum_{t=3}^{T} z_{t-2}z_{t-2}'\right)^{-1} \otimes \mathbf{R}.$$

In the next section, I outline the BLAME modification and showcase its technical advantages over the classical EM algorithm.

# 4 BLAME

This modified EM algorithm is in the spirit of the famous Hamiltonian Monte Carlo (HMC) method developed by Duane et al. (1987). Their algorithm uses Langevin dynamics to specify a thermal potential under which the system's particles are influenced. The dynamics of the particles are dictated (specifically, restricted) by the specification of the Hamiltonian, i.e. the energy of the system. In this research, the underlying states themselves are not deeply investigated, but rather are the parameters $\theta$. To correctly estimate the parameters, it is desirable to integrate the states out yielding the actual posterior of interest:

$$p(\theta|x_{1:T}) = \int\int\ldots\int p(\xi_{0:T},\theta|x_{1:T})\, d\xi_0 d\xi_1 \ldots d\xi_T. \tag{4.1}$$

This $T+1$ dimensional integral is impossible to compute analytically, and becomes increasingly difficult the more observations we have. Instead of attempting to numerically compute the above quantity directly, the marginal posterior is proportional to

$$p(\theta|x_{1:T}) \propto p(x_{1:T}|\theta)p(\theta).$$

For implementation of the proposed BLAME algorithm, it is easier to replace the posterior distribution with another quantity, the *energy function* (Särkkä (2013)). This is simply the

unnormalised negative logarithm of the above posterior:

$$\phi_T(\theta) = -\log p(\theta) - \log p(x_{1:T}|\theta). \tag{4.2}$$

The crux of parameter estimation in the BLAME algorithm therefore lies in the computation of the observation marginal log-likelihood observations $\log p(x_{1:T}|\theta)$ and the log-prior distribution. The beauty of the BLAME algorithm is that it can be applied when considering both unconstrained and constrained macroeconomic systems, without requiring much additional computation power over the regular (constrained) EM algorithm. We can thus approximate the posterior distribution as:

$$\phi_T(\theta_n) = -\log p(\theta_n) - \mathcal{Q}\left(\theta_n, \theta_n^{(k)}\right). \tag{4.3}$$

### 4.1 Prior Distributions

In order to move forward, it is necessary to now make assumptions about the prior distributions of the parameters $\theta_n$. In principle, any prior distribution can be handled by the BLAME algorithm, since only the analytic form of the probability density function is required, with the only requirement that its derivative exists. Given that the macrofinance state space model is linear Gaussian, for the purposes of this research I specify the system matrices $\mathbf{F}, \mathbf{H}, \mathbf{K}_1$ and $\mathbf{K}_2$ to follow matricvariate normal prior distributions, while the covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ will follow inverse Wishart distributions. More specifically, the prior distributions are given as follows:

$$\mathbf{\Theta}_s \sim \mathcal{MN}\left(\mathbf{M}_s, \mathbf{U}_s, \mathbf{V}_s\right) \propto e^{-\frac{1}{2}\mathrm{tr}\left[\mathbf{V}_s^{-1}(\mathbf{\Theta}_s - \mathbf{M}_s)'\mathbf{U}_s^{-1}(\mathbf{\Theta}_s - \mathbf{M}_s)\right]}, \text{ for } s \in \{\mathbf{F}, \mathbf{H}, \mathbf{K}_1, \mathbf{K}_2\}, \tag{4.4}$$

$$\mathbf{\Theta}_c \sim \mathcal{W}^{-1}\left(\mathbf{\Psi}, \nu\right) \propto |\mathbf{\Theta}_c|^{-(\nu+p+1)/2} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi}\mathbf{\Theta}_c^{-1})}, \text{ for } c \in \{\mathbf{Q}, \mathbf{R}\}. \tag{4.5}$$

### 4.2 The BLAME Algorithm

Using the previous information, we are now ready to state the BLAME algorithm. The result is an important modification of the classical EM algorithm, and I outline it in Proposition 1 at the top of the next page.

**Proposition 1** (BLAME). *Let the linear Gaussian state space system be given by Equations* (3.1) *and* (3.2)*:*

$$\begin{aligned}\xi_{t+1} =& \mathbf{F}\xi_t + v_{t+1},\\ x_{t+1} =& \mathbf{H}\xi_{t+1} + \mathbf{K}_1 z_t + \mathbf{K}_2 z_{t-1} + w_{t+1},\end{aligned}$$

*where* $v_t \sim \mathcal{N}(0, \mathbf{Q})$ *and* $w_t \sim \mathcal{N}(0, \mathbf{R})$*. The d-vector of unknown system and covariance parameters is thus given as* $\theta_n^{(k)} = \{\mu_0, \mathbf{P}_0, \mathbf{F}, \mathbf{H}, \mathbf{K}_1, \mathbf{K}_2, \mathbf{Q}, \mathbf{R}\}$*. Let the parameters have prior distribution:*

$$\theta_n^{(k)} \sim p\left(\theta_n^{(k)}\right),$$

*specifically, those given in Equations* (4.4) *and* (4.5)*. Let the parameters also have* $p \leq d$ *linear restrictions dictated by the linear equation*

$$\mathbf{B}\theta_n^{(k)} = b.$$

*Then, the 'M' step of the EM algorithm is given by the argmin of the energy function* (4.3)*:*

$$\theta_n^{(k+1)} = \underset{\theta_n \in \Theta, \lambda \in \mathbb{R}^p}{\operatorname{argmin}} \left[ -\log p\left(\theta_n^{(k)}\right) - \mathcal{Q}\left(\theta_n, \theta_n^{(k)}\right) + \lambda(\mathbf{B}\theta_n^{(k)} - b) \right]$$

Convergence of the BLAME algorithm is guaranteed on the basis of it having the same structure as the classical EM algorithm, with the only added modification of analytically tractable prior information and linear restrictions. Nuances regarding exact convergences in certain situations with specific priors are left to further research. Nevertheless, the choice of prior should be appropriate such that convergence is not hindered. In fact, we recover the classical EM algorithm in the limit that every parameter has an uninformative prior and no restrictions.

**Corollary 1.1.** *If all parameters have uninformative priors, that is,* $p(\theta_n) \propto 1$*, then the log-prior is 0. If there are no linear restrictions, then* $\mathbf{B} = 0$ *and* $b = 0$*. Therefore, the BLAME algorithm solves:*

$$\underset{\theta_n \in \Theta}{\operatorname{argmin}} \left[ -\mathcal{Q}\left(\theta_n, \theta_n^{(k)}\right) \right] \iff \underset{\theta_n \in \Theta}{\operatorname{argmax}}\, \mathcal{Q}\left(\theta_n, \theta_n^{(k)}\right),$$

*and thus we yield the classical EM algorithm.*

In order to move forward, and to make a distinction between constrained and unconstrained systems, the unconstrained version of the BLAME algorithm will simply be referred to as *BLAME*, whereas the (more general) constrained version will be referred to as *Con-BLAME*. This is to keep consistency with the distinction made between EM and Con-EM. To illustrate this, below I quote the (unconstrained) BLAME quantities for the latent state autoregressive system matrix and latent state covariance matrix. The full derivations of these, as well as the remaining parameter matrices, can be found in Appendix E. Let $\mathbf{F}$ be matricvariate normally distributed:

$$p(\mathbf{F}) \sim \mathcal{MN}_{q\times q}\left(\mathbf{M}, \mathbf{U}, \mathbf{V}\right).$$

Then, the energy minimising quantity is given by:

$$\underset{\theta_n=\mathbf{F}}{\operatorname{argmin}}\, \phi_T(\theta_n) = \operatorname{vec}\left(\mathbf{F}_{\text{blame}}\right) = \mathbb{E}\left[\operatorname{vec}(\hat{\mathbf{F}})|\mathcal{I}_T\right] = \mathbf{Y}^{-1}\operatorname{vec}\left(\mathbf{\Gamma}\right), \tag{4.6}$$

where

$$\mathbf{Y} = \mathbf{I}_q \otimes \left(\mathbf{Q}\mathbf{U}^{-1}\right) + \left[\mathbf{V}\left(\sum_{t=1}^{T}\hat{\xi}_{t-1|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T}\right)'\right] \otimes \mathbf{I}_q,$$

and

$$\mathbf{\Gamma} = \left(\sum_{t=1}^{T}\hat{\xi}_{t|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t,t-1|T}\right)\mathbf{V} + \mathbf{Q}\mathbf{U}^{-1}\mathbf{M}.$$

Now let $\mathbf{Q}$ be inverse Wishart distributed:

$$p(\mathbf{Q}) \sim \mathcal{W}^{-1}(\mathbf{T}, \nu).$$

Then, the energy minimising quantity for the state covariance matrix is given by:

$$\begin{aligned}\underset{\theta_n=\mathbf{Q}}{\operatorname{argmin}}\, \phi_T(\theta_n) =& \mathbb{E}\left[\hat{\mathbf{Q}}|\mathcal{I}_T\right] \\ =& \frac{1}{T+\nu+q}\sum_{t=1}^{T}\Big[\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T} - \left(\hat{\xi}_{t|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t,t-1|T}\right)\mathbf{F}' - \mathbf{F}\left(\hat{\xi}_{t-1|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t-1,t|T}\right) \\ &+ \mathbf{F}\left(\hat{\xi}_{t-1|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T}\right)\mathbf{F}'\Big] + \frac{1}{T+\nu+q}\mathbf{T}'.\end{aligned} \tag{4.7}$$

Equations (4.6), (4.7), and (E.3)-(E.8) combined give the BLAME loop used for analytically calculating the posterior distribution of the parameters. Most importantly, the energy minimising

quantities for the covariance matrices are of the form:

$$\underset{\theta_n \in \{\mathbf{Q},\mathbf{R}\}}{\operatorname{argmin}}\ \phi_T(\theta_n) = \text{Classical EM } \mathcal{Q}\text{-maximiser } + \text{ Scaled Prior Hyperparameters.}$$

Of course, it is obvious that the strength of the prior via the hyperparameters determines the amount of shrinkage one gets throughout the BLAME loops. The computational complexity of this modification is comparable to the classical EM algorithm, albeit slightly more strenuous for specific parameters. The resultant energy minimising quantities at most require the inversion of matrices that scale according to the square of the dimensionality of the states, or whichever the largest dimension is reached based on the multiplied dimensions between $q$, $g$ or $r$. In this application, 2 states imply the inversion of a $4 \times 4$ matrix for the minimising step of $\mathbf{F}$. This is not a substantial problem in general for a $q$-state system, especially when one considers the form of the matrix to be inverted. These constitute a sum of a positive (semi-)definite block covariance matrix and a block matrix of the outer product of the smoothed states. The problem of degeneracy of the outer product is solved by the addition of the block covariance matrix. Finally, for most macrofinance applications, ATSMs almost never exceed 3-factor models, hence, this dimensional scaling effect does not play a part for state system parameter estimation. Things might start to become computationally demanding should one fill the observation equation with the full term structure of interest rates along with exogenous macroeconomic information; then the dimensional scaling effect may need to be addressed more closely. Nonetheless, this does not overshadow the gain one has by implementing the mentioned modification and the aforementioned capabilities of the BLAME algorithm showcase the computational advantage and flexibility of this modified version over the classical EM algorithm.

### 4.3 Constrained BLAME

I now treat the general form of the BLAME algorithm from Proposition 1, henceforth referred to as *Con-BLAME*. To illustrate the Con-BLAME algorithm, I revisit the system matrix $\mathbf{F}$ and quote the constrained energy minimising quantity, where the full derivation can be found in Appendix E.2. Considering the case again where $\mathbf{F}$ follows a matricvariate normal distribution,

the solution is given as:

$$\begin{aligned}\operatorname{vec}(\mathbf{F}_{\text{con-blame}}) = &\left[\mathbf{I}_{q^2} - \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\right] \\ &\times \left[\left[\mathbf{I}_{q^2} - \boldsymbol{\Omega}^{-1}\mathbf{J}\left(\mathbf{D}^{-1} + \mathbf{J}'\boldsymbol{\Omega}^{-1}\mathbf{J}\right)^{-1}\mathbf{J}'\right]\operatorname{vec}(\mathbf{F}_{\text{unc}}) + \boldsymbol{\Phi}^{-1}(\mathbf{V}\otimes\mathbf{U})^{-1}\operatorname{vec}(\mathbf{M})\right] \\ &+ \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1} b. \end{aligned} \tag{4.8}$$

The above expression is identical for all system matrices with the key differences being the dimensions of the identity matrices, as well as the expressions for $\boldsymbol{\Omega}$, and thus $\boldsymbol{\Phi}$, and the corresponding vectorised unconstrained EM system parameters. If no constraints are present, i.e. $\mathbf{B} = 0$ and $b = 0$, then Equation (4.8) reduces to

$$\begin{aligned}\operatorname{vec}(\mathbf{F}_{\text{con-blame}}) &= \operatorname{vec}(\mathbf{F}_{\text{unc}}) \underbrace{-\boldsymbol{\Omega}^{-1}\mathbf{J}\left(\mathbf{D}^{-1} + \mathbf{J}'\boldsymbol{\Omega}^{-1}\mathbf{J}\right)^{-1}\mathbf{J}'}_{=\boldsymbol{\Phi}^{-1}\boldsymbol{\Omega}-\mathbf{I}_{q^2}} \operatorname{vec}(\mathbf{F}_{\text{unc}}) + \boldsymbol{\Phi}^{-1}(\mathbf{V}\otimes\mathbf{U})^{-1}\operatorname{vec}(\mathbf{M}) \\ &= \boldsymbol{\Phi}^{-1}\boldsymbol{\Omega}\operatorname{vec}(\mathbf{F}_{\text{unc}}) + \boldsymbol{\Phi}^{-1}(\mathbf{V}\otimes\mathbf{U})^{-1}\operatorname{vec}(\mathbf{M}),\end{aligned}$$

which is understood as a scaled unconstrained EM solution with a scaled shift associated with the prior hyperparameters. Of course, the smaller in magnitude the elements of $\mathbf{V}\otimes\mathbf{U}$ are, its inverse becomes increasingly large. This puts a much larger emphasis on the prior mean given to $\mathbf{F}$. The prior covariance is clearly understood as the strength of the prior belief on the location of $\mathbf{F}$, that is, the amount of shrinkage we apply to $\mathbf{F}$.

### 4.4 Hierarchical BLAME

The versatility of the energy function allows for a direct specification of a hierarchical Bayesian structure. The advantage of doing so lies in any sensitivity of the algorithm to the choice of prior hyperparameters. Should the system matrices become of sizeable dimensions, then choosing appropriate locations and scales for their priors becomes tricky. Therefore, it is more favourable to define a distribution of hyperparameters and minimise the energy function with respect to these as well, yielding energy minimising quantities in terms of hierarchical hyperparameters. This is a direct extension of the BLAME algorithm, as we simply need to add extra information in the form of a hierarchical log-prior to the energy function and solve the first order condition with respect to the prior hyperparameters. However, this implementation would lead to an

extra layer of prior samples for which the model needs to be estimated, and this increases the computational strain by a few orders of magnitude becoming questionable if its implementation would add anything significant. Exploring this area is left for further research, as it detracts from the results of the standalone (constrained) BLAME algorithm.

### 4.5 Extending to Non-Gaussian State Space Models

When either the states or the observations are modelled to be non-Gaussian, the classical Kalman filter breaks down. Depending on the nature of the model, straightforward modifications may be implemented to adjust for certain degrees of non-Gaussianity or non-linearity. These can be in the form of the extended Kalman filter (EKF), or the unscented Kalman filter (UKF), and are readily implementable at the start of each BLAME loop with the aim of capturing non-Gaussianities or nonlinearities in either the state or observation equations. In principle, the addition of Bayesian prior information within the BLAME loops provide the fix required that usually cause a lot of the problems seen in implementations of the above Kalman filter modifications, including the severe biases the EKF may experience due to the linearisation of heavily non-linear functions. Should the model be so complex as to not be fully captured by the BLAME algorithm alongside a UKF/EKF, then this falls outside the scope of the analytical BLAME algorithm altogether, as Monte Carlo simulation is needed to approximate the Bayesian filtering equations. In these cases, particle filters in the form of sequential importance resamplers and Markov chain Monte Carlo (MCMC) methods for parameter estimation are implemented, effectively replacing the BLAME loop entirely. Therefore, I do not explore this area any further, and the non-Gaussian and/or nonlinear implementation of the BLAME algorithm is also left to further research.

## 5 Simulation Study

In this section, I conduct a simulation study inspired by the Jobson and Korkie (1980) paper in order to test the in-sample fit of: (i) classical EM; (ii) constrained EM; (iii) BLAME; (iv) constrained BLAME. The true parameters are loosely based on the RW estimates. The hard constraints imposed follow the RW specification of the macrofinance model. First, I discuss how the data is generated and then the section concludes with the results of the simulation study.

## 5.1 Time Series Data Generation

The simulated observations are generated iteratively using generated underlying states. Since all variables are (multivariate) normally distributed, it suffices to generate independent normal random numbers, and then induce a correlation based on the desired covariance matrix. The dynamics of the system follow exactly that of the macrofinance model given by Equations (3.1) and (3.2), so it is known that the model is not misspecified. It should be noted, that the simulated system does not contain a simulated unemployment variable, so as to mimic the RW set up more closely. For more in-depth information of the constraints and initialisation of the study, please refer to Appendix G.

In this simulation study, the state error vector $v_t$ is $2 \times 1$ and the observation error vector is $3 \times 1$. Assume that we wish to generate a single correlated vector $y$ of dimension $d$ with zero mean and covariance matrix $\mathbf{\Sigma}$. This is done by first applying a Choleski decomposition on $\mathbf{\Sigma} = \mathbf{Z}\mathbf{Z}'$, and then computing $y = \mathbf{Z}x$, where $x \sim \mathcal{N}_d(0, \mathbf{I}_d)$ is a vector of independent normally distributed variates. When $d = 2$, as for the states, it is most computationally efficient to generate these using the polar Box and Muller (1958) method as pairs of standard normals are generated at each iteration. When $d = 3$, as for the observations, then the inversion method proposed by Rao et al. (2011) is preferred (and thus applied 3 times to generate one vector $w_t$). See Appendix F for the algorithms for these two normal random variate generators. The simulated time series is initialised at a specified initial mean $\mu_0$. The system matrices are all specified beforehand and are kept fixed at every iteration. The first state is then generated as:

$$\tilde{\xi}_1 = \mathbf{F}^s \xi_0 + \tilde{v}_1$$

where the superscript $s$ signifies that the matrix is a user-defined matrix used for simulations. This yields the first set of observations:

$$\tilde{x}_1 = \mathbf{H}^s \tilde{\xi}_1 + \tilde{w}_1 \implies \tilde{z}_1 = \tilde{x}_1^{2,:}$$

where the tilde explicitly specifies a simulated result and the superscript $2,:$ signifies that the contemporaneous exogenous variable $z$ is built from the 2nd to the last element of the observations $x$. This is to emulate the interaction of inflation and output gap on the subsequent observations, leaving the first element of $x$, the simulated short rate, a purely observed variate.

The full simulated time series data is thus iteratively specified by:

$$\begin{aligned}\tilde{\xi}_{t+1} =& \mathbf{F}^s \tilde{\xi}_t + \tilde{v}_{t+1} \\ \tilde{x}_{t+1} =& \mathbf{H}^s \tilde{\xi}_{t+1} + \mathbf{K}_1^s \tilde{z}_t + \mathbf{K}_2^s \tilde{z}_{t-1} + \tilde{w}_{t+1},\end{aligned}$$

and this process is repeated 10,000 times to yield enough data for the algorithms to estimate accurate parameter values. A plot of the first 400 simulated time series data is found in Figure 6 in Appendix J. As for prior parameter generation, the methods used are shown in Appendix H.

### 5.2 In-Sample Fitting

The performance of all four estimation techniques can be found alongside the "true" parameters in Table 1. The number of EM loops is set to $K = 100$, although one may of course choose a tolerance beyond which the system is considered converged. For the Bayesian algorithms, the number of prior samples are $N = 100$. In order to get comparable results, when implementing the Bayesian algorithms specifically for the matricvariate normal system matrices, the location hyperprior $\mathbf{M}$ is set equal to the initialised matrices for the classical and constrained EM algorithms. Thus, what is truly investigated is how the extra prior uncertainty affects the estimation process and whether or not this allows more freedom for the algorithm to learn the true locations of the parameters better than relying on a lucky initial guess.

From Table 1 it is clear that some techniques excel in certain areas and are weak in others. Overall, classical EM performs the best when estimating the latent coefficients and does reasonably well for the observed system coefficients. On average, the relative difference of the coefficients to the true ones is $-16\%$, as is shown in Table 9 in Appendix J. Con-EM performs slightly less well with the latent coefficients, but is strong for the observed system coefficients. Overall, the relative difference of Con-EM's estimated parameters is $-34\%$. Although it does not come closest to the true parameters of the system, BLAME's performance is strong with an average relative difference of $-16\%$, equal to its frequentist counterpart. Con-BLAME also equals its counterpart in terms of relative difference. As can be seen from Figures 1 and 2, for some parameters, even with as little as 100 samples and 100 loops, the Bayesian algorithms have already yielded narrow posterior distributions despite having relatively wide prior distributions. This is not entirely the case for other parameters, however. Ideally, more samples

| Coefficient | True Value | EM | Con-EM | BLAME | Con-BLAME |
|---|---|---|---|---|---|
| $\rho_L$ | 0.94 | **0.94** | **0.94** | 0.75 | 0.78 |
| $\rho_S$ | 0.12 | **0.03** | 0.00 | 0.01 | 0.02 |
| $g_\pi$ | 0.48 | **0.39** | 0.63 | 0.32 | 0.07 |
| $\delta_L$ | 0.05 | 0.09 | **0.03** | 0.14 | 0.07 |
| $\delta_S$ | $-0.05$ | $-0.07$ | $-$**0.05** | $-0.02$ | $-0.01$ |
| $\mu_\pi$ | 0.45 | 0.10 | 0.04 | 0.22 | **0.28** |
| $\alpha_1$ | 1.22 | 0.89 | **1.21** | 1.15 | 1.35 |
| $\alpha_2$ | $-0.42$ | $-0.28$ | -0.31 | $-0.31$ | **-0.36** |
| $\alpha_y$ | $-0.13$ | $-$**0.13** | $-0.05$ | $-0.09$ | $-0.06$ |
| $(1-\mu_y)\beta_1$ | 0.77 | 0.79 | **0.78** | 0.74 | 0.75 |
| $(1-\mu_y)\beta_2$ | 0.09 | **0.08** | **0.08** | **0.08** | **0.08** |
| $\beta_r$ | 0.08 | 0.02 | -0.02 | **0.06** | 0.02 |

Table 1: Estimated parameter values using all four studied estimation techniques on a single simulated sample of $t = 1, \ldots, 10,000$ generated time series data. The posterior mean is quoted for the Bayesian algorithms. The true coefficients are chosen to closely follow the outcome of the RW macrofinance model. Values in bold are the ones closest to the true value.

and longer loops would be required to get more accurate posterior mean estimates, but for the eventual purpose of macroeconomic forecasting, a showcase of the accuracy of (Con-)BLAME at these low sample/loop levels is far more informative. It is thus impressive that the overall performance of the Bayesian algorithms in terms of relative difference to the true parameters is exactly comparable to their classical counterparts at even such low sample/loop levels.

All four algorithms struggle to estimate the off-diagonal element of the latent autoregressive matrix, most likely due to the fact that the slope autoregressive parameter $(\rho_S - 1)g_\pi$ and the off-diagonal parameter $\rho_S$ are mutually dependent. Thus, it makes it difficult for these algorithms to correctly identify the exact values. However, as can be seen in the top right plot of Figure 1, while the individual parameters may not be correctly identified, their interaction within the system is well approximated. The reason why the posterior mean for $\rho_L$ is underestimated by the Bayesian algorithms is due to the bimodal posterior distribution, as is seen in the top left plot of Figure 1. The second peak at around 0.55 biases the posterior mean, when it can be clearly seen that if the maximum a posteriori (MAP) was taken as the posterior estimate, then the BLAME algorithm would exactly identify this coefficient as well. This would not be true, however, for the inflation dependence on the latent level factor through the $\mu_\pi$ coefficient, which is somewhat poorly estimated by all four algorithms and has a highly non-standard posterior distribution. Should the MAP be used in this case, the algorithm would yield an

estimate close to zero, which is not desirable. This is also the case for the Con-BLAME algorithm, as is seen in Figure 2. Nonetheless, compared to the (Con-)EM algorithms, the prior distribution allows more freedom for the algorithm to converge to a variety of values, of which the posterior mean gives a better estimate of the true $\mu_\pi$. The remaining autoregressive parameters are well approximated by both the BLAME and Con-BLAME algorithms, as is exemplified by the narrow, Dirac delta-like posteriors shown in the middle row of Figures 1 and 2.

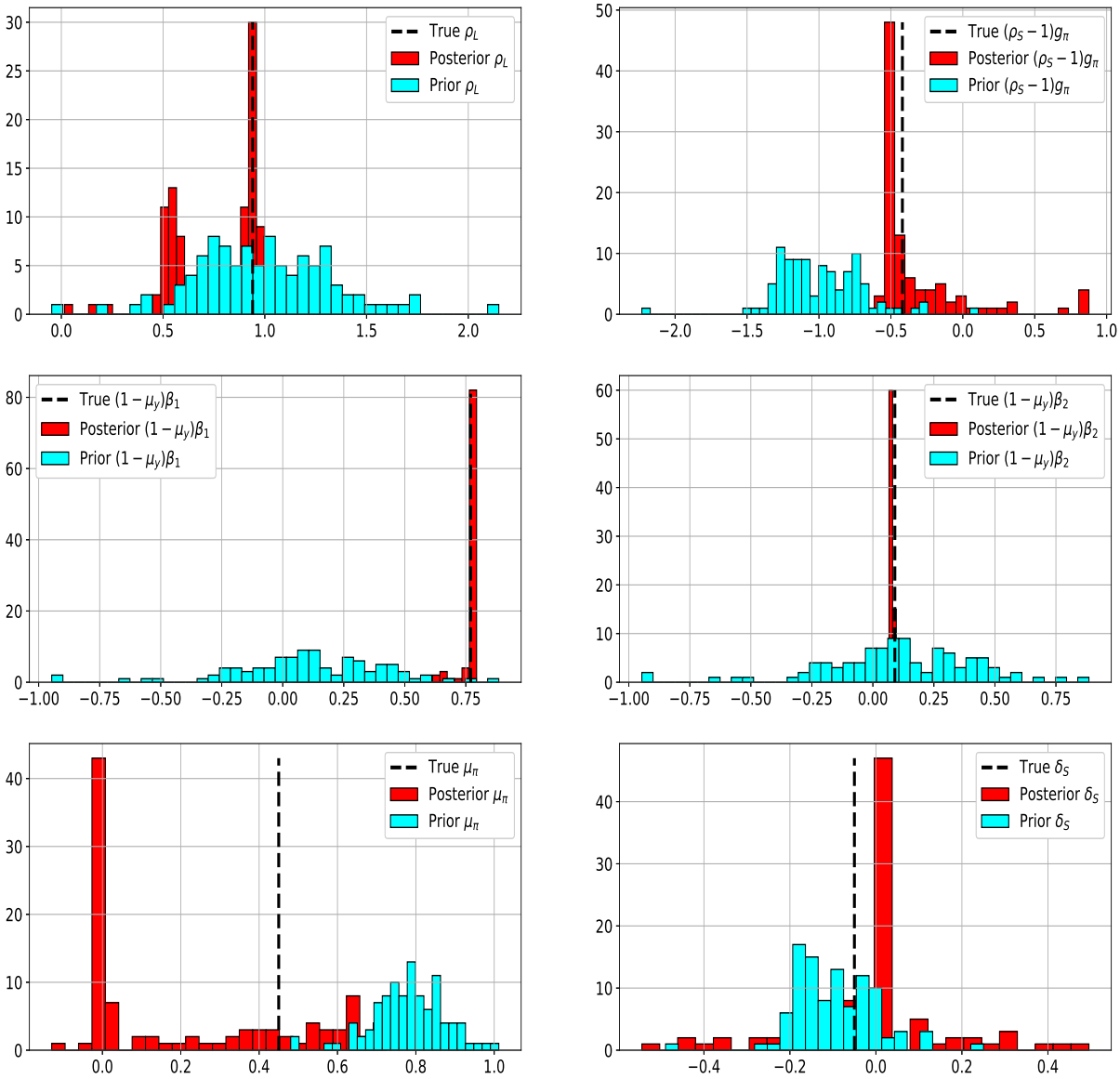



Figure 1: Prior (cyan) and posterior (red) densities of six coefficients taken from all four of the system matrices with the true values indicated by the black dashed vertical lines. These estimations have been done using the BLAME algorithm.

It is worth noting how the posterior distribution of $\rho_L$ does not exhibit the bimodal nature when estimated via Con-BLAME, as opposed to BLAME. It is not immediately clear why this occurs, however, this seems to come at the cost of accurately estimating $(\rho_S - 1)g_\pi$, as is depicted by the wider distribution in the top right plot of Figure 2. How this would affect forecasting performance is not easily understood, but it would seem natural to assume that these trade-offs are countered elsewhere in the estimation process. In general, the constrained algorithms perform best when estimating the observed system, whereas the unconstrained versions perform best when estimating the latent system. Therefore, it is expected that the constrained algorithms will have the best forecasting power when applied to macroeconomic data.

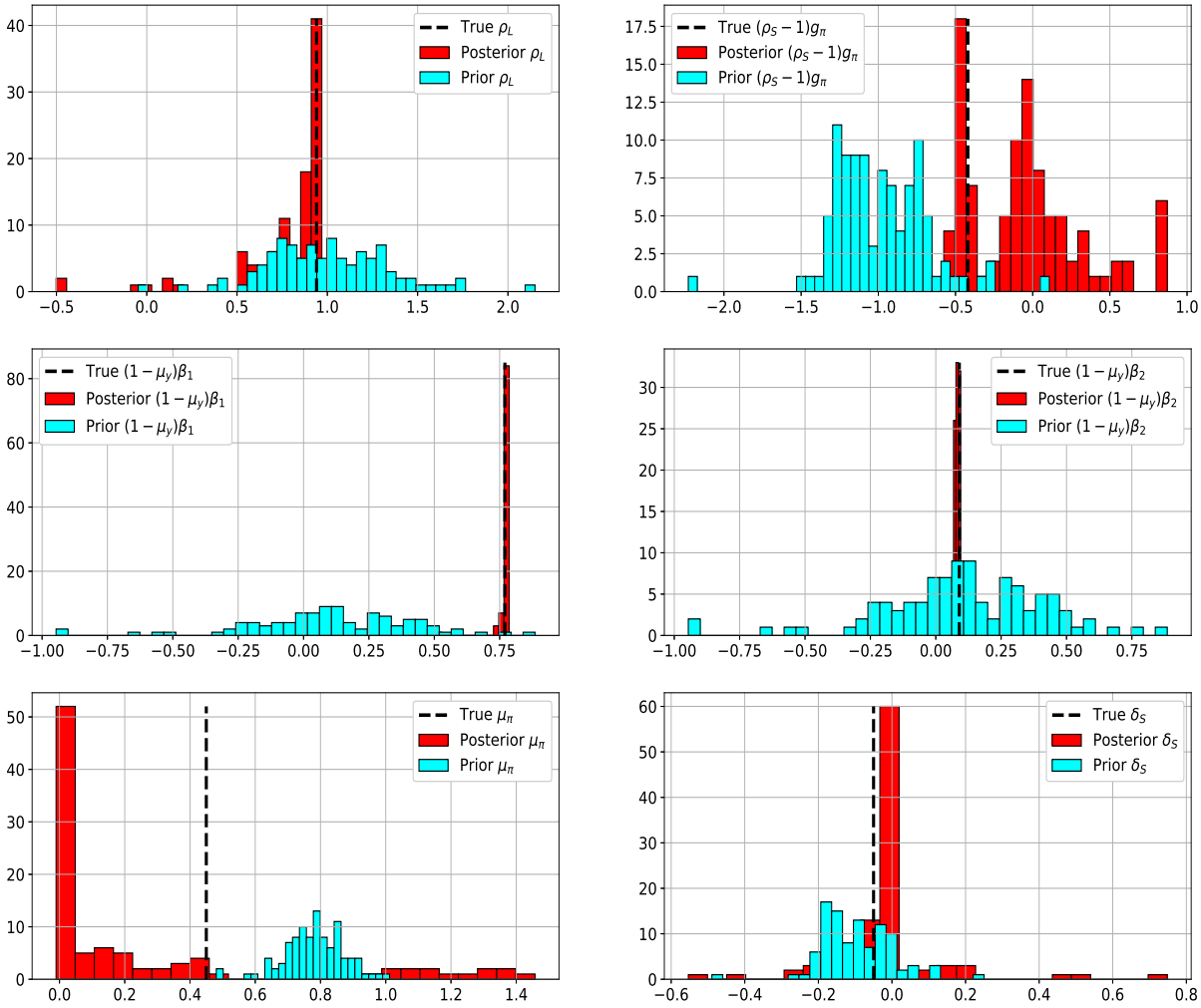



Figure 2: Prior (cyan) and posterior (red) densities of the same six coefficients taken from all four of the system matrices with the true values indicated by the black dashed vertical lines. These estimations have been done using the constrained BLAME algorithm.

# 6 Application to US Macroeconomic Data

Building on the simulation study and understanding the strengths and weaknesses of each estimation technique, I now apply these techniques to real US macroeconomic data. This section will begin with a preliminary analysis of the data itself, providing insight in the expected dynamics of the macrofinance model. Then, all four studied estimation procedures will be applied to the data and the results analysed in comparison to three benchmarks, a random walk (the control), a restricted VARIMA(2,1,0) model and an unrestricted VEC model (two macroeconomic models).

## 6.1 Preliminary Data Analysis

The monthly data on the US short rate and inflation are obtained from the OECD[1] online library, and data on the US output gap, unemployment, and NAIRU are obtained from the Federal Reserve of St. Louis (FRED)[2], where output gap is only available in quarterly frequency. In order to standardise the frequency across the entire data set, a cubic spline interpolation function is applied to the output gap data. The full data set spans from January 1983 until December 2018, thus 432 observations across 36 years. A graph of the data can be seen in Figure 3.

The output gap remains mostly negative, with a "seasonal" pattern associated with the business cycle of the economy, which is expected to be captured by the latent slope factor $S_t$. Inflation remains relatively stable, and the short rate has a downward sloping trend with the noticeable zero-lower-bound (ZLB) following the financial crash in 2008. Cyclical unemployment follows the same cycle as the output gap, but reacts in the opposite direction. This is exemplified in the cross covariance and correlations in Table 10 of Appendix J. Inflation and output gap have a fairly low full sample correlation, whereas the short rate is strongly correlated with inflation, and unemployment is strongly negatively correlated with the output gap. A full sample regression of the short rate on a constant, inflation and output gap yields the corresponding Taylor coefficients, assuming a target inflation rate of 2%. The resulting coefficients are found to be $r = 1.64\%$, $g_\pi = 0.397$ and $g_y = 0.362$, which is in line with the general Taylor rule. This regression is used as a part of the forecasting procedure for the macroeconomic benchmark models, emulating the contemporaneous dependence of the short rate on the latent factors in the SSMs.

[1] www.oecd.org
[2] fred.stlouisfed.org

Each variable's sample autocorrelation function (except cyclical unemployment) is plotted in Figure 7 in Appendix J, and is normalised such that the first lag has autocorrelation 1. The short rate is strongly persistent, with lagged rates of up to 8 years prior still having a significant effect on next month's rate. This is in line with the high persistence seen in previous works. Output gap is less persistent, and inflation the least of all. Therefore, since no lags of the short rate are modelled, it is expected that the underlying latent level factor $L_t$ will capture most this strong persistence and exhibit near-unit-root behaviour, whereas the latent slope factor $S_t$ should exhibit persistence similar to that of the output gap.

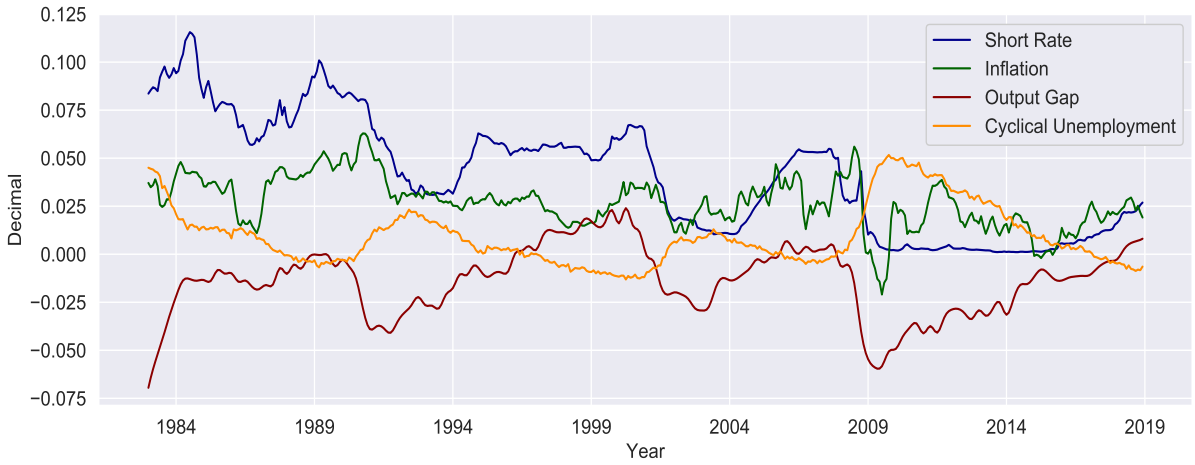



Figure 3: US macroeconomic data in monthly frequency, spanning from 1983:M1-2018:M12. The axis "Decimal" implies the decimal form of a percentage $[x\%/100]$.

Individual histograms for all four macroeconomic variables are presented in Figure 8 of Appendix J. Each histogram is overlayed with a maximum likelihood fitted Gaussian density (in black) along with each time series' kernel density estimator. The kernel densities of inflation and output gap are a close fit to the maximum likelihood Gaussian fit over the whole sample, implying that a Gaussian model for these macroeconomic variables is a suitable assumption. The kernel density of cyclical unemployment is less convincingly so, with a hint of bimodal peaks. The kernel density of the short rate is clearly bimodal owing to the ZLB. This implies that the short rate in this macrofinance model might be misspecified, however, on the basis of the maximum likelihood fitted Gaussian, this difference is not substantial and may only affect estimations close to the ZLB. Nonetheless, for future research, a Gaussian mixture model may be preferred for interest rate macrofinance modelling, which can easily be added within an EM-estimated framework via state augmentation.

To get a better feel for the nature of the latent factors, principal component analysis is applied to the US yield curve for the 3-, 6-, 12-, 24-, 36-, 60-, 84-, and 120-month interest rates. The data are retrieved from the US Treasury[3]. The constructed yield curve spans from 1990:M1 until 2018:M12. In Figure 4, the fitted level factor $\tilde{L}_t$ is plotted against the short rate and inflation, and the fitted slope factor $\tilde{S}_t$ is plotted against the output gap. It is clear that the slope factor is almost an exact replica of the output gap (except for the abnormal dip around 2009), which strongly supports the RW interpretation. On the other hand, the constructed level factor seems to switch between following the short rate and the inflation at most points in time, and thus may rather closely follow an amalgamation of the two. Nonetheless, the dynamics of $\tilde{L}_t$ tend to have the same pattern as the short rate, even though it does not always stay on the same level. This is clearly seen by its behaviour around the distinctive ZLB. The downward level shift of $\tilde{L}_t$ around the ZLB occurs at aroud the same time as the output gap dropping to its lowest, thus, this may be an indication of the identification problem these statistical factors face.

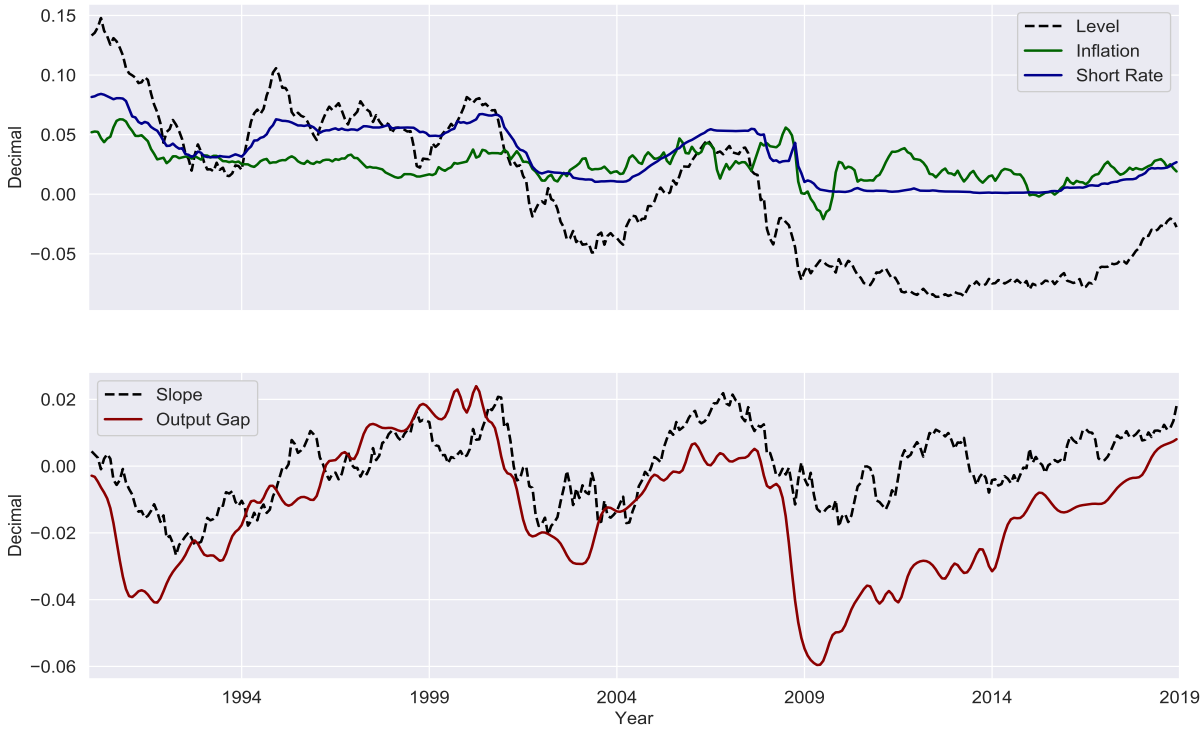



Figure 4: First two principal components of the US yield curve using data from 1990:M1-2018:M12. The slope factor is plotted against the output gap, and the level factor is plotted against both the inflation and the short rate. The axis "Decimal" implies the decimal form of a percentage $[x\%/100]$.

[3] www.treasury.gov

### 6.2 Macroeconomic Benchmark Models

In order to compare the forecasting power of the four models discussed in this paper, two macroeconomic models are estimated for comparison. The first is a restricted VARIMA(2,1,0) model, and the second is a VEC model. In these models, the observation vector does not contain the short rate, rather, only the macroeconomic variables $z_t$. This is because the short rate is handled separately in its forecasting. For the benchmark models, the corresponding macroeconomic model is estimated for the given window of observations, then a Taylor regression is run on the same window. The forecasted macroeconomic variables are then utilised in the fitted Taylor equation to yield the short rate forecast. This way, it is possible to directly compare the benchmarks to the state space models, whose restricted forms have affine latent factor dependencies for the short rate. Handling the Taylor regressions for the short rate separately aim to emulate this phenomenon in the state space models.

The restricted VARIMA(2,1,0) model is an augmented and restricted version of the RW macrofinance model. In this specification, all integrated macroeconomic variables are dependent on only their own lags. Therefore, the autoregressive matrices of both the first and second lags are diagonal, which directly implies that each macroeconomic variable is fitted by a simple ARIMA(2,1,0) process collected into a VAR specification. Results of a preliminary augmented Dickey-Fuller (ADF) test for each individual macroeconomic variable can be found in Table 11 in Appendix J. From these results, at 95% confidence, we are able only to marginally reject the null of a unit root for inflation, whereas the other three variables we do not reject non-stationarity. Thus, it can be safely assumed that at least three of the four macroeconomic variables have stationary roots, i.e. are I(1) time series. This provides the justification for the use of an integrated vector autoregressive model for forecasting purposes. Further tests on higher orders of integration are not implemented, as they are not particularly suitable for the purposes of this research.

It is well known that macroeconomic variables are also strongly cointegrated. The Johansen (1991) vector error correction (VEC) model is chosen to capture the long and short run equilibrium inter-dynamics of the macroeconomic variables, which are not captured in rudimentary VAR models:

$$\Delta z_t = \mu + \mathbf{B} d_t + \mathbf{\Pi} z_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta z_{t-i} + \eta_t.$$

In this specification, only one lagged difference is estimated with no deterministic term nor a constant term:

$$\Delta z_t = \mathbf{\Pi} z_{t-1} + \mathbf{\Gamma} \Delta z_{t-1} + \eta_t,$$

where $\mathbf{\Gamma} = \mathbf{\Pi} - \mathbf{I}_g$. To test the level of cointegration, I apply Johansen's trace and maximum eigenvalue tests (Johansen (1995)) on the above model, which are the multivariate generalisations of the ADF test. These tests estimate the rank of $\mathbf{\Pi}$, as this tells us the level of cointegration that exists in the model. In general, there are three interpretations of the test outcomes: (i) $\mathbf{\Pi} = 0$, which implies all its eigenvalues are zero. This means that no cointegration exists in the model and it reduces to a regular VAR model; (ii) $\mathbf{\Pi}$ has full rank $g$, which means all $z_t$ are stationary; (iii) $0 < \text{Rank}(\mathbf{\Pi}) = m < g$, which implies that $\det(\mathbf{\Pi}) = 0$, and hence cointegration exists. In the last case, $\mathbf{\Pi}$ can be written in terms of the $g \times m$ adjustment and cointegrating vectors $\alpha, \beta$, respectively: $\mathbf{\Pi} = \alpha\beta'$. Johansen's likelihood ratio cointegration maximum eigenvalue test statistic is given sequentially starting at the null hypothesis that $\text{Rank}(\mathbf{\Pi}) = r$ against the alternative $\text{Rank}(\mathbf{\Pi}) = r + 1$ for $r = 0, \ldots, g - 1$, where $g - 1$ is the maximum allowed number of cointegrating vectors. The LR test statistic is given as:

$$\text{LR}^{\text{MaxEig}}(r, r+1) = -T \log(1 - \lambda_{r+1}),$$

going in order from largest to smallest eigenvalue $\lambda$. The first test at which we fail to reject the null yields the estimator for $\text{Rank}(\mathbf{\Pi}) = \hat{r}$. In a similar fashion, the trace test has null hypothesis $\text{Rank}(\mathbf{\Pi}) = \hat{r}$ against the alternative $r < \text{Rank}(\mathbf{\Pi}) \leq g - 1$. The test statistic is given as:

$$\text{LR}^{\text{Tr}}(r, g-1) = -T \sum_{i=r+1}^{g-1} \log(1 - \lambda_i).$$

As with the maximum eigenvalue test, the first null that is not rejected yields the estimated rank of $\mathbf{\Pi}$. The results of these tests are given in Table 2 at the top of the next page. Based on the results of these tests, the VEC model containing inflation, the output gap and cyclical unemployment will be fitted with two cointegrating vectors. Similarly to the VARIMA model, the forecasts produced by the VEC model will be utilised to create a contemporaneous forecast of the short rate using the window-fitted Taylor regression.

| Coint. Vectors | Maximum Eigenvalue | | | | | Trace | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Test Stat. | $\lambda$ | 90% | 95% | 99% | Test Stat. | 90% | 95% | 99% |
| None | 285,79 | 0,49 | 15,72 | 17,80 | 22,25 | 527,29 | 21,78 | 24,28 | 29,51 |
| At most 1 | 167,29 | 0,32 | 9,47 | 11,22 | 15,09 | 241,50 | 10,47 | 12,32 | 16,36 |
| At most 2 | 74,21 | 0,16 | 2,98 | 4,13 | 6,94 | 74,21 | 2,98 | 4,13 | 6,94 |

Table 2: Johansen maximum eigenvalue (left) and trace (right) cointegration test on the full data sample. All tests reject the null and suggest Rank$(\mathbf{\Pi}) = 2$ cointegrating vectors.

### 6.3 State Space Initialisation and Forecasting

In this research, only one-month-ahead forecasts are considered. An in-sample estimation window of $w = 372$ months (31 years) is used for model calibration, which is then used to produce an $h = 1$ month ahead forecast for all macroeconomic variables considered. Then, the calibration window is rolled forward by one month and the model re-estimated for the new window. This yields 59 months' worth of forecasts, which is suitable for model comparison. In order to utilise the information gained from the previous estimation window and ensure convergence to a true optimum, for the (Con-)EM and (Con-)BLAME algorithms the initialisation of the system and covariance matrices after the first estimation are based on the optimum converged parameters of the previous window. The reason for this is because it is very likely that the full dynamical structure of each window is highly correlated to the previous window $\theta^{(K)}$. Specifically, the initialisation of the parameters $\theta = \{\mathbf{F}, \mathbf{H}, \mathbf{K}_1, \mathbf{K}_2, \mathbf{Q}, \mathbf{R}\}$ at each window $w_i = 1, \ldots, 131$ for the classical algorithms is given as:

$$\theta_{w_i}^{(0)} = \begin{cases} \left\{ \left(\mathbf{F}^{(0)}\right)_1, \left(\mathbf{H}^{(0)}\right)_1, \left(\mathbf{K}_1^{(0)}\right)_1, \left(\mathbf{K}_2^{(0)}\right)_1, \left(\mathbf{Q}^{(0)}\right)_1, \left(\mathbf{R}^{(0)}\right)_1 \right\}, & \text{if } w_i = 1, \\ \left\{ \left(\mathbf{F}^{(K)}\right)_{w_{i-1}}, \left(\mathbf{H}^{(K)}\right)_{w_{i-1}}, \left(\mathbf{K}_1^{(K)}\right)_{w_{i-1}}, \left(\mathbf{K}_2^{(K)}\right)_{w_{i-1}}, \left(\mathbf{Q}^{(K)}\right)_{w_{i-1}}, \left(\mathbf{R}^{(K)}\right)_{w_{i-1}} \right\}, & \text{otherwise.} \end{cases}$$

The reason why $\mu_0$ and $\mathbf{P}_0$ are not updated at each window is because they pertain to the beginning of the previous window and are not immediately relevant to the starting point of the new window. Hence, at each window, these are kept as a diffuse initialisation at each calibration window:

$$\left(\mu_0^{(0)}\right)_{w_i} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ and } \left(\mathbf{P}_0^{(0)}\right)_{w_i} = \begin{bmatrix} +\infty & 0 \\ 0 & +\infty \end{bmatrix}, \forall\, w_i = 1, \ldots, 131.$$

In practice, the prior variances are simply set to a large number, i.e. $10^6$. It would make sense to initialise $\mu_0$ and $\mathbf{P}_0$ each window with the first smoothed state and covariance of the previous window, $\hat{\xi}_{1|T}$ and $\mathbf{P}_{1|T}$, however, it is unlikely to have a great impact on the outcome, and so this was not implemented.

In a similar fashion for the Bayesian algorithms, the system matrices have location hyperparameters equal to the previous window's converged values and the covariance hyperparameters remain the same at every iteration window. In all windows, it is assumed that the prior hyperparameters are independent, thus these covariance hyperparameters are all diagonal. For the covariance matrices, the scale matrix and degrees of freedom also remain the same at each window. For more detailed information on the exact form of the prior hyperparameters, refer to Appendix I. The one-step-ahead forecast of the macroeconomic variables $x_{T+1}$ given the final observation of the window $x_T$ is given by:

$$\begin{aligned}\xi_{T+1|T} &= \mathbf{F}^{(K)}\xi_T,\\ x_{T+1|T} &= \mathbf{H}^{(K)}\xi_{T+1|T} + \mathbf{K}_1^{(K)}z_T + \mathbf{K}_2^{(K)}z_{T-1}.\end{aligned}$$

### 6.4 Model Performance

For all models, there are two metrics assessed in terms of forecasting performance, the mean absolute prediction error (MAPE) and the root mean squared prediction error (RMSPE). The reason both are chosen is because the RMSPE heavily penalises errors which are far off the true value, thus, if we do not see a large discrepancy between the two metrics, then it can be concluded that the model performs in a consistent manner throughout the estimation windows. For all seven estimation models considered, $m \in \{\text{R. Walk, VARIMA, VEC, EM, Con-EM, BLAME, Con-BLAME}\}$:

$$\text{MAPE}_m = \frac{1}{T-w-1}\sum_{t=1}^{T-w-1}\left|\hat{x}_{m,t+1} - x_{t+1}\right|, \quad \text{RMSPE}_m = \sqrt{\frac{1}{T-w-1}\sum_{t=1}^{T-w-1}(\hat{x}_{m,t+1} - x_{t+1})^2}.$$

The results of the out-of-sample forecasting performance in terms of MAPE can be found in Table 3 and in terms of RMSPE in Table 4. The random walk model performs strongly for the short rate with an MAPE of only 5 basis points. It predicts the output gap and cyclical unemployment very well, with an MAPE of 9 and 10 basis points, respectively. The random walk performs the least well for inflation, which is expected as inflation is the least persistent variable.

Overall, the random walk remains a difficult model to beat. However, it is not flawless. Several of the models studied here come very close to outperforming the random walk. Without considering the VARIMA and VEC models, the best performing algorithm for the short rate is Con-EM, whereas for the remaining macroeconomic variables, the best performing algorithms are EM and Con-EM with BLAME falling not far behind. Con-BLAME performs substantially worse on all accounts. The conclusions formed based on RMSPE do not differ at all from the ones formed based on MAPE.

| Variable | Estimation Method | | | | | | |
|---|---|---|---|---|---|---|---|
| | R. Walk | VARIMA | VEC | EM | Con-EM | BLAME | Con-BLAME |
| Short Rate | 0.0005 | 40.93 | 42.00 | 2.72 | 1.89 | 22.99 | 59.60 |
| Inflation | 0.0023 | 0.98 | 1.02 | 1.52 | 1.65 | 6.50 | 9.62 |
| Output Gap | 0.0009 | 0.22 | 0.23 | 1.40 | 1.45 | 7.63 | 16.30 |
| Cyc. Unempl. | 0.0010 | 1.02 | 1.08 | 1.28 | 1.41 | 3.99 | 10.52 |

Table 3: Mean absolute prediction errors (MAPEs) for one-step-ahead forecasts of a random walk, two macroeconomic models, two EM models, and two Bayesian EM models. Estimation window is 372 months long (one month rolling), with an out-of-sample testing window of 59 months. Values quoted to the right of the random walk are calculated with respect to the random walk's forecast error itself, so a value below 1 implies that the model beats the random walk by the quoted amount, and a value above 1 implies the opposite.

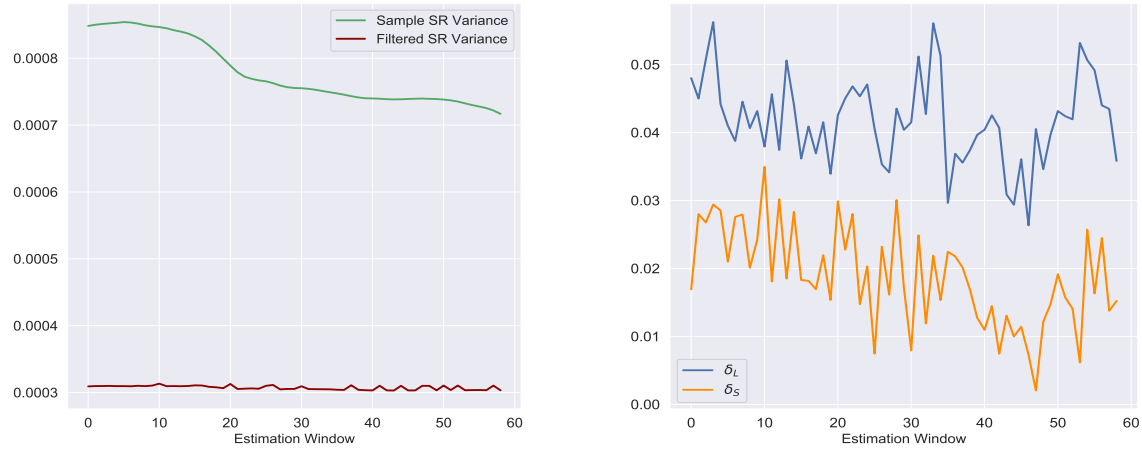
One particularly noticeable forecasting performance is that of the short rate for the VARIMA and VEC models, where they perform 40 times worse (in terms of MAPE) than the random walk control. This will have to do with the fact that the short rate was forecasted independently of the models via the Taylor regression for each independent window. Nonetheless, the VARIMA and VEC models perform very well when comparing the macroeconomic variable forecasts. The fitted macroeconomic models are especially strong when it comes to forecasting the output gap with roughly 4 times more forecasting power. However, this is because monthly output gap data was interpolated using a cubic spline based on the quarterly data, hence the resultant values are induced with a high autocorrelation, making them more predictable than regular data. The real performance is seen in the relative values for inflation and cyclical unemployment, where the VARIMA model outperforms the control only in inflation forecasting.

| Variable | Estimation Method | | | | | | |
|---|---|---|---|---|---|---|---|
| | R. Walk | VARIMA | VEC | EM | Con-EM | BLAME | Con-BLAME |
| Short Rate | 0.0008 | 26.61 | 27.11 | 2.14 | 1.73 | 16.18 | 45.88 |
| Inflation | 0.0028 | 0.99 | 1.03 | 1.59 | 1.67 | 5.65 | 9.39 |
| Output Gap | 0.0012 | 0.25 | 0.22 | 1.51 | 1.61 | 5.94 | 14.81 |
| Cyc. Unempl. | 0.0013 | 1.07 | 0.88 | 1.29 | 1.38 | 3.47 | 10.06 |

Table 4: Root mean squared prediction errors (RMSPEs) for one-step-ahead forecasts of a random walk, two macroeconomic models, two EM models, and two Bayesian EM models. Estimation window is 372 months long (one month rolling), with an out-of-sample testing window of 59 months. Values quoted to the right of the random walk are calculated with respect to the random walk's forecast error itself, so a value below 1 implies that the model beats the random walk by the quoted amount, and a value above 1 implies the opposite.

The EM algorithms show strong performance, only slightly beaten by the random walk for all variables. What is interesting to note is how Con-EM outperforms its unconstrained counterpart with respect to the short rate, which partly confirms the conclusion reached from the simulation study. Considering that unconstrained EM estimates 56 parameters, all macroeconomic variables are allowed to have an influence on the short rate on top of the latent factors, thus, an inevitable degree of overfitting has occurred. For all other macroeconomic variables, the unconstrained EM algorithm slightly beats its constrained counterpart, but not by a significant amount.

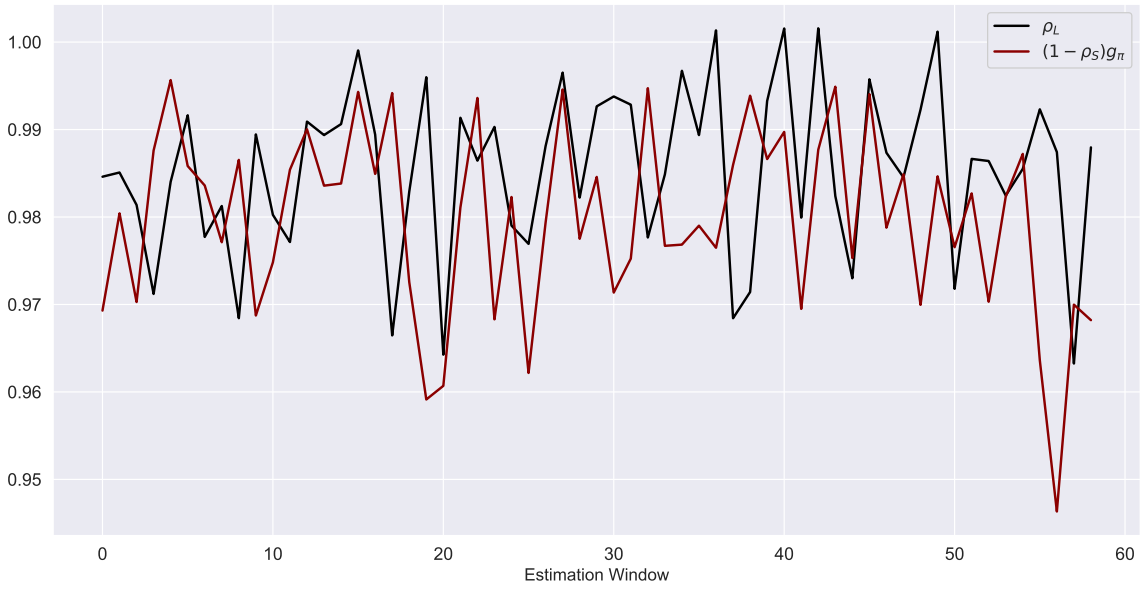
In general, BLAME and Con-BLAME perform poorly. BLAME performs reasonably for the macroeconomic variables being easily outperformed by its frequentist counterparts. Neither of the Bayesian methods perform as well as expected for the short rate. Con-BLAME is on average the worst performing model for forecasting, especially for the short rate where it is nearly 60 times worse than the random walk. To investigate the poorer-than-expected performance further, Figure 5 shows the evolution of the posterior means for some of the estimated parameters relevant to the short rate. The algorithm is seen to be stable, on account of the consistent convergence of the filtered short rate variance in 5a, but does not seem completely consistent when one looks at how the dependencies on the latent variables shift continually from window to window in 5b and 5c. This provides an indication as to why the short rate forecasts are as poor as they are; the algorithm has not properly converged to an optimum.




(a)




(b)




(c)

Figure 5: Rolling window results from the Con-BLAME algorithm. The plots specifically show: (5a) the sample short rate variance against the posterior mean short rate variance for each rolling window; (5b) the posterior mean of the affine dependencies of the short rate on the latent factors for each window; and (5c) the posterior mean of the latent autoregressive parameters for each rolling window.

# 7 Discussion of Results

The VEC and VARIMA models are the strongest forecasting models for the macroeconomic variables, but not for the short rate. This is because the short rate was not included in the model specification and was handled separately via the Taylor regression. Had the short rate

been included in these macroeconomic models, their forecasting power of the short rate would undoubtedly be significantly higher. However, this holistic approach of including the short rate in the VARIMA and VEC models detracts from the reason of including these models in the first place as macroeconomic benchmarks. These models are chosen to capture as many macroeconomic relations as possible, including Okun's law and the Phillips curve, hence the additional inclusion of cyclical unemployment in the model specification. Where these models excel is in their forecasting ability for the output gap, increasing predictive power four-fold over the random walk. As discussed earlier, this is likely due to the cubic spline interpolation applied on the originally attained quarterly data in order to yield monthly output gap data. In future research, it might be better to study quarterly data instead to avoid this effect.

In terms of forecasting, the classical EM algorithms are strong competitors to the macroeconomic models. The constrained EM algorithm is almost as powerful as unconstrained EM when it comes to forecasting macroeconomic variables, performing only very slightly worse than the random walk control. The constrained EM algorithm is the strongest at forecasting the short rate, which exemplifies the power of adding hard macroeconomic restrictions on the system. For future research, the classical EM algorithm remains a powerful analytical alternative for large factor model estimation.

The Bayesian algorithms perform adequately, but not up to the standard as hoped. This may be due to a variety of reasons. A prominent one may be the fact that only 100 samples are used for each estimation window, causing the posterior means of each parameter to be unreliable. This is clear from the discontinuous evolution of the posterior means in Figures 5b and 5c. Future research would need to increase the amount of samples used to at least 1000. Another potential downfall in the research is the number of EM loops used for the Bayesian algorithms. As certain sampled initialisations may stray far away from a given optimum of the likelihood function, 100 loops may in fact not be enough to guarantee convergence. Therefore, for future research, rather than setting the EM loops to a standard 100, a tolerance for convergence should be set such that the system is guaranteed to have reached an optimum, such as $K = 1000$. There is a caveat to these recommendations: implementing them will significantly increasing the computation time required to perform sound forecasting analysis, thus, a more efficient way of producing higher samples/faster convergence is desired (such as sampling the posterior parameters with replacement in the form of a bootstrap, for example).

There is one final concern I have with the implementation of BLAME on the macroeconomic data. Throughout several trial runs, I discovered that some (seemingly random) combinations of prior parameters yield near-singular Kalman gain matrices, which in turn yield huge predicted covariance matrices. These covariance matrices are prominent in the estimation of the latent covariance matrix $\mathbf{Q}$ and the latent autoregressive matrix $\mathbf{F}$, and in these trial runs it was the covariance matrix that took the brunt of the huge predicted covariances. This is simply because in the estimation of $\mathbf{F}$, the large nature of $\mathbf{P}_{t+1|t}$ is negated through the inverse in the 'M' step. Quick fixes may be implemented by using the Moore-Penrose pseudoinverse, however this does not solve the underlying problem. Therefore, I suggest extra care be taken in future research to test the convergence properties of the BLAME algorithm for a variety of prior distributions and hyperparameters.

On the whole, the BLAME algorithm performs only 3-4 times as bad as its classical counterpart, which is reasonable. However, this can only be said for the macroeconomic variables. For the short rate, the performance is worse by a factor of 10, and it is not immediately clear why this is the case, apart from those mentioned above. The Con-BLAME algorithm performs far below the standard expected from a forecasting model. One aspect of its implementation that may cause inaccuracies during the estimation procedure is imposing both hard constraints and prior distributions for all system matrices involved. With finite sample data, it might become cumbersome for the algorithm to handle optimising parameters within a system matrix that contains both hard constraints and shrinkage. For future research, it may be favourable to impose the macroeconomic (hard) constraints only on the observation system matrices, and thus only impose prior distributions on the latent system matrices. As can be seen in Figures 5b and 5c, the posterior means resulting from the Con-BLAME algorithm yield reasonable values, but fluctuate quite a lot throughout the estimation windows. It is, nonetheless, difficult to tell whether these fluctuations arise due to the naturally strongly persistent latent factors, or the estimation procedure itself. Therefore, further research is warranted into both Bayesian algorithms, with a possible mitigation of the types of constraints applied to the system and covariance matrices. Additionally, since the short rate is modelled to be influenced only by the latent factors, it should also have zero covariance with the remaining macroeconomic variables, however, these hard constraints were not imposed on the covariance matrices; this is also an aspect worthy of further research.

# 8 Concluding Remarks

This research has explored the Bayesian inferential method for multidimensional factor model estimation and forecasting. A self-contained macrofinance model containing the US short rate, inflation, output gap, and cyclical unemployment is constructed in the form of a latent 2-factor state space model, taking inspiration from the Rudebusch and Wu (2008) macrofinance model. Both Bayesian and classical estimation techniques have been applied to an in-sample calibration window for the purpose of evaluating one-step-ahead forecasting performance. The classical estimation techniques are the famous Expectation Maximisation (EM) algorithm, originally developed by Dempster et al. (1977), as well as a constrained version which utilises Lagrange multipliers. The Bayesian algorithms are newly developed analytical extensions of the EM algorithm, named the BLAME algorithm. The development of these extensions allow for a full analytical Bayesian treatment of multidimensional factor model estimation without the need for extensive computation in the form of particle filters or Markov chain Monte Carlo (MCMC) simulations. These extensions aim to solve one of the main problems faced by the classical EM algorithm, namely the degeneracy of covariance matrix estimation causing it to fail.

The linear Gaussian state space model allows complete freedom to the researcher to specify any prior distribution for the system and covariance matrices; for the purposes of this research, matricvariate normal prior distributions were chosen for the system matrices and inverse Wishart prior distributions for the covariance matrices. These specifications yield the analytical BLAME loop which provides exact posterior distributions of the parameters, on which one may apply a number of posterior analyses. In order to understand the strengths and weaknesses of the (Con-)EM and (Con-)BLAME algorithms, a simulation study is conducted on a simulated hidden Markov system based on a latent 2-factor model and three simulated independent Gaussian observations. The simulation study finds that the classical methods are more powerful at identifying latent factor coefficients, whereas the Bayesian methods excel at identifying observational coefficients, even when the number of prior samples are as low as $N = 100$. Nonetheless, I believe one single simulated dataset is not fully indicative of these algorithms' capabilities, thus I advocate the study of a more sophisticated simulation environment that would yield a distribution of posterior parameter means rather than a single simulated dataset, from which a more accurate depiction of how these algorithms estimate the parameters would be clear.

When applied to forecasting real macroeconomic data, the Bayesian algorithms do not perform as hoped. They tend to have difficulties identifying the system parameters when shifting from window to window, without a clear indication as to why this may be the case. This causes their one-step-ahead forecasts to be quite inaccurate, especially for the short rate, reaching as high as 60 times worse than a basic random walk. The reasons for this may be caused by lack of convergence or insufficient samples to yield a correct posterior distribution. On the other hand, this research has shown that the EM algorithm and its constrained version are powerful competitors to standard macroeconomic models. In terms of macroeconomic forecasting, they perform very well, being comparable in to the benchmark macroeconomic models, but still not beating a simple random walk.

Aspects of this research that remain unanswered exist both on the theoretical side of the BLAME algorithm, as well as on its practical implementation, including its effect in a univariate setting. Convergence of the BLAME estimates is clearly dependent on the choice of prior; there may well be a family of prior distributions and/or hyperparameters that produce divergent likelihoods. This remains to be explored in future research. As out-of-sample forecasting was a main topic of this research, the number of samples and EM loops where restricted to 100. It would be interesting to assess the effect of increasing these values by a few orders of magnitude on both the in-sample and out-of-sample estimations. It is expected, of course, that these will merely improve, making BLAME an even stronger candidate for future forecasting literature.

Additionally, only the short rate has been studied. In theory, the observation equation could be loaded with other interest rates spanning the entire yield curve. However, the high persistence of the yield curve may cause additional issues related to overfitting. Thus, should the entire yield curve be modelled, the constrained EM or constrained BLAME algorithms would be preferred, in order to avoid these complications. Additionally, I recommend that hard constraints need only be implemented for the observation system matrices, whereas the shrinkage techniques supplied by the BLAME algorithm need only apply to the covariance matrices and possibly the latent autoregressive matrix, as I am of the opinion that the mixing of the two types of constraints may interfere with the optimisation procedure as well.

In conclusion, this research advocates the usage of the BLAME algorithm as a powerful alternative to other computationally strenuous Bayesian methods. The BLAME algorithm (as well as its constrained version) bridges the gap between retrieving analytical results and a full

Bayesian treatment of a state space model. It allows for the injection of prior information on higher-dimensional systems, which are notorious for having sparse specifications, while also still remaining analytically tractable. In general, the use of the EM algorithm and its variants are powerful tools for forecasting macroeconomic variables and their usage in multidimensional factor models are warranted, especially for short rate modelling. In principle, one could estimate the short rate using either EM or BLAME, and then in combination with a non-Gaussian/nonlinear ATSM map the rest of the yield curve. One may be able to efficiently capture the entire yield curve by capturing only the dynamics of the short rate, driven solely by observed macroeconomic variables and filtered latent factors. Should this be successful, one could effectively link the yield curve to the macroeconomy, and the BLAME algorithm may be one of the steps allowing us to do so.

## References

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72(3):269–342.

Ang, A. and Piazzesi, M. (2003). A No Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics*, 50:745–787.

Batista, R. L. and Laurini, M. (2016). Bayesian estimation of term structure models: An application of the Hamiltonian Monte Carlo method. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 2:79–91.

Box, G. E. P. and Muller, M. E. (1958). A Note on the Generation of Random Normal Deviates. *Annals of Mathematical Statistics*, 29(2):610–611.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling. *Journal of the American Statistical Association*, 87(418):493–500.

Chib, S. and Greenberg, E. (1996). Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometric Theory*, 12(3):409–431.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316.

Christensen, J. H. E., Lopez, J. A., and Rudebusch, G. D. (2014). How Efficient is the Kalman Filter at Estimating Affine Term Structure Models? *Federal Reserve Bank of San Francisco*.

Christoffersen, P., Dorion, C., Jacobs, K., and Karoui, L. (2014). Nonlinear Kalman Filtering in Affine Term Structure Models. *Management Science*, 60(9).

Creal, D. D. and Wu, J. C. (2015). Estimation of affine term structure models with spanned or unspanned stochastic volatility. *Journal of Econometrics*, 185(1):60–81.

Dai, Q. and Singleton, K. J. (2000). Specification Analysis of Affine Term Structure Models. *The Journal of Finance*, 55(5):1943–1978.

de Jong, F. (2000). Time Series and Cross Section Information in Affine Term-Structure Models. *Journal of Business & Economic Statistics*, 18(3):300–314.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–22.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 113:309–338.

Dorfman, J. H. and Havenner, A. M. (1992). A Bayesian approach to state space multivariate time series modeling. *Journal of Econometrics*, 52(3):315–346.

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

Durbin, J. and Koopman, S. J. (2002). Time series analysis of non Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B*, 62(1):3–56.

Garutti, T., Papazoglou, I., Rallis, T., and Singh, S. (2019). Incorporating Unemployment in a Macrofinance Time Series Model. *Financial Case Studies, Erasmus University Rotterdam*. unpublished.

Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57(6):1317–1339.

Hamilton, J. D. (1994). State-space models, in R. F. Engle and D. McFadden (eds.). *Handbook of Econometrics*, 4:3039–3080.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212.

Jobson, J. D. and Korkie, B. (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association*, 75:544–554.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6):1551–1580.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng*, 82(1):35–45.

Lange, R.-J. (2019). A note on the EM algorithm with linear constraints.

Lopes, H. F. and Tsay, R. S. (2010). Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting*, 30(1):168–209.

Lund, J. (1997). Non Linear Kalman Filtering Technique for Term Structure Models. Finance working papers, University of Aarhus, Aarhus School of Business, Department of Business Studies.

Maanan, S., Dumitrescu, B., and Giurcaneanu, C. D. (2018). Maximum Entropy Expectation-Maximization Algorithm for Fitting Latent-Variable Graphical Models to Multivariate Time Series. *Entropy*, 20(1).

Mikkelsen, P. (2001). MCMC Based Estimation of Term Structure Models. Finance Working Papers 01-7, University of Aarhus, Aarhus School of Business, Department of Business Studies.

Rao, K. R., Boiroju, N. K., and Reddy, M. K. (2011). Generation of Standard Normal Random Variables. *Indian Journal of Scientific Research*, 2(4):83–85.

Rauch, H. E., Striebel, C. T., and Tung, F. (1965). Maximum likelihood estimates of linear dynamic systems. *American Institute of Aeronautics and Astronautics*, 3(8).

Rudebusch, G. D. and Wu, T. (2008). A Macrofinance Model of the Term Structure, Monetary Policy and the Economy. *The Economic Journal*, 118:906–926.

Sakaria, D. K. and Griffin, J. E. (2017). On efficient Bayesian inference for models with stochastic volatility. *Econometrics and Statistics*, 3:23–33.

Sanford, A. D. and Martin, G. M. (2005). Simulation based Bayesian estimation of an affine term structure model. *Computational Statistics & Data Analysis*, 49(2):527–554.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, 20:147–162.

Stock, J. H. and Watson, M. W. (2009). Phillips curve inflation forecasts. *Federal Reserve Bank of Boston*, 53.

Stock, J. H. and Watson, M. W. (2011). Dynamic Factor Models. *The Oxford Handbook of Economic Forecsting*.

Tanizaki, H. and Mariano, R. S. (1998). Nonlinear and non Gaussian state space modeling with Monte Carlo simulations. *Journal of Econometrics*, 83:263–290.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie Rochester Conference Series on Public Policy*, 39:195–214.

Thompson, S. B. (2008). Identifying Term Structure Volatility from the LIBOR-Swap Curve. *The Review of Financial Studies*, 21(2):819–854.

# Appendix

# A Probability Distributions

## A.1 Gaussian

**Definition A.1** (Gaussian distribution)**.** A random vector $x \in \mathbb{R}^n$ has a Gaussian distribution if its probability distribution has the form

$$x \sim \mathcal{N}\Big(\mu, \mathbf{\Sigma}\Big) = p(x|\mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\mathbf{\Sigma}^{-1}(x-\mu)} \tag{A.1}$$

given a mean $\mu \in \mathbb{R}^n$ and positive semi-definite covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$.

## A.2 Matricvariate Normal

**Definition A.2** (Matricvariate normal distribution)**.** A random matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ has a matricvariate normal distribution if its distribution has the form

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}\Big(\mathbf{M}, \mathbf{U}, \mathbf{V}\Big) = p(\mathbf{X}|\mathbf{M}, \mathbf{U}, \mathbf{V}) = \frac{1}{(2\pi)^{np/2}|\mathbf{V}|^{n/2}|\mathbf{U}|^{p/2}} e^{-\frac{1}{2}\mathrm{tr}\left[\mathbf{V}^{-1}(\mathbf{X}-\mathbf{M})'\mathbf{U}^{-1}(\mathbf{X}-\mathbf{M})\right]}, \tag{A.2}$$

where $\mathbf{M}$ is an $n \times p$ matrix, $\mathbf{U}$ is an $n \times n$ matrix, and $\mathbf{V}$ is a $p \times p$ matrix.

## A.3 Inverse Wishart

**Definition A.3** (Inverse Wishart distribution)**.** Let $\mathbf{X}$ be a $p \times p$ symmetric, positive definite, random matrix. Let $\mathbf{\Psi}$ be $p \times p$ a symmetric, positive semi-definite matrix. Let $\nu \geq p$ represent the degrees of freedom of $\mathbf{X}$. Then $\mathbf{X}$ has an inverse Wishart distribution if its probability density has the form

$$\mathbf{X} \sim \mathcal{W}_p^{-1}\Big(\mathbf{\Psi}, \nu\Big) = p(\mathbf{X}|\mathbf{\Psi}, \nu) = \frac{|\mathbf{\Psi}|^{\nu/2}}{2^{\nu p/2}\Gamma_p\left(\frac{\nu}{2}\right)} |\mathbf{X}|^{-(\nu+p+1)/2} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi}\mathbf{X}^{-1})}, \tag{A.3}$$

where $\Gamma_p$ is the multivariate gamma function as above.

# B Classical Filtering and Smoothing Results

## B.1 The Kalman Filter

Expressing the observation and state equation in this probabilistic language allows us to write:

$$p(\xi_t|\xi_{t-1},\theta) = \mathcal{N}\Big(\mathbf{F}\xi_{t-1}, \mathbf{Q}\Big),$$
$$p(x_t|\xi_t,\theta) = \mathcal{N}\Big(\mathbf{H}\xi_t + \mathbf{K}_1 z_{t-1} + \mathbf{K}_2 z_{t-2}, \mathbf{R}\Big).$$

The predicted state is thus given by

$$\mathbb{E}_t[\xi_{t+1}] \equiv \hat{\xi}_{t+1|t} = \mathbf{F}\hat{\xi}_{t|t}, \tag{B.1}$$

which will have covariance with the current-time state

$$\mathrm{Cov}_t[\hat{\xi}_{t+1|t}, \xi_t] \equiv \mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q} \tag{B.2}$$

The prediction step is given by the Chapman-Kolmogorov equation, which is now solvable analytically:

$$\begin{aligned} p(\xi_{t+1}|x_{1:t},\theta) &= \int p(\xi_{t+1}|\xi_t,\theta)p(\xi_t|x_{1:t},\theta)\, d\xi_t \\ &= \mathcal{N}\Big(\mathbf{F}\hat{\xi}_{t|t}, \mathbf{Q}\Big)\mathcal{N}\Big(\mathbf{F}\xi_{t-1}, \mathbf{Q}\Big) \\ &= \mathcal{N}_2\left(\begin{bmatrix}\mathbf{F}\hat{\xi}_{t|t} \\ \mathbf{F}\xi_{t-1}\end{bmatrix}, \begin{bmatrix}\mathbf{P}_{t+1|t} & \mathbf{F}\mathbf{P}_{t|t} \\ \mathbf{P}_{t|t}\mathbf{F}' & \mathbf{P}_{t|t}\end{bmatrix}\right). \end{aligned}$$

The joint distribution of the predicted states and the observations at $t+1$ is now derived in order to get the updating (filtering) step. Consider first the following:

$$\begin{aligned} \mathbb{E}_t[x_{t+1}] &= \mathbf{H}\hat{\xi}_{t+1|t} + \mathbf{K}_1 z_t + \mathbf{K}_2 z_{t-1} \\ \mathrm{Var}_t[x_{t+1}] &= \mathbf{H}\mathbf{P}_{t+1|t}\mathbf{H}' + \mathbf{R}. \end{aligned}$$

Therefore, the covariance of the states and observations is given by

$$\mathrm{Cov}_t[\xi_{t+1}, x_{t+1}] = \mathrm{Cov}_t[\xi_{t+1}, \mathbf{H}\xi_{t+1} + \mathbf{K}_1 z_t + \mathbf{K}_2 z_{t-1} + v_{t+1}]$$

$$= \text{Var}_t[\xi_{t+1}]\mathbf{H}'$$
$$= \mathbf{P}_{t+1|t}\mathbf{H}'.$$

Using the above relations, we are able to express the joint distribution of observations and states as a joint normal:

$$\begin{pmatrix} x_{t+1} \\ \xi_{t+1} \end{pmatrix} \Bigg| \mathcal{F}_t \sim \mathcal{N}_2 \left( \begin{bmatrix} \mathbf{H}\hat{\xi}_{t+1|t} + \mathbf{K}_1 z_t + \mathbf{K}_2 z_{t-1} \\ \mathbf{F}\hat{\xi}_{t|t} \end{bmatrix}, \begin{bmatrix} \mathbf{H}\mathbf{P}_{t+1|t}\mathbf{H}' + \mathbf{R} & \mathbf{H}\mathbf{P}_{t+1|t} \\ \mathbf{P}_{t+1|t}\mathbf{H}' & \mathbf{P}_{t+1|t} \end{bmatrix} \right).$$

We can express the Bayesian filtering distribution as the conditional distribution of $\xi_{t+1}$ given the new observation $x_{t+1}$, which is readily available using the joint distribution given above:

$$p(\xi_{t+1}|x_{1:t+1},\theta) = \frac{p(x_{t+1}|\xi_{t+1},\theta)p(\xi_{t+1}|x_{1:t},\theta)}{\int p(x_{t+1}|\xi_{t+1},\theta)p(\xi_{t+1}|x_{1:t},\theta)\, d\xi_{t+1}} = \mathcal{N}\left(\hat{\xi}_{t+1|t+1}, \mathbf{P}_{t+1|t+1}\right).$$

By defining the Kalman gain and prediction error as:

$$\mathbf{J} = \mathbf{P}_{t+1|t}\mathbf{H}' \left(\mathbf{H}\mathbf{P}_{t+1|t}\mathbf{H}' + \mathbf{R}\right)^{-1}$$
$$e_{t+1} = x_{t+1} - \mathbf{H}\hat{\xi}_{t+1|t} - \mathbf{K}_1 z_t - \mathbf{K}_2 z_{t-1},$$

we can thus express the mean and covariance of the filtering distribution as:

$$\hat{\xi}_{t+1|t+1} = \hat{\xi}_{t+1|t} + \mathbf{J}e_{t+1} \tag{B.3}$$
$$\mathbf{P}_{t+1|t+1} = (\mathbf{I} - \mathbf{J}\mathbf{H})\, \mathbf{P}_{t+1|t}. \tag{B.4}$$

The set of relations defined by Equations (B.1)-(B.4) complete the Kalman filtering equations. In the next subsection, we derive the Kalman smoothing distribution, specifically, the RTS smoother. This is a vital stepping stone to deriving the EM algorithm for calculating the posterior distribution of the parameters.

## B.2 The Rauch-Tung-Striebel Smoother

The Bayesian smoothing distribution is the joint distribution of any two time-adjacent states given the full observation set $\mathcal{F}_T$. This allows us to perform a "backward-pass" of the data given

the filtered states and smooth out the state estimates. This joint distribution is given as:

$$\begin{aligned}
p(\xi_{t+1}, \xi_t | x_{1:T}, \theta) &= p(\xi_t | \xi_{t+1}, x_{1:T}, \theta) \times p(\xi_{t+1} | x_{1:T}, \theta) \\
&= p(\xi_t | \xi_{t+1}, x_{1:t}, \theta) \times p(\xi_{t+1} | x_{1:T}, \theta) \\
&= \mathcal{N}\Big(\hat{\xi}_{t|t} + \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}(\xi_{t+1} - \hat{\xi}_{t+1|t}), \mathbf{P}_{t|t} - \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}\mathbf{F}\mathbf{P}_{t|t}\Big) \times p(\xi_{t+1} | x_{1:T}, \theta).
\end{aligned}$$

Now we integrate over all possible "future" states $\xi_{t+1}$, and thus the RTS smoothing distribution is given as:

$$p(\xi_t | x_{1:T}, \theta) = \mathcal{N}\Big(\hat{\xi}_{t|T}, \mathbf{P}_{t|T}\Big),$$

where the Kalman smoothing steps are:

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} + \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}) \tag{B.5}$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} - \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}(\mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|T})\mathbf{P}_{t+1|t}^{-1}\mathbf{F}\mathbf{P}_{t|t}. \tag{B.6}$$

Collected altogether and in full form, the Kalman filtering, updating and RTS smoothing quantities for the LGau-SSM system are:

$$\begin{aligned}
\text{Kalman prediction step: } & \hat{\xi}_{t+1|t} = \mathbf{F}\hat{\xi}_{t|t} \\
& : \mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q} \\
\text{Kalman update step: } & \hat{\xi}_{t+1|t+1} = \hat{\xi}_{t+1|t} + \mathbf{P}_{t+1|t}\mathbf{H}'\left(\mathbf{H}\mathbf{P}_{t+1|t}\mathbf{H}' + \mathbf{R}\right)^{-1}\left(x_{t+1} - \mathbf{H}\hat{\xi}_{t+1|t} - \mathbf{K}_1 z_t - \mathbf{K}_2 z_{t-1}\right) \\
& : \mathbf{P}_{t+1|t+1} = \left(\mathbf{I}_q - \mathbf{P}_{t+1|t}\mathbf{H}'\left(\mathbf{H}\mathbf{P}_{t+1|t}\mathbf{H}' + \mathbf{R}\right)^{-1}\mathbf{H}\right)\mathbf{P}_{t+1|t} \\
\text{RTS smoothing step: } & \hat{\xi}_{t|T} = \hat{\xi}_{t|t} + \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}\left(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}\right) \\
& : \mathbf{P}_{t|T} = \mathbf{P}_{t|t} - \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}_{t+1|t}^{-1}\left(\mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|T}\right)\mathbf{P}_{t+1|t}^{-1}\mathbf{F}\mathbf{P}_{t|t} \\
& : \mathbf{P}_{t+1,t|T} = \mathbf{P}_{t+1|T}\mathbf{P}_{t+1|t}^{-1}\mathbf{F}\mathbf{P}_{t+1|t+1},
\end{aligned}$$

where $\mathbf{P}_{t+1,t|T}$ is the cross covariance between the latent states $\hat{\xi}_{t+1}$ and $\hat{\xi}_t$, given by their joint normal distribution.

# C Classical EM Results

The quantity to be maximised given a set of chosen parameters is:

$$\begin{aligned}
\mathcal{Q}\left(\theta, \theta^{(k)}\right) &= \mathbb{E}\left[\log p(\xi_{0:T}, x_{1:T}|\theta)|\, \theta^{(k)}\right] = \int p\left(\xi_{0:T}|x_{1:T}, \theta^{(k)}\right) \log p(\xi_{0:T}, x_{1:T}|\theta)\, d\xi_{0:T} \\
&= \int p\left(\xi_{0:T}|x_{1:T}, \theta^{(k)}\right) \log p\left(\xi_0|\theta^{(k)}\right) d\xi_0 \\
&\quad + \sum_{t=1}^{T} \int p\left(\xi_t, \xi_{t-1}|x_{1:T}, \theta^{(k)}\right) \log p\left(\xi_t|\xi_{t-1}, \theta^{(k)}\right) d\xi_t d\xi_{t-1} \\
&\quad + \sum_{t=1}^{T} \int p\left(\xi_t|x_{1:T}, \theta^{(k)}\right) \log p\left(x_t|\xi_t, \theta^{(k)}\right) d\xi_t \\
&= -\frac{2T-2}{2} \log(2\pi) + \frac{1}{2} \log\left|\mathbf{P}_0^{-1}\right| - \frac{1}{2} \left(\xi_0 - \mu_0\right)' \mathbf{P}_0^{-1} \left(\xi_0 - \mu_0\right) \\
&\quad + \frac{T-1}{2} \log\left|\mathbf{Q}^{-1}\right| - \frac{1}{2} \sum_{t=1}^{T} \left(\xi_t - \mathbf{F}\xi_{t-1}\right)' \mathbf{Q}^{-1} \left(\xi_t - \mathbf{F}\xi_{t-1}\right) \\
&\quad + \frac{T-2}{2} \log\left|\mathbf{R}^{-1}\right| - \frac{1}{2} \sum_{t=3}^{T} \left(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\right)' \mathbf{R}^{-1} \left(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\right).
\end{aligned} \tag{C.1}$$

The algorithm is initialised given a set of prior parameters $\theta^{(k)}$, which may or may not be based on the nature of the data or problem. Equation (C.1) is effectively the 'E' step of the EM algorithm. What remains is to maximise this quantity with respect to the parameter vector $\theta = \{\mu_0, \mathbf{P}_0, \mathbf{F}, \mathbf{H}, \mathbf{K}_1, \mathbf{K}_2, \mathbf{Q}, \mathbf{R}\}$, which is the 'M' step. The $\mathcal{Q}$-maximising quantities are given below. I omit their derivations as they are standard procedure (Särkkä (2013)) and form a part of the BLAME algorithm seen later. For the system matrices we have:

$$\mathbf{F}^{(k+1)} = \left(\sum_{t=1}^{T} \hat{\xi}_{t|T} \hat{\xi}'_{t-1|T} + \mathbf{P}_{t,t-1|T}\right) \left(\sum_{t=1}^{T} \hat{\xi}_{t-1|T} \hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T}\right)^{-1}, \tag{C.2}$$

$$\mathbf{H}^{(k+1)}|\mathbf{K}_1^{(k)}, \mathbf{K}_2^{(k)} = \left(\sum_{t=3}^{T} \left(x_t - \mathbf{K}_1^{(k)} z_{t-1} - \mathbf{K}_2^{(k)} z_{t-2}\right) \hat{\xi}'_{t|T}\right) \left(\sum_{t=3}^{T} \hat{\xi}_{t|T} \hat{\xi}'_{t|T} + \mathbf{P}_{t|T}\right)^{-1}, \tag{C.3}$$

$$\mathbf{K}_1^{(k+1)}|\mathbf{H}^{(k+1)}, \mathbf{K}_2^{(k)} = \left(\sum_{t=3}^{T} \left(x_t - \mathbf{H}^{(k+1)} \hat{\xi}_{t|T} - \mathbf{K}_2^{(k)} z_{t-2}\right) z'_{t-1}\right) \left(\sum_{t=3}^{T} z_{t-1} z'_{t-1}\right)^{-1}, \tag{C.4}$$

$$\mathbf{K}_2^{(k+1)}|\mathbf{H}^{(k+1)}, \mathbf{K}_1^{(k+1)} = \left(\sum_{t=3}^{T} \left(x_t - \mathbf{H}^{(k+1)} \hat{\xi}_{t|T} - \mathbf{K}_1^{(k)} z_{t-1}\right) z'_{t-2}\right) \left(\sum_{t=3}^{T} z_{t-2} z'_{t-2}\right)^{-1}, \tag{C.5}$$

and for the covariance matrices we have:

$$\mathbf{Q}^{(k+1)}|\mathbf{F}^{(k+1)} = \sum_{t=1}^{T}\Big[\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T} - \Big(\hat{\xi}_{t|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t,t-1|T}\Big)\mathbf{F}'^{(k+1)} \tag{C.6}$$

$$- \mathbf{F}^{(k+1)}\Big(\hat{\xi}_{t-1|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t-1,t|T}\Big)$$

$$+ \mathbf{F}^{(k+1)}\Big(\hat{\xi}_{t-1|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T}\Big)\mathbf{F}'^{(k+1)}\Big] \tag{C.7}$$

$$\mathbf{R}^{(k+1)}|\mathbf{H}^{(k+1)}, \mathbf{K}_1^{(k+1)}, \mathbf{K}_2^{(k+1)} = \sum_{t=3}^{T}\Bigg[\Big(x_t - \mathbf{K}_1^{(k+1)}z_{t-1} - \mathbf{K}_2^{(k+1)}z_{t-2}\Big)\Big(x_t - \mathbf{K}_1^{(k+1)}z_{t-1} - \mathbf{K}_2^{(k+1)}z_{t-2}\Big)'$$

$$- \mathbf{H}^{(k+1)}\hat{\xi}_{t|T}\Big(x_t - \mathbf{K}_1^{(k+1)}z_{t-1} - \mathbf{K}_2^{(k+1)}z_{t-2}\Big)'$$

$$- \Big(x_t - \mathbf{K}_1^{(k+1)}z_{t-1} - \mathbf{K}_2^{(k+1)}z_{t-2}\Big)\hat{\xi}'_{t|T}\mathbf{H}'^{(k+1)}$$

$$+ \mathbf{H}^{(k+1)}\Big[\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T}\Big]\mathbf{H}'^{(k+1)}\Bigg]. \tag{C.8}$$

# D Constrained EM Derivation

In what follows, I explicitly derive the $\mathcal{Q}_{\text{con}}$-maximising quantity for the system matrix $\mathbf{F}$. This derivation is based on a note written by Lange (2019), who does so for the $\mathbf{H}$ system matrix of a more general state space model.

When applying constraints to a system, the most appropriate way of doing so is in the form of Lagrange multipliers. To do so for a matrix, it is required to express its elements in vector format, i.e. we need to *vectorise* it. Suppose we impose $p \leq q^2$ restrictions on the values of the elements of the state autoregressive matrix $\mathbf{F}$, which will thus be written in the form:

$$\mathbf{B}\text{vec}\left(\mathbf{F}\right) = b,$$

where $\mathbf{B}$ is the $p \times q^2$ matrix of ones and zeros picking out the specific elements of vec $(\mathbf{F})$ to equal the corresponding elements of the $p \times 1$ vector $b$. Specifically, for $\mathbf{F}$, we have only one restriction: $F_{1,2} = b = 0$, hence we have in fact $B = [0\,0\,1\,0]$. For better illustration, the constraint

matrix and values for vec $(\mathbf{K}_1)$ would be

$$\underset{[p\times rg]}{\mathbf{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}; \quad \underset{[p\times 1]}{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Returning to the specification for vec $(\mathbf{F})$, the state transition equation in vectorised form is thus written as:

$$\begin{aligned} \xi_{t+1} &= \text{vec}\left(\mathbf{F}\xi_t\right) + v_{t+1} \\ &= \text{vec}\left(\mathbf{I}_q\mathbf{F}\xi_t\right) + v_{t+1} \\ &= \left(\xi_t' \otimes \mathbf{I}_q\right)\text{vec}\left(\mathbf{F}\right) + v_{t+1} \end{aligned}$$

where we use vec $(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})$ vec $(\mathbf{B})$. Using this formulation, the constrained expected log-likelihood (with only the terms relevant to vec $(\mathbf{F})$) is therefore written as:

$$\mathcal{Q}_{\text{con}}|_{\theta=\text{vec}(\mathbf{F})} = -\frac{1}{2}\sum_{t=1}^{T}\left[\xi_t - \left(\xi_t' \otimes \mathbf{I}_q\right)\text{vec}\left(\mathbf{F}\right)\right]'\mathbf{Q}^{-1}\left[\xi_t - \left(\xi_t' \otimes \mathbf{I}_q\right)\text{vec}\left(\mathbf{F}\right)\right] - \lambda'\left(\mathbf{B}\text{vec}\left(\mathbf{F}\right) - b\right),$$

where $\lambda$ is the Lagrange multiplier of appropriate dimensions. The maximisation of this quantity requires setting the first order condition with respect to both vec $(\mathbf{F})$ and $\lambda$ to zero:

$$\begin{aligned} \frac{\partial \mathcal{Q}_{\text{con}}}{\partial \lambda} &= \mathbf{B}\text{vec}\left(\mathbf{F}\right) - b = 0 \\ \frac{\partial \mathcal{Q}_{\text{con}}}{\partial \text{vec}\left(\mathbf{F}\right)} &= \sum_{t=1}^{T}\left(\xi_t' \otimes \mathbf{I}_q\right)'\mathbf{Q}^{-1}\left[\xi_t - \left(\xi_t' \otimes \mathbf{I}_q\right)\text{vec}\left(\mathbf{F}\right)\right] - \mathbf{B}'\lambda \\ &= -\sum_{t=1}^{T}\left(\xi_t' \otimes \mathbf{I}_q\right)'\mathbf{Q}^{-1}\left(\xi_t' \otimes \mathbf{I}_q\right)\text{vec}\left(\mathbf{F}\right) + \sum_{t=2}^{T}\left(\xi_t' \otimes \mathbf{I}_q\right)'\mathbf{Q}^{-1}\xi_t - \mathbf{B}'\lambda = 0. \end{aligned}$$

It is thus possible to combine the two conditions into one system solution expressed as:

$$\begin{bmatrix} \mathbf{\Omega} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0}_p \end{bmatrix}\begin{bmatrix} \text{vec}\left(\mathbf{F}\right) \\ \lambda \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T}\left(\xi_t' \otimes \mathbf{I}_q\right)'\mathbf{Q}^{-1}\xi_t \\ b \end{bmatrix},$$

where $\mathbf{\Omega} = \sum_{t=1}^{T} \left(\xi_t' \otimes \mathbf{I}_q\right)' \mathbf{Q}^{-1} \left(\xi_t' \otimes \mathbf{I}_q\right)$. Assuming that the block matrix is invertible and that $\mathbf{B}$ has independent rows, its inversion can be expressed using the property:

$$\begin{bmatrix} \mathbf{\Omega} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0}_p \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1}\mathbf{B}' \left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \mathbf{B}\mathbf{\Omega}^{-1} & \mathbf{\Omega}^{-1}\mathbf{B}' \left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \\ \left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \mathbf{B}\mathbf{\Omega}^{-1} & -\left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \end{bmatrix},$$

and thus the solution to the system of equations above is given as:

$$\begin{bmatrix} \text{vec}\left(\mathbf{F}\right) \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1}\mathbf{B}' \left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \mathbf{B}\mathbf{\Omega}^{-1} & \mathbf{\Omega}^{-1}\mathbf{B}' \left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \\ \left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \mathbf{B}\mathbf{\Omega}^{-1} & -\left(\mathbf{B}\mathbf{\Omega}^{-1}\mathbf{B}'\right)^{-1} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^{T} \left(\xi_t' \otimes \mathbf{I}_q\right)' \mathbf{Q}^{-1} \xi_t \\ b \end{bmatrix}$$

It is clear that if no restrictions were in place, that is $\mathbf{B} = b = 0$, then the solution yields the unconstrained classical EM result in vectorised form:

$$\begin{aligned} \text{vec}\left(\mathbf{F}_{\text{unc}}\right) &= \mathbf{\Omega}^{-1} \sum_{t=1}^{T} \left(\xi_t' \otimes \mathbf{I}_q\right)' \mathbf{Q}^{-1} \xi_t \\ &= \left[\sum_{t=1}^{T} \left(\xi_t' \otimes \mathbf{I}_q\right)' \mathbf{Q}^{-1} \left(\xi_t' \otimes \mathbf{I}_q\right)\right]^{-1} \sum_{t=1}^{T} \left(\xi_t' \otimes \mathbf{I}_q\right)' \mathbf{Q}^{-1} \xi_t \end{aligned}$$

At this point, it becomes useful to recall the unconstrained $\mathcal{Q}$-maximising quantity for $\mathbf{F}$ and express this in vectorised form to see the relation with smoothed quantities. The goal is to express the constrained solution, $\text{vec}\left(\mathbf{F}_{\text{con}}\right)$, in terms of the unconstrained vectorised solution, $\text{vec}\left(\mathbf{F}_{\text{unc}}\right)$. Recall Equation (C.2):

$$\mathbf{F}_{\text{unc}} = \left(\sum_{t=1}^{T} \hat{\xi}_{t|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t,t-1|T}\right) \left(\sum_{t=1}^{T} \hat{\xi}_{t-1|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t-1|T}\right)^{-1},$$

which implies:

$$\begin{aligned} \text{vec}\left(\mathbf{F}_{\text{unc}}\right) &= \text{vec}\left[\left(\sum_{t=1}^{T} \hat{\xi}_{t|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t,t-1|T}\right) \left(\sum_{t=1}^{T} \hat{\xi}_{t-1|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t-1|T}\right)^{-1}\right] \\ &= \sum_{t=1}^{T} \text{vec}\left[\mathbf{I}_q \left[\hat{\xi}_{t|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t,t-1|T}\right] \left(\sum_{t=1}^{T} \hat{\xi}_{t-1|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t-1|T}\right)^{-1}\right] \\ &= \sum_{t=1}^{T} \left[\left(\sum_{t=1}^{T} \hat{\xi}_{t-1|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t-1|T}\right)^{-1} \otimes \mathbf{I}_q\right] \text{vec}\left(\hat{\xi}_{t|T} \hat{\xi}_{t-1|T}' + \mathbf{P}_{t,t-1|T}\right) \end{aligned}$$

$$= \sum_{t=1}^{T} \left[ \left( \sum_{t=1}^{T} \hat{\xi}_{t-1|T} \hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T} \right)^{-1} \otimes \mathbf{I}_q \right] \left[ \hat{\xi}_{t-1|T} \otimes \hat{\xi}_{t|T} + \text{vec}\left( \mathbf{P}_{t,t-1|T} \right) \right].$$

It is also possible to express $\mathbf{\Omega}$ in terms of smoothed quantities:

$$\begin{aligned} \underset{[q^2 \times q^2]}{\mathbf{\Omega}} &= \sum_{t=1}^{T} \left( \xi'_t \otimes \mathbf{I}_q \right)' \mathbf{Q}^{-1} \left( \xi'_t \otimes \mathbf{I}_q \right) = \sum_{t=1}^{T} \left( \xi_t \otimes \mathbf{I}_q \right) \mathbf{Q}^{-1} \left( \xi'_t \otimes \mathbf{I}_q \right) \\ &= \left( \sum_{t=1}^{T} \xi_t \xi'_t \right) \otimes \mathbf{Q}^{-1} = \left( \sum_{t=1}^{T} \hat{\xi}_{t|T} \hat{\xi}'_{t|T} + \mathbf{P}_{t|T} \right) \otimes \mathbf{Q}^{-1}. \end{aligned}$$

Using all of the above information, the constrained solution can thus be expressed as:

$$\begin{aligned} \text{vec}\left( \mathbf{F}_{\text{con}} \right) &= \left[ \mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1} \mathbf{B}' \left( \mathbf{B} \mathbf{\Omega}^{-1} \mathbf{B}' \right)^{-1} \mathbf{B} \mathbf{\Omega}^{-1} \right] \left( \sum_{t=1}^{T} \left( \xi'_t \otimes \mathbf{I}_q \right)' \mathbf{Q}^{-1} \xi_t \right) + \mathbf{\Omega}^{-1} \mathbf{B}' \left( \mathbf{B} \mathbf{\Omega}^{-1} \mathbf{B}' \right)^{-1} b \\ &= \left[ \mathbf{I}_{q^2} - \mathbf{\Omega}^{-1} \mathbf{B}' \left( \mathbf{B} \mathbf{\Omega}^{-1} \mathbf{B}' \right)^{-1} \mathbf{B} \right] \text{vec}\left( \mathbf{F}_{\text{unc}} \right) + \mathbf{\Omega}^{-1} \mathbf{B}' \left( \mathbf{B} \mathbf{\Omega}^{-1} \mathbf{B}' \right)^{-1} b \\ &= \text{vec}\left( \mathbf{F}_{\text{unc}} \right) + \left( \mathbf{\Omega}^{-1} \mathbf{B}' \left( \mathbf{B} \mathbf{\Omega}^{-1} \mathbf{B}' \right)^{-1} \mathbf{B} \right) \left( b - \mathbf{B} \text{vec}\left( \mathbf{F}_{\text{unc}} \right) \right), \end{aligned} \tag{D.1}$$

where $\mathbf{\Omega}^{-1} = \left( \sum_{t=1}^{T} \hat{\xi}_{t|T} \hat{\xi}'_{t|T} + \mathbf{P}_{t|T} \right)^{-1} \otimes \mathbf{Q}$.

# E Bayesian Analysis

## E.1 Derivation of BLAME Quantities

We wish to minimise the the quantity:

$$\begin{aligned} \phi_T(\theta_n) &= -\log p(\theta_n) - \mathcal{Q}\left( \theta_n, \theta_n^{(k)} \right) \\ &= -\log p(\theta_n) + \frac{2T-2}{2} \log(2\pi) - \frac{1}{2} \log \left| \mathbf{P}_0^{-1} \right| + \frac{1}{2} (\xi_0 - \mu_0)' \mathbf{P}_0^{-1} (\xi_0 - \mu_0) \\ &\quad - \frac{T-1}{2} \log \left| \mathbf{Q}^{-1} \right| + \frac{1}{2} \sum_{t=1}^{T} \left( \xi_t - \mathbf{F} \xi_{t-1} \right)' \mathbf{Q}^{-1} \left( \xi_t - \mathbf{F} \xi_{t-1} \right) \\ &\quad - \frac{T-2}{2} \log \left| \mathbf{R}^{-1} \right| + \frac{1}{2} \sum_{t=3}^{T} \left( x_t - \mathbf{H} \xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2} \right)' \mathbf{R}^{-1} \left( x_t - \mathbf{H} \xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2} \right) \end{aligned}$$

with respect to the vector of parameters $\theta_n = \{\mu_0, \mathbf{P}_0, \mathbf{F}, \mathbf{H}, \mathbf{K}_1, \mathbf{K}_2, \mathbf{Q}, \mathbf{R}\}$. The full energy function comprises of the expected log-likelihood, as well as the collection of prior parameter

distributions:

$$p(\theta_n) = p(\mu_0) \times p(\mathbf{P}_0) \times p(\mathbf{F}) \times p(\mathbf{H}) \times p(\mathbf{K}_1) \times p(\mathbf{K}_2) \times p(\mathbf{Q}) \times p(\mathbf{R}),$$

which yields the log-prior

$$\log p(\theta_n) = \log p(\mu_0) + \log p(\mathbf{P}_0) + \log p(\mathbf{F}) + \log p(\mathbf{H}) + \log p(\mathbf{K}_1) + \log p(\mathbf{K}_2) + \log p(\mathbf{Q}) + \log p(\mathbf{R}).$$

The full energy function, therefore, reads:

$$\begin{aligned}
\phi_T(\theta_n) =& \frac{2T-2}{2}\log(2\pi) - \frac{1}{2}\log\left|\mathbf{P}_0^{-1}\right| + \frac{1}{2}\left(\xi_0 - \mu_0\right)'\mathbf{P}_0^{-1}\left(\xi_0 - \mu_0\right) \\
&- \frac{T-1}{2}\log\left|\mathbf{Q}^{-1}\right| + \frac{1}{2}\sum_{t=1}^{T}\left(\xi_t - \mathbf{F}\xi_{t-1}\right)'\mathbf{Q}^{-1}\left(\xi_t - \mathbf{F}\xi_{t-1}\right) \\
&- \frac{T-2}{2}\log\left|\mathbf{R}^{-1}\right| + \frac{1}{2}\sum_{t=3}^{T}\left(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\right)'\mathbf{R}^{-1}\left(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\right) \\
&- \log p(\mu_0) - \log p(\mathbf{P}_0) - \log p(\mathbf{F}) - \log p(\mathbf{H}) - \log p(\mathbf{K}_1) - \log p(\mathbf{K}_2) - \log p(\mathbf{Q}) - \log p(\mathbf{R}).
\end{aligned}$$

The resultant energy minimising quantities are derived below. For ease of computation, I initially write the first order condition, solve for the parameter of interest, and then take the expectation conditional on the full information set $\mathcal{I}_T$.

### E.1.1 BLAME the System

I first apply the BLAME analysis to the state autoregressive system matrix. Let $\theta_n = \mathbf{F}$, where $\mathbf{F}$ is a $q \times q$ matrix with a matricvariate normal distribution, then

$$\begin{aligned}
p(\mathbf{F}) &\sim \mathcal{MN}_{q\times q}\left(\mathbf{M}, \mathbf{U}, \mathbf{V}\right), \\
p(\mathbf{F}) &\propto e^{-\frac{1}{2}\mathrm{tr}\left[\mathbf{V}^{-1}(\mathbf{F}-\mathbf{M})'\mathbf{U}^{-1}(\mathbf{F}-\mathbf{M})\right]}, \\
\implies \log p(\mathbf{F}) &\propto -\frac{1}{2}\mathrm{tr}\left[\mathbf{V}^{-1}(\mathbf{F}-\mathbf{M})'\mathbf{U}^{-1}(\mathbf{F}-\mathbf{M})\right].
\end{aligned}$$

The first order condition is therefore written as

$$\frac{\partial \phi_T(\theta_n)}{\partial \mathbf{F}} = \frac{1}{2}\frac{\partial}{\partial \mathbf{F}}\left[\sum_{t=1}^{T}\left(\xi_t - \mathbf{F}\xi_{t-1}\right)'\mathbf{Q}^{-1}\left(\xi_t - \mathbf{F}\xi_{t-1}\right)\right] + \frac{1}{2}\frac{\partial}{\partial \mathbf{F}}\mathrm{tr}\left[\mathbf{V}^{-1}(\mathbf{F}-\mathbf{M})'\mathbf{U}^{-1}(\mathbf{F}-\mathbf{M})\right]$$

$$=\frac{1}{2}\frac{\partial}{\partial \mathbf{F}}\text{tr}\left[\mathbf{Q}^{-1}\sum_{t=1}^{T}\Big(\xi_t-\mathbf{F}\xi_{t-1}\Big)\Big(\xi_t-\mathbf{F}\xi_{t-1}\Big)'\right]+\frac{1}{2}\frac{\partial}{\partial \mathbf{F}}\text{tr}\left[\mathbf{V}^{-1}(\mathbf{F}-\mathbf{M})'\mathbf{U}^{-1}(\mathbf{F}-\mathbf{M})\right],$$

which I break up into two parts. The expression for the derivative of the prior reduces to

$$\frac{1}{2}\frac{\partial}{\partial \mathbf{F}}\text{tr}\left[\mathbf{V}^{-1}(\mathbf{F}-\mathbf{M})'\mathbf{U}^{-1}(\mathbf{F}-\mathbf{M})\right]=\mathbf{U}^{-1}\mathbf{F}\mathbf{V}^{-1}-\mathbf{U}^{-1}\mathbf{M}\mathbf{V}^{-1}=\mathbf{U}^{-1}\Big(\mathbf{F}-\mathbf{M}\Big)\mathbf{V}^{-1}.$$

Similarly for the expected log-likelihood term, we have:

$$\frac{1}{2}\frac{\partial}{\partial \mathbf{F}}\text{tr}\left[\mathbf{Q}^{-1}\sum_{t=1}^{T}\Big(\xi_t-\mathbf{F}\xi_{t-1}\Big)\Big(\xi_t-\mathbf{F}\xi_{t-1}\Big)'\right]=\mathbf{Q}^{-1}\sum_{t=1}^{T}\Big(-\xi_t\xi_{t-1}'+\mathbf{F}\xi_{t-1}\xi_{t-1}'\Big).$$

The first order condition therefore reads:

$$\frac{\partial \phi_T(\theta_n)}{\partial \mathbf{F}}=\mathbf{Q}^{-1}\sum_{t=1}^{T}\Big(-\xi_t\xi_{t-1}'+\mathbf{F}\xi_{t-1}\xi_{t-1}'\Big)+\mathbf{U}^{-1}\Big(\mathbf{F}-\mathbf{M}\Big)\mathbf{V}^{-1}.$$

This equation is now solved for $\frac{\partial \phi_T(\theta_n)}{\partial \mathbf{F}}=0$. Pre-multiply by $\mathbf{Q}$ to yield

$$-\sum_{t=1}^{T}\xi_t\xi_{t-1}'+\mathbf{F}\sum_{t=1}^{T}\xi_{t-1}\xi_{t-1}'+\mathbf{Q}\mathbf{U}^{-1}\mathbf{F}\mathbf{V}^{-1}-\mathbf{Q}\mathbf{U}^{-1}\mathbf{M}\mathbf{V}^{-1}=0,$$

then post-multiply both sides by $\mathbf{V}$ and rearrange:

$$\mathbf{F}\underbrace{\left(\sum_{t=1}^{T}\xi_{t-1}\xi_{t-1}'\right)\mathbf{V}}_{\mathbf{\Psi}}+\overbrace{\mathbf{Q}\mathbf{U}^{-1}}^{\mathbf{\Xi}}\mathbf{F}=\underbrace{\left(\sum_{t=1}^{T}\xi_t\xi_{t-1}'\right)\mathbf{V}+\mathbf{Q}\mathbf{U}^{-1}\mathbf{M}}_{\mathbf{\Lambda}}$$

$$\mathbf{F}\mathbf{\Psi}+\mathbf{\Xi}\mathbf{F}=\mathbf{\Lambda}.$$

The final expression is solved for $\mathbf{F}$ using the Lyapunov equation, which, for a general $q\times q$ system, yields

$$\begin{aligned}\text{vec}(\hat{\mathbf{F}})&=\left(\mathbf{I}_q\otimes\mathbf{\Xi}+\mathbf{\Psi}'\otimes\mathbf{I}_q\right)^{-1}\text{vec}(\mathbf{\Lambda})\\&=\left[\mathbf{I}_q\otimes\Big(\mathbf{Q}\mathbf{U}^{-1}\Big)+\left[\mathbf{V}\left(\sum_{t=1}^{T}\xi_{t-1}\xi_{t-1}'\right)'\right]\otimes\mathbf{I}_q\right]^{-1}\text{vec}\left[\left(\sum_{t=1}^{T}\xi_t\xi_{t-1}'\right)\mathbf{V}+\mathbf{Q}\mathbf{U}^{-1}\mathbf{M}\right],\end{aligned}$$

where $\mathbf{I}_q$ is the $q \times q$ identity matrix. Finally, I apply the conditional expectation on all the variables to get the minimising quantity:

$$\underset{\theta_n=\mathbf{F}}{\operatorname{argmin}}\, \phi_T(\theta_n) = \operatorname{vec}(\mathbf{F}_{\text{blame}}) = \mathbb{E}\left[\operatorname{vec}(\hat{\mathbf{F}})|\mathcal{I}_T\right] = \mathbf{Y}^{-1}\operatorname{vec}(\mathbf{\Gamma}), \tag{E.1}$$

where

$$\mathbf{Y} = \mathbf{I}_q \otimes \left(\mathbf{Q}\mathbf{U}^{-1}\right) + \left[\mathbf{V}\left(\sum_{t=1}^{T} \hat{\xi}_{t-1|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T}\right)'\right] \otimes \mathbf{I}_q,$$

and

$$\mathbf{\Gamma} = \left(\sum_{t=1}^{T} \hat{\xi}_{t|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t,t-1|T}\right)\mathbf{V} + \mathbf{Q}\mathbf{U}^{-1}\mathbf{M}.$$

### E.1.2 BLAME the Covariance

I now apply the BLAME analysis to the state covariance matrix. The derivation for $\mathbf{R}$ is similar, and can be found in Appendix E. Let $\theta_n = \mathbf{Q}$, where $\mathbf{Q}$ is a $q \times q$ matrix with an inverse Wishart distribution, then

$$\begin{aligned} p(\mathbf{Q}) &\sim \mathcal{W}^{-1}(\mathbf{T}, \nu), \\ p(\mathbf{Q}) &\propto |\mathbf{Q}|^{-\frac{\nu+q+1}{2}} e^{-\frac{1}{2}\operatorname{tr}(\mathbf{T}\mathbf{Q}^{-1})}, \\ \implies \log p(\mathbf{Q}) &\propto \frac{\nu+q+1}{2}\log\left|\mathbf{Q}^{-1}\right| - \frac{1}{2}\operatorname{tr}(\mathbf{T}\mathbf{Q}^{-1}), \end{aligned}$$

where I have used that $\left|\mathbf{Q}^{-1}\right| = 1/|\mathbf{Q}|$. In this situation, it is more convenient to minimise the energy function with respect to $\mathbf{Q}^{-1}$, rather than the regular $\mathbf{Q}$. The first order condition reads

$$\frac{\partial \phi_T(\theta_n)}{\partial \mathbf{Q}^{-1}} = -\frac{T-1}{2}\mathbf{Q} + \frac{1}{2}\sum_{t=1}^{T}\left(\xi_t - \mathbf{F}\xi_{t-1}\right)\left(\xi_t - \mathbf{F}\xi_{t-1}\right)' - \frac{\nu+q+1}{2}\mathbf{Q} + \frac{1}{2}\mathbf{T}',$$

where I have used that:

$$\frac{\partial}{\partial \mathbf{A}}\operatorname{tr}(\mathbf{B}\mathbf{A}) = \mathbf{B}'.$$

Setting this quantity to zero and solving for $\mathbf{Q}$ yields

$$\hat{\mathbf{Q}} = \frac{1}{T+\nu+q}\sum_{t=1}^{T}\left(\xi_t - \mathbf{F}\xi_{t-1}\right)\left(\xi_t - \mathbf{F}\xi_{t-1}\right)' + \frac{1}{T+\nu+q}\mathbf{T}'.$$

We can see how the prior hyperparameter scale matrix $\mathbf{T}$ has the ability to fix the usual problem faced by classical EM by shifting the covariance matrix to becoming full rank whenever necessary. Taking the conditional expectation of this quantity gives

$$\begin{aligned}\underset{\theta_n=\mathbf{Q}}{\operatorname{argmin}}\, \phi_T(\theta_n) =& \mathbb{E}\left[\hat{\mathbf{Q}}|\mathcal{I}_T\right] \\ =& \frac{1}{T+\nu+q}\sum_{t=1}^{T}\Big[\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T} - \Big(\hat{\xi}_{t|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t,t-1|T}\Big)\mathbf{F}' - \mathbf{F}\Big(\hat{\xi}_{t-1|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t-1,t|T}\Big) \\ &+ \mathbf{F}\Big(\hat{\xi}_{t-1|T}\hat{\xi}'_{t-1|T} + \mathbf{P}_{t-1|T}\Big)\mathbf{F}'\Big] + \frac{1}{T+\nu+q}\mathbf{T}'. \end{aligned} \tag{E.2}$$

### E.1.3 BLAME the Rest

Following the same steps as above, the remaining matrices are analysed in turn.

- If $\theta_n = \mu_0$, then $p(\mu_0) \propto 1$. Therefore,

$$\begin{aligned}\frac{\partial \phi_T(\theta_n)}{\partial \mu_0} &= -\mathbf{P}_0^{-1}\xi_0 + \mathbf{P}_0^{-1}\mu_0 = 0 \\ &\implies \hat{\mu}_0 = \xi_0.\end{aligned}$$

  Taking the conditional expectation gives

$$\underset{\theta_n=\mu_0}{\operatorname{argmin}}\, \phi_T(\theta_n) = \mathbb{E}\left[\hat{\mu}_0|\mathcal{I}_T\right] = \hat{\xi}_{0|T}, \tag{E.3}$$

  which is readily given by the RTS smoother.

- If $\theta_n = \mathbf{P}_0$, then $p(\mathbf{P}_0) \propto 1$. Therefore,

$$\begin{aligned}\frac{\partial \phi_T(\theta_n)}{\partial \mathbf{P}_0} &= -\frac{1}{2}\mathbf{P}_0 + \frac{1}{2}(\xi_0-\mu_0)(\xi_0-\mu_0)' = 0 \\ &\implies \hat{\mathbf{P}}_0 = (\xi_0-\mu_0)(\xi_0-\mu_0)'.\end{aligned}$$

  Taking the conditional expectation gives

$$\underset{\theta_n=\mathbf{P}_0}{\operatorname{argmin}}\, \phi_T(\theta_n) = \mathbb{E}\left[\hat{\mathbf{P}}_0|\mathcal{I}_T\right] = \mathbf{P}_{0|T}, \tag{E.4}$$

  which is also readily given by the RTS smoother.

- For $\theta_n = \mathbf{H}$ with $r \times q$ dimensions and a matricvariate normal prior, we follow the same procedure as before, and we yield:

$$\underset{\theta_n=\mathbf{H}}{\operatorname{argmin}}\, \phi_T(\theta_n) = \mathbb{E}\left[\hat{\mathbf{H}}|\mathcal{I}_T\right] = \operatorname{vec}_{r\times q}^{-1}\left(\mathbf{Y_H}^{-1}\operatorname{vec}_{rq}(\mathbf{\Gamma}_{\boldsymbol{H}})\right), \tag{E.5}$$

where

$$\mathbf{Y_H} = \mathbf{I}_q \otimes \left(\mathbf{R}\mathbf{U}^{-1}\right) + \left(\mathbf{V}\left(\sum_{t=3}^{T} \hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T}\right)'\right) \otimes \mathbf{I}_r,$$

and

$$\mathbf{\Gamma}_{\boldsymbol{H}} = \left(\sum_{t=3}^{T}\left(x_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\right)\hat{\xi}'_{t|T}\right)\mathbf{V} + \mathbf{R}\mathbf{U}^{-1}\mathbf{M}.$$

- Similarly for a matricvariate normal $\mathbf{K}_1$ with $r \times g$ dimensions:

$$\underset{\theta_n=\mathbf{K_1}}{\operatorname{argmin}}\, \phi_T(\theta_n) = \mathbb{E}\left[\hat{\mathbf{K}}_1|\mathcal{I}_T\right] = \operatorname{vec}_{r\times g}^{-1}\left(\mathbf{Y_{K_1}}^{-1}\operatorname{vec}_{rg}(\mathbf{\Gamma}_{\mathbf{K_1}})\right), \tag{E.6}$$

where

$$\mathbf{Y_{K_1}} = \mathbf{I}_g \otimes \left(\mathbf{R}\mathbf{U}^{-1}\right) + \left(\mathbf{V}\left(\sum_{t=3}^{T} z_{t-1}z'_{t-1}\right)'\right) \otimes \mathbf{I}_r,$$

and

$$\mathbf{\Gamma}_{\boldsymbol{K_1}} = \left(\sum_{t=3}^{T}\left(x_t - \mathbf{H}\hat{\xi}_{t|T} - \mathbf{K}_2 z_{t-2}\right)z'_{t-1}\right)\mathbf{V} + \mathbf{R}\mathbf{U}^{-1}\mathbf{M}.$$

- Similarly for a matricvariate normal $\mathbf{K}_2$ with $r \times g$ dimensions:

$$\underset{\theta_n=\mathbf{K_2}}{\operatorname{argmin}}\, \phi_T(\theta_n) = \mathbb{E}\left[\hat{\mathbf{K}}_2|\mathcal{I}_T\right] = \operatorname{vec}_{r\times g}^{-1}\left(\mathbf{Y_{K_2}}^{-1}\operatorname{vec}_{rg}(\mathbf{\Gamma}_{\mathbf{K_2}})\right), \tag{E.7}$$

where

$$\mathbf{Y_{K_2}} = \mathbf{I}_g \otimes \left(\mathbf{R}\mathbf{U}^{-1}\right) + \left(\mathbf{V}\left(\sum_{t=3}^{T} z_{t-2}z'_{t-2}\right)'\right) \otimes \mathbf{I}_r,$$

and

$$\mathbf{\Gamma}_{\boldsymbol{K_2}} = \left(\sum_{t=3}^{T}\left(x_t - \mathbf{H}\hat{\xi}_{t|T} - \mathbf{K}_1 z_{t-1}\right)z'_{t-2}\right)\mathbf{V} + \mathbf{R}\mathbf{U}^{-1}\mathbf{M}.$$

- If $\theta_n = \mathbf{R}$, where $\mathbf{R}$ is a $r \times r$ matrix with an inverse Wishart distribution, then

$$p(\mathbf{R}) \sim \mathcal{W}^{-1}(\mathbf{S}, \rho),$$

$$p(\mathbf{R}) \propto |\mathbf{R}|^{-\frac{\rho+r+1}{2}} e^{-\frac{1}{2}\text{tr}(\mathbf{S}\mathbf{R}^{-1})},$$
$$\implies \log p(\mathbf{R}) \propto \frac{\rho+r+1}{2} \log \left|\mathbf{R}^{-1}\right| - \frac{1}{2}\text{tr}(\mathbf{S}\mathbf{R}^{-1}),$$

Following the same steps as for $\mathbf{Q}$, the first order conditions reads

$$\begin{aligned}\frac{\partial \phi_T(\theta_n)}{\partial \mathbf{R}^{-1}} =& -\frac{T-2}{2}\mathbf{R} + \frac{1}{2}\sum_{t=3}^{T}\Big(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)\Big(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)' \\ & - \frac{\rho+r+1}{2}\mathbf{R} + \frac{1}{2}\mathbf{S}',\end{aligned}$$

which yields

$$\hat{\mathbf{R}} = \frac{1}{T+\rho+r-1}\left[\sum_{t=3}^{T}\Big(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)\Big(x_t - \mathbf{H}\xi_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)' + \mathbf{S}'\right].$$

In a similar fashion to $\mathbf{Q}$, the minimising quantity is given as:

$$\begin{aligned}\underset{\theta_n=\mathbf{R}}{\text{argmin}}\, \phi_T(\theta_n) = \mathbb{E}\left[\hat{\mathbf{R}}|\mathcal{I}_T\right] =& \frac{1}{T+\rho+r-1}\sum_{t=3}^{T}\Bigg[\Big(x_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)\Big(x_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)' \\ & - \mathbf{H}\hat{\xi}_{t|T}\Big(x_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)' - \Big(x_t - \mathbf{K}_1 z_{t-1} - \mathbf{K}_2 z_{t-2}\Big)\hat{\xi}'_{t|T}\mathbf{H}' \\ & + \mathbf{H}\Big[\hat{\xi}_{t|T}\hat{\xi}'_{t|T} + \mathbf{P}_{t|T}\Big]\mathbf{H}'\Bigg] \\ & + \frac{1}{T+\rho+r-1}\mathbf{S}'. \end{aligned} \tag{E.8}$$

## E.2 Derivation of Con-BLAME Quantities

It will be assumed again that $\mathbf{F}$ follows a matricvariate normal, which implies that $\text{vec}(\mathbf{F})$ follows a multivariate normal:

$$p(\mathbf{F}) \sim \mathcal{MN}_{q\times q}(\mathbf{M}, \mathbf{U}, \mathbf{V}) \iff p(\text{vec}(\mathbf{F})) \sim \mathcal{N}_{q^2}(\text{vec}(\mathbf{M}), \mathbf{V}\otimes\mathbf{U}).$$

With this prior information, the constrained energy function (considering only variables relevant to $\mathbf{F}$) reads:

$$\begin{aligned}\phi_{T,\text{con}}|_{\theta=\text{vec}(\mathbf{F})} \propto & -\mathcal{Q}_{\text{con}}|_{\theta=\text{vec}(\mathbf{F})} + \frac{1}{2}\left[\text{vec}\left(\mathbf{F}\right) - \text{vec}\left(\mathbf{M}\right)\right]' \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}\left[\text{vec}\left(\mathbf{F}\right) - \text{vec}\left(\mathbf{M}\right)\right] \\ \propto & \frac{1}{2}\sum_{t=1}^{T}\left[\xi_t - \left(\xi_t'\otimes\mathbf{I}_q\right)\text{vec}\left(\mathbf{F}\right)\right]' \mathbf{Q}^{-1}\left[\xi_t - \left(\xi_t'\otimes\mathbf{I}_q\right)\text{vec}\left(\mathbf{F}\right)\right] + \lambda'\left(\mathbf{B}\text{vec}\left(\mathbf{F}\right) - b\right) \\ & + \frac{1}{2}\left[\text{vec}\left(\mathbf{F}\right) - \text{vec}\left(\mathbf{M}\right)\right]' \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}\left[\text{vec}\left(\mathbf{F}\right) - \text{vec}\left(\mathbf{M}\right)\right]\end{aligned}$$

The first order condition thus yields

$$\begin{aligned}\frac{\partial\phi_{T,\text{con}}}{\partial\text{vec}\left(\mathbf{F}\right)} = & -\frac{\partial\mathcal{Q}_{\text{con}}}{\partial\text{vec}\left(\mathbf{F}\right)} + \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}\left[\text{vec}\left(\mathbf{F}\right) - \text{vec}\left(\mathbf{M}\right)\right] \\ = & \begin{bmatrix}\boldsymbol{\Omega} + \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0}_p\end{bmatrix}\begin{bmatrix}\text{vec}\left(\mathbf{F}\right) \\ \lambda\end{bmatrix} - \begin{bmatrix}\left(\sum_{t=1}^{T}\left(\xi_t'\otimes\mathbf{I}_q\right)'\mathbf{Q}^{-1}\xi_t\right) + \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}\text{vec}\left(\mathbf{M}\right) \\ b\end{bmatrix} = 0,\end{aligned}$$

from which we can easily infer the constrained EM solution with added information from the prior distribution. Denote $\boldsymbol{\Phi} = \boldsymbol{\Omega} + \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}$, then $\boldsymbol{\Phi}^{-1} = \left(\boldsymbol{\Omega} + \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}\right)^{-1}$. It is now possible to express the solution to the above system of equations in a similar fashion to the constrained EM solution:

$$\begin{aligned}\begin{bmatrix}\text{vec}\left(\mathbf{F}_{\text{con-blame}}\right) \\ \lambda\end{bmatrix} = & \begin{bmatrix}\boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\boldsymbol{\Phi}^{-1} & \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1} \\ \left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\boldsymbol{\Phi}^{-1} & -\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\end{bmatrix} \\ & \times \begin{bmatrix}\left(\sum_{t=1}^{T}\left(\xi_t'\otimes\mathbf{I}_q\right)'\mathbf{Q}^{-1}\xi_t\right) + \left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}\text{vec}\left(\mathbf{M}\right) \\ b\end{bmatrix}\end{aligned}$$

It is also possible to express the vectorised solution of constrained vec $\left(\mathbf{F}_{\text{con-blame}}\right)$ in terms of the unconstrained classical version vec $\left(\mathbf{F}_{\text{unc}}\right)$, as had been done for the constrained EM algorithm. Since $\left(\mathbf{V}\otimes\mathbf{U}\right)^{-1}$ is a $q^2\times q^2$ symmetric matrix, single value decomposition (SVD) gives:

$$\left(\mathbf{V}\otimes\mathbf{U}\right)^{-1} = \mathbf{J}\mathbf{D}\mathbf{J}'$$

where $\mathbf{J}$ contains all eigenvectors of $(\mathbf{V} \otimes \mathbf{U})^{-1}$ as its columns and $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $(\mathbf{V} \otimes \mathbf{U})^{-1}$ along the diagonal. Therefore, I can write

$$\begin{aligned}\boldsymbol{\Phi}^{-1} &= \left(\boldsymbol{\Omega} + \mathbf{J}\mathbf{D}\mathbf{J}'\right)^{-1} \\ &= \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{J}\left(\mathbf{D}^{-1} + \mathbf{J}'\boldsymbol{\Omega}^{-1}\mathbf{J}\right)^{-1}\mathbf{J}'\boldsymbol{\Omega}^{-1},\end{aligned}$$

where I invoke the Woodbury identity to expand this inverse. By recalling the form of vec $(\mathbf{F}_{\text{unc}})$:

$$\begin{aligned}\text{vec}\left(\mathbf{F}_{\text{con-blame}}\right) &= \left[\boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\boldsymbol{\Phi}^{-1}\right]\left[\left(\sum_{t=1}^{T}\left(\xi_t' \otimes \mathbf{I}_q\right)'\mathbf{Q}^{-1}\xi_t\right) + \left(\mathbf{V} \otimes \mathbf{U}\right)^{-1}\text{vec}\left(\mathbf{M}\right)\right] \\ &\quad + \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}b \\ &= \left[\mathbf{I}_{q^2} - \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\right]\left[\boldsymbol{\Phi}^{-1}\left(\sum_{t=1}^{T}\left(\xi_t' \otimes \mathbf{I}_q\right)'\mathbf{Q}^{-1}\xi_t\right) + \boldsymbol{\Phi}^{-1}\left(\mathbf{V} \otimes \mathbf{U}\right)^{-1}\text{vec}\left(\mathbf{M}\right)\right] \\ &\quad + \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}b,\end{aligned}$$

where I pull one $\boldsymbol{\Phi}^{-1}$ into the second square bracket. I then expand the $\boldsymbol{\Phi}^{-1}$ in the first term within the second square bracket to yield an expression in terms of vec $(\mathbf{F}_{\text{unc}})$:

$$\begin{aligned}\text{vec}\left(\mathbf{F}_{\text{con-blame}}\right) &= \left[\mathbf{I}_{q^2} - \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\right] \\ &\quad \times \left[\left[\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{J}\left(\mathbf{D}^{-1} + \mathbf{J}'\boldsymbol{\Omega}^{-1}\mathbf{J}\right)^{-1}\mathbf{J}'\boldsymbol{\Omega}^{-1}\right]\sum_{t=1}^{T}\left(\xi_t' \otimes \mathbf{I}_q\right)'\mathbf{Q}^{-1}\xi_t + \boldsymbol{\Phi}^{-1}\left(\mathbf{V} \otimes \mathbf{U}\right)^{-1}\text{vec}\left(\mathbf{M}\right)\right] \\ &\quad + \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}b \\ &= \left[\mathbf{I}_{q^2} - \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}\mathbf{B}\right] \\ &\quad \times \left[\left[\mathbf{I}_{q^2} - \boldsymbol{\Omega}^{-1}\mathbf{J}\left(\mathbf{D}^{-1} + \mathbf{J}'\boldsymbol{\Omega}^{-1}\mathbf{J}\right)^{-1}\mathbf{J}'\right]\text{vec}\left(\mathbf{F}_{\text{unc}}\right) + \boldsymbol{\Phi}^{-1}\left(\mathbf{V} \otimes \mathbf{U}\right)^{-1}\text{vec}\left(\mathbf{M}\right)\right] \\ &\quad + \boldsymbol{\Phi}^{-1}\mathbf{B}'\left(\mathbf{B}\boldsymbol{\Phi}^{-1}\mathbf{B}'\right)^{-1}b. \qquad \text{(E.9)}\end{aligned}$$

# F Standard Normal Random Variate Generators

## F.1 Polar Box-Muller Method

The polar Box-Muller method returns a pair of independent normal random variates based on two independent uniform variates and is given in Algorithm 1 below.

**Algorithm 1** Polar Box-Muller

1: Generate $u_1 \sim U(0,1)$ and $u_2 \sim U(0,1)$;
2: Set $v_1 = 2u_1 - 1$ and $v_2 = 2u_2 - 1$;
3: Compute $s = v_1^2 + v_2^2$;
4: **if** ($s = 0$ **or** $s \geq 1$) **then**
5: Discard $u_1$ and $u_2$, go to Step 1;
6: **else**
7: Return $z_1 \leftarrow u_1 \sqrt{\frac{-2\log s}{s}}$;
8: Return $z_2 \leftarrow u_2 \sqrt{\frac{-2\log s}{s}}$;
9: **end if**

## F.2 Inversion Method

The inversion method by Rao et al. (2011) utilises the "best fit" logistic regression on the cumulative distribution function of a standard normal random variable $z \sim \mathcal{N}(0,1)$, which they give as:

$$\phi(z) = \frac{1}{1 + e^{-1.702z}}.$$

They then invert this formula and yield the method given in Algorithm 2 below.

**Algorithm 2** Inversion Method

1: Generate $u \sim U(0,1)$;
2: Return $z \leftarrow \frac{-\log\left(\frac{1}{u} - 1\right)}{1.702}$

# G Simulation Study Architecture

Here I outline the setup of the simulation study in terms of the constraints imposed and the initial (hyper)parameters used for the estimation process. One simulated dataset of 10,000 data points is produced for analysis.

## G.1 Initialisation

Below are tables showing the initialisation of all the parameters for the production of the results in the simulation study. For the classical algorithms, the initial point coincides with the location hyperprior of the Bayesian algorithms. For the covariance matrices in the classical algorithms, the initialisation is $\mathbf{\Psi}^{-1}$. The state is given a diffuse initialisation for all algorithms.

| System Matrix | **M** | **U** | **V** |
|---|---|---|---|
| **F** | $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ |
| **H** | $\begin{bmatrix} 0.1 & -0.1 \\ 0.8 & 0 \\ 0 & -0.01 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$ |
| $\mathbf{K}_1$ | $\begin{bmatrix} 0 & 0 \\ 0.1 & -0.1 \\ 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ |
| $\mathbf{K}_2$ | $\begin{bmatrix} 0 & 0 \\ -0.1 & 0 \\ 0 & 0.05 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ |

Table 5: Prior hyperparameters for all the system matrices used for both the BLAME and the Con-BLAME estimation methods in the simulation study.

| Covariance Matrix | $\mathbf{\Psi}$ | $\nu$ |
|---|---|---|
| **Q** | $\begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$ | 4 |
| **R** | $\begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{bmatrix}$ | 4 |

Table 6: Prior hyperparameters for all the covariance matrices used for both the BLAME and the Con-BLAME estimation methods in the simulation study.

### G.2 Constraints

The constraints imposed in both the constrained EM and constrained BLAME algorithms are a direct imposition of the RW macrofinance model. Thus, the restricted system matrices that are enforced through the linear constraints have the form:

$$\mathbf{F} = \begin{bmatrix} \rho_L & 0 \\ \rho_S & (\rho_S - 1)g_\pi \end{bmatrix}$$

for the state autoregressive matrix, and

$$\mathbf{H} = \begin{bmatrix} \delta_L & \delta_S \\ \mu_\pi & 0 \\ 0 & -\beta_r \end{bmatrix}, \mathbf{K}_1 = \begin{bmatrix} 0 & 0 \\ (1-\mu_\pi)\alpha_1 & \alpha_y \\ 0 & (1-\mu_y)\beta_1 \end{bmatrix}, \mathbf{K}_2 = \begin{bmatrix} 0 & 0 \\ (1-\mu_\pi)\alpha_2 & 0 \\ 0 & (1-\mu_y)\beta_2 \end{bmatrix}$$

for the observed system matrices.

## H Prior Parameter Generation

Implementing any form of the BLAME algorithm whether on simulated or real data requires the generation of prior estimates of the system and covariance matrices given sets of hyperparameters. Here, I digress and outline the methods used to generate matricvariate normal and inverse Wishart random variates for the system and covariance matrices, respectively.

### H.1 Generating Matricvariate Normal Matrices

As an example of the method, I use the system autoregressive matrix $\mathbf{F}$. Let $p(\mathbf{F}) \sim \mathcal{MN}_{q\times q}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ for given hyperparameters $\mathbf{M}, \mathbf{U}$ and $\mathbf{V}$. Then, $p(\text{vec}(\mathbf{F})) \sim \mathcal{N}_{q^2}(\text{vec}(\mathbf{M}), \mathbf{V} \otimes \mathbf{U})$. The method that generates the matricvariate normal $\mathbf{F}$ is given in Algorithm 3 below.

**Algorithm 3** Generate Matricvariate Normal $\mathbf{F}$

1: Vectorise $\mathbf{M}$;
2: Perform Choleski decomposition on $\mathbf{V} \otimes \mathbf{U} = \mathbf{C}\mathbf{C}'$;
3: Generate $q^2$ independent standard normal random variables via Box-Muller or Inversion;
4: Collect them into a $q^2 \times 1$ vector $a$;
5: Set $f = \text{vec}(\mathbf{M}) + \mathbf{C}a$;
6: Return reshaped $\mathbf{F} \leftarrow \text{vec}^{-1}(f)$

### H.2 Generating Inverse Wishart Matrices

To exemplify the method on generating inverse Wishart random matrices, I use the state covariance matrix. Let $p(\mathbf{Q}) \sim \mathcal{W}^{-1}(\mathbf{T}, \tau)$. In order to generate an initial estimate of $\mathbf{Q}$, we begin by generating a $d-$dimensional correlated multivariate normal random vector $z \sim \mathcal{N}_d(0, \mathbf{\Sigma})$. Then, for $\tau$ degrees of freedom, we have the property that

$$\mathbf{W} = \sum_{i=1}^{\tau} zz' = \tau zz' \sim \mathcal{W}(\mathbf{\Sigma}, \tau).$$

Therefore, $\mathbf{W}$ has an inverse Wishart distribution with scale matrix $\mathbf{\Sigma}^{-1}$ and $\tau$ degrees of freedom. Given hyperparameters $\mathbf{T}$ and $\nu$ degrees of freedom, the algorithm for generating inverse Wishart matrices is given in Algorithm 4 below.

**Algorithm 4** Generate Inverse Wishart $\mathbf{Q}$

1: Invert $\mathbf{T}$, $\mathbf{\Sigma} = \mathbf{T}^{-1}$;
2: Perform Choleski decomposition on $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}'$;
3: Generate $q^2$ independent standard normal random variables via Box-Muller or Inversion;
4: Collect them into a $q^2 \times 1$ vector $a$;
5: Set $z = \mathbf{C}a$;
6: Calculate $\mathbf{W} = \nu zz'$;
7: Return $\mathbf{Q} \leftarrow \mathbf{W}$

# I Prior Hyperparameters for US Data Study

This section outlines the exact hyperparameters used in the estimation procedure of the Bayesian algorithms for the real data study. For the system matrices, the location parameters at subsequent estimation windows after the first are based on the converged matrices of the previous window. Thus, given below in Table 7 are only the location parameters of the first initial estimation window. In Table 8 the hyperparameters for the covariance matrices are given, which remain the same at each estimation window.

| System Matrix | **M** | **U** | **V** |
|---|---|---|---|
| **F** | $\begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ |
| **H** | $\begin{bmatrix} 0.1 & 0.1 \\ 0.01 & 0 \\ 0 & -0.01 \\ 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$ |
| $\mathbf{K}_1$ | $\begin{bmatrix} 0 & 0 & 0 \\ 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ |
| $\mathbf{K}_2$ | $\begin{bmatrix} 0 & 0 & 0 \\ 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$ | $\begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ |

Table 7: Prior hyperparameters for all the system matrices used for both the BLAME and the Con-BLAME estimation methods on the first initial estimation window.

| Covariance Matrix | $\mathbf{\Psi}$ | $\nu$ |
|---|---|---|
| **Q** | $\begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$ | 5 |
| **R** | $\begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}$ | 5 |

Table 8: Prior hyperparameters for all the covariance matrices used for both the BLAME and the Con-BLAME estimation methods at each estimation window.

# J Supplementary Figures and Tables

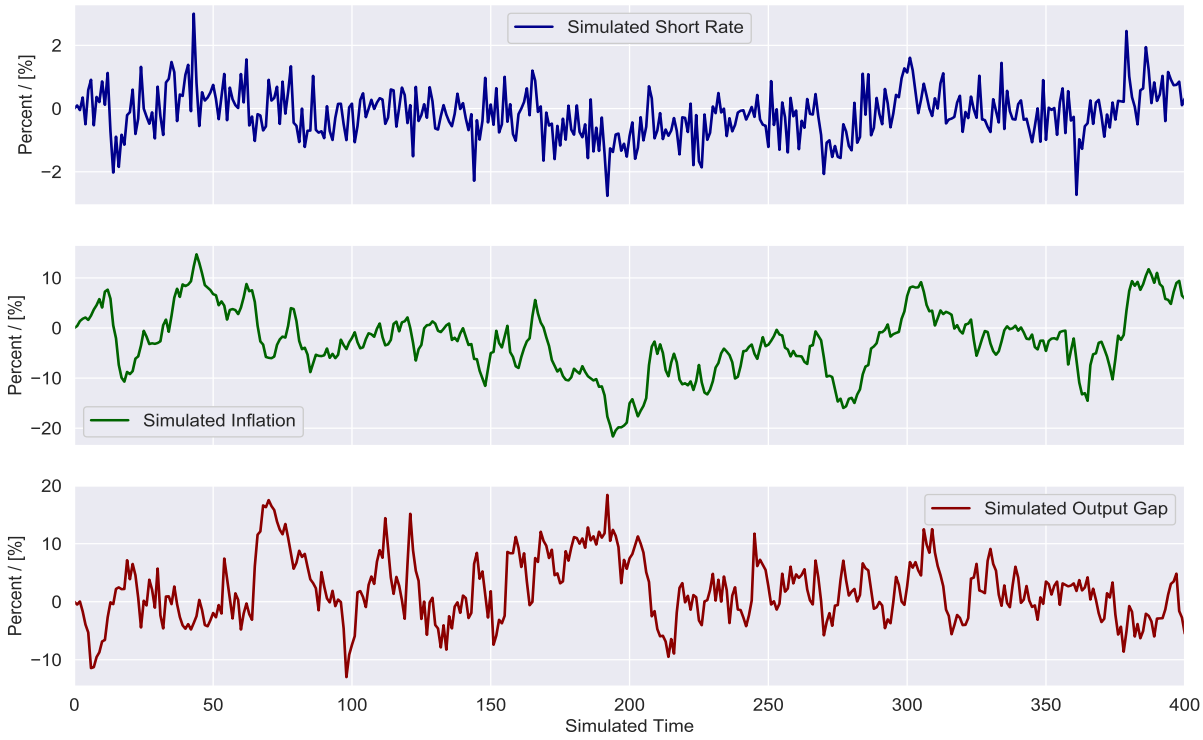



Figure 6: The first 400 out of 10,000 simulated data for short rate, inflation and output gap variables with simulated dynamics similar to the estimates found by RW.

| Coefficient | EM | Con-EM | BLAME | Con-BLAME |
|---|---|---|---|---|
| $\rho_L$ | 0.00 | 0.00 | $-0.20$ | $-0.17$ |
| $\rho_S$ | $-0.71$ | $-0.97$ | $-0.94$ | $-0.84$ |
| $g_\pi$ | $-0.19$ | 0.32 | $-0.32$ | $-0.85$ |
| $\delta_L$ | 0.90 | $-0.32$ | 0.77 | 0.44 |
| $\delta_S$ | 0.38 | 0.01 | $-0.66$ | $-0.83$ |
| $\mu_\pi$ | $-0.78$ | $-0.92$ | $-0.50$ | $-0.38$ |
| $\alpha_1$ | $-0.27$ | $-0.01$ | $-0.06$ | 0.11 |
| $\alpha_2$ | $-0.32$ | $-0.27$ | $-0.26$ | $-0.14$ |
| $\alpha_y$ | $-0.01$ | $-0.63$ | $-0.31$ | $-0.52$ |
| $(1-\mu_y)\beta_1$ | 0.03 | 0.01 | $-0.03$ | $-0.02$ |
| $(1-\mu_y)\beta_2$ | $-0.15$ | $-0.06$ | $-0.14$ | $-0.09$ |
| $\beta_r$ | $-0.78$ | $-1.25$ | $-0.27$ | $-0.80$ |
| Mean Relative Diff | $-0.16$ | $-0.34$ | $-0.16$ | $-0.34$ |

Table 9: These are the (average) relative differences of the estimated parameters with respect to the true ones from the simulation study in Table 1. All four techniques underestimate the parameters on average, with BLAME performing the best on average.

| | Short Rate | Inflation | Output Gap | Unemployment |
|---|---|---|---|---|
| Short Rate | 0.00091 | 0.65 | 0.32 | −0.12 |
| Inflation | 0.00026 | 0.00017 | 0.17 | −0.11 |
| Output Gap | 0.00017 | 0.00004 | 0.00032 | −0.86 |
| Unemployment | −0.00006 | −0.00002 | −0.00003 | 0.00026 |

Table 10: Corr-covariance matrix of US macroeconomic data over the full sample. The diagonal represents the sample variance, the lower triangular matrix contains sample covariances, and the upper triangular matrix contains the sample correlations.

| Macroeconomic Variable | ADF Statistic | 5% p-value |
|---|---|---|
| Short Rate | −2.79 | 0.060 |
| Inflation | −2.92 | 0.043 |
| Output Gap | −2.52 | 0.110 |
| Cyclical Unemployment | −2.83 | 0.055 |

Table 11: Prior hyperparameters for all the covariance matrices used for both the BLAME and the Con-BLAME estimation methods at each estimation window.

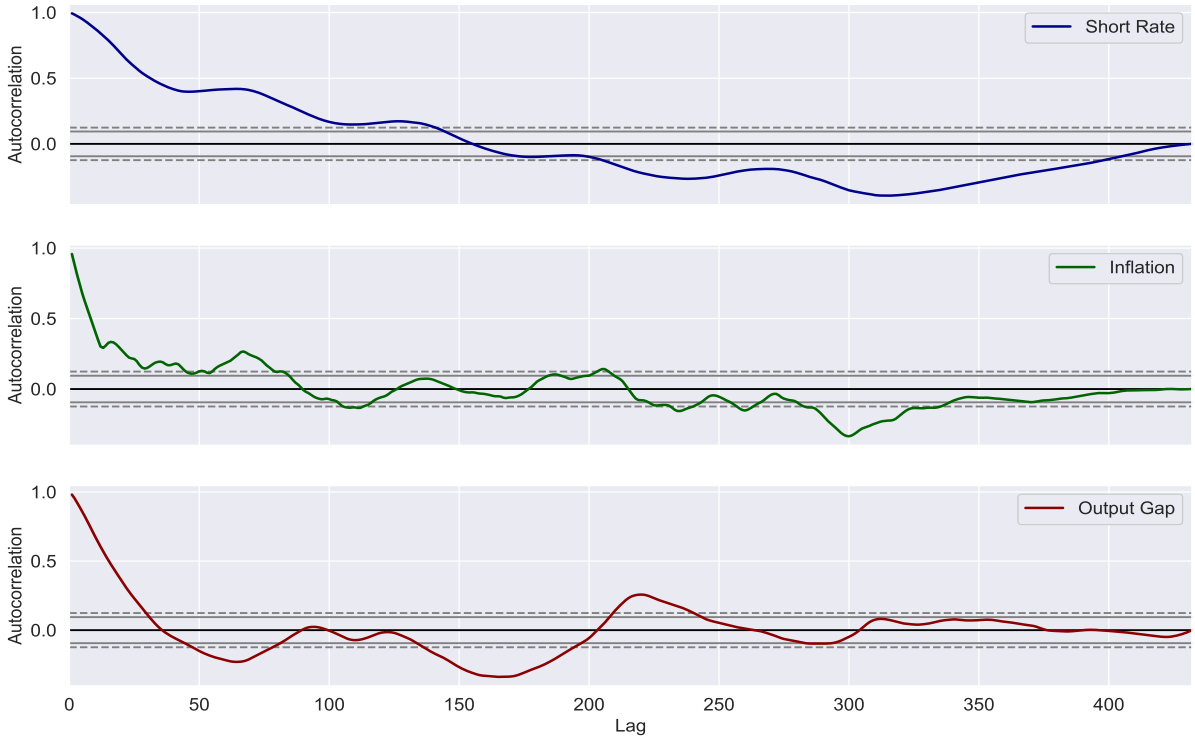



Figure 7: Normalised sample autocorrelation functions of US short rate, inflation and output gap with standardised 95% confidence intervals.

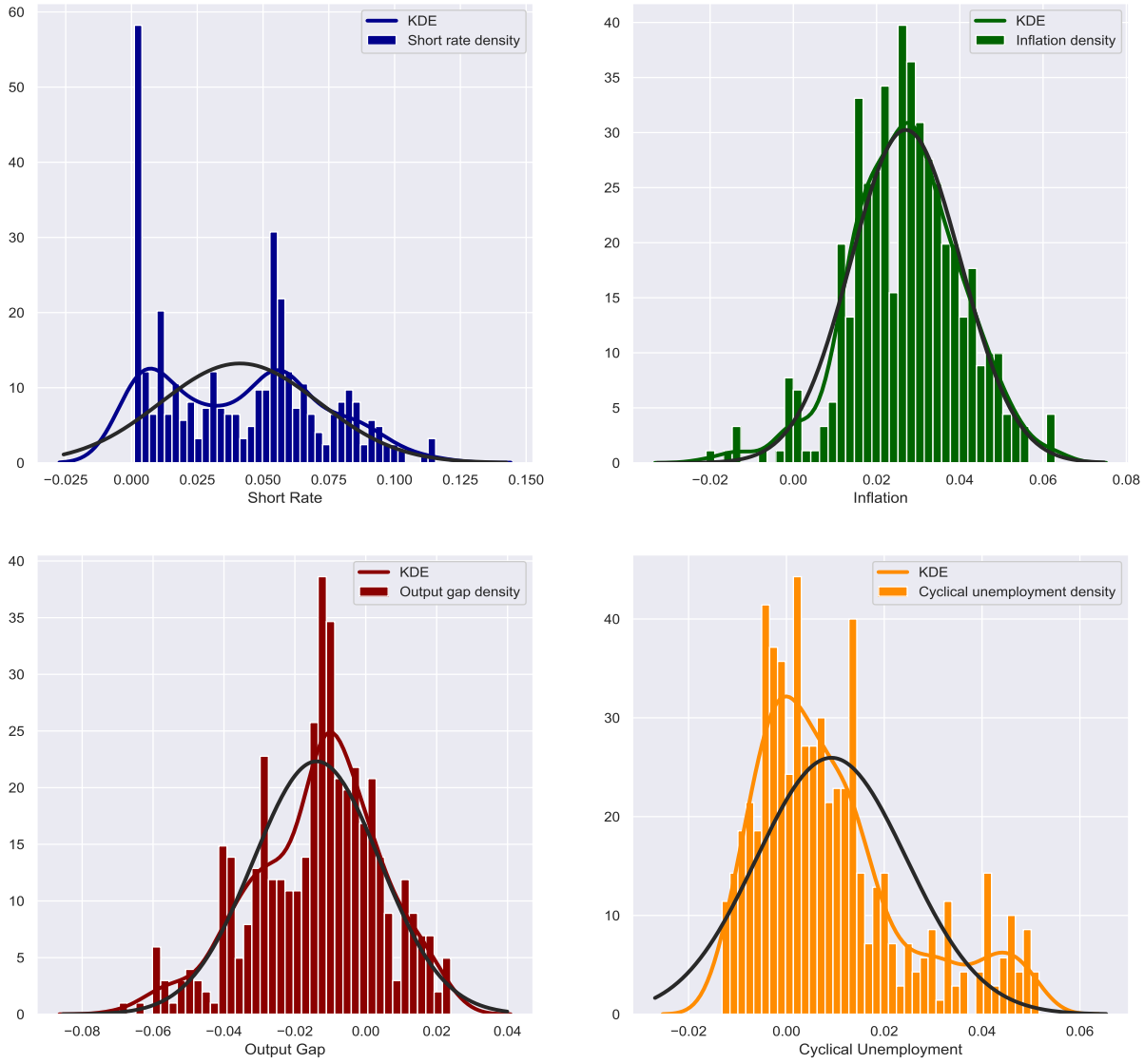



Figure 8: Histograms of US short rate, inflation, output gap and cyclical unemployment data from 1983:M1-2018:M12. For each histogram, the kernel density function is plotted in the same color as the histogram, and overlayed is a fitted Gaussian kernel done via maximum likelihood. The data density kernels for output gap and inflation are similar to the fitted Gaussian kernel, whereas that of cyclical unemployment and the short rate density show a bimodal nature and are not well represented by the Gaussian kernel.