University of Strathclyde

Department of Economics

Bayesian inference in high-dimensional state space models with time-varying parameters and stochastic volatility

Seyma Vahap

Supervisors

Professor Gary Koop

Professor Julia Darby

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

September 22, 2025

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Seyma Vahap

Date: September 22, 2025

Abstract

This thesis begins with an introductory chapter that outlines the inferential challenges of high-dimensional time series models with parameter changes, multiple dependent variables, numerous predictors, and substantial computational complexity, and motivates the methodological contributions that follow.

The first objective is to evaluate a conventional Bayesian estimation approach in modelling the real effects of credit market disruption through a measure of financial distress, the financial external premium, which may vary over time. For this purpose, we specify an eight-variable structural vector autoregressive model with time-varying parameters and stochastic volatility (TVP-VAR-SV). We use the model to examine the nature and evolving features of the links between macroeconomic and financial variables in the U.S. economy.

The second objective is to develop a *Bayesian pairwise composite likelihood method* to address a high-dimensional inference problem in time series models with parameter changes, many dependent variables, and computational complexity, in order to conduct structural analysis. While a larger macroeconomic and financial dataset could be analyzed using a suitable TVP-VAR-SV model, the computational burden of such a model becomes prohibitive in high dimensions. To address this, the method replaces the full likelihood function with a product of pairwise marginal likelihoods and then combines the results to make inference from the composite model. To efficiently aggregate information across bivariate models, we introduce the *Direct Averaging Method*, a novel approach that

provides a computationally tractable approximation to the multivariate structure without requiring simulation from the joint model. An empirical study of time-varying pairwise composite impulse responses demonstrates the impact of an unexpected financial shock on fifty quarterly U.S. macroeconomic and financial variables, yielding economically meaningful results.

The third objective is to develop a *Bayesian dynamic graphical model approach* to overcome a high-dimensional sparse inference problem in TVP-VAR models with volatility discounting, many predictors, and computational complexity, in order to conduct forecasting analysis. The approach incorporates pairwise conditional independence structures in both the coefficient states and the off-diagonal elements of the covariance states. The Bayesian dynamic graphical framework improves forecast combinations of multiple quarterly U.S. macroeconomic and financial variables.

To my mother Wasfeya Habib Ibrahem,

and to the loving memories of my father Joma Wahab Younis

Contents

List of Figures

List of Tables

Acknowledgement

I would like to express my gratitude to Gary Koop for his inspirational supervision and intellectual guidance over the past few years. He has profoundly influenced my understanding and appreciation of the beauty of Bayesian econometrics.

I am especially indebted to Julia Darby for her exemplary enthusiasm and her deeply human-centered approach to supervision. Her encouragement sustained my motivation during the most challenging stages of this journey. The completion of this thesis would not have been possible without her support, for which I will remain grateful throughout my life.

I also wish to extend my heartfelt thanks to Alex Dickson and Fiona McIntosh for their generosity and consistent readiness to offer guidance, which enriched both my work and my personal resilience in difficult times. I am equally grateful to the Strathclyde Business School IT Support team, particularly Josephine Sergeant, Colin Ewing, and Alistair Laidlaw, for their technical expertise and kindness. I also acknowledge the staff of the Andersonian Library, whose commitment and efficiency ensured that I could access every resource necessary to complete my studies.

On a personal note, I am grateful to Ali Taleb, Shazmeen Maroof, Orion Meyer, and Ayse Yaylali for their enduring friendship and warmth. I owe special thanks to Pinar Feyzioglu Akkoyunlu and Milena Hoyos, whose encouragement, wisdom, and support have been a constant source of strength and inspiration.

Above all, my eternal gratitude is to my mother, Wasfeya Habib Ibrahem, for surrounding me with her unique love, support and hope.

Chapter 1

Introduction

Contemporary inferential problems in macroeconomics and finance with ever-increasing comprehensive time series indices have recently given rise to high-dimensional models that may capture the interaction between economies' real and financial sectors. Multidimensional indices are helpful for analysing business cycle dynamics and are particularly important for comparing monetary policy behaviour across very different causes and economic effects of financial crises. High-dimensional multivariate models accounting for parameter changes over time are particularly interesting in this context. Those models involve a vast number of unknown time-varying parameters. However, standard time series modelling has been insufficient in quantifying time-varying relationships between the macroeconomy and the financial sector when large datasets are available. Our inferential goal is to develop innovative Bayesian estimation methods for high-dimensional state space models. In particular, we address two key challenges in high-dimensional time series modelling: (i) high-dimensional inference problems and (ii) high-dimensional sparse inference problems, both posed by models with parameter changes, many dependent variables, many predictors, and substantial computational complexity. An initial empirical study motivates these challenges by examining the effects of credit market disruptions in a TVP-VAR-SV model.

1.1 Bayesian analysis of high-dimensional state space models

Following the seminal work of Sims (1980), vector autoregressive (VAR) models have been recognised as a well-established macroeconometric framework. Since then, VARs have provided "a coherent and credible approach to data description, forecasting, structural inference and policy analysis" as noted by Stock and Watson (2001). Extending over the VAR models with constant parameters, a general methodology with evolving features in parameters was developed by Cogley and Sargent (2002) and later extended by Cogley and Sargent (2005) and a class of model incorporating multivariate stochastic volatility was proposed by Primiceri (2005). However, typically the estimation of the latter model has been carried out by using a conventional Bayesian approach based on small datasets.

This class of models belongs to a special case of state space systems with two features of particular economic importance. First, the parameters are allowed to vary smoothly over time, enabling us to trace how relationships between the real and financial sectors evolve across different economic environments, such as expansions, recessions, and periods of financial distress. Second, by allowing the error covariance matrices to vary over time, the model captures changes in volatility and nonlinear dynamics, which are essential for understanding fluctuations in uncertainty, risk transmission, and the amplification of shocks in the economy.

To evaluate the conventional Bayesian estimation approach, we specify an eight-variable structural vector autoregressive model with time-varying parameters and stochastic volatility (TVP-VAR-SV model). Our focus is on the impact of the financial distress to the real economy. We do so by extending the study by Gilchrist and Zakrajšek (2012) who construct a proxy measure of external finance premium and use a structural VAR model that uses this measure as a variable. However, their analysis lacks examining the evolving features of the coefficients

and the volatilities. In this spirit, we extend their study by evaluating the TVP-VAR-SV model and conduct structural analysis and formal model assessment. In the second chapter, we investigate how a surprise innovation in credit spreads affects real economic activity. We use the model to examine the nature and evolving features of the links between the macroeconomic and financial variables. Experimental results on a quarterly U.S. dataset, including impulse response analysis and formal model comparisons with alternative models, validate our assumptions on time-varying coefficient and covariance states.

1.2 Pairwise composite likelihood methods for TVP-VAR-SV model

A high-dimensional inference problem may arise when the posterior density is demanding to simulate with large datasets as noted by Koop (2013). In a conventional Bayesian analysis, we devote a considerable time in designing and assessing simulation methods based on Markov chain Monte Carlo (MCMC) algorithms as an option to approximate or simulate the posterior density. However, MCMC algorithms are proved to be inefficient, in particular, when a multivariate time series model with time-varying parameters include a large number of dependent variables and time-varying parameters.

Our primary objective is to resolve the computational complexity of this class of model with the conventional wisdom of Bayesian analysis in a statistically parsimonious way. A methodologically appealing approach is to split a high-dimensional likelihood function to low dimensional computationally feasible likelihood components using composite likelihood method that was first developed by Besag (1974) under the term pseudo likelihood approach. Afterwards, Lindsay (1988) proposed composite likelihood method that was based on logarithms of marginal or conditional densities. Bayesian estimation based on composite like-

lihoods in macroeconometrics is an area of research has been partially explored with two examples so far including Chan et al. (2020) and Canova and Matthes (2021).

We use a class of composite likelihood formed by pairwise likelihoods constructed from bivariate marginal densities. More precisely, we adapt a special case of pairwise likelihood methods that was proposed by Verbeke and Molenberghs (2005) and Fieuws and Verbeke (2006) in a novel way to be applied for the first time in the context of TVP-VAR-SV models.

In contrast to most composite likelihood approaches that assume common parameters across components, our method allows for distinct pair-specific parameters. The proposed approach combines the pairwise composite likelihood framework with a novel aggregation scheme, the *Direct Averaging Method* (DAM). The main idea is to avoid the computational burden of sampling from the posterior of the full joint TVP-VAR-SV model by instead fitting all bivariate models separately. In the first step, we simulate from the bivariate posterior distributions associated with each bivariate model. In the second step, we apply the DAM to map and aggregate the pair-specific parameter estimates into a coherent set of pairwise composite parameters for the high-dimensional system. This local parameterization, together with DAM, allows us to decouple the estimation of bivariate components from the global aggregation step, thereby providing a computationally feasible approximation to joint inference in high-dimensional settings.

We evaluate the empirical merits of the Bayesian PCL method. Our evidence is based on a fifty-variable PCL-TVP-VAR-SV model as a combination of 1225 submodels applied to quarterly U.S. macroeconomic and financial data spanning from 1959Q1 to 2018Q1. Time-varying pairwise composite impulse responses show the effects of a surprise increase in credit spreads on 50 variables with encouraging results.

1.3 Dynamic graphical models for TVP-VAR-VD model

Another high-dimensional problem in time series models containing time-varying parameters and time-varying covariance matrix may arise when the model has many predictors with minimal effect on the model, but it is unclear which predictors are relatively more important as noted by Koop (2017). This problem is known as high-dimensional sparse inference problem.

Gaussian graphical models have been used to shed light on the relative importance of predictors. The key ideas of graphical models build on foundational theory in Lauritzen (1996) and Whittaker (2008) that define a rather broad framework with multivariate Gaussian graphical models. However, most graphical models either do not consider time-variation features in temporal coefficients combining a lagged variable and a current variable, and time-variation in covariance coefficients combining two current variable or allow for a fully arbitrary mean. We propose a dynamic graphical model framework that allows combining traditional graphical models described in Whittaker (2008), Chapter 10, and recently introduced dynamic dependence networks in Zhao et al. (2016).

In chapter 4, we propose a Bayesian dynamic graphical model approach incorporating pairwise conditional independence structures in both the coefficient states and the off-diagonal elements of the covariance states, ordered recursively. We achieve this by developing an efficient Bayesian graphical variable selection method that can be applied recursively equation-by-equation in parallel using a Gray code algorithm. Then we perform Bayesian model averaging of the top selected models with high posterior probabilities to forecast ten quarterly U.S. macroeconomic and financial variables over a pseudo-out-of-sample forecast period 1984:Q2-2022:Q3. Comparing out-of-sample forecast performances shows that the joint model with BMA outperforms the joint model with highest poste-

rior probability over the most considered horizons.

1.4 Contribution to knowledge

The main contribution of this thesis is to develop a Bayesian pairwise composite likelihood method to address the high-dimensional inference problem posed by models with parameter changes, many dependent variables, and computational complexities, and a Bayesian dynamic graphical model approach to overcome the high-dimensional sparse inference problem posed by models with parameter changes, many predictors, and computational complexities.

In chapter 2, we analyse the predictive ability of credit spread as a financial distress indicator on the real economy within a VAR model with time-varying temporal coefficients, time-varying contemporaneous relations and time-varying volatilities. With this highly flexible framework and extension, decision makers are able to dynamically assess effects on the real economy and financial sector of a surprise increase of the identified financial shocks, which emphasize evolving features in the business cycle dynamics with a financial accelerator present.

In Chapter 3, we develop a novel class of pairwise composite likelihood approach that departs significantly from existing methods. This approach introduces the DAM, an innovative aggregation mechanism that combines parameter estimates from all bivariate models into a coherent set of pairwise composite parameters for the high-dimensional system. Together, these components form a new class of estimation techniques for the high-dimensional TVP-VAR-SV model, applied here for the first time in the field of macroeconometrics. The proposed framework not only addresses the computational challenges of the full likelihood estimation but also offers substantial potential in methodology, and empirical application. In particular, it enables the analysis of responses of a broad range of macroeconomic and financial variables to financial shocks within a composite

framework that is both scalable and interpretable.

In chapter 4, we extend a traditional graphical model and recently introduced a dynamic graphical model approach to account for pairwise conditional independence uncertainties on the time-varying temporal coefficients in addition to the uncertainties on the independence structures of the time-varying covariance matrix using a Gray code algorithm. Our Bayesian dynamic graphical framework improves forecast combinations and provides insight into the unobservable structures of multiple macroeconomic and financial variables.

Chapter 2

Bayesian vector autoregressive model with time-varying parameters and stochastic volatility in identification of financial shocks

2.1 Introduction

There is a growing consensus that as a source of financial distress, widening of credit spreads, form a reliable class of financial indicators in predicting real economic activity emphasized by Bernanke (2018) and Gertler and Gilchrist (2018). However, a standard time series modelling has been insufficient in quantifying a time-varying relationship between the tightness in the financial conditions and its transmission into the real economy on a high-dimensional setup. In this context, we examine the nature and evolving features of the links between the macroeconomy and the financial sector. Our goal is to assess the extent to which the response of macroeconomic and financial variables to a financial shock is consistent with the financial accelerator mechanism within a model featuring time-varying parameters and stochastic volatility.

The credit spread, first introduced by Bernanke and Gertler (1989) as a proxy for the external finance premium, plays a central role in the financial accelerator

mechanism. According to this framework, the external finance premium exhibits a countercyclical pattern: it rises when borrowers' net worth deteriorates during economic contractions and falls when net worth improves in expansions. Within the financial accelerator, shocks to the external finance premium amplify economic fluctuations, leading to sharp contractions in real activity and declines in asset prices, as emphasized by Bernanke and Gertler (1989), Kiyotaki and Moore (1997), and Bernanke et al. (1999).

Our study links to modern literature on analysing the effects of financial indicators on the real economy, e.g. Gilchrist et al. (2009), Helbling et al. (2011), Gilchrist and Zakrajšek (2012), Stock and Watson (2012), Gilchrist et al. (2014), and Boivin et al. (2020). By exploring the interaction between the macroeconomy and the financial market, Gilchrist and Zakrajšek (2012) constructed a proxy measure of the external finance premium and used a structural vector autoregressive model (VAR model) that uses this measure as an endogenous variable. However, their analysis lacks examining the evolving features of the coefficients and the volatilities. In this spirit, we extend their work by making the assumption that the transmission of a financial shock into the real economy has been evolving over time.[1] This motivates our analysis of possible time variations in coefficients, time variations in contemporaneous relations, and time variations in volatilities. Two research questions then arise as to whether the sources of financial and real relationships come from financial shocks and as to whether the financial and real relationships evolve over time.

For this purpose, we specify an eight-variable structural vector autoregressive model with time-varying parameters and stochastic volatility (TVP-VAR-SV) model to examine the nature and evolving features of the links between the macroeconomic and financial variables in the U.S. economy. The eight-variable TVP-VAR-SV model consists of quarterly U.S. data for consumption, investment,

[1]We use the term "financial shock" to emphasize a shock to the credit spread, a measure of the external finance premium.

GDP growth, GDP deflator, corporate bond spread, stock market returns (S&P 500), the ten-year Treasury yield, and the Federal funds rate.

From a methodological stand point, our work is inspired by a class of VAR model with evolving parameters and multivariate stochastic volatility that was developed by Primiceri (2005). Several authors used the TVP-VAR-SV model to capture the time variation features of macroeconomic and financial time series in alternative contexts including Galí and Gambetti (2015), Prieto et al. (2016), and Gambetti and Musso (2017).

Our key findings are as follows. First, shocks in the financial market lead to a severe economic contraction and slow recoveries, with some detectable differences with the financial accelerator framework. Shocks to the credit spreads appear to cause a much stronger effect on the real economic activity during the Great Recession. This is supported by the associated contraction in real economic activity especially during the Great Recession as the responses to the financial shock exceed the average.

Second, the resulting economic downturn leads to an ambiguous response of inflation over time. In line with the financial accelerator effect, there is an appreciation of disinflation until late 1980s. However, after then a persistent increase in inflation takes place. With the adverse economic effects, the Federal Reserve reduced the Federal funds rate. There is ambiguity about the response of the stock market returns, which experiences a significant drop in recent years.

Third, a model comparison assessment with a discrete set of competing models shows that using four competing models and assessing pairwise Bayes factor, the TVP-VAR-SV model is supported by the data and the prior information. However, most of the support of the TVP-VAR-SV model fit seems to come from allowing time variations in the volatilities.

Fourth, Bayes factors can miss out small amounts of time variation features in parameters in high-dimensional TVP-VAR-SV models. In this regard, we assess

the TVP-VAR-SV model fit by a model comparison using a hybrid approach to analyse whether some equations in our eight-variable TVP-VAR-SV model have time-varying parameters or not. Indeed, the results of comparing $1 + N(N + 1)/2 = 37$ models with $N = 8$ show that some equations may have time-varying parameter features and some other may miss those features.

The rest of the chapter is structured as follows. Section 2.2 describes the TVP-VAR-SV model design and specifications. Section 2.3 describes the identification scheme and the selected proxies of the credit spreads and the term spread. Section 2.4 shows the main empirical findings based on a baseline eight–variable TVP-VAR-SV model, explores robustness to alternative measures of corporate credit spreads, and performs model assessment with two alternative formal model comparison methods using Bayes factors. A summary is provided in Section 2.5. Finally, Section 2.6 provides an appendix for estimating a constant coefficient and constant volatility VAR model, a summary of MCMC algorithm for TVP-VAR-SV model and a data appendix.

2.2 The Model

In this section, we introduce the TVP-VAR-SV model, and complete model specification by choosing prior distributions for all parameters of interest.

2.2.1 Structural TVP-VAR-SV model

Let $\mathbf{y}_t$ denote an $N \times 1$ vector of time series variables. For each cross-sectional variable $i = 1 : N$, $t = 1 : T$ denotes the time series observations. The dynamics of $\mathbf{y}_t$ follows a structural vector autoregressive process with model order p,

$$\mathbf{A}_t \mathbf{y}_t = \mathbf{A}_{0,t} + \sum_{k=1}^{p} \mathbf{A}_{k,t} \mathbf{y}_{t-k} + \mathbf{u}_t, \tag{2.1}$$

where $\mathbf{A}_{0,t}$ is an $(N\times 1)$ is a vector of time-varying intercepts, $\mathbf{A}_{k,t} = (\mathbf{A}_{1,t}, \ldots, \mathbf{A}_{p,t})$ are $(N \times N)$ matrices of the time-varying parameters with $k = 1, \ldots, p$, and $\mathbf{u}_t$ are the structural disturbances with diagonal covariance matrix. The reduced form representation of the structural VAR model in Equation (2.1) can be written in a special case state space system as follows

$$\mathbf{y}_t = \mathbf{x}_t \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \tag{2.2}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t^{\beta}, \tag{2.3}$$

where $\mathbf{x}_t = \mathbf{I}_N \otimes (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})$ is an $N \times LN$ matrix and $\boldsymbol{\beta}_t = \text{vec}([\mathbf{B}_{0,1}, \ldots, \mathbf{B}_{p,t}]')$ is $LN \times 1$ vector with $\mathbf{B}_{k,t} = \mathbf{A}_t^{-1}\mathbf{A}_{k,t}$, $k = 0, \ldots, p$, and $L = 1 + Np$. We assume the state coefficients, $\boldsymbol{\beta}_t$, evolve in time according to a random walk and the evolution error terms are assumed to be a zero-mean process with state evolution variance matrix, $\boldsymbol{\eta}_t^{\beta} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\beta})$. $\boldsymbol{\epsilon}_t$ is an $(N \times 1)$ vector of reduced form disturbances assumed to be a zero-mean process with time-varying covariance matrix, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\epsilon,t})$, and the reduced from disturbances, $\boldsymbol{\epsilon}_t$, are obtained from the structural shocks by a linear transformation as

$$\boldsymbol{\epsilon}_t \equiv \mathbf{P}_t \mathbf{u}_t, \tag{2.4}$$

where $E(\mathbf{u}_t \mathbf{u}'_t) = \mathbf{I}_N$ and $E(\mathbf{u}_t \mathbf{u}'_{t-k}) = 0$ for all t and $k = 1, 2, 3, \ldots$.

Let's define $\boldsymbol{\Sigma}_{\epsilon,t}$ as a positive definite symmetric $(N \times N)$ matrix, which has the following representation

$$\boldsymbol{\Sigma}_{\epsilon,t} = \mathbf{A}_t \mathbf{D}_t \mathbf{A}'_t. \tag{2.5}$$

A necessary condition to offer a unique solution to the system $\boldsymbol{\Sigma}_{\epsilon,t} = \mathbf{A}_t \mathbf{D}_t \mathbf{A}'_t$ requires an order condition for identification. The order condition implies that $\mathbf{A}_t$ and $\mathbf{D}_t$ have $N(N+1)/2$ distinct elements, which is equivalent to the elements

of $\mathbf{\Sigma}_{\epsilon,t}$. If $\mathbf{D}_t$ is diagonal, it requires N elements, meaning that $\mathbf{A}_t$ can have maximum $N(N-1)/2$ free elements, as described by Hamilton (1994). This leads to specify $\mathbf{A}_t$ as a lower triangular matrix with 1's along the principal diagonal. We define $\boldsymbol{\alpha}_t$ as the parameters of the lower triangular elements of $\mathbf{A}_t$, and define $\mathbf{d}_t = (d_{1,t}, \ldots, d_{i,t}, \ldots, d_{N,t})'$.

$$\mathbf{A}_t = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \alpha_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \alpha_{N1,t} & \ldots & \alpha_{NN-1,t} & 1 \end{pmatrix}, \quad \text{and}$$

$$\mathbf{D}_t = \begin{pmatrix} d_{1,t} & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & d_{N,t} \end{pmatrix},$$

where the $\alpha_{i,j,t}$ elements are real-valued and the $d_{i,t}$ are positive.[2]

In this framework, volatility is captured by the time-varying standard deviations of the reduced-form disturbances, governed by the square roots of the diagonal elements of $\mathbf{D}_t$, which result from the time variation in the transformation applied to the structural shocks. Let $\boldsymbol{\sigma}_t$ denote the vector of standard deviations, i.e., the square roots of the diagonal elements of $\mathbf{D}_t$. Hence, the elements of the time-varying covariance matrix are specified as random walks

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t^{\alpha}, \tag{2.6}$$

[2]This decomposition follows Primiceri (2005) and Galí and Gambetti (2015). It is equivalent to the more common representation $\mathbf{\Sigma}_{\epsilon,t} = \mathbf{A}_t^{-1}\mathbf{D}_t(\mathbf{A}_t^{-1})'$ (e.g., Hamilton (1994)), but differs by a reparameterisation: here $\mathbf{A}_t$ is defined as the lower-triangular Cholesky factor with unit diagonal, while $\mathbf{D}_t$ collects the time-varying variances.

$$\mathbf{h}_t = \mathbf{h}_{t-1} + \boldsymbol{\eta}_t^h, \tag{2.7}$$

where $\mathbf{h}_t = \log \boldsymbol{\sigma}_t$, $E(\boldsymbol{\eta}_t^\alpha) = 0$ and $E(\boldsymbol{\eta}_t^h) = 0$, $E(\boldsymbol{\eta}_t^\alpha \boldsymbol{\eta}_t'^\alpha) = \boldsymbol{\Sigma}_\alpha$ and $E(\boldsymbol{\eta}_t^h \boldsymbol{\eta}_t'^h) = \boldsymbol{\Sigma}_h$, and $\boldsymbol{\eta}_t^\alpha \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\alpha)$ and $\boldsymbol{\eta}_t^h \sim \mathcal{N}(0, \boldsymbol{\Sigma}_h)$.

2.2.2 Priors and initial values

To complete the model specifications, we calibrate the values of prior hyperparameters by applying a "training sample prior". This prior choice reflects the belief that there is absence of information to select a prior distribution but presence of a large number of observations as in Primiceri (2005).

To calibrate the prior distributions for $\boldsymbol{\beta}_0$, $\boldsymbol{\alpha}_0$, and $\mathbf{h}_0$, we estimate a time invariant version of the model in Equation (2.2) based on an initial training sample of 40 observations. The priors for the initial states $\boldsymbol{\beta}_0$, $\boldsymbol{\alpha}_0$, and $\mathbf{h}_0$ are Normal with parameters $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and are defined as

$$\begin{aligned} \boldsymbol{\beta}_0 &\sim \mathcal{N}(\hat{\boldsymbol{\beta}}_0, \tau_\beta \times \hat{\mathbf{P}}_{\beta_0}), \\ \boldsymbol{\alpha}_0 &\sim \mathcal{N}(\hat{\boldsymbol{\alpha}}_0, \tau_\alpha \times \hat{\mathbf{P}}_{\alpha_0}), \\ \mathbf{h}_0 &\sim \mathcal{N}(\hat{\mathbf{h}}_0, \tau_h \times \mathbf{I}_N), \end{aligned} \tag{2.8}$$

where τ_β, τ_α, and τ_h are scale factors, $\boldsymbol{\mu}_0$ may represent any of the parameters $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\alpha}}_0$ or $\hat{\mathbf{h}}_0$, and $\boldsymbol{\Sigma}_0$ may refer to any of the parameters $\tau_\beta \times \hat{\mathbf{P}}_{\beta_0}$, $\tau_\alpha \times \hat{\mathbf{P}}_{\alpha_0}$ or $\tau_h \times \hat{\mathbf{P}}_{h_0}$. The matrices, $\boldsymbol{\Sigma}_{\beta,0}$, $\boldsymbol{\Sigma}_{\alpha,0}$, and $\boldsymbol{\Sigma}_{h,0}$, follow inverse Wishart distributions with parameters $(\mathbf{U}_0, \nu)$. Then, $\boldsymbol{\Sigma}_{\beta,0}$ is defined as

$$\boldsymbol{\Sigma}_{\beta,0} \sim \mathcal{IW}_{N(Np+1)}(\kappa^2_{\Sigma_\beta} \times (N(Np+1)+1) \times \hat{\mathbf{P}}_{\beta_0}, (N(Np+1)+1)), \tag{2.9}$$

with scale matrix $U_{\beta,0} = \kappa^2_{\Sigma_\beta} \times \nu_\beta \times \hat{\mathbf{P}}_{\beta_0}$, and prior degrees of freedom $\nu_\beta = (N(Np+1)+1)$.

We calibrate the elements of $\boldsymbol{\Sigma}_{\alpha,0}$ by assuming block-diagonal with blocks

corresponding to the equations of the system. In other words, for the lower triangular matrix $\mathbf{A}_t$, the blocks have size $\mathbf{S} = \{(1\times1), (2\times2), \ldots, [(N-1)\times(N-1)]\}$, where $\mathbf{S}_i$ denotes the ith block of $\mathbf{S}$ that follow inverse Wishart distributions

$$\begin{aligned}
\mathbf{S}_1 &\sim \mathcal{IW}_1(\kappa^2_{S_1} \times 2 \times \hat{\mathbf{P}}_{\hat{\alpha}_{21}}, 2), \\
\mathbf{S}_2 &\sim \mathcal{IW}_2(\kappa^2_{S_2} \times 3 \times \hat{\mathbf{P}}_{\hat{\alpha}_{31},\hat{\alpha}_{32}}, 3), \\
&\vdots \\
\mathbf{S}_{N-1} &\sim \mathcal{IW}_{N-1}(\kappa^2_{S_{N-1}} \times N \times \hat{\mathbf{P}}_{\hat{\alpha}_{N1},\ldots,\hat{\alpha}_{NN-1}}, N),
\end{aligned} \tag{2.10}$$

where $\hat{\alpha}_{i,j}$ refers to elements of blocks of the $\hat{\boldsymbol{\alpha}}_0$ for $i \neq j$.

Finally, the variances of the stochastic volatility innovations follow inverse Wishart distribution of the form

$$\boldsymbol{\Sigma}_{h,0} \sim \mathcal{IW}_N(\kappa^2_{\Sigma_h} \times (N+1) \times \mathbf{I}_N, (N+1)). \tag{2.11}$$

To sum up, the priors for the eight-variable TVP-VAR-SV model are specified as follows

$$
\begin{aligned}
\boldsymbol{\beta}_0 &\sim \mathcal{N}(\hat{\boldsymbol{\beta}}_0, 9 \times \mathbf{P}_{\hat{\beta}_0}), \\
\boldsymbol{\alpha}_0 &\sim \mathcal{N}(\hat{\boldsymbol{\alpha}}_0, 9 \times \mathbf{P}_{\hat{\alpha}_0}), \\
\mathbf{h}_0 &\sim \mathcal{N}(\hat{\mathbf{h}}_0, 1 \times \mathbf{I}_N), \\
\boldsymbol{\Sigma}_{\beta,0} &\sim \mathcal{IW}_{(pN^2+N)}(0.01^2 \times 137 \times \hat{\mathbf{P}}_{\hat{\beta}_0}, 137), \\
\boldsymbol{\Sigma}_{h,0} &\sim \mathcal{IW}_N(0.01^2 \times 9 \times \mathbf{I}_N, 9), \\
\mathbf{S}_1 &\sim \mathcal{IW}_1(0.001 \times 2 \times \mathbf{P}_{(\hat{\alpha}_{21,0})}, 2), \\
\mathbf{S}_2 &\sim \mathcal{IW}_2(0.001 \times 3 \times \mathbf{P}_{(\hat{\alpha}_{31,0},\hat{\alpha}_{32,0})}, 3), \\
\mathbf{S}_3 &\sim \mathcal{IW}_3(0.001 \times 4 \times \mathbf{P}_{(\hat{\alpha}_{41,0},\hat{\alpha}_{42,0}),\hat{\alpha}_{43,0}}, 4), \\
\mathbf{S}_4 &\sim \mathcal{IW}_4(0.001 \times 5 \times \mathbf{P}_{(\hat{\alpha}_{51,0},\hat{\alpha}_{52,0},\hat{\alpha}_{53,0},\hat{\alpha}_{54,0})}, 5), \\
\mathbf{S}_5 &\sim \mathcal{IW}_5(0.001 \times 6 \times \mathbf{P}_{(\hat{\alpha}_{61,0},\hat{\alpha}_{62,0},\hat{\alpha}_{63,0},\hat{\alpha}_{64,0},\hat{\alpha}_{65,0})}, 6), \\
\mathbf{S}_6 &\sim \mathcal{IW}_6(0.001 \times 7 \times \mathbf{P}_{(\hat{\alpha}_{71,0},\hat{\alpha}_{72,0},\hat{\alpha}_{73,0},\hat{\alpha}_{74,0},\hat{\alpha}_{75,0},\hat{\alpha}_{76,0})}, 7), \\
\mathbf{S}_7 &\sim \mathcal{IW}_7(0.001 \times 8 \times \mathbf{P}_{(\hat{\alpha}_{81,0},\hat{\alpha}_{82,0},\hat{\alpha}_{83,0},\hat{\alpha}_{84,0},\hat{\alpha}_{85,0},\hat{\alpha}_{86,0},\hat{\alpha}_{87,0})}, 8).
\end{aligned}
$$

2.3 Identification and proxies to external finance premium

In this section, we discuss a recursive identification scheme and consider alternative proxies to the external finance premium.

2.3.1 Identification

A way to achieve identification of the financial shocks is to assume that the structural innovations $\mathbf{u}_t$ in Equation (2.4) can be obtained from the reduced form disturbances $\boldsymbol{\epsilon}_t$ by orthogonalising the reduced form errors, this follows Kilian and Lütkepohl (2017). We accomplish this task by defining a lower triangular $N \times N$

matrix $\mathbf{P}_t\mathbf{P}_t' = \mathbf{\Sigma}_{\epsilon,t}$, such that $\mathbf{P}_t$ is the Cholesky decomposition of $\mathbf{\Sigma}_{\epsilon,t}$. It follows from Equation (2.5) that one possible solution to the problem is $\mathbf{P}_t = \mathbf{A}_t^{-1}\mathbf{D}_t^{1/2}$. We compute the impulse responses as described in Galí and Gambetti (2015). As mentioned by Kilian and Lütkepohl (2017), the recursive condition in $\mathbf{P}_t$ requires ordering of variables based on economic theory. Furthermore, we assume that the transmission channel of the financial shocks is in line with the financial accelerator mechanism. Thus, an unexpected increase in credit spreads leads to an immediate reaction of the financial sector to the surprise news while movements in real economic activities and prices occur only within a period. This is an example of a recursive identification scheme on a TVP-VAR-SV model.[3]

We construct an eight–variable TVP-VAR-SV model that includes both macroeconomic and financial variables. The eight variables in this specification are, consumption, investment, GDP growth, GDP deflator, corporate bond spread (A-spread), stock market returns (S&P 500), the ten-year Treasury yield, and the Federal funds rate. The identifying assumption implied by this recursive ordering is that shocks to the A–spread affect consumption, investment, GDP growth, and inflation within a period, while the stock returns, the ten-year Treasury yield, and the Federal funds rate can react contemporaneously to such a financial disturbance. The TVP-VAR-SV model is estimated over the full sample period, using 2 lags of each variable.

2.3.2 Proxies of external finance premium

We consider a series of proxies to the external finance premium. Namely, default risk premium, the excess bond premium, and the GZ-spread. We also use the term spread as an alternative indicator.

[3]Impulse responses are normalised to ensure comparability across shocks. The normalisation procedure follows the approach described in detail in Section 3.3.3.

Default risk premium

Our benchmark measure is A–spread (the difference between BAA corporate bond yields and AAA corporate bond yields). As alternatives, we also consider the 10–year A–spread (the difference between AAA corporate bond yields and 10–year Treasury bond yields) and the 10–year B–spread (the difference between BAA corporate bond yields and 10–year Treasury bond yields). We take those measures as proxies to the external finance premium, i.e. measures of indicators of financial distress. We can see their sensitivities to episodes of recessions in the U.S. economy in Figure 2.1.

Excess bond premium and the GZ-spread

Other measures of the external finance premium include the excess bond premium and the GZ spread, widely used indicators of financial market tightness developed by Gilchrist and Zakrajšek (2012). They construct a credit spread index from a panel of bond prices and show that it has strong predictive power for real economic activity. This index is decomposed into a predicted component reflecting expected default risk and a residual component capturing the excess bond premium.

The term spread

There is extensive evidence on the predictive power of the term spread (the slope of the yield curve) for fluctuations in real economic activity and inflation. Although different proxies are used in the literature, the most common definition is the difference between the ten-year Treasury bond yield and the three-month Treasury bill yield.

The term spread has proven useful in forecasting real economic activity, as shown by Hamilton and Kim (2002) and Andrew et al. (2006). In a comprehensive review, Stock and Watson (2003) also emphasize its ability to predict

U.S. economic growth and inflation. As illustrated in Figure 2.1, the term spread typically declines to low or even negative values before recessions, which explains its role as a leading indicator of downturns. Accordingly, we consider the term spread as an alternative financial indicator to assess its effect on real economic activity and whether this relationship changes over time.

Figure 2.1 also displays alternative measures of the external finance premium. The GZ spread, for example, rises sharply during recessions and peaked above 6% in the Great Recession. More generally, all credit spreads move countercyclically, increasing in economic downturns, whereas the inverted term spread tends to rise before downturns occur.

To compare the relative information content of these indicators, we extend the analysis by adding them one at a time to a six-variable TVP-VAR-SV model. In this specification, y_t includes GDP growth, the GDP deflator, the unemployment rate, the A-spread, stock market returns, and the federal funds rate. The recursive ordering assumes that shocks to the A-spread affect output, inflation, and unemployment contemporaneously, while stock returns and the federal funds rate can also respond immediately to such shocks. In alternative specifications, we replace the A-spread with the 10-year A-spread, the 10-year B-spread, the GZ spread, the excess bond premium, or the term spread as financial shocks. Data definitions and sources are provided in the Data Appendix.

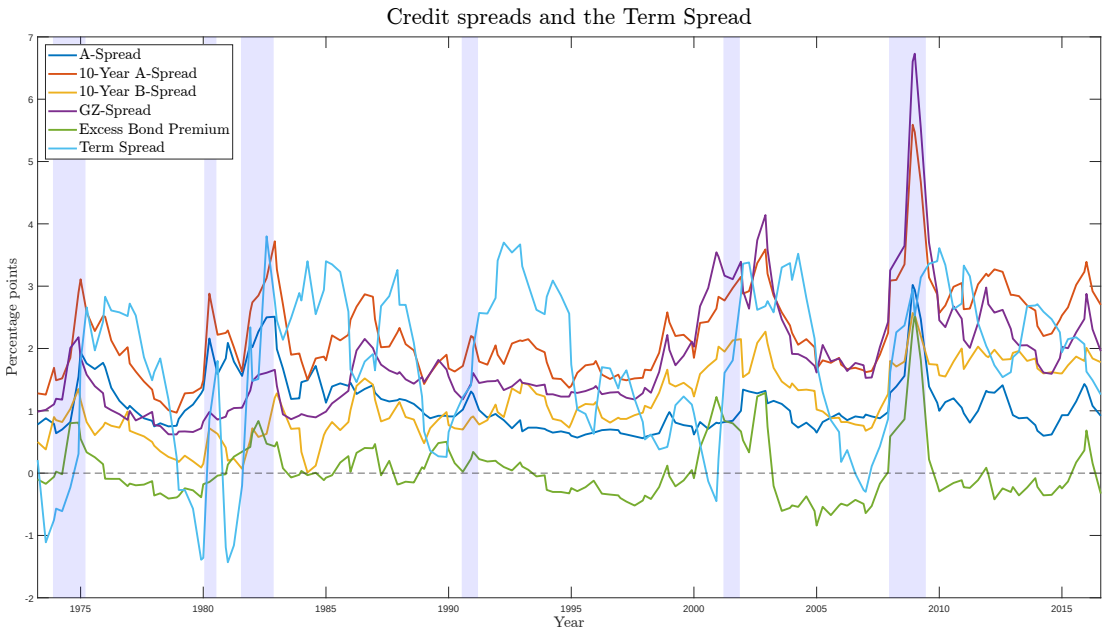

Figure 2.1: The figure depicts the selected credit spreads and the term spread. Quarterly time series plots span from 1973:Q2 to 2016:Q3. The shaded vertical bars represent the National Bureau of Economic Research's recession dates in the U.S. economy.

2.4 Empirical Results

In this section, we report the consequences of a surprise innovation to the credit spread on the real economic activity. All data are obtained from the Fred database of the Federal Reserve Bank of St Louis except the GZ-spread and the excess bond premium of Gilchrist and Zakrajšek (2012), which are obtained from the authors' website. The time span for some variables starts in 1959:Q1 and ends in 2018:Q1 and for the variables, GZ–spread, excess bond premium, and term spread, it starts in 1973:Q2 and ends in 2016:Q3.[4] The TVP-VAR-SV models are estimated over the full sample period, using two lags of each endogenous variable and 10,000 Markov chain Monte Carlo (MCMC) algorithm replications.

[4]The Data appendix provides a complete listing of all variables.

2.4.1 Results from the recursive identification scheme

Figures 2.2 through Figure 2.9 plot the impulse responses of the endogenous variables to a financial shock to the A–spread. An unexpected increase of in the A-spread causes a reduction in the real economic activity, with consumption, investment and output decreasing over the next quarters. In response to this adverse financial shock, the macroeconomic implications are severe; the decrease in the investment outweighs the sharp drop in the real GDP growth, which is about 3 percentage point.

However, the resulting economic downturn leads to an ambiguous response of inflation over time. In line with the financial accelerator effect, there is an appreciation of disinflation until late 1980s.[5] However, after then there is a persistent increase in inflation. In response to these adverse macroeconomic effects of the financial shock, the Federal Reserve reduced the Federal funds rate as depicted by the time-varying impulse response of the relevant variable in Figure 2.9 that occurs about one quarter after the initial impact of the shock. There is ambiguity about the response of the stock market returns, which experiences a significant drop in recent years.

[5]Bernanke (2003) defines disinflation as "a decline in the rate of inflation".

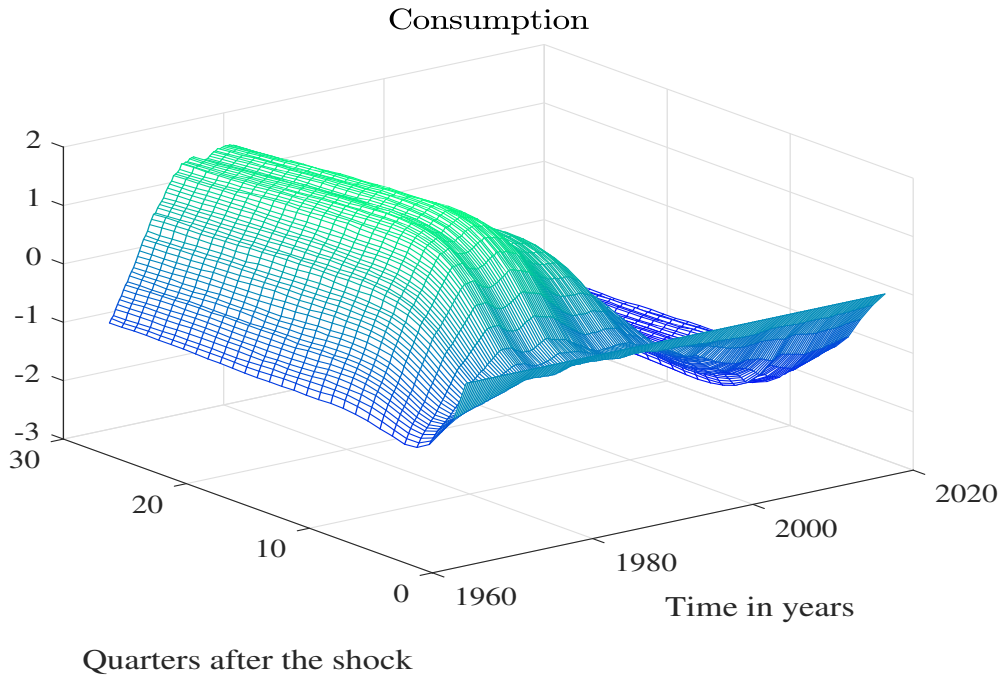

Figure 2.2: The figure depicts the impulse response of Real Consumption to the identified shock of the A-spread estimated from an eight–variable structural TVP-VAR-SV model.

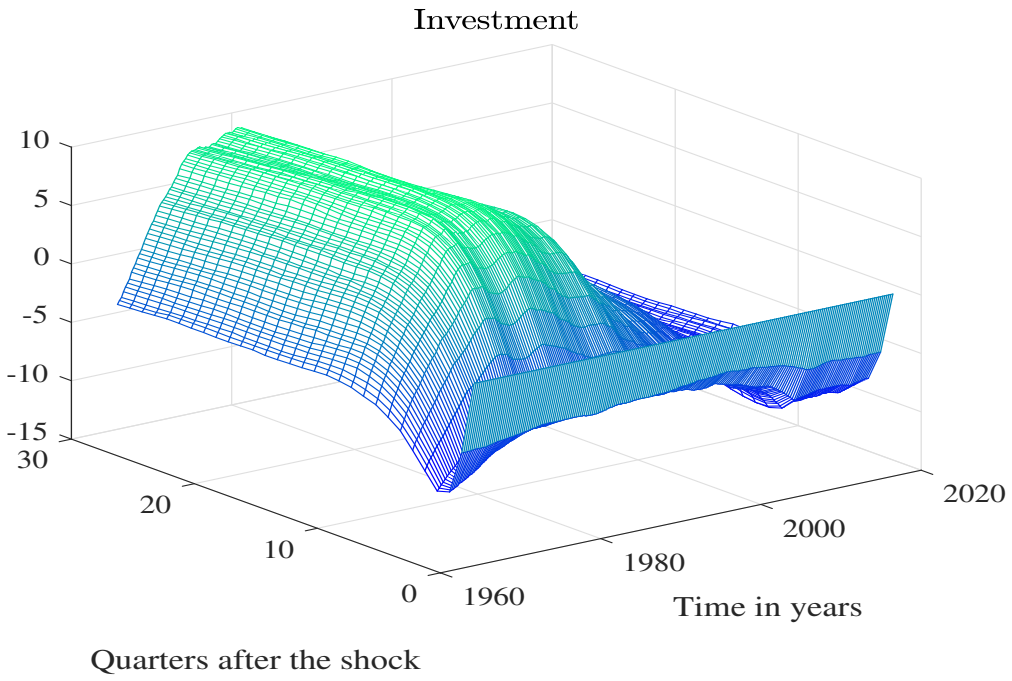

Figure 2.3: The figure depicts the impulse response of Real Investment to the identified shock of the A-Spread estimated from an eight–variable structural TVP-VAR-SV model.

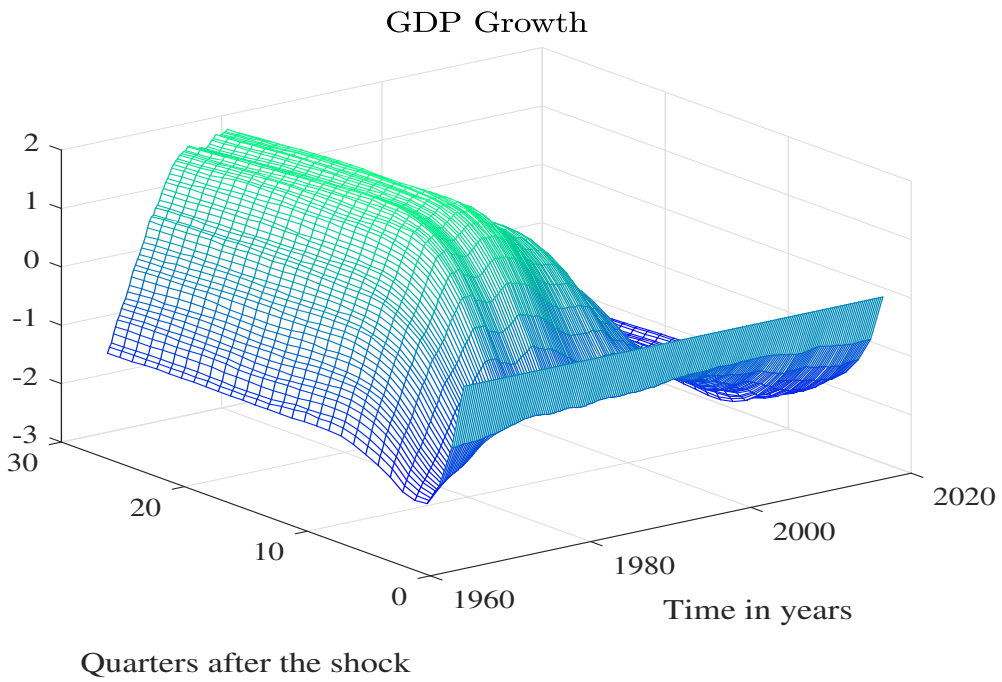

Figure 2.4: The figure depicts the impulse response of Real GDP Growth to the identified shock of the A-Spread estimated from an eight–variable structural TVP-VAR-SV model.

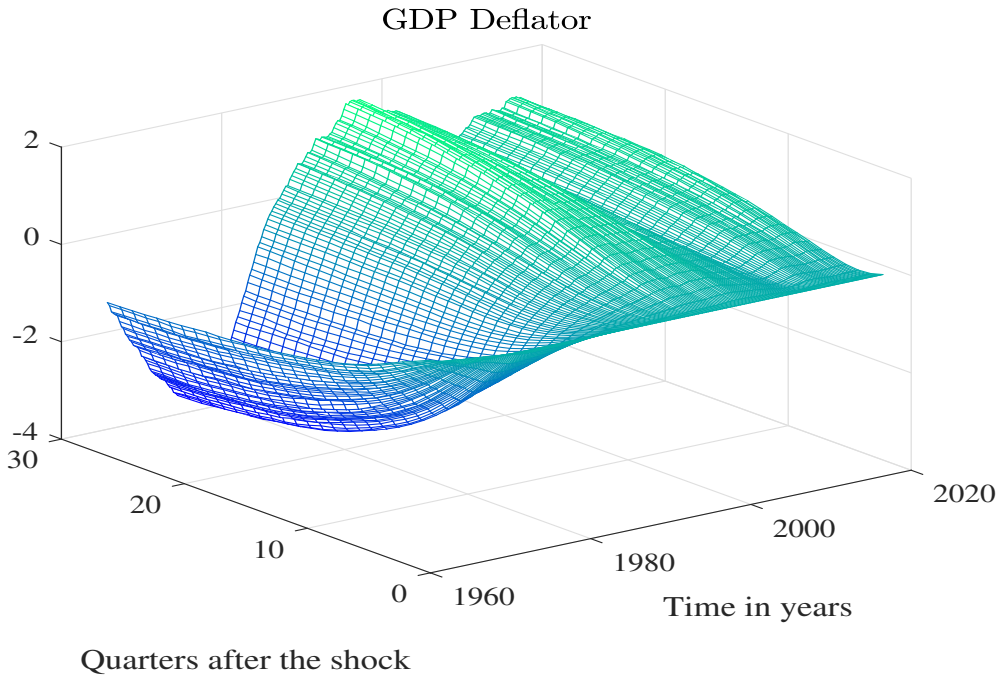

Figure 2.5: The figure depicts the impulse response of GDP Deflator to the identified shock of the A-Spread estimated from an eight–variable structural TVP-VAR-SV model.

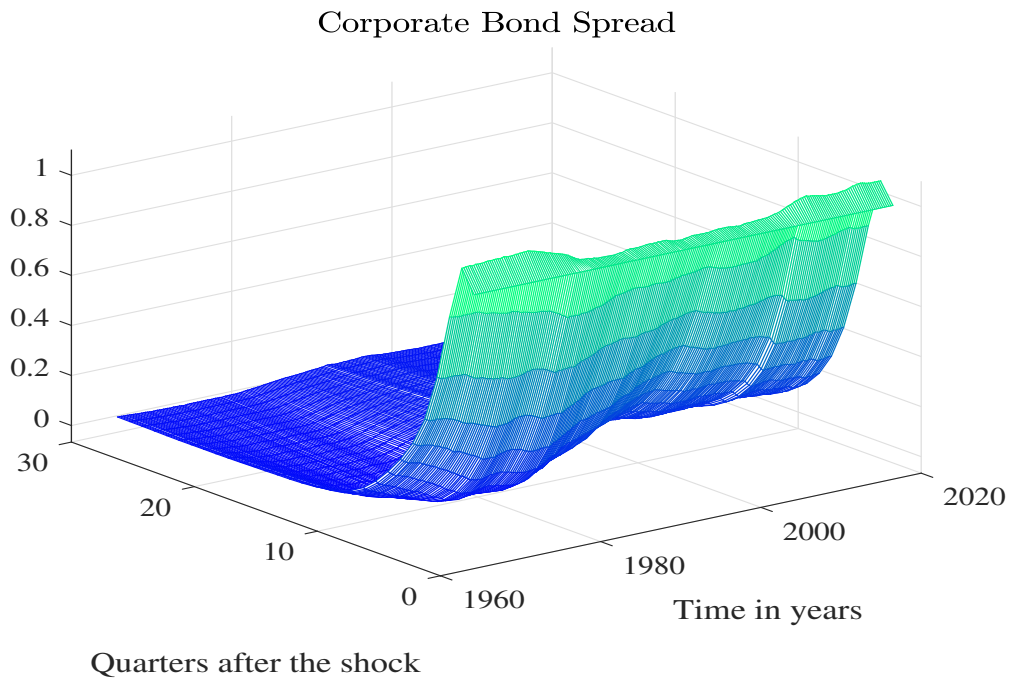

Figure 2.6: The figure depicts the impulse response of A-Spread to the identified shock of the A-Spread estimated from an eight–variable structural TVP-VAR-SV model.

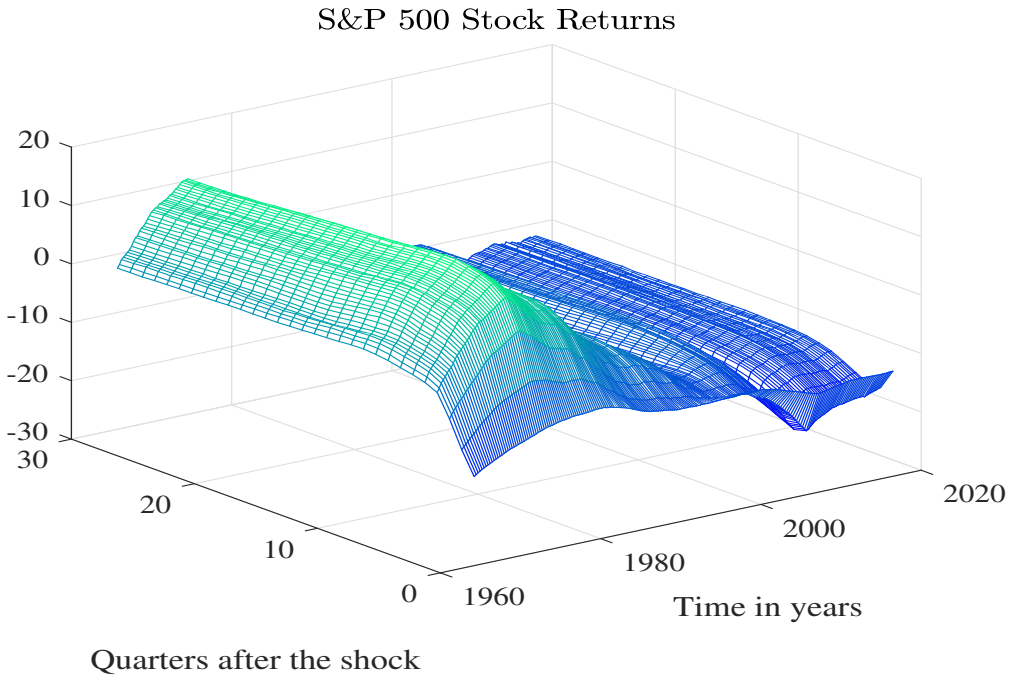

Figure 2.7: The figure depicts the impulse response of Stock Returns to the identified shock of the A-Spread estimated from an eight–variable structural TVP-VAR-SV model.

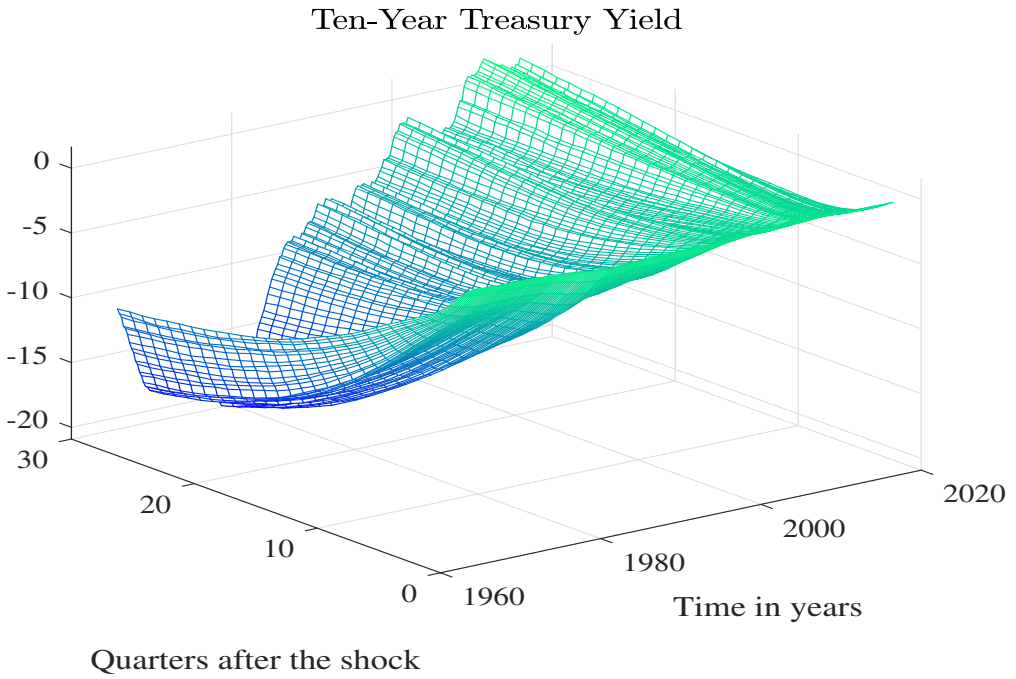

Figure 2.8: The figure depicts the impulse response of Ten-Year Treasury Yield to the shock of the A-Spread estimated from an eight–variable structural TVP-VAR-SV model.

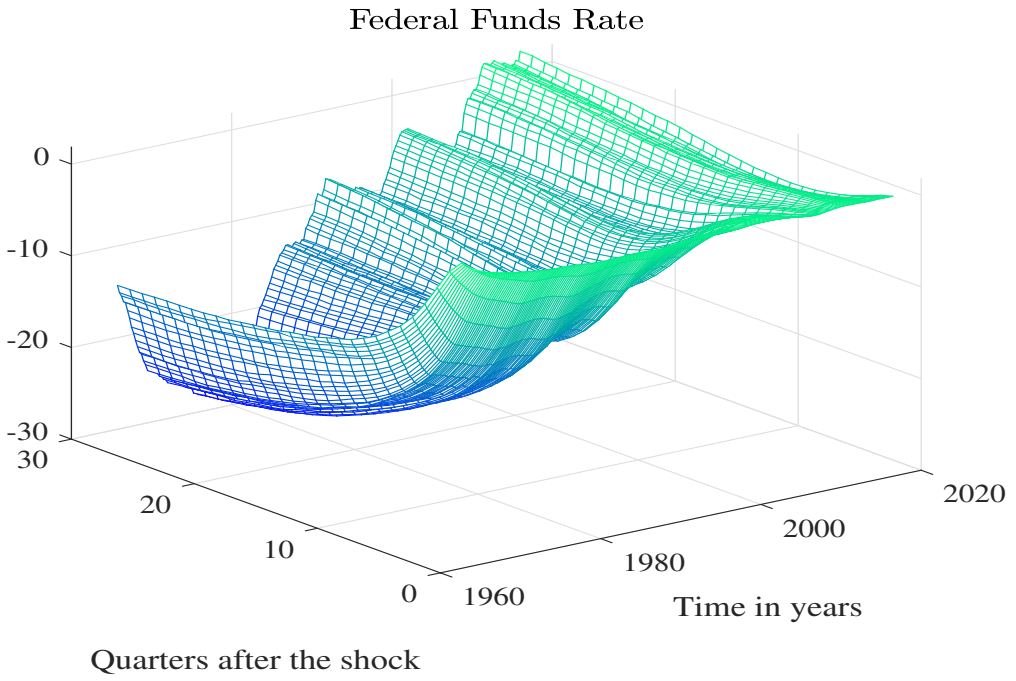

Figure 2.9: The figure depicts the impulse response of Federal Funds Rate to the identified shock of the A-Spread estimated from an eight–variable structural TVP-VAR-SV model.

2.4.2 Stochastic volatility

We present the posterior median, 5-th and 95-th percentiles of the time-varying standard deviation of residuals of consumption, investment, GDP growth, GDP deflator, A-spread, S&P 500 stock returns, ten-year Treasury yield, and Federal funds rate in Figure 2.10. In most cases, there is evidence of a considerable time variation in the standard deviation of the innovations with spikes appearing during different episodes of business cycle dynamics.

The first column and the third row of Figure 2.10 displays the time-varying standard deviation of the A-spread shocks. The periods between 1981:Q3- 1982:Q4 and 2007:Q4-2009:Q2, which are periods of contractions, present a higher variance of A-spread shocks. Another interesting fact from the analysis of the stochastic volatilities is related to time-varying standard deviations of the ten-year Treasury yield and the Federal funds rate. The former is displayed in the first column and the fourth row and the latter is presented in the second column and the fourth row of Figure 2.10. The variances of these variables peak between 1979 and 1983. The result for the Federal funds rate is in line with the findings of Primiceri (2005).

Overall, the evidence supports the use of stochastic volatility specifications.

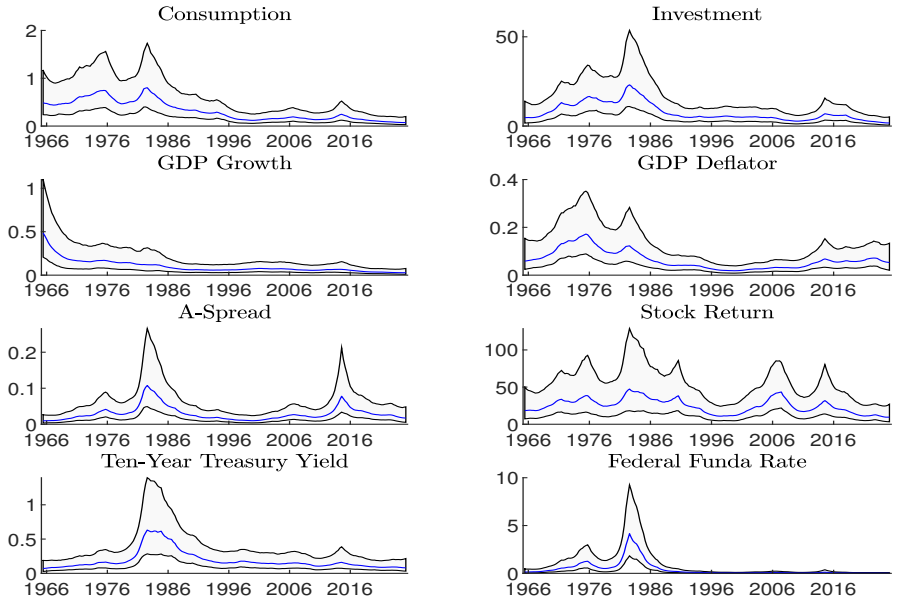

Figure 2.10: The figure depicts posterior median, 5-th and 95-th percentiles of the standard deviation of residuals of Consumption, Investment, GDP Growth, GDP Deflator, A-spread, S&P 500 Stock Returns, Ten-Year Treasury Yield, and Federal Funds Rate equations.

2.4.3 The effect of alternative financial shocks

To investigate the sensitivity of responses of real economic activity to financial shocks, we document the responses of the growth rate of GDP to a surprise increase in the A–spread, the 10–year A–spread, the 10–year B–spread, the GZ–spread, the excess bond premium, and the term spread. We consider alternative credit spreads by adding one at a time to a six-variable TVP-VAR-SV model. In other words, y_t contains only one of the credit spreads at a time and other macroeconomic and financial series.

The posterior median of the impulse responses of the growth rate of real GDP over the whole sample period appear in most cases are quite similar. For example, a positive financial shock to the A–spread, 10–year A–spread and 10–year B–spread (panel (a), (b) and (c) in Figure 2.11) seems to have a large negative impact on real GDP growth. There is a sharp drop in output growth that begins after several quarters. Figure 2.11, panels (d) and (e) display the impulse

responses of output growth that shows a persistent negative decline after the widening of the GZ–spread and the excess bond premium.

Panel (f) of Figure 2.11 shows the changing response of output growth, which generally rises on impact in response to a term spread shock. Until the early 1990s that rise is persistent. By contrast, starting in the early 1990s, the initial rise is rapidly reversed with GDP growth rising above their initial value until early 2000s.

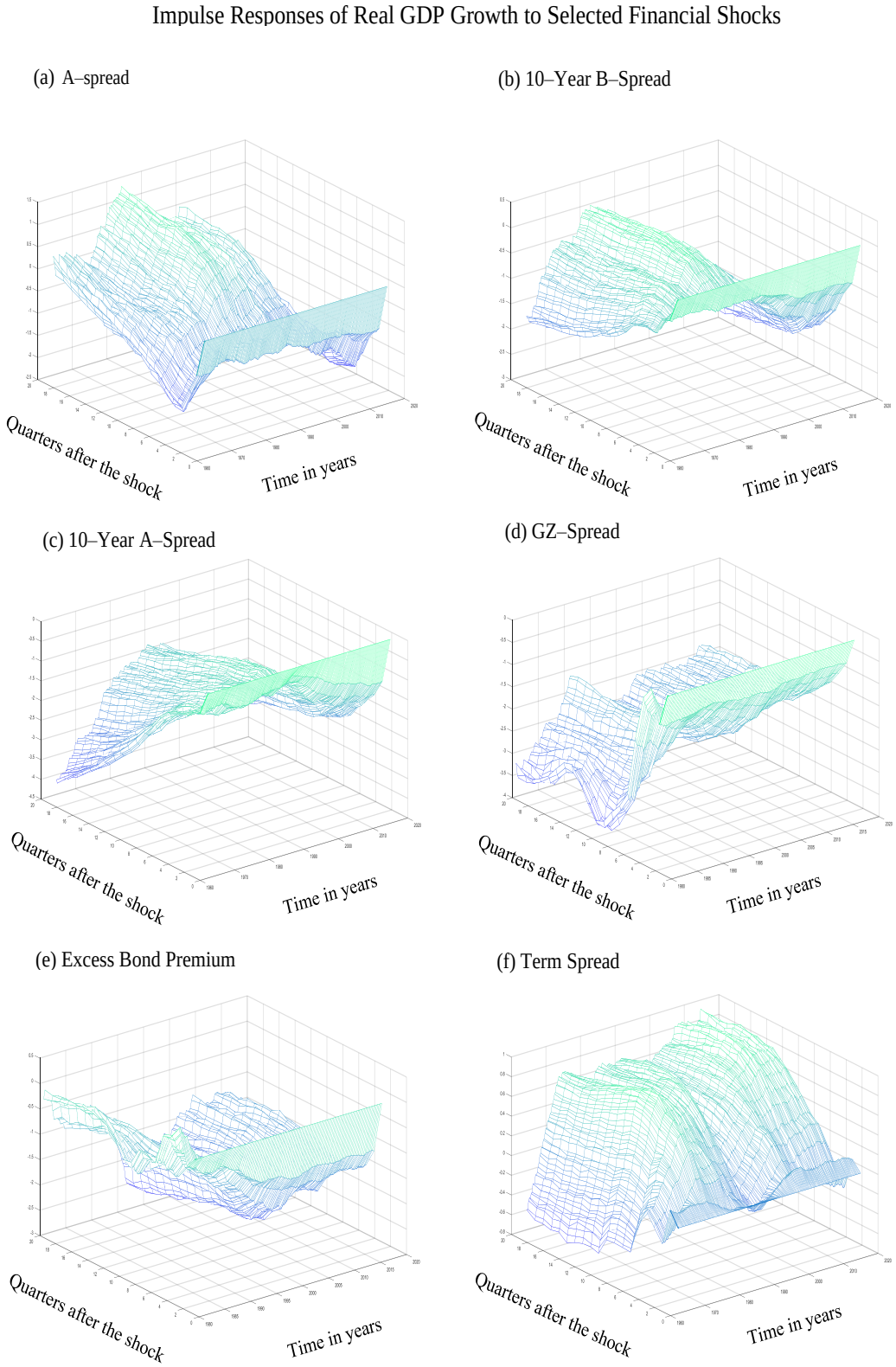

Figure 2.11: The figure shows the median of the impulse responses of real GDP growth to the identified shock of the A–Spread, 10–Year B–Spread, 10–Year A–Spread, GZ–Spread, Excess Bond Premium, and Term Spread. For each case, a six-variable structural TVP-VAR-SV model has been estimated.

2.4.4 Model comparisons

We perform a formal Bayesian model assessment by comparing alternative VAR models. We propose a discrete set of four competing models and compute Bayes factor, which is defined as a ratio of marginal likelihood of a model relative to the marginal likelihood of a second model, Chan and Eisenstat (2018a). Those four labelled models are M_1: TVP-VAR-SV model, M_2: TVP-VAR model with time-varying parameters and constant volatilities, a M_3: VAR-SV with constant parameters and time-varying volatilities, and M_4: VAR model with constant parameters and volatilities. In this case, if we compare two competing models, M_1 and M_4, the ratio of their posterior probabilities is

$$\frac{p(M_1|\mathbf{y}_t)}{p(M_4|\mathbf{y}_t)} = \frac{p(M_1)}{p(M_4)} \times \text{Bayes factor}(M_1; M_4),$$

where $p(M_1)/p(M_4)$ are the prior odds and the Bayes factor of the data is defined as $p(\mathbf{y}_t|M_1)/p(\mathbf{y}_t|M_4)$. If we set the prior odds equal to one, then the Bayes factor is equal to the posterior odds. Table 2.1 and Table 2.2 report the logarithms of marginal likelihoods and standard deviations under two alternative specifications of the TVP-VAR-SV model and other competing VAR models.[6] Firstly, the Bayes factor of the data of the TVP-VAR-SV over the constant parameter VAR in Table 2.1, is $p(\mathbf{y}_t|M_1)/p(\mathbf{y}_t|M_4) = \exp(-1744.10 - (-1980.60))$. Then, the posterior odds are $p(M_1|\mathbf{y}_t)/p(M_4|\mathbf{y}_t) = 5.1 \times 10^{102}$. From this experiment we conclude that the TVP-VAR-SV is supported by the data. However, we may ask the question whether the relative strength of the model M_1 over the model M_4 comes from allowing the time-varying parameters or the time-varying volatilities? To answer this question we compare M_2 and M_4 models in Table 2.1. We obtain a posterior odds ratio 6.3×10^{-54}, which does not favor the TVP-VAR over the VAR model. Hence, we can conclude that the support of the stochastic volatility

[6]We calculate the marginal likelihoods and the standard deviations reported in Table 2.1 and 2.2 using the computational algorithms written by Chan and Eisenstat (2018a).

seems to outperform the time-varying parameters in the VAR model fit. Those results are in line with the findings of Chan and Eisenstat (2018a) who use a three-variable alternative VAR models. Similar results follow from Table 2.2.

Logarithms of marginal likelihoods for eight-variable VAR(2) models				
Model	TVP-VAR-SV	TVP-VAR	VAR-SV	VAR
Log-ML	-1744.10	-2103.10	-1596.20	-1980.60
Standard errors	(0.61)	(0.60)	(0.14)	(0.05)

Table 2.1: The logarithms of the marginal likelihood estimates and standard deviations (in parenthesis) are calculated using 10000 posterior draws after a burn-in period of 0.1×10000 draws. The eight variables used in the estimation are consumption, investment, GDP growth, GDP deflator, A-spread, stock returns, ten-year- Treasury yield, and Federal funds rate for a period spanning from 1959:Q1 to 2018:Q1.

Logarithms of marginal likelihoods for six-variable VAR(2) models				
Model	TVP-VAR-SV	TVP-VAR	VAR-SV	VAR
Log-ML	-1382.50	-1612.00	-1322.20	-1591.90
Standard errors	(0.61)	(0.43)	(0.11)	(0.01)

Table 2.2: The logarithms of the marginal likelihood estimates and standard deviations (in parenthesis) are calculated using 10000 posterior draws after a burn-in period of 0.1×10000 draws. The six variables used in the estimation are GDP growth, GDP deflator, unemployment rate, A-spread, stock returns, and Federal funds rate for a period spanning from 1973:Q2 to 2016:Q3.

An overview of the use of Bayes factors for comparing models can be found in Koop (2003), Gelman et al. (2013), Kroese and Chan (2014) and for the use of Bayes factors with TVP-VAR-SV models see Chan and Eisenstat (2018a).

2.4.5 Model comparison using a hybrid approach

The model comparison of the previous section shows that there is a strong support for stochastic volatility. However, any conclusions of a constant parameter model can be misleading. Multiple sources of evidence show that the Bayes factors can miss out small amounts of time variation in parameters in

high-dimensional models. For instance, if most equations have constant parameters but one or two do not, Bayes factors can miss it. Hence, we go ahead with the TVP-VAR-SV model in our empirical analysis. However, note that the time variation in the impulse responses could be due to stochastic volatility and not time-varying parameters in the VAR coefficients.

The framework of Chan and Eisenstat (2018b) can be used to analyse whether some equations in our benchmark eight-variable TVP-VAR-SV model have constant parameters and some have time-varying parameters. Unlike Chan and Eisenstat (2018b) who use all subsets of models, we adapt an alternative approach. We begin with a null model containing all the constant parameter VAR equations, and then iteratively add the time-varying parameter VAR equations, one-at-a-time. We restrict the search to include only $1+N(N+1)/2 = 37$ models if $N = 8$, and so can be applied in our setting where N is relatively large to apply the full subset selection.[7]

The results of the estimates of marginal likelihoods and standard deviations are reported in Table 2.3.[8] Zero (one) binary numbers in Table 2.3 indicate constant parameters (time-varying parameters) in the relevant equations.

In general, we find that the VAR model with constant parameters and stochastic volatility performs well when compared to alternative hybrid specifications of the parameters and the full TVP-VAR-SV model. Adding one time-varying parameter specification at-a-time reveals that equations of consumption, GDP growth and stock returns performs well when time variation is allowed in parameters. On the other hand, allowing time variation in the equations of investment, inflation, A-spread, ten-year-Treasury yield, and Federal funds rate does not reveal dramatic improvements in model fit.

[7]The full subset selection of Chan and Eisenstat (2018b) leads to $2^N = 256$ with $N = 8$. It is computationally infeasible to solve the full subset selection with 256 eight-variable TVP-VAR-SV models. We leave investigating the full subset selection problem to future.

[8]We calculate the marginal likelihoods and the standard deviations reported in Table 2.3 using the computational algorithms written by Chan and Eisenstat (2018b).

When pairs of parameters are allowed to vary over time, the performance of the VAR models depends on the performance of individual parameters. For example, if a VAR model with parameters of pair of variables are allowed to vary over time performs well when compared to the VAR model with constant parameters, this must be due to the performance of each parameter individually. This can be seen from the performance of the VAR models with equations of consumption and GDP growth, consumption and stock returns, and GDP growth and stock returns when the parameters are allowed to vary over time. Those VAR models outperform the all-constant parameters VAR model.

An interesting fact of our results is that there is no evidence of support of time variation in the equation of inflation, which contrasts the findings of Chan and Eisenstat (2018b) and Karlsson and Österholm (2023).

Eight-variable hybrid TVP-VAR-SV models										
Choice	$y_{1,t}$	$y_{2,t}$	$y_{3,t}$	$y_{4,t}$	$y_{5,t}$	$y_{6,t}$	$y_{7,t}$	$y_{8,t}$	Log ML	SE
1	0	0	0	0	0	0	0	0	$-1,571.50$	(0.63)
2	1	0	0	0	0	0	0	0	$\mathbf{-1,565.50}$	(0.73)
3	0	1	0	0	0	0	0	0	$-1,578.20$	(0.41)
4	0	0	1	0	0	0	0	0	$\mathbf{-1,570.90}$	(0.56)
5	0	0	0	1	0	0	0	0	$-1,577.30$	(0.53)
6	0	0	0	0	1	0	0	0	$-1,592.20$	(0.50)
7	0	0	0	0	0	1	0	0	$\mathbf{-1,567.20}$	(0.54)
8	0	0	0	0	0	0	1	0	$-1,606.90$	(0.78)
9	0	0	0	0	0	0	0	1	$-1,627.60$	(0.48)
10	1	1	0	0	0	0	0	0	$-1,572.80$	(0.68)
11	1	0	1	0	0	0	0	0	$\mathbf{-1,566.50}$	(0.52)
12	1	0	0	1	0	0	0	0	$-1,574.80$	(0.91)
13	1	0	0	0	1	0	0	0	$-1,586.30$	(0.78)
14	1	0	0	0	0	1	0	0	$\mathbf{-1,564.10}$	(0.53)
15	1	0	0	0	0	0	1	0	$-1,602.40$	(0.47)
16	1	0	0	0	0	0	0	1	$-1,623.00$	(0.61)
17	0	1	1	0	0	0	0	0	$-1,577.10$	(0.53)
18	0	1	0	1	0	0	0	0	$-1,584.90$	(0.71)
19	0	1	0	0	1	0	0	0	$-1,599.60$	(0.71)
20	0	1	0	0	0	1	0	0	$-1,574.80$	(0.54)
21	0	1	0	0	0	0	1	0	$-1,614.20$	(0.85)
22	0	1	0	0	0	0	0	1	$-1,634.80$	(0.35)
23	0	0	1	1	0	0	0	0	$-1,577.00$	(0.59)
24	0	0	1	0	1	0	0	0	$-1,590.50$	(0.56)
25	0	0	1	0	0	1	0	0	$\mathbf{-1,567.60}$	(0.42)
26	0	0	1	0	0	0	1	0	$-1,605.70$	(0.55)
27	0	0	1	0	0	0	0	1	$-1,626.40$	(0.42)
28	0	0	0	1	1	0	0	0	$-1,599.40$	(0.85)
29	0	0	0	1	0	1	0	0	$-1,574.00$	(0.92)
30	0	0	0	1	0	0	1	0	$-1,614.60$	(0.47)
31	0	0	0	1	0	0	0	1	$-1,634.50$	(0.29)
32	0	0	0	0	1	1	0	0	$-1,587.40$	(0.72)
33	0	0	0	0	1	0	1	0	$-1,627.50$	(0.62)
34	0	0	0	0	1	0	0	1	$-1,648.60$	(0.54)
35	0	0	0	0	0	1	1	0	$-1,602.70$	(0.67)
36	0	0	0	0	0	1	0	1	$-1,624.20$	(0.34)
37	1	1	1	1	1	1	1	1	$-1,710.20$	(0.87)

Table 2.3: The models are estimated using 10000 posterior draws after a burn-in period of 0.1×10000 draws. The logarithms of the marginal likelihood estimates and standard errors (in parenthesis) are calculated using 10000 replications. Zero (one) stands for constant parameters (time-varying parameters) in the relevant equations. The eight variables used in the estimation are $y_{1,t}$ = consumption, $y_{2,t}$ = investment, $y_{3,t}$ = GDP growth, $y_{4,t}$ = GDP deflator, $y_{5,t}$ = A-spread, $y_{6,t}$ = stock returns, $y_{7,t}$ = ten-year-Treasury yield, and $y_{8,t}$ = Federal funds rate for a period spanning from 1959:Q1 to 2018:Q1.

2.5 Summary

We examined the nature and the evolving features of the real economic activity, the prices and the financial indicators in response to the financial shocks using several specifications of a Bayesian TVP-VAR-SV model. We identified the shocks by imposing a recursive identification scheme on the impact responses of multiple macroeconomic and financial indicators.

The evolution of the impulse responses over time for different horizons suggests that potential time variation can be detected in several cases. Financial shocks appear to having a substantial effect on the real economic activity during the Great Recession. This is supported by the associated contraction in real economic activity especially during the Great Recession as the responses to the financial shock exceed the average.

Our assumptions that allow for the time-varying vector autoregressive coefficient states, the time-varying covariance states, and the time-varying volatilities yield a more realistic picture of the effects of financial shocks on the real economic activity and provide a useful resource in understanding the evolution of transmission of these shocks into the real economy. Supported by the financial accelerator framework, the results show that a surprise increase of a measure of the external finance premium generates economically meaningful results with some detectable differences; a decrease in real consumption, real investment, and real output. However, it creates a puzzle in inflation and stock market returns.

2.6 Appendix

2.6.1 A constant coefficient VAR model

We first calibrate the prior hyperparameters of $\boldsymbol{\beta}_0$, $\boldsymbol{\alpha}_0$, and $\mathbf{h}_0$ by estimating a time invariant version of the model in Equation (2.2) based on an initial training sample of 40 observations using the method of ordinary least squares (OLS). Let's consider writing Equation (2.2) in a compact form as

$$\mathbf{y} = (\mathbf{I}_N \otimes \mathbf{X})\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.12}$$

where we define $\mathbf{y}$ as an $NT \times 1$ vector, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)'$ as a $T \times L$ matrix by setting $L = (1+Np)$ and rearranging $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})$, $\boldsymbol{\beta} = \text{vec}([\mathbf{B}_0, \ldots, \mathbf{B}_p]')$ as an $LN \times 1$ vector, $\boldsymbol{\epsilon}$ is an $(NT \times 1)$ vector of reduced form disturbances assumed to be a zero-mean process with covariance matrix $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\epsilon \otimes \mathbf{I}_T)$. We may write the covariance matrix as $\boldsymbol{\Sigma}_\epsilon = \mathbf{A}^{-1}\mathbf{D}\mathbf{A}'^{-1}$.

The OLS estimates and the covariance matrix of the slope coefficient vector may be obtained by $\hat{\boldsymbol{\beta}}_0 = [I_N \otimes (X'X)^{-1}X']y$ and $\hat{\mathbf{P}}_{\hat{\beta}_0} = \boldsymbol{\Sigma}_0 \otimes (\mathbf{X}'\mathbf{X})^{-1}$. The OLS estimates of the elements of $\hat{\boldsymbol{\alpha}}_0$ are obtained by regressing $\hat{\epsilon}_{i,t}$ on $-\hat{\boldsymbol{\epsilon}}_{1:i-1,t}$ and obtaining the estimated variance $\hat{\mathbf{P}}_{\hat{\alpha}_0}$. Finally, the elements of $\hat{\mathbf{h}}_0$ are generated by the logarithm of the diagonal elements of $\boldsymbol{\Sigma}_0^{1/2} = \mathbf{D}^{1/2}$ and the implied variances are calculated as $\mathbf{I}_N$.

2.6.2 The FFBS algorithm

Bayesian inference for the state space representation in Equations (2.2), (2.3), (2.6), and (2.7) are based on the joint posterior density $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\theta}|\mathbf{y})$ of all states $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_T)$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_T)$, and $\mathbf{h} = (\mathbf{h}_0, \ldots, \mathbf{h}_T)$, and the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\Sigma}_\beta, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_h)$. We estimate the model by simulating the distribution of states and hyperparameters. We proceed by applying a particular Gibbs Sampler

to the system in Equation (2.3) obtained as

$$p(\boldsymbol{\beta}_{1:T}|\boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{1:T}, \boldsymbol{\theta}) = p(\boldsymbol{\beta}_T|\boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{1:T}, \boldsymbol{\theta}) \times \prod_{t=1}^{T-1} p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1}, \boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{1:t}, \boldsymbol{\theta}). \quad (2.13)$$

Equation (2.13) provides an efficient way to sample $\boldsymbol{\beta}_{1:T}$ for a linear Normal state space model with forward filtering and backward sampling (FFBS) algorithm that was proposed by Carter and Kohn (1994) and named as the FFBS algorithm by Frühwirth-Schnatter (1994). The FFBS algorithm samples the whole state vector $\boldsymbol{\beta}_{1:T}$ from the joint probability density given the sample $\mathbf{y}_{1:T}$ and the relevant hyperparameters in $\boldsymbol{\theta}$.

Forward filtering

1. Sampling $\boldsymbol{\beta}_t$:

The Kalman filter consists of the following equations which proceed sequentially in time starting from initial values $\hat{\boldsymbol{\beta}}_0$ and $\hat{\mathbf{P}}_{\hat{\beta}_0}$. Consider the posterior density of the state $\boldsymbol{\beta}_{t-1}$ given the information up to time $t-1$ is

$$\boldsymbol{\beta}_{t-1}|\boldsymbol{\alpha}_{t-1}, \mathbf{h}_{t-1}, \mathbf{y}_{1:t-1}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}).$$

Then the following statements hold

1. The one-step-ahead predictive distribution of $\boldsymbol{\beta}_t$ conditional on $\boldsymbol{\alpha}_t$, $\mathbf{h}_t$, and $\mathbf{y}_{1:t-1}$ is Normal with parameters

$$\mathbf{a}_t = E(\boldsymbol{\beta}_t|\boldsymbol{\alpha}_t, \mathbf{h}_t, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}) = \mathbf{m}_{t-1}, \quad (2.14)$$

$$\mathbf{R}_t = V(\boldsymbol{\beta}_t|\boldsymbol{\alpha}_t, \mathbf{h}_t, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}) = \mathbf{C}_{t-1} + \boldsymbol{\Sigma}_\beta. \quad (2.15)$$

2. The one-step-ahead predictive distribution of $\mathbf{y}_t$ conditional on $\mathbf{y}_{1:t-1}$ is

Normal with parameters

$$\mathbf{f}_t = E(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \mathbf{x}_t\mathbf{a}_t, \tag{2.16}$$

$$\mathbf{Q}_t = V(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \mathbf{x}_t\mathbf{R}_t\mathbf{x}_t' + \mathbf{\Sigma}_{\epsilon,t}. \tag{2.17}$$

3. The filtering distribution of $\boldsymbol{\beta}_t$ conditional on $\boldsymbol{\alpha}_t$, $\mathbf{h}_t$, and $\mathbf{y}_{1:t}$ is Normal, with parameters

$$\mathbf{m}_t = E(\boldsymbol{\beta}_t|\boldsymbol{\alpha}_t, \mathbf{h}_t, \mathbf{y}_{1:t}, \boldsymbol{\theta}) = \mathbf{a}_t + \mathbf{K}_t(\mathbf{y}_t - \mathbf{x}_t\mathbf{a}_t), \tag{2.18}$$

$$\mathbf{C}_t = V(\boldsymbol{\beta}_t|\boldsymbol{\alpha}_t, \mathbf{h}_t, \mathbf{y}_{1:t}, \boldsymbol{\theta}) = \mathbf{R}_t - \mathbf{K}_t\mathbf{x}_t\mathbf{R}_t, \tag{2.19}$$

where $\mathbf{K}_t = \mathbf{R}_t\mathbf{x}_t'\mathbf{Q}_t^{-1}$ is Kalman gain.

Backward smoothing

At the end of sampling, the forward filtering delivers the mean and the variance for $\boldsymbol{\beta}_T$ of the form

$$p(\boldsymbol{\beta}_T|\boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{1:T}, \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{m}_T, \mathbf{C}_T).$$

The backward sampling step computes the remaining terms in Equation (2.13). Because the state space model in Equations (2.2), (2.3) are linear and Normal, the distribution of $\boldsymbol{\beta}_T$ given $\mathbf{y}_{1:T}$ and that of $\boldsymbol{\beta}_t$ given $\boldsymbol{\beta}_{t+1}$ and $\mathbf{y}_{1:t}$ for $t = T-1, \ldots, 1$ are also Normal. Conditional on generating $\boldsymbol{\beta}_T$, we can compute $\boldsymbol{\beta}_t$ from

$$p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1}, \boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \mathbf{y}_{1:T}, \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{g}_t, \mathbf{G}_t),$$

with moments

$$\begin{aligned} \mathbf{g}_t &= \mathbf{m}_t + \mathbf{B}_t(\beta_{t+1} - \mathbf{m}_t), \\ \mathbf{G}_t &= \mathbf{C}_t - \mathbf{B}_t\mathbf{R}_{t+1}\mathbf{B}_t', \end{aligned} \tag{2.20}$$

where $\mathbf{B}_t = \mathbf{C}_t\mathbf{R}_{t+1}^{-1}$.[9]

2. Sampling $\boldsymbol{\alpha}_t$:

To sample $\boldsymbol{\alpha}_t$ elements of the lower triangular matrix $\mathbf{A}_t$, let's denote

$$\mathbf{A}_t^{-1}(\mathbf{y}_t - \mathbf{x}\boldsymbol{\beta}_t) = \mathbf{A}_t^{-1}\hat{\mathbf{y}}_t = \mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t. \tag{2.21}$$

Then, we may write the following equation for each row of a lower triangular matrix for $i = 2 : N$

$$\hat{y}_{i,t} = -\hat{\mathbf{y}}_{1:i-1,t}\boldsymbol{\alpha}_{i,t} + \sigma_{i,t}\epsilon_{i,t}, \tag{2.22}$$

where $\sigma_{i,t}$ and $\epsilon_{i,t}$ represent the ith elements of $\boldsymbol{\sigma}_t$ and $\boldsymbol{\epsilon}_t$, respectively and $\hat{\mathbf{y}}_{1:i-1,t} = (\hat{y}_{1,t}, \ldots, \hat{y}_{i-1,t})$. Conditional on $\boldsymbol{\beta}_T$ and $\mathbf{h}_T$, Equation (2.22), is the measurement equation of a state space model where the state is defined in Equation (2.6) with components of the state $\boldsymbol{\alpha}_{i,t}$. We apply the FFBS algorithm and draw $\boldsymbol{\alpha}_{i,t}$ from $N(\boldsymbol{\alpha}_{i,t|t+1}, \mathbf{V}^{\alpha}_{i,t|t+1})$, where the smoothing mean and the variance are denoted as

$$\begin{aligned} \boldsymbol{\alpha}_{i,t|t+1} &= E(\boldsymbol{\alpha}_{i,t}|\boldsymbol{\alpha}_{i,t+1}, \mathbf{y}_{1:t}, \boldsymbol{\beta}_{1:T}, \mathbf{h}_{1:T}, \boldsymbol{\theta}), \\ \mathbf{V}^{\alpha}_{i,t|t+1} &= V(\boldsymbol{\alpha}_{i,t}|\boldsymbol{\alpha}_{i,t+1}, \mathbf{y}_{1:t}, \boldsymbol{\beta}_{1:T}, \mathbf{h}_{1:T}, \boldsymbol{\theta}). \end{aligned} \tag{2.23}$$

3. Sampling $\mathbf{h}_t$:

Sampling the volatilities is not straightforward as the stochastic volatility is a nonlinear state space model. The challenge arises because the joint conditional density of $p(\mathbf{h}_{1:T}|\mathbf{y}_{1:t}, \boldsymbol{\beta}_T, \boldsymbol{\alpha}_T, \boldsymbol{\Sigma}_\beta, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_h)$ is not Normal. We use auxiliary mixture sampling algorithm to estimate the stcohastic volatility. This estimation method was developed by Kim et al. (1998) and was implemented by Primiceri (2005), Galí and Gambetti (2015), and for a book length see Kroese and Chan (2014). Let's start implementing this approach by transforming $\mathbf{y}_t$ in order to

[9]For details, see West and Harrison (1997) Chapter 15, Kim and Nelson (1999) Chapter 8 and Blake and Mumtaz (2012) Chapter 3.

obtain a measurement that is linear in $\mathbf{h}_t$

$$\mathbf{y}_t^* = (\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\beta}_t) = e^{\mathbf{h}_t}\boldsymbol{\upsilon}_t, \tag{2.24}$$

where $e^{\mathbf{h}_t}\boldsymbol{\upsilon}_t = \boldsymbol{\eta}_t^h$ and $\boldsymbol{\upsilon}_t \sim \mathcal{N}(0,1)$. Squaring both sides of Equation (2.24) and taking the logarithm, we obtain the following form

$$\mathbf{y}_t^{**} = 2\mathbf{h}_t + \boldsymbol{\upsilon}_t^*, \tag{2.25}$$

where $\mathbf{y}_t^{**} = \log \mathbf{y}_t^{*2}$, $\mathbf{h}_t = \log \boldsymbol{\sigma}_t$ and $\boldsymbol{\upsilon}_t^* = \log \boldsymbol{\upsilon}_t^2$. Hence, Equation (2.25) together with Equation (2.7) represent a state space model. However, the issue with the above representation is that the error term $\boldsymbol{\upsilon}_t^*$ does not have a Normal distribution. The auxiliary mixture sampling finds a Normal mixture that approximates the pdf of the $\boldsymbol{\upsilon}_t^*$

$$f(\boldsymbol{\upsilon}_t^*) \approx \sum_{i=1}^{n} w_i \phi(\boldsymbol{\upsilon}_t^*; \boldsymbol{\mu}_{h,i}, \boldsymbol{\sigma}_{h,i}^2), \tag{2.26}$$

where $\phi(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ is the Normal density with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$, w_i are the mixture probabilities for the ith component and n is the number of components. We can define an auxiliary random variable $\mathbf{s}_{h,t} \in \{1, \ldots, n\}$, which can be used as a mixture component indicator

$$\begin{aligned}(\boldsymbol{\upsilon}_t^* | \mathbf{s}_{h,t} = i) &\sim \mathcal{N}(\boldsymbol{\mu}_{h,i} - 1.2704, \boldsymbol{\sigma}_{h,i}^2),\\ \Pr(\mathbf{s}_{h,t} = i) &= w_i.\end{aligned}$$

Hence, the desired linear Normal model can be defined conditional on the component indicator $\mathbf{s}_{h,t}$ as $(\boldsymbol{\upsilon}_t^* | \mathbf{s}_{h,t}) \sim \mathcal{N}(\mathbf{d}_t^*, \boldsymbol{\Sigma}_t^*)$, where $\mathbf{d}_t^* = \boldsymbol{\mu}_{h,i} - 1.2704$ and $\boldsymbol{\Sigma}_t^* = \boldsymbol{\sigma}_{h,i}^2$ for $t = 1 : T$. $\boldsymbol{\mu}_{h,i}$ and $\boldsymbol{\sigma}_{h,i}$ have fixed values represented as a seven

component Normal mixture in Table 4 of Kim et al. (1998). Then it follows that

$$(\mathbf{y}_t^{**}|\mathbf{s}_{h,t}, \mathbf{h}_t) \sim \mathcal{N}(2\mathbf{h}_t + \mathbf{d}_t^*, \boldsymbol{\Sigma}_t^*), \tag{2.27}$$

for $t = 1 : T$. Then, we can use the FFBS algorithm and draw $\mathbf{h}_t$ from $N(\mathbf{h}_{t|t+1}, \mathbf{V}_{t|t+1}^h)$, where the smoothing mean and the variance are denoted as

$$\begin{aligned} \mathbf{h}_{t|t+1} &= E(\mathbf{h}_t|\mathbf{h}_{t+1}, \mathbf{y}_{1:t}, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \mathbf{s}_{h,1:T}, \boldsymbol{\theta}), \\ \mathbf{V}_{t|t+1}^h &= V(\mathbf{h}_t|\mathbf{h}_{t+1}, \mathbf{y}_{1:t}, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \mathbf{s}_{h,1:T}, \boldsymbol{\theta}). \end{aligned} \tag{2.28}$$

4. Sampling $\mathbf{s}_{h,t}$

We can sample $\mathbf{s}_{h,t}$ as follows

$$f(\mathbf{s}_h|\mathbf{y}^{**}, \mathbf{h}) = \prod_{t=1}^{T} f(\mathbf{s}_{h,t}|\mathbf{y}_t^{**}, \mathbf{h}_t), \tag{2.29}$$

where each $\mathbf{s}_{h,t}$ can be sampled independently for $t = 1 : T$. We can compute $\Pr(\mathbf{s}_{h,t} = i|\mathbf{y}_t^{**}, \mathbf{h}_t)$ for $i = 1 : 7$ as

$$\Pr(\mathbf{s}_{h,t} = i|\mathbf{y}_t^{**}, \mathbf{h}_t) = \frac{w_i \phi(\mathbf{y}_t^{**}; 2\mathbf{h}_t + \boldsymbol{\mu}_{h,i} - 1.2704, \boldsymbol{\sigma}_{h,i}^2)}{\sum_{i=1}^{7} w_i \phi(\mathbf{y}_t^{**}; 2\mathbf{h}_t + \boldsymbol{\mu}_{h,i} - 1.2704, \boldsymbol{\sigma}_{h,i}^2)}. \tag{2.30}$$

5. Sampling $\boldsymbol{\Sigma}_\beta$

Sampling the hyperparameter $\boldsymbol{\Sigma}_\beta$ as

$$p(\boldsymbol{\Sigma}_\beta|\mathbf{y}_{1:t}, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_h) \sim \mathcal{IW}\big(\mathbf{U}_{\beta_0} + \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{x}_t'\boldsymbol{\beta}_t)(\mathbf{y}_t - \mathbf{x}_t'\boldsymbol{\beta}_t)', \nu_\beta + T\big).$$

6. Sampling $\boldsymbol{\Sigma}_\alpha$

Sampling the hyperparameter $\boldsymbol{\Sigma}_\alpha$ blockwise $\mathbf{S}$ for $n = 1 : N - 1$

$$p(\mathbf{S}|\mathbf{y}_{1:t}, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \boldsymbol{\Sigma}_\beta, \boldsymbol{\Sigma}_h) \sim \mathcal{IW}\big(\mathbf{U}_{\alpha_0} + \sum_{t=1}^{T} (\boldsymbol{\alpha}_{i,j,t} - \boldsymbol{\alpha}_{i,j,t-1})(\boldsymbol{\alpha}_{i,j,t} - \boldsymbol{\alpha}_{i,j,t-1})', \nu_\alpha + T\big).$$

7. Sampling $\boldsymbol{\Sigma}_h$

Sampling the hyperparameter $\boldsymbol{\Sigma}_h$ as

$$p(\boldsymbol{\Sigma}_h|\mathbf{y}_{1:t}, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_h) \sim \mathcal{IW}\big(\mathbf{U}_{h_0} + \sum_{t=1}^{T}(\mathbf{h}_t - \mathbf{h}_{t-1})(\mathbf{h}_t - \mathbf{h}_{t-1})', \nu_h + T\big).$$

2.6.3 Data appendix

The quarterly time series variables used in the TVP-VAR-SV models are taken from the FRED database of the Federal Reserve Bank of St Louis spanning from 1959Q1 to 2018Q1. The columns of Table 2.4, denote the series numbers, Tcode denotes the data transformations based on McCracken and Ng (2020) and Gilchrist and Zakrajšek (2012), series denotes the FRED mnemonic except for GZS and EBP, and description denotes a brief definition of the series.

A–Spread is defined as Moody's BAA corporate bond yield minus Moody's AAA corporate bond yield, the 10–year B–spread is Moody's BAA corporate bond yield minus 10–year Treasury bond yield, the 10–year A–spread is Moody's AAA corporate bond yield minus 10–year Treasury bond yield, the GZ–spread and excess bond premium are the spreads of Gilchrist and Zakrajšek (2012), and the Term Spread is defined as the difference between the 10–year Treasury bond yield and the 3–month Treasury bill as suggested by Stock and Watson (2003). The time span for the variables, GZ–spread, excess bond premium, and term spread, starts in 1973:Q2 and ends in 2016:Q3.

The modified Tcode, 1^*, stands for no transformation of the series.

Time series used in the TVP-VAR-SV models			
ID	Series	Tcode	Description
1	PCECC96	5	Real Personal Consumption Expenditures
2	GPDIC1	5	Real Gross Domestic Product
3	GDPC1	5	Real Gross Private Domestic Investment
4	GDPCTPI	6	Gross Domestic Product: Chain-type Price Index
5	AAA	1*	Moody's Seasoned AAA Corporate Bond Yield
6	BAA	1*	Moody's Seasoned BAA Corporate Bond Yield
7	S&P 500	5	S&P's Common Stock Price Index: Composite
8	GS10	1*	10-Year Treasury Constant Maturity Rate
9	FEDFUNDS	1*	Effective Federal Funds Rate
10	UNRATE	2	Civilian Unemployment Rate
11	TB3MS	1*	3-Month Treasury Bill
12	GZS	1*	GZ-spread
13	EBP	1*	Excess bond premium

Table 2.4: The quarterly time series variables used in the TVP-VAR-SV models.

Chapter 3

Bayesian pairwise composite likelihood method for large vector autoregressive models with time-varying parameters and stochastic volatility

3.1 Introduction

A vector autoregressive model with time-varying parameters and stochastic volatility (TVP-VAR-SV model) is a popular member of a family of time series models in examining the complexity of macroeconomic phenomena. Estimation of this beautiful model is typically carried out by using a conventional Bayesian approach based on small datasets in a class of models that have been used in a range of papers, including Cogley and Sargent (2005), Primiceri (2005), Koop et al. (2009), Koop and Korobilis (2010), Galí and Gambetti (2015), Prieto et al. (2016), Gambetti and Musso (2017) and Chan and Eisenstat (2018a). Estimating models with large number of variables is challenging since they include a large number of parameters relative to the number of observations. When estimating high-dimensional complex models with large datasets, the full likelihood function and the full posterior density may become impractical to compute. Our primary objective is to resolve the computational complexity of the TVP-VAR-SV models

using large datasets in a parsimonious way.

We contribute to this field by developing an innovative Bayesian inferential framework to address the high-dimensional inference problem posed by the TVP-VAR-SV model, which features parameter changes, many dependent variables, and substantial computational complexity. Although the full likelihood is analytically known, its evaluation becomes impractical in high dimensions. To overcome this challenge, we construct a *Pairwise Composite Likelihood* (PCL) by combining pairwise likelihood functions, each based on bivariate densities from a collection of bivariate TVP-VAR-SV models. These bivariate models are estimated independently via Gibbs sampler, enabling parallel computation and scalability.

A central innovation of this chapter is the introduction of the *Direct Averaging Method* (DAM), which systematically aggregates parameter draws from the bivariate posteriors into estimates for the joint pairwise composite model. This is achieved through a structured mapping that accounts for the multiple appearances of each variable across the bivariate models via a matrix-weighting scheme. The resulting approximation effectively replaces the intractable joint likelihood with a product of lower-dimensional components, while preserving key dynamic features of the data. A critical research question addressed by DAM is ensuring that the aggregation of pairwise contributions yields accurate and interpretable pairwise composite estimates, while maintaining computational tractability in high-dimensional settings.

Composite likelihood methods share some theoretical properties with the theory of misspecified models. Thus, a wealth of literature has proposed many composite likelihoods under various different names.[1] All of these can be seen as special cases of the composite likelihoods defined by Lindsay (1988). He defined

[1]Those are pseudo likelihood by Besag (1974) for spatial data and Verbeke and Molenberghs (2005) for longitudinal data, partial likelihood by Cox (1975) for proportional hazards model, the mth-order likelihood by Azzalini (1983), pairwise likelihood by Cox and Reid (2004), split-data likelihood by Vandekerkhove (2005), and quasi-likelihood by Hjort and Varin (2008) and Pakel et al. (2014).

composite likelihood as a likelihood obtained by the sum of low-dimensional log likelihood functions. Those component likelihoods are based on either marginal or conditional density functions.

Bayesian estimation using composite likelihood methods is an area of research that has been partially explored. Examples include Pauli et al. (2011), Ribatet et al. (2012), Friel (2012), Roche (2016), Chan et al. (2020) and Canova and Matthes (2021).

We evaluate the empirical performance of the Bayesian pairwise composite likelihood method by studying the changing dynamics in a TVP-VAR-SV model with key U.S. macroeconomic variables. In particular, we study the macroeconomic consequences of shocks to the corporate bond spread by adding the corporate bond spread to a pairwise composite likelihood TVP-VAR-SV (PCL-TVP-VAR-SV) model that includes $N = 50$ endogenous variables. The framework provides a practical way to make inference about the joint fifty-variable PCL-TVP-VAR-SV model combined from a set of $\binom{N}{2} = 1225$ bivariate models.[2] Time-varying pairwise composite impulse responses show the effects of a surprise increase in the corporate bond spread on fifty variables, with sensible and economically meaningful results.

The remainder of the chapter is structured as follows. Section 3.2 presents a specialized version of the pairwise composite likelihood approach tailored for time series analysis. Section 3.3 outlines both componentwise and joint modelling strategies within the TVP-VAR-SV framework. Section 3.4 introduces the proposed Direct Averaging Method for aggregating pairwise estimates into pairwise composite parameter structures. Section 3.5 reports empirical findings based on pairwise composite impulse response analysis. Section 3.6 concludes with a summary of the main results and their implications. The Appendix provides additional algorithmic details, extended empirical results, and a data description.

[2]The $\binom{N}{k} = \frac{N!}{k!(N-k)!}$ denotes the binomial coefficient for counting the number of k combinations of a set with N distinct cross-sectional variables, where $k = 2$ and $N \geq k$.

3.2 Bayesian pairwise composite likelihood

We use a class of pairwise composite likelihood approach formed by pairwise likelihoods constructed from bivariate marginal densities. More precisely, we adapt a special case of pairwise likelihood methods proposed by Verbeke and Molenberghs (2005) and Fieuws and Verbeke (2006) in a novel way to be applied for the first time in the context of TVP-VAR-SV models.

3.2.1 Pairwise composite model

Let $\mathbf{y}_t$ denote a $N \times 1$ vector of time series variables. For each cross-sectional variable $j = 1 : N$, $t = 1 : T$ denotes the time series observations. Suppose the system consists of multiple state parameters $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t, \boldsymbol{\alpha}_t, \mathbf{h}_t)'$ that are indexed by time and that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T)'$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T)'$, and $\mathbf{h} = (\mathbf{h}_1, \ldots, \mathbf{h}_T)'$ on a parameter space $\boldsymbol{\Theta}$. A key assumption is that an observation density conditional on state parameters $p(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\beta}_t, \boldsymbol{\alpha}_t, \mathbf{h}_t)$ is correctly specified for all $t = 1, \ldots, T$. The vector $\mathbf{x}_t$ contains lagged values of $\mathbf{y}_t$ with $k = 1 : p$ and p defining the order of the VAR(p) model, and the vectors of parameters, $\boldsymbol{\beta}_t$, $\boldsymbol{\alpha}_t$, and $\mathbf{h}_t$, appear in the density for every time t. Then, the full likelihood function of this system may be obtained as

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}; \mathbf{y}, \mathbf{x}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\beta}_t, \boldsymbol{\alpha}_t, \mathbf{h}_t),$$

for the contribution of all variables to the full joint likelihood.

Assuming the joint probability density function $p(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\beta}_t, \boldsymbol{\alpha}_t, \mathbf{h}_t)$ is correctly specified, suppose it is a challenge to specify the full N dimensional distribution at time t but that it is possible to specify all two dimensional distributions. Let a combination of the N dimensional random vectors taken k at a time, be any subset of k elements. Assuming $k = 2$, take 2 elements at a time and randomly select $\binom{N}{2}$ all 'pairs of data' in the cross section. Thus, each pair of y_i and y_j is an $(N_i \times T)$ matrix, with $N_i = 2$ and tth elements $y_{i,t}$ and $y_{j,t}$. Let

$\mathbf{y}_t = \{(y_{1,t}, y_{2,t})', (y_{1,t}, y_{3,t})', \ldots, (y_{N-1,t}, y_{N,t})'\}$ denote a set of all pairs of variables to be modelled componentwise, with $i = 1, \ldots, N-1$ and $j = (i+1), \ldots, N$. To simplify notation, we may write $\mathbf{y}_{i,t}$ for a pair of variables with $i = 1, \ldots, M$, where $M = \binom{N}{2}$. For instance, $\mathbf{y}_{1,t} = (y_{1,t}, y_{2,t})'$. Then, the pairwise composite posterior density can be obtained as

$$\begin{aligned} p(\boldsymbol{\beta}_c, \boldsymbol{\alpha}_c, \mathbf{h}_c|\mathbf{y}) \propto & \left(\prod_{i=1}^{M} \prod_{t=1}^{T} p(\mathbf{y}_{i,t}|\boldsymbol{\beta}_{i,t}, \boldsymbol{\alpha}_{i,t}, \mathbf{h}_{i,t}) \right) \\ & \times \left(\prod_{i=1}^{M} p(\boldsymbol{\beta}_{i,0}) \prod_{t=1}^{T} p(\boldsymbol{\beta}_{i,t}|\boldsymbol{\beta}_{i,t-1}) \right) \\ & \times \left(\prod_{i=1}^{M} p(\boldsymbol{\alpha}_{i,0}|\boldsymbol{\beta}_{i,0}) \prod_{t=1}^{T} p(\boldsymbol{\alpha}_{i,t}|\boldsymbol{\alpha}_{i,t-1}, \boldsymbol{\beta}_{i,t}) \right) \\ & \times \left(\prod_{i=1}^{M} p(\mathbf{h}_{i,0}|\boldsymbol{\beta}_{i,0}, \boldsymbol{\alpha}_{i,0}) \prod_{t=1}^{T} p(\mathbf{h}_t|\mathbf{h}_{i,t-1}, \boldsymbol{\beta}_{i,t}, \boldsymbol{\alpha}_{i,t}) \right), \end{aligned} \tag{3.1}$$

where the pairwise composite likelihood constructed from bivariate marginal densities is a function of $\boldsymbol{\beta}_c$, $\boldsymbol{\alpha}_c$, and $\mathbf{h}_c$ defined as

$$L^c(\boldsymbol{\beta}_c, \boldsymbol{\alpha}_c, \mathbf{h}_c; \mathbf{y}) = \prod_{i=1}^{M} \prod_{t=1}^{T} p(\mathbf{y}_{i,t}; \boldsymbol{\beta}_{i,t}, \boldsymbol{\alpha}_{i,t}, \mathbf{h}_{i,t}), \tag{3.2}$$

and the composite log likelihood

$$\ell^c(\boldsymbol{\beta}_c, \boldsymbol{\alpha}_c, \mathbf{h}_c; \mathbf{y}) = \sum_{i=1}^{M} \sum_{t=1}^{T} \log p(\mathbf{y}_{i,t}; \boldsymbol{\beta}_{i,t}, \boldsymbol{\alpha}_{i,t}, \mathbf{h}_{i,t}), \tag{3.3}$$

where we drop the conditioning argument $\mathbf{x}_t$ for notational simplicity, and $\boldsymbol{\beta}_{c,t} = (\boldsymbol{\beta}_{1,2,t}, \ldots, \boldsymbol{\beta}_{N-1,N,t})'$, $\boldsymbol{\alpha}_{c,t} = (\boldsymbol{\alpha}_{1,2,t}, \ldots, \boldsymbol{\alpha}_{N-1,N,t})'$, and $\mathbf{h}_{c,t} = (\mathbf{h}_{1,2,t}, \ldots, \mathbf{h}_{N-1,N,t})'$ are the vectors combining all the M pair-specific parameter vectors. Pair-specific parameters $\boldsymbol{\beta}_{i,t}$, $\boldsymbol{\alpha}_{i,t}$, and $\mathbf{h}_{i,t}$ are assumed to be subsets of $\boldsymbol{\beta}_{c,t}$, $\boldsymbol{\alpha}_{c,t}$, and $\mathbf{h}_{c,t}$, respectively, related to the bivariate distribution of $\mathbf{Y}_{i,t} = (Y_{i,t}, Y_{j,t})'$.

3.2.2 Mapping bivariate parameters to pairwise composite parameters

One complication of the pairwise composite likelihood method is the uneven frequency with which parameters appear. Each dependent variable appears in $N-1$ bivariate models, and thus certain parameters occur only once, while others appear $N-1$ times.

The stacked parameters $\boldsymbol{\beta}_{c,t}$, $\boldsymbol{\alpha}_{c,t}$, and $\mathbf{h}_{c,t}$ obtained from the bivariate models are not, in general, equal to the pairwise composite parameters $\boldsymbol{\beta}_t$, $\boldsymbol{\alpha}_t$, and $\mathbf{h}_t$. Elements of $\boldsymbol{\beta}_t$ may correspond to multiple entries in $\boldsymbol{\beta}_{c,t}$, depending on how often they are identified across submodels. Elements of $\boldsymbol{\alpha}_t$, in contrast, are often uniquely identified from a single submodel. The log-volatilities $\mathbf{h}_t$ generally appear multiple times across submodels due to the repeated inclusion of individual series.

To construct the full set of time-varying pairwise composite parameters, we define a linear mapping from the bivariate parameter vector to the pairwise composite parameter vector

$$\boldsymbol{\theta}_t = \boldsymbol{\Omega}\, \boldsymbol{\theta}_{c,t}, \tag{3.4}$$

where $\boldsymbol{\theta}_{c,t}$ collects all parameters from the bivariate models, and $\boldsymbol{\Omega}$ is a block-diagonal matrix that encodes the appropriate selection and averaging scheme to construct joint pairwise composite parameters.

This weighting scheme differs from traditional Bayesian model averaging, where weights are probabilistic and sum to one. Instead, the weights here are deterministic, inspired by the approach of Verbeke and Molenberghs (2005), reflecting the combinatorial structure of pairwise decomposition in the PCL-TVP-VAR-SV model. While our weighting scheme results in weights that sum to one within each $(N-1)$-dimensional marginal structure, it fundamentally differs from probabilistic model averaging. The weights are not derived from posterior model

probabilities, but rather from the deterministic frequency with which each parameter appears across the M bivariate models. This distinction is crucial as our approach does not attempt to average over model uncertainty in the Bayesian sense, but rather to correct for structural overlap inherent in the pairwise composite likelihood framework.

3.2.3 Theoretical review

Several key questions arise: first, how should we treat asymptotics of the posterior distribution of the pairwise composite parameters? Is a pairwise composite posterior estimate (the posterior mean) asymptotically approximate a pairwise composite maximum likelihood estimate? How can we overcome the problem that there is a one-to-many matches between the joint and the pairwise composite parameters? And how can we deal with the fact that the parameter space $\boldsymbol{\Theta}$ is not necessarily finite dimensional? To answer these questions, we need to describe the inferential validity of the pairwise composite likelihood approach from a Bayesian perspective. Currently, a general theory to the Bayesian inference in time series models using the pairwise composite posterior distribution is not available, so a careful treatment of every problem is essential.

The validity of the pairwise composite posterior distribution can be based on asymptotic results. When performing Bayesian analysis, Central Limit Theorem (C.L.T.) can justify approximations for large sample sizes. A rather general behaviour of the posterior distribution can be analysed through the C.L.T. and the asymptotic theory. We prefer to discuss concepts in a somewhat informal way and motivate key results.

It was shown by Kent (1982), Lindsay (1988), White (1994) and Verbeke and Molenberghs (2005) that some general results on consistency and asymptotic normality of composite likelihood estimate exist for fixed N and $T \to \infty$ to obtain

the limiting distribution

$$\sqrt{T}\big[\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}_c\big] \xrightarrow{d} \mathcal{N}(0, \mathbf{G}^{-1}(\boldsymbol{\theta}_c)), \tag{3.5}$$

where $\hat{\boldsymbol{\theta}}_c$ is the maximum composite likelihood estimate (MCLE) and $\mathbf{G}(\boldsymbol{\theta}_c) = \mathbf{H}^{-1}(\boldsymbol{\theta}_c)\mathbf{J}(\boldsymbol{\theta}_c)\mathbf{H}'^{-1}(\boldsymbol{\theta}_c)$ is the Godambe information matrix, Godambe (1960), or the so-called sandwich information matrix. $\mathbf{H}(\boldsymbol{\theta}_c) = E[-\nabla u_c(\boldsymbol{\theta}_c, \mathbf{y})]$ is defined as sensitivity matrix and $\mathbf{J}(\boldsymbol{\theta}_c) = var[u_c(\boldsymbol{\theta}_c, \mathbf{y})]$ is defined as variability matrix with $u_c(\boldsymbol{\theta}_c, \mathbf{y})$ is the pairwise composite score function. For a detailed discussion, see Varin (2008) and Varin et al. (2011). A natural question is whether a pairwise composite posterior summary is asymptotically close to above frequentist summary. Under appropriate regularity conditions, the C.L.T. of probability theory can be described in a Bayesian language to show the pairwise composite posterior distribution

$$\sqrt{T}\big[E(\boldsymbol{\theta}_c|\mathbf{y}) - \boldsymbol{\theta}_c\big] \xrightarrow{d} \mathcal{N}(0, \mathbf{G}^{-1}(\boldsymbol{\theta}_c)), \tag{3.6}$$

where $E(\boldsymbol{\theta}_c|y)$ is the mean of the pairwise composite posterior. This property shows that as the sample size increases, the Bayesian pairwise composite posterior estimate is approximately the same as the MCLE. Wakefield (2013) adds that if model misspecification is present, its effect on both inferential tools is the same. However, this strong statement should be used with caution. There is an effect of the prior in these results; the prior may not vanish, for a discussion, see DasGupta (2008), Chapter 20. As Wakefield (2013) notes, "While sandwich estimation can be used to correct the variance estimate for the maximum likelihood estimate, there is no such simple solution for the posterior mean, or other Bayesian summaries." Previous attempts at Bayesian composite likelihood include using a constant weight so as to best adjust the composite posterior variance matrix to the asymptotic variance matrix of the MCLE Pauli et al. (2011), or performing an adjustment method Ribatet et al. (2012). Both of these methods share the

fact that they are applied within a Markov chain Mote Carlo (MCMC) algorithm. Canova and Matthes (2021) follow the approach developed by Ribatet et al. (2012) and conclude that it produces estimates with good finite sample properties.

Note, however, that in those examples of the composite likelihood method, there is a one-to-one parameterisation in the different parts of the composite likelihood function, whereas in our example the set of parameters in $\boldsymbol{\theta}_c$ is considered to be a combination of pair-specific parameters. Another drawback of the approaches in Pauli et al. (2011) and Ribatet et al. (2012) is that they do not assume the weights of different components of the composite likelihood function sum to one which contrasts with the Bayesian notion. To overcome the problem that there is a one-to-many matches between the joint parameters $\boldsymbol{\theta}$ and the pairwise composite parameters $\boldsymbol{\theta}_c$, we follow Verbeke and Molenberghs (2005), chapter 25 and obtain a single estimate of $\boldsymbol{\theta}$ by averaging all the matching pair-specific parameters in $\boldsymbol{\theta}_c$. Then, the asymptotic distribution can be written as

$$\sqrt{T}\big[E(\boldsymbol{\theta}|\mathbf{y}) - \boldsymbol{\theta}\big] \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Omega}\mathbf{G}^{-1}(\boldsymbol{\theta}_c)\boldsymbol{\Omega}'). \tag{3.7}$$

Hence, under very mild regularity conditions, we assume that the basic asymptotic theory of an approximate pairwise composite posterior distribution may hold with an asymptotic multivariate Normal distribution for parameters using a general framework of the composite likelihood (under the term pseudo likelihood) theory presented in Verbeke and Molenberghs (2005), Chapters 9. An additional condition, in the context of time series, is that the pairwise composite posterior mean converges to $\boldsymbol{\theta}$ at a usual rate, assuming the data are stationary time series as described by Cox and Reid (2004).

In this setup, we deal with a very large number of unknown parameters. To approach this problem, we follows DasGupta (2008), Chapter 34, and assume that a parameter space $\boldsymbol{\Theta}$ in finite-dimension is available and the dimension

depends on T. We may let $\boldsymbol{\Theta} = \boldsymbol{\Theta}_t \to \infty$ as $T \to \infty$, where T is the time series dimension. Our interest is to answer the question of how to achieve consistency and asymptotic normality of the pairwise composite posterior when the number of parameters goes to infinity with the number of observations. One theoretical question left for future study is that in an infinite dimensional parameter space, we need to know whether the pairwise composite posterior converges near the truth. To be precise, we must ensure that the pairwise composite posterior mean is consistent, that is, it converges to the true value of the parameters when the number of parameters increases with the number of observations. We also need to establish asymptotic properties of the approximate pairwise composite posterior, that is, the approximate pairwise composite posterior distribution, under some appropriate regularity conditions for time series models, converges asymptotically to the Normal distribution.

It is important to clarify the role of the pairwise composite likelihood approach in the proposed methodology. The estimates obtained from individual bivariate models are not assumed to recover the parameters of the unrestricted joint model. Rather, the composite likelihood approach uses these marginal contributions as building blocks in a pseudo-likelihood that approximates the full likelihood. Under the assumption that the full joint model is correctly specified, each bivariate submodel corresponds to a valid marginal distribution, and the resulting composite likelihood estimator targets the true joint parameter vector (Lindsay (1988) and Varin (2008)). When this assumption does not hold, the estimator converges to a pseudo-true value that best approximates the joint model under the composite likelihood criterion. We acknowledge that this procedure does not yield the unrestricted joint density or an unrestricted VAR specification but provides a practical and theoretically grounded approximation that remains computationally feasible in high-dimensional settings.

3.3 Bivariate and composite models

In this section we outline the bivariate TVP-VAR-SV models and the pairwise composite PCL-TVP-VAR-SV model. We fit all bivariate models containing all the M pair-specific parameters to construct the joint PCL-TVP-VAR-SV model.

3.3.1 The bivariate models

For each component model $i = 1, \ldots, M$, the ith bivariate VAR process can be written in a form of measurement equation of a state space model as follows

$$\mathbf{y}_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_{i,t} + \boldsymbol{\epsilon}_{i,t}, \tag{3.8}$$

and $\boldsymbol{\beta}_{i,t}$ is the ith component state evolution equation following a random walk process as

$$\boldsymbol{\beta}_{i,t} = \boldsymbol{\beta}_{i,t-1} + \boldsymbol{\eta}^{\beta}_{i,t}, \tag{3.9}$$

where $\mathbf{y}_{i,t}$ is an $N_i \times 1$ vector, $\mathbf{x}_{i,t} = \mathbf{I}_{N_i} \otimes (1, \mathbf{y}'_{i,t-1}, \ldots, \mathbf{y}'_{i,t-p})$ is an $N_i \times L_i N_i$ matrix, $\boldsymbol{\epsilon}_{i,t}$ is an $N_i \times 1$ vector of reduced form disturbances assumed to be a zero-mean process with time-varying covariance matrix $\boldsymbol{\epsilon}_{i,t} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{i,t})$ and $\boldsymbol{\eta}^{\beta}_{i,t} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{i,\beta})$. $\boldsymbol{\beta}_{i,t} = \text{vec}([\mathbf{B}_{i,0,t}, \ldots, \mathbf{B}_{i,p,t}]')$ is an $L_i N_i \times 1$ vector of pair-specific parameters, with the model order p is being preselected for $k = 0, \ldots, p$ and $L_i = (1 + N_i p)$. Let's define $\boldsymbol{\Sigma}_{i,t}$ as a positive definite symmetric $N_i \times N_i$ matrix, which has the following representation

$$\boldsymbol{\Sigma}_{i,t} = \mathbf{A}_{i,t}\mathbf{D}_{i,t}\mathbf{A}'_{i,t}. \tag{3.10}$$

To establish a unique solution to the system in Equation (3.10), we further assume $\mathbf{A}_{i,t}$ is a lower triangular matrix with elements $\alpha_{ji,t}$ below the diagonal and 1s along the principal diagonal, and $\mathbf{D}_{i,t}$ is a diagonal matrix. In this framework, volatility is captured by the time-varying standard deviations of the reduced-form

disturbances, governed by the square roots of the diagonal elements of $\mathbf{D}_{i,t}$, which result from the time variation in the transformation applied to the structural shocks. Let $\boldsymbol{\sigma}_{i,t}$ denote the vector of standard deviations, i.e., the square roots of the diagonal elements of $\mathbf{D}_{i,t}$. Hence, the elements of the time-varying covariance matrix are specified as random walks

$$\begin{aligned} \boldsymbol{\alpha}_{i,t} &= \boldsymbol{\alpha}_{i,t-1} + \boldsymbol{\eta}^{\alpha}_{i,t}, \quad \text{and} \\ \mathbf{h}_{i,t} &= \mathbf{h}_{i,t-1} + \boldsymbol{\eta}^{h}_{i,t}, \end{aligned} \tag{3.11}$$

where $\mathbf{h}_{i,t} = \log \boldsymbol{\sigma}_{i,t}$, $E(\boldsymbol{\eta}^{\alpha}_{i,t}) = 0$ and $E(\boldsymbol{\eta}^{h}_{i,t}) = 0$, $E(\boldsymbol{\eta}^{\alpha}_{i,t}\boldsymbol{\eta}'^{\alpha}_{i,t}) = \boldsymbol{\Sigma}_{i,\alpha}$ and $E(\boldsymbol{\eta}^{h}_{i,t}\boldsymbol{\eta}'^{h}_{i,t}) = \boldsymbol{\Sigma}_{i,h}$, and $\boldsymbol{\eta}^{\alpha}_{i,t} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{i,\alpha})$ and $\boldsymbol{\eta}^{h}_{i,t} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{i,h})$.

3.3.2 The pairwise composite model

Let's define the *pairwise composite parameters* as $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t, \boldsymbol{\alpha}_t, \mathbf{h}_t, \boldsymbol{\Sigma}_{\beta}, \boldsymbol{\Sigma}_{\alpha}, \boldsymbol{\Sigma}_{h})$, and the *stacked bivariate parameters* as $\boldsymbol{\theta}_{c,t} = \big(\boldsymbol{\beta}_{c,t}, \boldsymbol{\alpha}_{c,t}, \mathbf{h}_{c,t}, \boldsymbol{\Sigma}_{c,\beta}, \boldsymbol{\Sigma}_{c,\alpha}, \boldsymbol{\Sigma}_{c,h}\big)$.

The pairwise composite model of interest, the PCL-TVP-VAR-SV, is given by

$$\mathbf{y}_t = \mathbf{x}_t\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \tag{3.12}$$

where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t)$, and the pairwise composite time-varying covariance matrix is decomposed as

$$\boldsymbol{\Sigma}_t = \mathbf{A}_t\mathbf{D}_t\left(\mathbf{A}_t\right)'. \tag{3.13}$$

The parameter vector $\boldsymbol{\theta}_t$ consists of time-varying VAR coefficients $\boldsymbol{\beta}_t$, the nonzero elements of the lower-triangular contemporaneous matrix $\mathbf{A}_t$, and log-volatilities $\mathbf{h}_t$. These quantities are not estimated jointly but instead are reconstructed from M bivariate models using the pairwise composite likelihood method. This approach creates the stacked parameter vectors $\boldsymbol{\beta}_{c,t}, \boldsymbol{\alpha}_{c,t}, \mathbf{h}_{c,t}$. To reconcile the stacked bivariate parameters into a coherent set of pairwise composite parameters, we use the linear mapping as described in Section 3.2.2.

3.3.3 Identification and pairwise composite impulse responses

We identify the pairwise composite (joint) structural shocks $\mathbf{u}_t$ from the pairwise composite reduced form disturbances $\boldsymbol{\epsilon}_t$ by assuming a linear transformation $\boldsymbol{\epsilon}_t \equiv \mathbf{P}_t\mathbf{u}_t$, where $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{I}_N$ and $E(\mathbf{u}_t\mathbf{u}_{t-k}') = 0$ for all t and $k = 1, 2, 3, \ldots$. Let's define $\boldsymbol{\Sigma}_t$ as a positive definite symmetric $N \times N$ matrix, which has the representation of Equation (3.13). To establish a solution to the system in Equation (3.13), we define $\mathbf{P}_t$ as an $N \times N$ lower triangular matrix with positive main diagonal elements such that $\mathbf{P}_t\mathbf{P}_t' = \boldsymbol{\Sigma}_t$, where $\mathbf{P}_t$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_t$. Then, $\mathbf{P}_t = \mathbf{A}_t\mathbf{D}_t^{1/2}$.

We may carry out dynamic impulse response analysis based on the pairwise composite likelihood method. A central question in the analysis of pairwise composite structural impulse responses is that how much sensible the results are when compared to the results from a conventional estimation approach? We try to measure the expected dynamic responses of future realisations of $\mathbf{y}_{t+h}$ for $h = 1, \ldots, H$ to a pairwise composite structural shock at time t. Then, we may write the joint model in Equation (3.12) in companion form following Kilian and Lütkepohl (2017), and Galí and Gambetti (2015) as

$$\mathbf{y}_t^* = \boldsymbol{\mu}_t^* + \mathbf{B}_t^*\mathbf{y}_{t-1}^* + \boldsymbol{\epsilon}_t^*,$$

where $\mathbf{y}_t^* \equiv (\mathbf{y}_t', \ldots, \mathbf{y}_{t-p+1}')'$, $\boldsymbol{\mu}_t^* \equiv (\mathbf{B}_{0,t}', 0, \ldots, 0)'$, $\boldsymbol{\epsilon}_t^* \equiv (\boldsymbol{\epsilon}_t', 0, \ldots, 0)'$, and the companion matrix

$$\mathbf{B}_t^* \equiv \begin{pmatrix} \mathbf{B}_t & \\ \mathbf{I}_{N(p-1)} & 0_{N(p-1),N} \end{pmatrix}.$$

Then, a dynamic response to a time t shock is denoted as

$$\frac{\partial \mathbf{y}_{t+h}}{\partial \boldsymbol{\epsilon}_t'} = \left[\mathbf{B}_t^{*h}\right]_{N,N} \equiv \mathbf{Z}_{t,h},$$

for $h = 1, \ldots, H$ and $[\mathbf{B}_t^{*h}]_{N,N}$ denotes the first $N \times N$ matrix and $\mathbf{Z}_{t,0} \equiv \mathbf{I}$. Our interest is in the dynamic responses of each element in $\mathbf{y}_t$ to a financial shock (impulse response) in $\mathbf{u}_t^f$ at time t,

$$\frac{\partial \mathbf{y}_{t+h}}{\partial \mathbf{u}_t^f} = \frac{\partial \mathbf{y}_{t+h}}{\partial \boldsymbol{\epsilon}_t'} \frac{\partial \boldsymbol{\epsilon}_t}{\partial \mathbf{u}_t^f} \equiv \mathbf{Z}_{t,h} \mathbf{P}_t^{(40)} \equiv \mathbf{C}_{t,h},$$

for $h = 1, \ldots, H$ and $\mathbf{P}_t^{(40)}$ represents the fortieth column of $\mathbf{P}_t$ assuming the financial variable is ordered on the fortieth column of $\mathbf{y}_t$.

Normalisation of the pairwise composite impulse responses

Since the stochastic volatility introduces time variation in the magnitude of the financial shocks, the raw impulse responses $\mathbf{C}_{h,t}$ reflect not only the transmission of shocks but also the time-varying standard deviation of the identified shock. To ensure comparability over time, we apply a normalisation step, given by

$$\mathbf{C}_{h,t}^{\text{norm}} = \frac{\mathbf{C}_{h,t}}{\mathbf{C}_{0,t}^{(j)}},$$

where $\mathbf{C}_{0,t}^{(j)}$ is the scalar contemporaneous impact of the identified shock on the system (i.e., the first element in the impulse response function, which reflects the direct effect of the shock at time t). This is not an element-wise division in the typical sense (i.e., not a Hadamard element-wise division where each element of a matrix is divided by a corresponding element in another matrix). Instead, it is a column-wise normalisation where each element of the impulse response vector is divided by a single scalar value.

Following Galí and Gambetti (2015), we normalise the impulse responses by dividing them by the contemporaneous effect of the identified financial shock. This ensures that the responses are expressed in terms of a unit shock rather than a time-varying standard deviation shock. While this normalisation removes the direct influence of stochastic volatility on the scale of the impulse responses,

the stochastic volatility component still influences the dynamics of the model through time-varying coefficients and the covariance matrix. This normalisation step ensures that the impulse responses are expressed consistently, making it easier to visually interpret the dynamic effects over time, especially in the 3D graphs.

3.3.4 Priors and initial values used in the empirical study

To complete the model specifications, consider independent priors over $i = 1, \ldots, M$ component bivariate state parameters and hyperparameters $\boldsymbol{\beta}_{i,0}$, $\boldsymbol{\alpha}_{i,0}$, $\mathbf{h}_{i,0}$, $\boldsymbol{\Sigma}_{i,\beta}$, $\boldsymbol{\Sigma}_{i,\alpha}$ and $\boldsymbol{\Sigma}_{i,h}$. To calibrate the priors for the state parameters, we estimate M time invariant VAR models on an initial training sample of size $T_{i,0} = 40$. The priors for the initial component states $\boldsymbol{\beta}_{i,0}$, $\boldsymbol{\alpha}_{i,0}$ and $\mathbf{h}_{i,0}$ are Normal with parameters $(\boldsymbol{\mu}_{i,0}, \boldsymbol{\Sigma}_{i,0})$ and are defined as

$$\begin{aligned}
\boldsymbol{\beta}_{i,0} &\sim \mathcal{N}(\hat{\boldsymbol{\beta}}_{i,0}, \tau_\beta \times \hat{\mathbf{V}}_{\beta_{i,0}}), \\
\boldsymbol{\alpha}_{i,0} &\sim \mathcal{N}(\hat{\boldsymbol{\alpha}}_{i,0}, \tau_\alpha \times \hat{\mathbf{V}}_{\alpha_{i,0}}), \\
\mathbf{h}_{i,0} &\sim \mathcal{N}(\hat{\mathbf{h}}_{i,0}, \tau_h \times \hat{\mathbf{V}}_{h_{i,0}}),
\end{aligned}$$

where τ_β, τ_α, and τ_h are scale factors. Furthermore, we assume that the prior hyperparameters of the component covariance matrices for the state equations follow inverse Wishart with parameters $(\mathbf{U}_{i,0}, \nu_i)$ for $\boldsymbol{\Sigma}_{\beta_{i,0}}$, $\boldsymbol{\Sigma}_{\alpha_{i,0}}$ and $\boldsymbol{\Sigma}_{h_{i,0}}$

$$\begin{aligned}
\boldsymbol{\Sigma}_{\beta_{i,0}} &\sim \mathcal{IW}_{N_i(N_ip+1)}(\kappa_{i,\beta} \times (N_i(N_ip+1)+1) \times \hat{\mathbf{V}}_{\beta_{i,0}}, (N_i(N_ip+1)+1)), \\
\boldsymbol{\Sigma}_{\alpha_{i,0}} &\sim \mathcal{IW}_{N_i(N_i-1)/2}(\kappa_{i,\alpha} \times (n_i+1) \times \hat{\mathbf{V}}_{\alpha_{i,0}}, (n_i+1)), \\
\boldsymbol{\Sigma}_{h_{i,0}} &\sim \mathcal{IW}_{N_i}(\kappa_{i,h} \times (N_i+1) \times \hat{\mathbf{V}}_{h_{i,0}}, (N_i+1)),
\end{aligned}$$

where ν_i is the degrees of freedom, κ_i^2 is the scale factor, n_i+1 defines the number of elements on each row of the lower triangular elements of $\mathbf{A}_{i,t}$ for $n_i = 1 : N_i - 1$.

For the empirical exercise, we set the scale factors as $\tau_\beta = N_i + 1$, $\tau_\alpha = N_i + 1$

and $\tau_h = 1$, and set the scale factors for $\kappa_{i,\beta} = 1.0 \times 10^{-3}$, $\kappa_{i,\alpha} = 1.0 \times 10^{-2}$ and $\kappa_{i,h} = 1.0 \times 10^{-2}$.

3.4 Direct Averaging Method

The model is estimated using a two-step simulation approach based on Markov chain Monte Carlo (MCMC) methods: the Gibbs sampler and a second-stage procedure we refer to as the Direct Averaging Method. The rationale for this two-step structure is that the first step, which operates on a collection of bivariate submodels, does not directly yield draws from the joint pairwise composite posterior, which is our target density. Instead, it provides draws from M bivariate posterior densities. The second step combines these draws into approximations of the target posterior density, and can be interpreted as a computationally convenient approximation to a Metropolis-Hastings (MH) algorithm in which all proposals are implicitly accepted with probability one.

In the first step, we estimate each bivariate TVP-VAR-SV model using the efficient MCMC method of Carter and Kohn (1994), implemented via forward filtering and backward sampling (FFBS) as described in Frühwirth-Schnatter (1994), and applied to this context in Primiceri (2005). Specifically, we run a seven-block Gibbs sampler for each of the M bivariate submodels in parallel, producing R posterior draws for each set of bivariate parameters $\boldsymbol{\theta}_{i,t}$ for $i = 1, \ldots, M$ submodels.

The second step aggregates the draws from all submodels to approximate moments and other summaries of the pairwise composite parameter vector $\boldsymbol{\theta}_t$. The idea is to compute each element of $\boldsymbol{\theta}_t$ by averaging over the submodel draws in which it appears. In practice, for a parameter like $\beta_{11,t}$, which only appears in submodels involving the first variable, we average only over those draws that include it yielding $R \times (N-1)$ draws instead of the full $R \times M$.

This aggregation can be interpreted as an approximation to a MH algorithm, where the submodel posteriors collectively form a proposal distribution for the joint parameter space. Crucially, in this interpretation, the proposed values are implicitly accepted with probability one. That is, instead of explicitly computing MH acceptance ratios, we treat the draws as if they were accepted and average them to estimate posterior quantities. This simplification is valid under the assumption that the bivariate submodels provide consistent marginal information about the composite parameters, and that parameter sharing across submodels is appropriately handled by the averaging scheme.

We refer to this approach as the DAM, following similar motivations in Verbeke and Molenberghs (2005) and Fieuws and Verbeke (2006). While it does not involve a formal MH acceptance step, it retains the spirit of MH by combining local draws into pairwise composite posterior approximations through a deterministic transformation.

Formally, we organize the pairwise composite parameters into three components: the time-varying VAR coefficient states $\boldsymbol{\beta}_t$, the off-diagonal elements of the Cholesky factor $\boldsymbol{\alpha}_t$, and the log-volatilities $\mathbf{h}_t$. The mapping matrix is correspondingly defined as

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_\beta & 0 & 0 \\ 0 & \boldsymbol{\Omega}_\alpha & 0 \\ 0 & 0 & \boldsymbol{\Omega}_h \end{bmatrix}. \tag{3.14}$$

The mapping matrix $\boldsymbol{\Omega}$ serves as a structured linear operator that transforms the vector of collected bivariate parameters into the pairwise composite parameter vector of the joint model. Each block of $\boldsymbol{\Omega}$ is responsible for one parameter group: $\boldsymbol{\Omega}_\beta$ maps the bivariate coefficient states to the pairwise composite coefficient states, $\boldsymbol{\Omega}_\alpha$ constructs the pairwise composite covariance states, and $\boldsymbol{\Omega}_h$ aggregates bivariate volatility states into the pairwise composite volatility states. The specific structure of each block reflects the frequency and position of variables

across bivariate models and determines how pair-specific estimates are combined under the DAM. Detailed descriptions for building these blocks are provided in subsequent section

3.4.1 Averaging coefficient states

Let's define the set of all unique bivariate combinations of the N variables as

$$\mathcal{P} = \left\{(i,j) \in \mathbb{N}^2 : 1 \leq i < j \leq N\right\}, \quad \text{with number of models} \quad M.$$

For each pair $(i,j) \in \mathcal{P}$, we estimate a bivariate VAR(p) model that includes an intercept in each equation. In the bivariate system, each equation includes $N_i p$ lag coefficients and one intercept, giving $N_i p + 1$ total parameters per equation. Therefore, the full coefficient vector for the bivariate model has dimension $K = N_i(N_i p + 1)$.

We denote the coefficient vector from the bivariate model for pair (i,j) at time t as $\boldsymbol{\beta}_t^{(i,j)} \in \mathbb{R}^K$, and stack these vectors for all pairs into a vector $\boldsymbol{\beta}_{c,t} \in \mathbb{R}^{K_c}$, where $K_c = KM$.

The objective is to map these bivariate estimates into a single, pairwise composite parameter vector representing approximately the pairwise composite N-dimensional VAR(p) system, which includes all N variables and their Np lagged values along with an intercept for each equation. The total number of coefficients in the full system is $K' = N(Np + 1)$. Let's define the pairwise composite coefficient vector as $\boldsymbol{\beta}_t \in \mathbb{R}^{K'}$.

We construct a sparse selection-and-averaging matrix $\boldsymbol{\Omega}_\beta \in \mathbb{R}^{K' \times K_c}$ such that

$$\boldsymbol{\beta}_t = \boldsymbol{\Omega}_\beta \cdot \boldsymbol{\beta}_{c,t},$$

where $\boldsymbol{\beta}_t$ contains both overlapping and uniquely identified parameters of the composite system. Since the pairwise decomposition is designed so that each pa-

rameter may appear in up to $N-1$ submodels, $\mathbf{\Omega}_\beta$ contains entries such as $\frac{1}{(N-1)}$ or 1 per row, and zeros elsewhere, depending on the frequency of occurrence. This aggregation ensures internal consistency and avoids duplication across overlapping parameter estimates.

The mapping structure $\mathbf{\Omega}_\beta$ is fixed and deterministic, depending solely on the inclusion pattern of each coefficient across bivariate models. It enforces an averaging rule that ensures the pairwise composite model parameters are identified and stable across time, based on the disaggregated bivariate estimates.

3.4.2 Averaging covariance states

In time-varying covariance decompositions, a lower-triangular Cholesky factor $\mathbf{A}_t \in \mathbb{R}^{N\times N}$ is used to capture contemporaneous dependencies. The $L = \frac{N(N-1)}{2}$ time-varying off-diagonal elements collected in the vector $\boldsymbol{\alpha}_t \in \mathbb{R}^L$. Each element $\alpha_t^{(i,j)}$ (for $i > j$) is estimated from the unique bivariate model involving variables (i, j), in the ordering implied by the Cholesky decomposition. Since each of these parameters appears only once across bivariate decompositions, no averaging is necessary. We define the composite vector $\boldsymbol{\alpha}_{c,t}$, which collects the relevant off-diagonal coefficients from all bivariate models in accordance with the recursive ordering. The mapping matrix $\mathbf{\Omega}_\alpha \in \mathbb{R}^{L\times L}$ is simply the identity

$$\boldsymbol{\alpha}_t = \mathbf{\Omega}_\alpha \cdot \boldsymbol{\alpha}_{c,t}, \quad \text{with } \mathbf{\Omega}_\alpha = \mathbf{I}_L,$$

where $\boldsymbol{\alpha}_t \in \mathbb{R}^L$ contains the uniquely identified parameters of the full system. Since each structural parameter is estimated from exactly one submodel, $\mathbf{\Omega}_\alpha$ typically contains a single '1' per row and zeros elsewhere — acting effectively as a selector matrix. This implies that the vector $\boldsymbol{\alpha}_t$ is a direct extraction from the bivariate estimates, requiring no transformation beyond consistent indexing.

3.4.3 Averaging volatility states

We now describe how the pairwise composite vector of log-volatilities $\mathbf{h}_t \in \mathbb{R}^N$ is constructed from bivariate estimates. Suppose we have N variables and consider all M bivariate models indexed by pairs $(i, j) \in \mathcal{P}$, where each model collects the log-volatilities from each bivariate covariance matrix into a vector $\mathbf{h}_{c,t} \in \mathbb{R}^{N_i M}$.

We define a weighting matrix $\mathbf{\Omega}_h \in \mathbb{R}^{N \times N_i M}$, where each row corresponds to a variable, and collects its associated log-volatility estimates across bivariate models. The entries of $\mathbf{\Omega}_h$ are binary weights scaled by $\frac{1}{N-1}$, since each log-volatility appears in exactly $N - 1$ bivariate models

$$\mathbf{h}_t = \mathbf{\Omega}_h \cdot \mathbf{h}_{c,t},$$

where $\mathbf{h}_t$ contains the overlapping log-volatility parameters of the full system. The matrix $\mathbf{\Omega}_h$ typically contains entries like $\frac{1}{(N-1)}$ per row and zeros elsewhere, depending on the number of times each log-volatility is observed.

The structure of $\mathbf{\Omega}_h$ is deterministic and depends only on the inclusion pattern of variables in the bivariate subsets. This operation performs a simple average of the available log-volatility estimates from the bivariate models to produce a coherent pairwise composite vector of time-varying log-volatilities.

The matrices $\mathbf{\Omega}_\beta$, $\mathbf{\Omega}_\alpha$, and $\mathbf{\Omega}_h$ are all highly sparse. Most of their entries are zero since each joint parameter typically appears in only a few bivariate models. This sparsity is a key computational advantage. In practice, $\mathbf{\Omega}$ can be stored and applied using sparse matrix formats, yielding significant gains in memory efficiency and computational speed, particularly for large systems (i.e., large N).

In summary, the DAM relies on three core mapping procedures that transform bivariate estimates into pairwise composite parameters. Specifically, these mappings construct: (i) the pairwise composite coefficient vector $\boldsymbol{\beta}_t$ from all bivariate

coefficient estimates, (ii) the pairwise composite vector of Cholesky parameters $\boldsymbol{\alpha}_t$ from the corresponding off-diagonal elements of the lower triangular factor, and (iii) the pairwise composite log-volatility vector $\mathbf{h}_t$ from bivariate log-volatility estimates.

For completeness and reproducibility, detailed pseudocode for these three algorithms is provided in the Appendix: Algorithm 1 maps bivariate VAR coefficients to the pairwise composite VAR coefficient vector, Algorithm 2 maps bivariate Cholesky parameters to the pairwise composite covariance structure, and Algorithm 3 maps pairwise log-volatilities to the pairwise composite log-volatility vector.

3.4.4 Advantages and considerations of the Direct Averaging Method

The DAM offers several key advantages that make it an attractive inferential framework. First and foremost, it is simple and computationally efficient. Unlike a full joint estimation method, which requires intensive sampling and optimization, direct averaging bypasses these computational burdens by leveraging existing posterior draws from bivariate models. This makes it significantly faster and easier to implement, particularly in high-dimensional models where full Bayesian inference is often infeasible.

Another major advantage is that the method provides a quick and direct point estimate for each parameter in the full joint model. Instead of requiring additional iterations, accept-reject steps, or complex sampling algorithms, this approach allows researchers to obtain a single, well-defined estimate for each parameter by averaging across bivariate posterior draws. This simplicity ensures transparency and interpretability while still capturing the key statistical relationships present in the data.

Despite these advantages, there are certain considerations to acknowledge

when using the DAM. One key consideration when using the DAM is the lack of formal uncertainty quantification. Since this approach aggregates independent bivariate posterior draws, it does not explicitly estimate the full joint posterior uncertainty. However, this limitation can be effectively addressed by computing the impulse responses of the PCL-TVP-VAR-SV model using the joint pairwise composite parameters obtained from the DAM and quantifying the uncertainty by Monte Carlo simulation method.

To assess the robustness of these estimates, we employ Monte Carlo simulation to compute joint pairwise composite impulse responses, which allows us to capture the underlying uncertainty in the dynamic relationships between variables. By repeatedly simulating the model's response to shocks using parameter draws from the averaged estimates, we can construct confidence bands around the impulse responses, providing an intuitive measure of the uncertainty in the full composite model. This approach ensures that the method remains statistically rigorous while benefiting from its computational efficiency.

Additionally, in highly nonlinear settings, where the joint posterior may exhibit complex dependencies, averaging across bivariate models without further updates could introduce potential model misspecification. However, in many practical applications—especially in macroeconomic and financial modelling, this approach remains a computationally efficient and well, supported alternative to full joint estimation. The method is grounded in established composite likelihood theory (Verbeke and Molenberghs (2005) and Fieuws and Verbeke (2006)), ensuring that the resulting estimates retain desirable statistical properties.

Overall, the DAM strikes a balance between efficiency and accuracy. It provides a fast, transparent, and computationally feasible alternative to full MH algorithm, making it particularly useful for high-dimensional models. While there are certain trade-offs in terms of uncertainty quantification and introducing model misspecification, empirical results demonstrate that it performs well in practice.

Given these strengths, this approach represents a pragmatic and well-justified alternative for parameter estimation in complex models, such as TVP-VAR-SV models.

3.5 Empirical results

In this section, we study the macroeconomic consequences of a shock to the corporate bond spread using the PCL-TVP-VAR-SV model that includes 50 macroeconomic and financial variables. Our identification strategy is inspired by that proposed by Gilchrist and Zakrajšek (2012). As extensions to their work, we consider including a larger number of variables to address the omitted variable problem of low-dimensional VAR models and allow time-varying parameters and stochastic volatility in our novel setup. The implied identification assumption allows for financial conditions to react contemporaneously to the corporate bond spread shock while the economic activities response within a period. This is an example of a recursive identification scheme that implies a recursive ordering of variables. Furthermore, we split the variables into two groups: slow moving variables, i.e. macroeconomic series, and fast moving variables, i.e. financial series. We order the variables in a similar fashion as described by Bernanke et al. (2005), Bańbura et al. (2010), Gilchrist and Zakrajšek (2012) and Stock and Watson (2016). For instance, a shock to the corporate bond spread is assumed to affect real economic activity within a period, while the financial variables can react contemporaneously to such a financial disturbance.

Our data comprises a broad set of quarterly U.S. macroeconomic and financial variables spanning from 1959:Q1 to 2018:Q2. The variables in the slow moving group can be classified into six broad categories: National Income and Product Accounts (NIPA), prices, industrial production, sales, labour market, earnings and productivity, and financial indicators representing fast moving variables can

be classified into three broad categories: interest rates, money and credit, and exchange rates.[3]

We work with a PCL-TVP-VAR-SV with 50 variables and transform the majority of variables to stationarity using the benchmark transformation code of McCracken and Ng (2020) and transform the corporate bond spread, the effective Federal funds rate and 10-Year Treasury Constant Maturity Rate as recommended by Gilchrist and Zakrajšek (2012). The Data appendix in Section 7.5 provides a complete listing and definition of the variables. The results are based on taking 10,000 draws from each component bivariate posterior. 10% of 10,000 draws from each component bivariate posterior are burn-in draws which are dropped from the analysis. The time-varying pairwise composite impulse responses are calculated using the compositional form of the joint model.

3.5.1 Time-varying pairwise composite impulse responses

Figure 3.1 through Figure 3.5 depict the pairwise composite impulse responses of all the variables in our specification to the financial shock, defined as worsening business credit conditions (identified via the widening of corporate bond spreads). An unanticipated increase in the corporate bond spread causes a decline in real economic activity, with consumption, investment, output, exports, imports, and business sector output all decline sharply as shown in Figure 3.1. On the other hand, government consumption and investment rises slightly on impact but then falls after a quarter. Real disposable personal income has a modest declining pattern after the initial impact of the shock.

Figure 3.2 shows the impact of the shock on the variables under the classification of industrial production. The industrial production falls slightly on impact, but then bottoms out within the first 10 quarters. Similarly, the effects of the

[3]Board of Governors of the Federal Reserve System defines the industrial production index as real output (percentage of real output in a base year 2012). We include two variables from the sales category, i.e. real manufacturing and trade industries sales and real retail and food services sales, deflated by core personal consumption expenditures.

shock on the variables under the classification of the labour market is provided in Figure 3.2 through Figure 3.4. The impact of the shock to the corporate bond spread on the unemployment rate is stronger starting from mid-2009.

In response to these adverse macroeconomic effects of the financial shock, the Federal Reserve eased monetary policy as shown in Figure 3.5. While the Federal funds rate is prevented from declining immediately, by assumption, it does fall in the subsequent quarters. There is ambiguity about the response of S&P500 stock market returns, which generally increase on impact. By contrast, starting in the early 2000s, the initial rise is rapidly reversed with stock market returns declining quickly below their initial value. 3-month and 6-month Treasury bill rates fall on impact and in years following the financial shock. Note that as interest rates declines, the demand for monetary aggregate M1 increases, while M2 show a modest rise.

Some of these responses, in particular those involving real economic activity and interest rates, are in line with those of Gilchrist and Zakrajšek (2012), who have assumed that macroeconomic variables could respond on impact to corporate bond spread shocks within a period, and financial variables could respond contemporaneously using constant parameter VAR model involving a small number of variables. However, although the associated slowdown in the economy lead to a decline in inflation until late 1980s, this pattern discontinues after that period with inflation rising toward zero, i.e. the period including the Great Recession. This contrasts the results of Gilchrist and Zakrajšek (2012). They make a different finding that the economic slowdown implies a continuing disinflation over time. We also compare our results to the findings of Boivin et al. (2020) who have estimated a large factor-augmented VAR model using U.S. data. They have found that price indices show modest changes on impact to a credit spread shock and a gradual decline afterwards. The findings of Prieto et al. (2016), on the other hand, tend to contrast the main results of Gilchrist and Zakrajšek

(2012) and Boivin et al. (2020), however, they make a similar finding in line to the results in this study.

To benchmark the performance of the PCL-TVP-VAR-SV model, a comparative analysis is conducted using a conventional six-variable TVP-VAR-SV model estimated on a subset of key macroeconomic and financial indicators: GDP growth, inflation, unemployment rate, corporate bond spread, stock market returns, and the Federal funds rate. This specification adopts a recursive identification scheme, under which shocks to the corporate bond spread are allowed to contemporaneously affect real economic activity and inflation, while financial variables can respond within the same period. The comparison, detailed in the Appendix, highlights the added value of the high-dimensional pairwise composite likelihood framework in capturing richer dynamics without sacrificing computational feasibility.

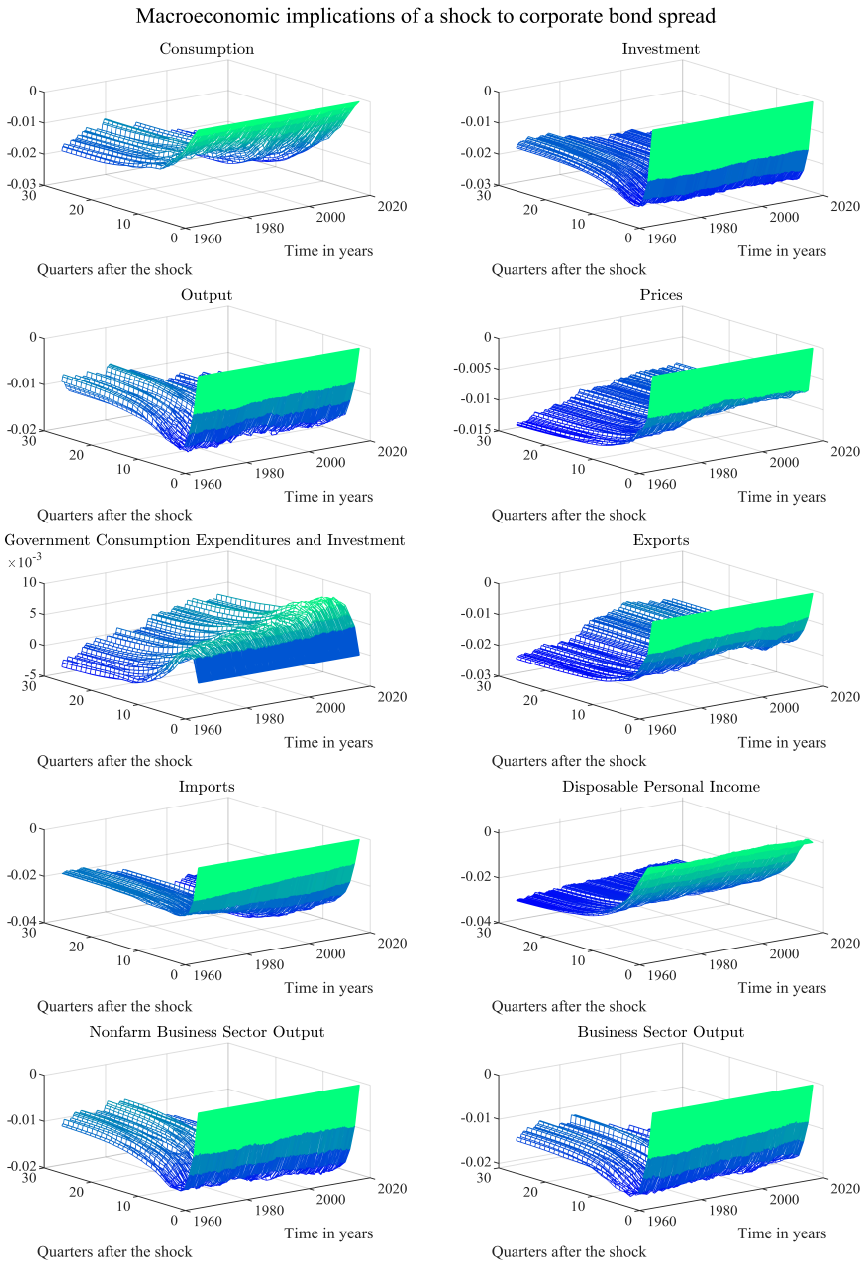

Figure 3.1: The figure depicts the pairwise composite impulse responses of Consumption, Investment, Output, Prices, Government Consumption Expenditures and Investment, Exports, Imports, Disposable Personal Income, Nonfarm Business Sector Output, and Business Sector Output to the identified shock of the Corporate Bond Spread estimated from a fifty–variable PCL-TVP-VAR-SV model.

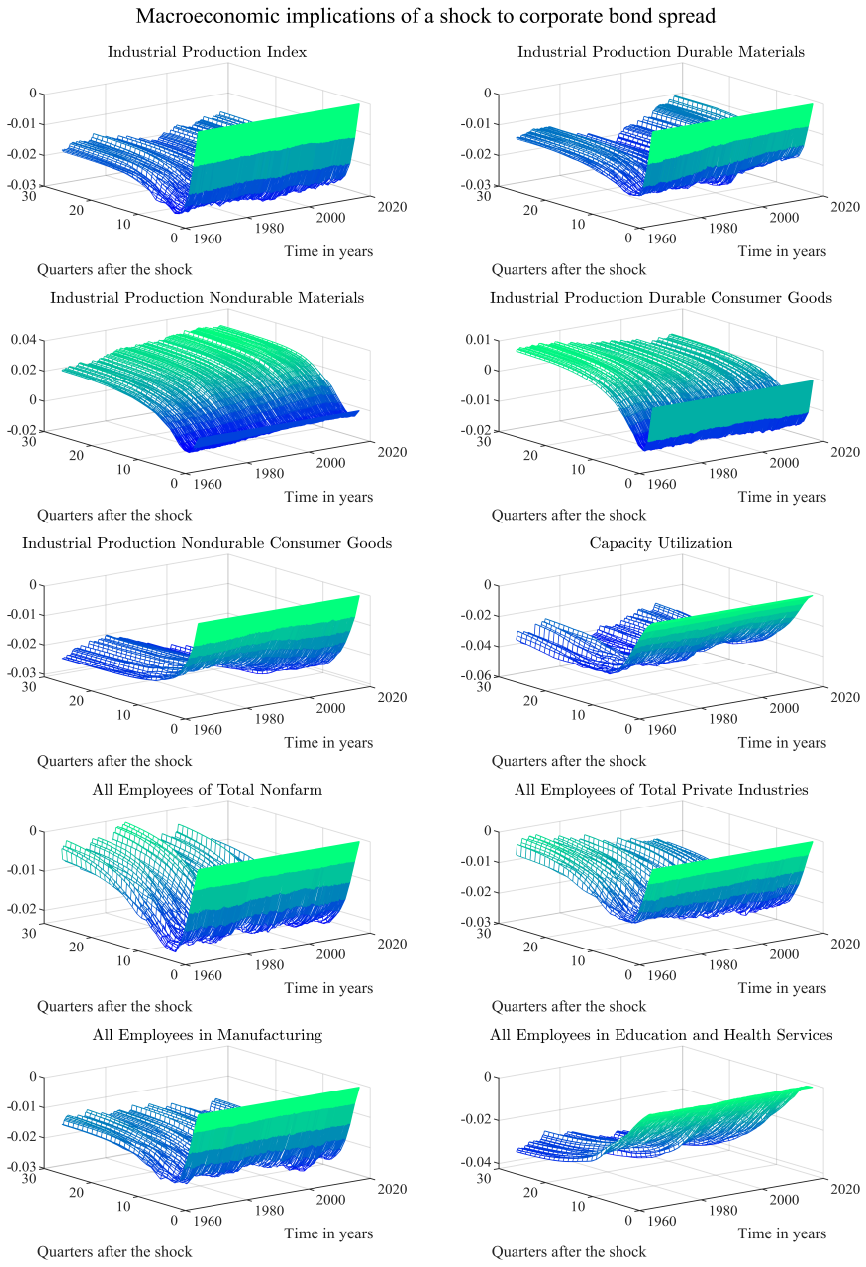

Figure 3.2: The figure depicts the pairwise composite impulse responses of Industrial Production Index, Industrial Production Durable Materials, Industrial Production Nondurable Materials, Industrial Production Durable Consumer Goods, Industrial Production Nondurable Consumer Goods, Capacity Utilization, All Employees of Total Nonfarm, All Employees of Total Private Industries, All Employees in Manufacturing, and All Employees in Education and Health Services to the identified shock of the Corporate Bond Spread estimated from a fifty–variable PCL-TVP-VAR-SV model.

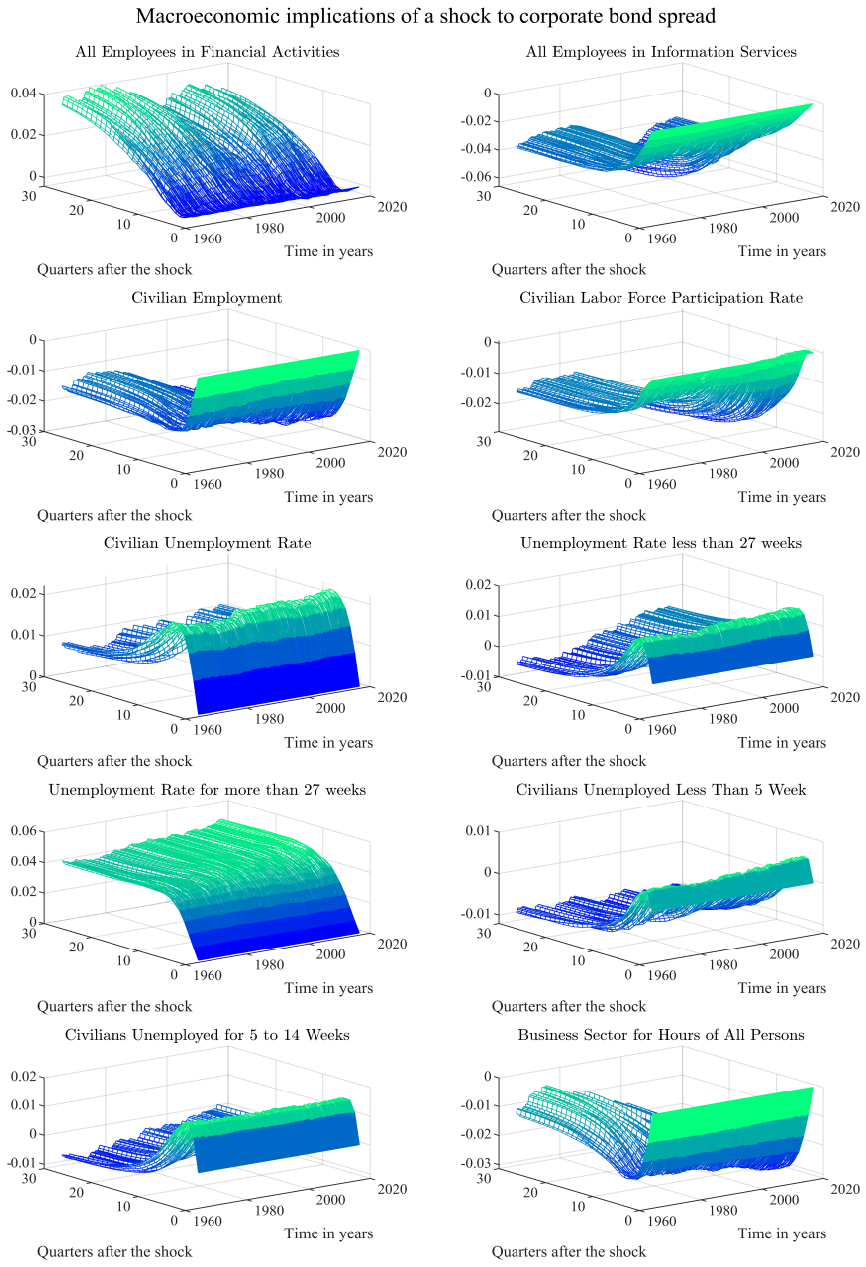

Figure 3.3: The figure depicts the pairwise composite impulse responses of All Employees in Financial Activities, All Employees in Information Services, Civilian Employment, Civilian Labor Force Participation Rate, Civilian Unemployment Rate, Unemployment Rate less than 27 weeks, Unemployment Rate for more than 27 weeks, Civilians Unemployed Less Than 5 Week, Civilians Unemployed for 5 to 14 Weeks, and Business Sector for Hours of All Persons to the identified shock of the Corporate Bond Spread estimated from a fifty–variable PCL-TVP-VAR-SV model.

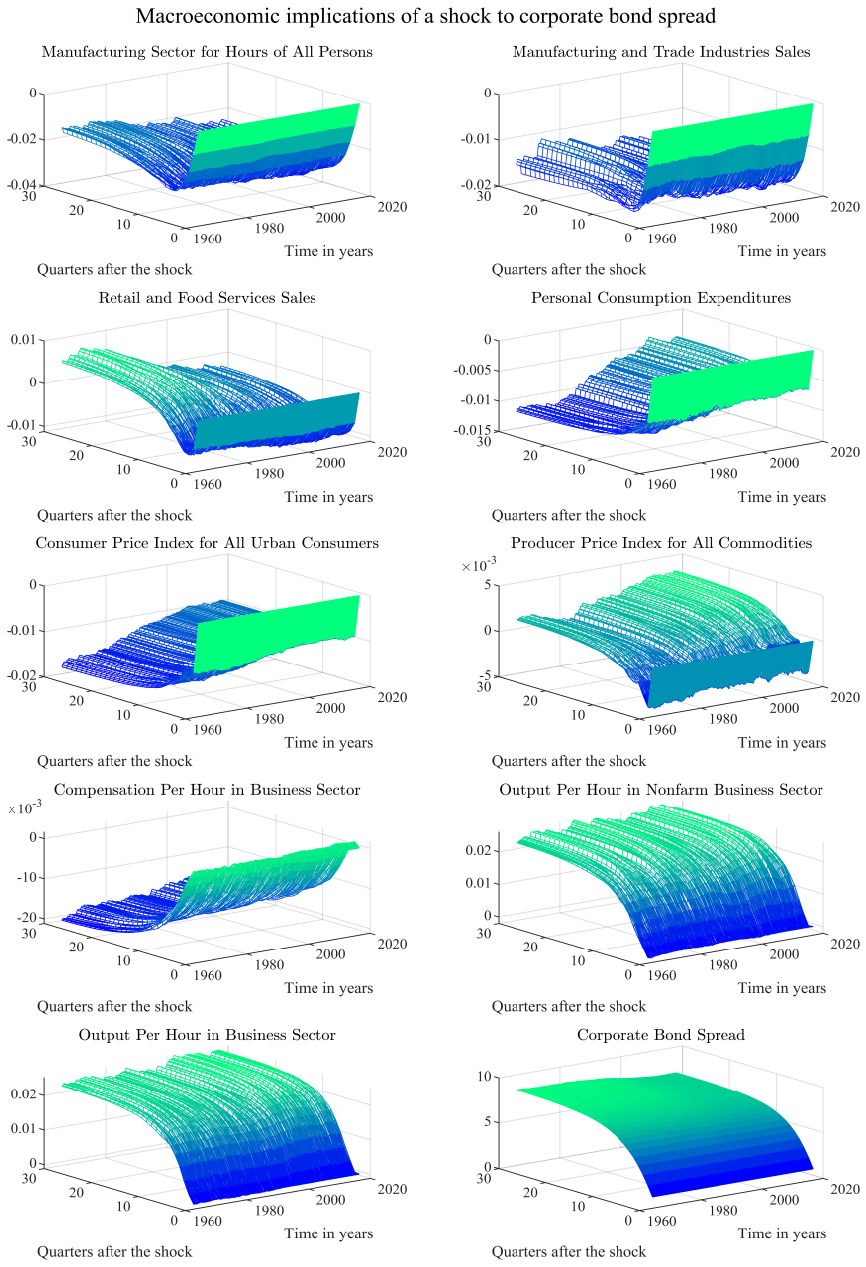

Figure 3.4: The figure depicts the pairwise composite impulse responses of Manufacturing Sector for Hours of All Persons, Manufacturing and Trade Industries Sales, Retail and Food Services Sales, Personal Consumption Expenditures, Consumer Price Index for All Urban Consumers, Producer Price Index for All Commodities, Compensation Per Hour in Business Sector, Output Per Hour in Nonfarm Business Sector, Output Per Hour in Business Sector, and Corporate Bond Spread to the identified shock of the Corporate Bond Spread estimated from a fifty–variable PCL-TVP-VAR-SV model.

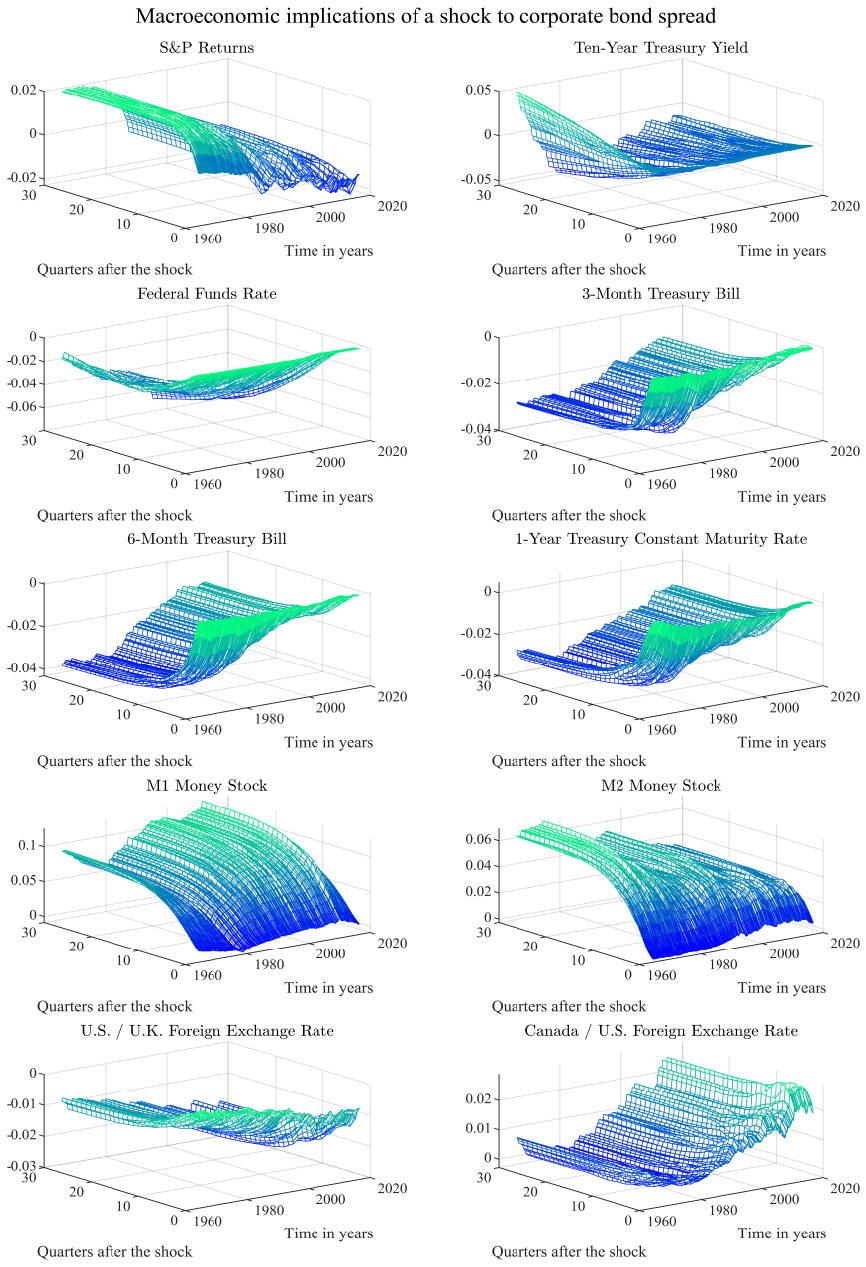

Figure 3.5: The figure depicts the pairwise composite impulse responses of S&P Returns, Ten-Year Treasury Yield, Federal Funds Rate, 3-Month Treasury Bill, 6-Month Treasury Bill, 1-Year Treasury Constant Maturity Rate, M1 Money Stock, M2 Money Stock, U.S. / U.K. Foreign Exchange Rate, and Canada / U.S. Foreign Exchange Rate to the identified shock of the Corporate Bond Spread estimated from a fifty–variable PCL-TVP-VAR-SV model.

3.5.2 Comparing impulse responses of PCL-TVP-VAR-SV and TVP-VAR-SV models

We compare the results from a PCL-TVP-VAR-SV model to those obtained from a simple approach, i.e. a TVP-VAR-SV model with a small number of variables. Hence, we consider a six-variable TVP-VAR-SV model. The six variables in this specification are, GDP growth, inflation, unemployment rate, corporate bond spread, stock market returns, and the Federal funds rate. The identifying assumption implied by this recursive ordering is that shocks to the corporate bond spread affect economic activity, inflation and unemployment rate within a period, while the stock returns and the Federal funds rate can react contemporaneously to such a financial disturbance.

Figure 3.6 and Figure 3.8 (reported in the Appendix) display the impulse responses of GDP growth, inflation, unemployment rate, corporate bond spread, S&P stock returns, and Federal funds rate to the identified shock to the corporate bond spread estimated from a six–variable PCL-TVP-VAR-SV and TVP-VAR-SV models, respectively, for a representative time 2010Q4. The date 2010:Q4 is chosen to represent a period after the Great Recession. In these figures the posterior median is shown by the red line and the blue lines are the 16th and 84th percentiles. A shock to the corporate bond spread causes a decline in the GDP growth over the next quarter. The effect of the shock leads to a fall in inflation in the short run and a rise afterwards. Unemployment rate rises on impact. The response pattern for the stock returns is also seen to rise in the short run, but ends up declining towards zero.

Figure 3.7 and Figure 3.9 depict the dynamic impulse responses of all six variables to the identified shock of the corporate bond spread estimated from a six–variable PCL-TVP-VAR-SV and TVP-VAR-SV models, respectively. Obviously, in both figures, the initial decline reverses with the inflation rising over time. Similarly, the effect of a shock to the corporate bond spread on the stock

returns is ambiguous; the initial rise is reversed with a decline for the period covering the Great Recession.

In summary, the PCL-TVP-VAR-SV model delivers sensible and economically meaningful results comparable to those obtained from the standard TVP-VAR-SV model. Nevertheless, some differences are detectable. For example, the initial impact of the shock on inflation shows almost no change in Figure 3.9. On the other hand, the initial response of inflation displays a sharp decline in Figure 3.7. Furthermore, there are some differences in the magnitude values of the percentage points of impulse responses delivered from each inferential tool. We leave investigating those differences to future.

3.6 Summary

This chapter develops a practical and computationally tractable estimation strategy based on the Direct Averaging Method, which leverages bivariate models estimated via Gibbs sampler. The method provides a scalable solution for the high-dimensional TVP-VAR-SV models, avoiding the computational burden of full multivariate estimation by aggregating partial posteriors.

We evaluate the empirical performance of the Bayesian pairwise composite likelihood method using quarterly U.S. macroeconomic and financial data. The estimated PCL-TVP-VAR-SV model reveals that an unexpected increase in the corporate bond spread induces a notable slowdown in real economic activity and a decline in interest rates. However, the model also generates puzzling responses in inflation and stock market returns, suggesting areas for further refinement.

The pairwise composite approach is especially appealing in settings where full multivariate estimation is infeasible. A sufficient condition for its use is that estimation of each bivariate TVP-VAR-SV model is computationally manageable. Nonetheless, implementing the FFBS algorithm across M models in parallel for

$N > 50$ presents memory bottlenecks, due to the need to store the full posterior draws of $\boldsymbol{\beta}_{c,t}$, $\boldsymbol{\alpha}_{c,t}$ and $\mathbf{h}_{c,t}$ over all models, time periods, and MCMC iterations. Future work should explore strategies to reduce memory usage.

Although the impulse responses we obtain are economically plausible, questions remain regarding their magnitudes. Understanding the factors driving these results, and whether they reflect model limitations or data properties, is a key priority for further analysis.

We interpret DAM as a computational approximation to a MH algorithm, where parameter draws from each bivariate model are implicitly accepted with probability one. While this perspective is intuitively appealing, a formal derivation of the second-step MH procedure, including a proper acceptance ratio, remains open. Future research will aim to establish these conditions.

3.7 Appendix

3.7.1 First Step: The FFBS algorithm

Prior hyperparameters and initial values

We first proceed by obtaining the ordinary least squares (OLS) estimates of the prior hyperparameter values of $\boldsymbol{\beta}_{i,0}$, $\boldsymbol{\alpha}_{i,0}$, and $\mathbf{h}_{i,0}$ by estimating the model $\mathbf{y}_i = (\mathbf{I}_{N_i} \otimes \mathbf{X}_i)\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ based on an initial training sample of 40 observations, where we define $\mathbf{y}_i$ as an $N_iT \times 1$ vector, $\mathbf{X}_i = (\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,T})'$ as a $T \times L_i$ matrix by setting $L_i = (1 + N_ip)$ and rearranging $\mathbf{x}_{i,t} = (1, \mathbf{y}'_{i,t-1}, \ldots, \mathbf{y}'_{i,t-p})$, $\boldsymbol{\beta}_i = \text{vec}([\mathbf{B}_{i,0}, \ldots, \mathbf{B}_{i,p}]')$ as an $L_iN_i \times 1$ vector, $\boldsymbol{\epsilon}_i$ is an $(N_iT \times 1)$ vector of reduced form disturbances assumed to be a zero-mean process with covariance matrix $\boldsymbol{\epsilon}_{i,t} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{i,\epsilon} \otimes \mathbf{I}_{i,T})$. We may write the covariance matrix as $\boldsymbol{\Sigma}_{i,\epsilon} = \mathbf{A}_i^{-1}\mathbf{H}_i\mathbf{A}_i'^{-1}$.

We set $\hat{\boldsymbol{\beta}}_{i,0} = [I_{N_i} \otimes (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i']\mathbf{y}_i$ and $\hat{\mathbf{V}}_{\beta_{i,0}} = \boldsymbol{\Sigma}_{i,0} \otimes (\mathbf{X}_i'\mathbf{X}_i)^{-1}$ as the OLS coefficient vector and the covariance matrix, respectively. The OLS estimates of

the elements $\hat{\boldsymbol{\alpha}}_{i,0}$ are generated by regressing $\hat{\boldsymbol{\epsilon}}_{i,j,t}$ on $-\hat{\boldsymbol{\epsilon}}_{i,1:j-1,t}$ for $j = 2 : N_i$, and the implied estimated variance is $\hat{\mathbf{V}}_{\alpha_{i,0}}$. Let $\hat{\boldsymbol{\Sigma}}_{i,\epsilon} = \hat{\mathbf{A}}_i^{-1}\hat{\mathbf{H}}_i\hat{\mathbf{A}}_i^{\prime-1}$ be the estimated covariance matrix of $\boldsymbol{\epsilon}_{i,t}$ from the time invariant VAR model, where we apply the same decomposition as in Equation (3.13). Then, we may set $\hat{\mathbf{h}}_{i,0}$ equal to the logarithms of the square root elements on the diagonal of $\hat{\mathbf{H}}_i^{1/2} = \boldsymbol{\Sigma}_{i,0}^{1/2}$ and $\hat{\mathbf{V}}_{h_{i,0}} = \mathbf{I}_{N_i}$.

Sampling bivariate coefficient states

Assume that initial prior information is given by $\mathbf{y}_{i,0}$ at $t = 0$, at any future time t the available information set is $\mathbf{y}_{i,1:t} = \{\mathbf{y}_{i,t}, \mathbf{y}_{i,1:t-1}\}$, where $\mathbf{y}_{i,t}$ is the observed value of the pairs of series at time t.

The FFBS algorithm is implemented as follows

Forward filtering

For each model $i = 1 : M$, consider the posterior distribution of the state $\boldsymbol{\beta}_{i,t-1}$ given the information up to time $t - 1$ is

$$\boldsymbol{\beta}_{i,t-1}|\boldsymbol{\alpha}_{i,t-1}, \mathbf{h}_{i,t-1}, \mathbf{y}_{i,1:t-1}, \boldsymbol{\Sigma}_i \sim \mathcal{N}(\mathbf{m}_{i,t-1}, \mathbf{C}_{i,t-1}).$$

Then the following statements hold

1. The one-step-ahead predictive distribution of $\boldsymbol{\beta}_{i,t}$ conditional on $\boldsymbol{\alpha}_{i,t}$, $\mathbf{h}_{i,t}$, and $\mathbf{y}_{i,1:t-1}$ is Gaussian with parameters

$$\mathbf{a}_{i,t} = E(\boldsymbol{\beta}_{i,t}|\boldsymbol{\alpha}_{i,t}, \mathbf{h}_{i,t}, \mathbf{y}_{i,1:t-1}, \boldsymbol{\Sigma}_i) = \mathbf{m}_{i,t-1}, \tag{3.15}$$

$$\mathbf{R}_{i,t} = V(\boldsymbol{\beta}_{i,t}|\boldsymbol{\alpha}_{i,t}, \mathbf{h}_{i,t}, \mathbf{y}_{i,1:t-1}, \boldsymbol{\Sigma}_i) = \mathbf{C}_{i,t-1} + \boldsymbol{\Sigma}_{i,\beta}. \tag{3.16}$$

2. The one-step-ahead predictive distribution of $\mathbf{y}_{i,t}$ conditional on $\mathbf{y}_{i,1:t-1}$ is

Gaussian with parameters

$$\begin{aligned} \mathbf{f}_{i,t} = E(\mathbf{y}_{i,t}|\mathbf{y}_{i,1:t-1}) &= \mathbf{x}_{i,t}\mathbf{a}_{i,t}, & (3.17)\\ \mathbf{Q}_{i,t} = V(\mathbf{y}_{i,t}|\mathbf{y}_{i,1:t-1}) &= \mathbf{x}_{i,t}\mathbf{R}_{i,t}\mathbf{x}'_{i,t} + \mathbf{\Sigma}_{i,t}. & (3.18) \end{aligned}$$

3. The filtering distribution of $\boldsymbol{\beta}_{i,t}$ conditional on $\boldsymbol{\alpha}_{i,t}$, $\mathbf{h}_{i,t}$, and $\mathbf{y}_{i,1:t}$ is Gaussian, with parameters

$$\begin{aligned} \mathbf{m}_{i,t} = E(\boldsymbol{\beta}_{i,t}|\boldsymbol{\alpha}_{i,t}, \mathbf{h}_{i,t}, \mathbf{y}_{i,1:t}, \mathbf{\Sigma}_i) &= \mathbf{a}_{i,t} + \mathbf{K}_{i,t}(\mathbf{y}_{i,t} - \mathbf{x}_{i,t}\mathbf{a}_{i,t}), & (3.19)\\ \mathbf{C}_{i,t} = V(\boldsymbol{\beta}_{i,t}|\boldsymbol{\alpha}_{i,t}, \mathbf{h}_{i,t}, \mathbf{y}_{i,1:t}, \mathbf{\Sigma}_i) &= R_{i,t} - \mathbf{K}_{i,t}\mathbf{x}_{i,t}R_{i,t}, & (3.20) \end{aligned}$$

where $\mathbf{K}_{i,t} = \mathbf{R}_{i,t}\mathbf{x}'_{i,t}\mathbf{Q}_{i,t}^{-1}$.

Backward sampling

At the end of forward filtering, the mean and the variance for $\boldsymbol{\beta}_{i,t}$ of the form

$$p(\boldsymbol{\beta}_{i,t}|\boldsymbol{\alpha}_{i,1:T}, \mathbf{h}_{i,1:T}, \mathbf{y}_{i,1:t}, \mathbf{\Sigma}_i) \sim \mathcal{N}(\mathbf{m}_{i,t}, \mathbf{C}_{i,t}).$$

The backward sampling updates the conditional means and the variances to reflect the additional information about $\boldsymbol{\beta}_{i,t}$ contained in $\beta_{i,t+1}$. Because the state space model in Equations (3.8) and (3.9) are linear and Gaussian, the distribution of $\boldsymbol{\beta}_{i,t}$ given $\mathbf{y}_{i,1:t}$ and that of $\boldsymbol{\beta}_{i,t}$ given $\boldsymbol{\beta}_{i,t+1}$ and $\mathbf{y}_{i,1:t}$ for $t = T-1, \ldots, 1$ are also Gaussian. Once $\boldsymbol{\beta}_{i,t}$ is generated, we generate $\boldsymbol{\beta}_{i,t}$ from

$$p(\boldsymbol{\beta}_{i,t}|\boldsymbol{\beta}_{i,t+1}, \boldsymbol{\alpha}_{i,1:T}, \mathbf{h}_{i,1:T}, \mathbf{y}_{i,1:t}, \mathbf{\Sigma}_i) \sim \mathcal{N}(\mathbf{g}_{i,t}, \mathbf{G}_{i,t}),$$

with moments

$$\begin{aligned} \mathbf{g}_{i,t} &= \mathbf{m}_{i,t} + \mathbf{B}_{i,t}(\boldsymbol{\beta}_{t+1} - \mathbf{m}_{i,t}),\\ \mathbf{G}_{i,t} &= \mathbf{C}_{i,t} - \mathbf{B}_{i,t}\mathbf{R}_{i,t+1}\mathbf{B}'_{i,t}, \end{aligned} \quad (3.21)$$

where $\mathbf{B}_{i,t} = \mathbf{C}_{i,t}\mathbf{R}_{i,t+1}^{-1}$.

Sampling bivariate covariance states

To sample $\boldsymbol{\alpha}_{i,t}$ elements of the lower triangular matrix $\mathbf{a}_{i,t}$, let's denote

$$\mathbf{A}_{i,t}^{-1}(\mathbf{y}_{i,t} - \mathbf{x}_{i,t}\boldsymbol{\beta}_{i,t}) = \mathbf{A}_{i,t}^{-1}\hat{\mathbf{y}}_{i,t} = \mathbf{H}_{i,t}^{1/2}\boldsymbol{\epsilon}_{i,t}. \tag{3.22}$$

Then, we may write the following equation for each row of the lower triangular matrix $\mathbf{A}_{i,t}^{-1}$ for each $j = 2 : N_i$ and for each bivariate model $i = 1 : M$ as

$$\hat{\mathbf{y}}_{i,j,t} = \hat{\mathbf{y}}_{i,1:j-1,t}\boldsymbol{\alpha}_{i,j,t} + \boldsymbol{\sigma}_{i,j,t}\boldsymbol{\epsilon}_{i,j,t}, \tag{3.23}$$

where $\boldsymbol{\sigma}_{i,j,t}$ and $\boldsymbol{\epsilon}_{i,j,t}$ are model ith, jth elements of $\boldsymbol{\Sigma}_{i,t}$ and $\boldsymbol{\epsilon}_{i,t}$, respectively and $\hat{\mathbf{y}}_{i,1:j-1,t} = (\hat{\mathbf{y}}_{i,1,t}, \ldots, \hat{\mathbf{y}}_{i,j-1,t})$.

Conditional on $\boldsymbol{\beta}_{i,t}$ and $\mathbf{h}_{i,t}$, Equation (3.23), is the measurement equation of a state space model where the state is defined in the first line of Equation (3.11) with components of the state $\boldsymbol{\alpha}_{i,j,t}$. Using the FFBS algorithm, we can draw

$$p(\boldsymbol{\alpha}_{i,j,t}|\boldsymbol{\alpha}_{i,j,t+1}, \boldsymbol{\beta}_{i,1:T}, \mathbf{h}_{i,1:T}, \mathbf{y}_{i,1:t}, \boldsymbol{\Sigma}_i) \;\; \sim \;\; \mathcal{N}(\mathbf{g}_{i,\alpha,t}, \mathbf{G}_{i,\alpha,t}),$$

with moments

$$\begin{aligned} \mathbf{g}_{i,\alpha,t} &= \mathbf{m}_{i,\alpha,t} + \mathbf{B}_{i,\alpha,t}(\boldsymbol{\beta}_{i,\alpha,t+1} - \mathbf{m}_{i,\alpha,t}), \\ \mathbf{G}_{i,\alpha,t} &= \mathbf{C}_{i,\alpha,t} - \mathbf{B}_{i,\alpha,t}\mathbf{R}_{i,\alpha,t+1}\mathbf{B}'_{i,\alpha,t}, \end{aligned} \tag{3.24}$$

where $\mathbf{B}_{i,\alpha,t} = \mathbf{C}_{i,\alpha,t}\mathbf{R}_{i,\alpha,t+1}^{-1}$.

Sampling bivariate volatility states

Let's start implementing this approach by transforming $\mathbf{y}_{i,t}$ in order to obtain a measurement that is linear in $\mathbf{h}_{i,t}$

$$\mathbf{y}^*_{i,t} = (\mathbf{y}_{i,t} - \mathbf{x}_{i,t}\boldsymbol{\beta}_{i,t}) = e^{\mathbf{h}_{i,t}}\boldsymbol{\upsilon}_{i,t}, \tag{3.25}$$

where $e^{\mathbf{h}_{i,t}}\boldsymbol{\upsilon}_{i,t} = \boldsymbol{\eta}^h_{i,t}$ and $\boldsymbol{\upsilon}_{i,t} \sim \mathcal{N}(0,1)$. Squaring both sides of Equation (3.25) and taking the logarithm, we obtain the following form

$$\mathbf{y}^{**}_{i,t} = 2\mathbf{h}_{i,t} + \boldsymbol{\upsilon}^*_{i,t}, \tag{3.26}$$

where $\mathbf{y}^{**}_{i,t} = \log \mathbf{y}^2_{i,t}$ and $\boldsymbol{\upsilon}^*_{i,t} = \log \boldsymbol{\upsilon}^2_{i,t}$. Hence, Equation (3.26) together with second line of Equation (3.11) represent a state space model. However, the issue with the above representation is that the error term $\boldsymbol{\upsilon}^*_{i,t}$ does not have a Gaussian distribution. The auxiliary mixture sampling finds a Gaussian mixture that approximates the pdf of the $\boldsymbol{\upsilon}^*_{i,t}$

$$p(\boldsymbol{\upsilon}^*_{i,t}) \approx \sum_{n=1}^{7} w_{i,n}\phi(\boldsymbol{\upsilon}^*_{i,t}; \boldsymbol{\mu}_{h,i,n} - 1.2704, \boldsymbol{\sigma}^2_{h,i,n}), \tag{3.27}$$

where $\phi(\mathbf{y}_i; \boldsymbol{\mu}_i, \boldsymbol{\sigma}^2_i)$ is the Gaussian density with mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{\sigma}^2_i$, $w_{i,n}$ are the mixture probabilities for the nth component and 7 is the number of components. We can define an auxiliary random variable $s_{h,i,t} \in \{1, \ldots, 7\}$, which can be used as a mixture component indicator

$$\begin{aligned}(\boldsymbol{\upsilon}^*_{i,t}|s_{h,i,t} = n) &\sim \mathcal{N}(\boldsymbol{\mu}_{h,i,n} - 1.2704, \boldsymbol{\sigma}^2_{h,i,n}),\\ \Pr(s_{h,i,t} = n) &= w_{i,n}.\end{aligned}$$

Hence, the desired linear Gaussian model can be defined conditional on the component indicator $s_{h,i,t}$ as $(\boldsymbol{\upsilon}^*_{i,t}|s_{h,i,t}) \sim \mathcal{N}(\mathbf{d}^*_{i,t}, \boldsymbol{\Sigma}^*_{i,t})$, where $\mathbf{d}^*_{i,t} = \boldsymbol{\mu}_{h,i,n} - 1.2704$

and $\mathbf{\Sigma}^*_{i,t} = \boldsymbol{\sigma}^2_{h,i,n}$ for each $t = 1 : T$. $\boldsymbol{\mu}_{h,i,n}$ and $\boldsymbol{\sigma}_{h,i,n}$ have fixed values represented as a seven component Gaussian mixture in Table 4 of Kim et al. (1998). Then it follows that

$$(\mathbf{y}^{**}_{i,t}|s_{h,i,t}, \mathbf{h}_{i,t}) \sim \mathcal{N}(2\mathbf{h}_{i,t} + \mathbf{d}^*_{i,t}, \mathbf{\Sigma}^*_{i,t}), \tag{3.28}$$

for each $t = 1 : T$.

Using the FFBS algorithm, we can draw $\mathbf{h}_{i,t}$ conditional on $\boldsymbol{\beta}_{i,t}$ and $\boldsymbol{\alpha}_{i,t}$ from

$$p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t+1}, \boldsymbol{\beta}_{i,1:T}, \boldsymbol{\alpha}_{i,1:T}, \mathbf{y}_{i,1:t}, s_{h,i,t}, \mathbf{\Sigma}_i) \quad \sim \quad \mathcal{N}(\mathbf{g}_{i,h,t}, \mathbf{G}_{i,h,t}),$$

with moments

$$\begin{aligned} \mathbf{g}_{i,h,t} &= \mathbf{m}_{i,h,t} + \mathbf{B}_{i,h,t}(\boldsymbol{\beta}_{i,h,t+1} - \mathbf{m}_{i,h,t}), \\ \mathbf{G}_{i,h,t} &= \mathbf{C}_{i,h,t} - \mathbf{B}_{i,h,t} R_{i,h,t+1} \mathbf{B}'_{i,h,t}, \end{aligned} \tag{3.29}$$

where $\mathbf{B}_{i,h,t} = \mathbf{C}_{i,h,t}\mathbf{R}^{-1}_{i,h,t+1}$.

Sampling $s_{h,i,t}$

For each $i = 1 : M$, we can sample $s_{h,i,t}$ as follows

$$p(\mathbf{s}_{h,i}|\mathbf{y}^{**}_i, \mathbf{h}_i) = \prod_{t=1}^{T} p(s_{h,i,t}|\mathbf{y}^{**}_{i,t}, \mathbf{h}_{i,t}), \tag{3.30}$$

where each $s_{h,i,t}$ can be sampled independently for $t = 1 : T$. We can compute $\Pr(s_{h,i,t} = n|\mathbf{y}^{**}_{i,t}, \mathbf{h}_{i,t})$ for $n = 1 : 7$ as

$$\Pr(s_{h,i,t} = n|\mathbf{y}^{**}_{i,t}, \mathbf{h}_{i,t}) = \frac{w_{i,n}\phi(\mathbf{y}^{**}_{i,t}; 2\mathbf{h}_{i,t} + \boldsymbol{\mu}_{h,i,n} - 1.2704, \boldsymbol{\sigma}^2_{h,i,n})}{\sum_{n=1}^{7} w_{i,n}\phi(\mathbf{y}^{**}_{i,t}; 2\mathbf{h}_{i,t} + \boldsymbol{\mu}_{h,i,n} - 1.2704, \boldsymbol{\sigma}^2_{h,i,n})}. \tag{3.31}$$

Sampling $\boldsymbol{\Sigma}_{i,\beta}$

Sampling the hyperparameter $\boldsymbol{\Sigma}_{i,\beta}$ as

$$p(\boldsymbol{\Sigma}_{i,\beta}|\mathbf{y}_{i,1:t}, \boldsymbol{\beta}_{i,1:T}, \boldsymbol{\alpha}_{i,1:T}, \mathbf{h}_{i,1:T}, \boldsymbol{\Sigma}_{i,\alpha}, \boldsymbol{\Sigma}_{i,h}) \\ \sim \mathcal{IW}\big(\mathbf{U}_{\beta_{i,0}} + \sum_{t=1}^{T}(\mathbf{y}_{i,t} - \mathbf{x}'_{i,t}\boldsymbol{\beta}_{i,t})(\mathbf{y}_{i,t} - \mathbf{x}'_{i,t}\boldsymbol{\beta}_{i,t})', \nu_{i,\beta} + T\big).$$

Sampling $\boldsymbol{\Sigma}_{i,\alpha}$

Sampling the hyperparameter $\boldsymbol{\Sigma}_{i,\alpha}$ blockwise with a single block for $n_i = 1 : N_i - 1$

$$p(\boldsymbol{\Sigma}_{i,\alpha}|\mathbf{y}_{i,1:t}, \boldsymbol{\beta}_{i,1:T}, \boldsymbol{\alpha}_{i,1:T}, \mathbf{h}_{i,1:T}, \boldsymbol{\Sigma}_{i,\beta}, \boldsymbol{\Sigma}_{i,h}) \\ \sim \mathcal{IW}\big(\mathbf{U}_{\alpha_{i,0}} + \sum_{t=1}^{T}(\boldsymbol{\alpha}_{i,j,t} - \boldsymbol{\alpha}_{i,j,t-1})(\boldsymbol{\alpha}_{i,j,t} - \boldsymbol{\alpha}_{i,j,t-1})', \nu_{i,\alpha} + T\big).$$

Sampling $\boldsymbol{\Sigma}_{i,h}$

Sampling the hyperparameter $\boldsymbol{\Sigma}_{i,h}$ as

$$p(\boldsymbol{\Sigma}_{i,h}|\mathbf{y}_{i,1:t}, \boldsymbol{\beta}_{i,1:T}, \boldsymbol{\alpha}_{i,1:T}, \boldsymbol{\Sigma}_{i,\alpha}, \boldsymbol{\Sigma}_{i,h}) \\ \sim \mathcal{IW}\big(\mathbf{U}_{h_{i,0}} + \sum_{t=1}^{T}(\mathbf{h}_{i,t} - \mathbf{h}_{i,t-1})(\mathbf{h}_{i,t} - \mathbf{h}_{i,t-1})', \nu_{i,h} + T\big).$$

3.7.2 Second Step: The Direct Averaging Method

To implement the proposed Bayesian pairwise composite likelihood method, a sequence of mapping algorithms is developed to recover parameter vectors from the posterior draws of bivariate models. These include: (i) a mapping algorithm for aggregating bivariate VAR coefficients into the full vector of VAR parameters, (ii) an algorithm for assembling the Cholesky parameter vector that characterizes the lower-triangular structure of the time-varying covariance matrix, and (iii) a

procedure for constructing the log-volatility vector from bivariate log-volatility estimates. Each algorithm ensures that the composite information extracted from the bivariate models is consistently aligned with the model structure.

In Algorithm 1, we implement a structured mapping procedure that consolidates the time-varying coefficients from all bivariate VAR(p) models into a single parameter vector $\boldsymbol{\beta}_t^{(r)}$ for the full N-dimensional VAR system. Specifically, for each MCMC draw $r = 1, \ldots, R$ and time point $t = 1, \ldots, T$, we stack the bivariate coefficient vectors $\boldsymbol{\beta}_t^{(i,j,r)} \in \mathbb{R}^K$ into a composite vector $\boldsymbol{\beta}_{c,t}^{(r)} \in \mathbb{R}^{K_c}$. A sparse, deterministic mapping matrix $\boldsymbol{\Omega}_\beta \in \mathbb{R}^{K' \times K_c}$ then transforms this composite vector into the parameter vector:

$$\boldsymbol{\beta}_t^{(r)} = \boldsymbol{\Omega}_\beta \cdot \boldsymbol{\beta}_{c,t}^{(r)}.$$

This matrix encodes the selection and averaging structure across overlapping bivariate models, ensuring that parameters appearing in multiple systems are aggregated consistently, typically using uniform weights of $1/(N-1)$. This step preserves internal consistency, avoids duplication, and ensures valid identification of the model's coefficients.

Contemporaneous dependencies are captured by a lower-triangular Cholesky matrix $\mathbf{A}_t \in \mathbb{R}^{N \times N}$ with time-varying off-diagonal elements $\alpha_t^{(i,j,r)}$ for $i > j$ in Algorithm 3. These elements are uniquely estimated from the bivariate models involving each pair (i, j), and stacked into a composite vector $\boldsymbol{\alpha}_{c,t}^{(r)} \in \mathbb{R}^L$ according to a fixed recursive ordering. Since each $\alpha_{ij,t}^{(r)}$ appears only once across the bivariate systems, no averaging is required. The vector is recovered via the identity mapping:

$$\boldsymbol{\alpha}_t^{(r)} = \mathbf{I} \cdot \boldsymbol{\alpha}_{c,t}^{(r)} = \boldsymbol{\alpha}_{c,t}^{(r)},$$

where $\mathbf{I}$ is the identity matrix of appropriate dimension. This mapping is purely notational and relies only on consistent indexing across MCMC draws and time points.

For each bivariate model $(i, j) \in \mathcal{P}$, we estimate a time-varying log-volatility vector $\mathbf{h}_t^{(i,j,r)} = \left(h_{i,t}^{(i,j,r)}, h_{j,t}^{(i,j,r)}\right)' \in \mathbb{R}^{N_i}$ across all $t = 1, \ldots, T$ and MCMC iterations $r = 1, \ldots, R$ in Algorithm 2. These vectors are stacked into a composite vector $\mathbf{h}_{c,t}^{(r)} \in \mathbb{R}^{N_i M}$. To recover the volatility vector $\mathbf{h}_t^{(r)} \in \mathbb{R}^N$, we apply a fixed sparse mapping matrix $\mathbf{\Omega}_h \in \mathbb{R}^{N \times N_i M}$ such that

$$\mathbf{h}_t^{(r)} = \mathbf{\Omega}_h \cdot \mathbf{h}_{c,t}^{(r)}.$$

Each row of $\mathbf{\Omega}_h$ averages log-volatility estimates for a given variable across its $N - 1$ pairwise appearances. This mapping ensures coherent aggregation across the system. The procedure is repeated across all time steps and MCMC draws.

Algorithm 1 Mapping bivariate coefficient states to pairwise composite coefficient states

Input:

- Posterior draws $\boldsymbol{\beta}_t^{(i,j,r)} \in \mathbb{R}^K$ for all pairs (i, j), time $t = 1, \ldots, T$, and iterations $r = 1, \ldots, R$
- Fixed mapping matrix $\boldsymbol{\Omega}_\beta \in \mathbb{R}^{K' \times K_c}$, determined by the inclusion structure of coefficients in bivariate models

Define:

$$K = N_i(N_i p + 1), \quad K' = N(Np + 1), \quad \text{and} \quad M$$

For each MCMC iteration and time step:
for $r = 1$ to R **do**
  for $t = 1$ to T **do**
    Initialize stacked coefficient vector $\boldsymbol{\beta}_{c,t}^{(r)} \in \mathbb{R}^{K_c}$
    Step 1: Stack bivariate coefficients
    Set block index $m = 0$
    for $i = 1$ to $N - 1$ **do**
      for $j = i + 1$ to N **do**
        Insert $\boldsymbol{\beta}_t^{(i,j,r)}$ into block m of $\boldsymbol{\beta}_{c,t}^{(r)}$
        Increment $m \leftarrow m + 1$
      end for
    end for
    Step 2: Map to pairwise composite coefficients

$$\boldsymbol{\beta}_t^{(r)} = \boldsymbol{\Omega}_\beta \cdot \boldsymbol{\beta}_{c,t}^{(r)} \in \mathbb{R}^{K'}$$

  end for
end for
Output: Posterior draws $\{\boldsymbol{\beta}_t^{(r)}\}_{r=1}^R$ for $t = 1, \ldots, T$, or posterior means:

$$\hat{\boldsymbol{\beta}}_t = \frac{1}{R} \sum_{r=1}^{R} \boldsymbol{\beta}_t^{(r)}$$

Algorithm 2 Mapping bivariate covariance states to pairwise composite covariance states

Input:

- Posterior draws $\boldsymbol{\alpha}_t^{(i,j,r)} \in \mathbb{R}$ from bivariate model for (i, j) where $i > j$, for time $t = 1, \ldots, T$, and iterations $r = 1, \ldots, R$
- Recursive ordering over N variables

Define:

$$\mathcal{A} = \{(i, j) : 1 \leq j < i \leq N\}, \quad L = \frac{N(N-1)}{2}$$

For each MCMC iteration and time step:
for $r = 1$ to R **do**
  for $t = 1$ to T **do**
    Initialize empty vector $\boldsymbol{\alpha}_{c,t}^{(r)} \in \mathbb{R}^L$
    Set index $\ell = 0$
    for $i = 2$ to N **do**
      for $j = 1$ to $i - 1$ **do**
        Extract scalar $\alpha_t^{(i,j,r)}$ from bivariate model (i, j)
        Set $(\boldsymbol{\alpha}_{c,t}^{(r)})_\ell \leftarrow \alpha_t^{(i,j,r)}$
        Increment $\ell \leftarrow \ell + 1$
      end for
    end for
    Step 2: Map to pairwise composite vector

$$\boldsymbol{\alpha}_t^{(r)} = \mathbf{I}_L \cdot \boldsymbol{\alpha}_{c,t}^{(r)} \in \mathbb{R}^L$$

  end for
end for
Output: Posterior draws $\{\boldsymbol{\alpha}_t^{(r)}\}_{r=1}^R$ for $t = 1, \ldots, T$, or posterior means:

$$\hat{\boldsymbol{\alpha}}_t = \frac{1}{R} \sum_{r=1}^{R} \boldsymbol{\alpha}_t^{(r)}$$

Algorithm 3 Mapping bivariate volatility states to pairwise composite volatility states

Input:

- Posterior draws $\mathbf{h}_t^{(i,j,r)} \in \mathbb{R}^{N_i}$ for all pairs (i, j), time $t = 1, \ldots, T$, and iterations $r = 1, \ldots, R$
- Fixed mapping matrix $\mathbf{\Omega}_h \in \mathbb{R}^{N \times N_i M}$, which maps stacked bivariate volatilities to the log-volatility vector

Define:

$$N_i = 2, \quad M, \quad \mathcal{P} = \{(i, j) \in \mathbb{N}^2 : 1 \leq i < j \leq N\}$$

For each MCMC iteration and time step:
for $r = 1$ to R **do**
  for $t = 1$ to T **do**
    Initialize stacked log-volatility vector $\mathbf{h}_{c,t}^{(r)} \in \mathbb{R}^{N_i M}$
    Step 1: Stack bivariate volatilities
    Set block index $m = 0$
    for $i = 1$ to $N - 1$ **do**
      for $j = i + 1$ to N **do**
        Insert $\mathbf{h}_t^{(i,j,r)}$ into block m of $\mathbf{h}_{c,t}^{(r)}$
        Increment $m \leftarrow m + 1$
      end for
    end for
    Step 2: Map to pairwise composite volatilities

$$\mathbf{h}_t^{(r)} = \mathbf{\Omega}_h \cdot \mathbf{h}_{c,t}^{(r)} \in \mathbb{R}^N$$

  end for
end for
Output: Posterior draws $\{\mathbf{h}_t^{(r)}\}_{r=1}^R$ for $t = 1, \ldots, T$, or posterior means:

$$\hat{\mathbf{h}}_t = \frac{1}{R} \sum_{r=1}^{R} \mathbf{h}_t^{(r)}$$

3.7.3 Comparing impulse responses of PCL-TVP-VAR-SV and TVP-VAR-SV models

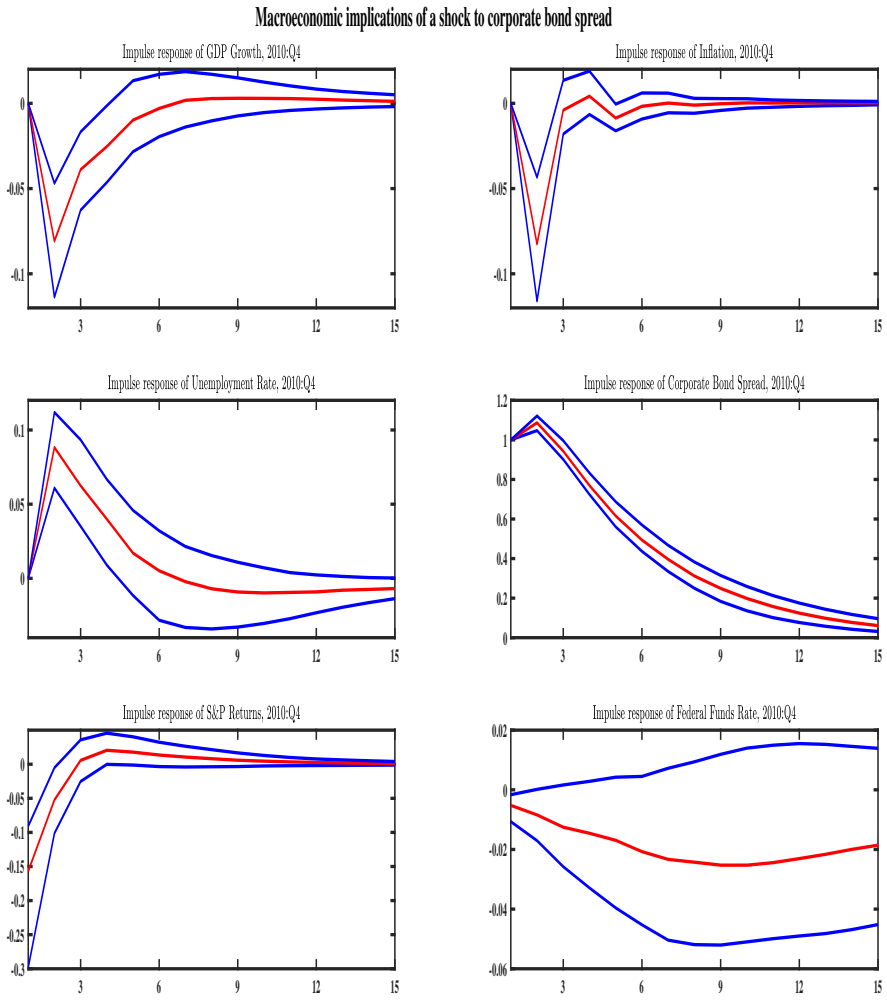

Figure 3.6: The figure depicts the pairwise composite impulse responses of GDP Growth, Inflation, Unemployment Rate, Corporate Bond Spread, S&P Returns, and Federal Funds Rate to the identified shock to the Corporate Bond Spread estimated from a six–variable PCL-TVP-VAR-SV model for a representative time 2010Q4. In this figure the posterior median is the red line and the blue lines are the 16th and 84th percentiles.

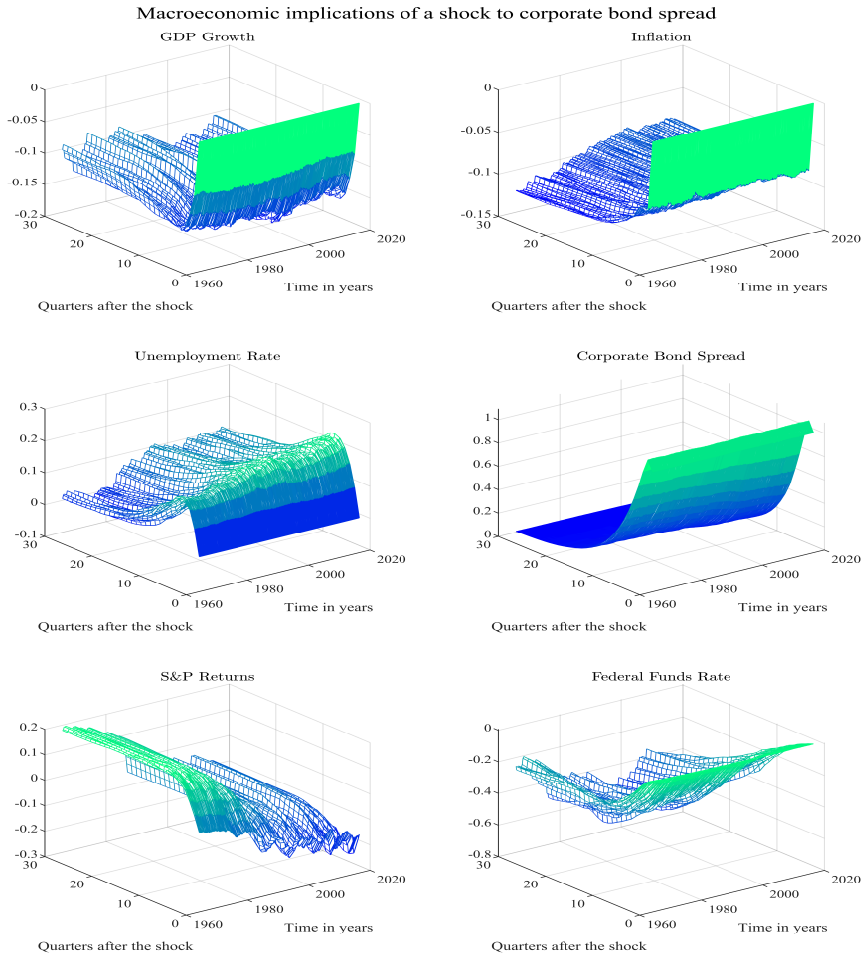

Figure 3.7: The figure depicts the pairwise composite impulse responses of GDP Growth, Inflation, Unemployment Rate, Corporate Bond Spread, S&P Returns, and Federal Funds Rate to the identified shock of the Corporate Bond Spread estimated from a six–variable PCL-TVP-VAR-SV model.

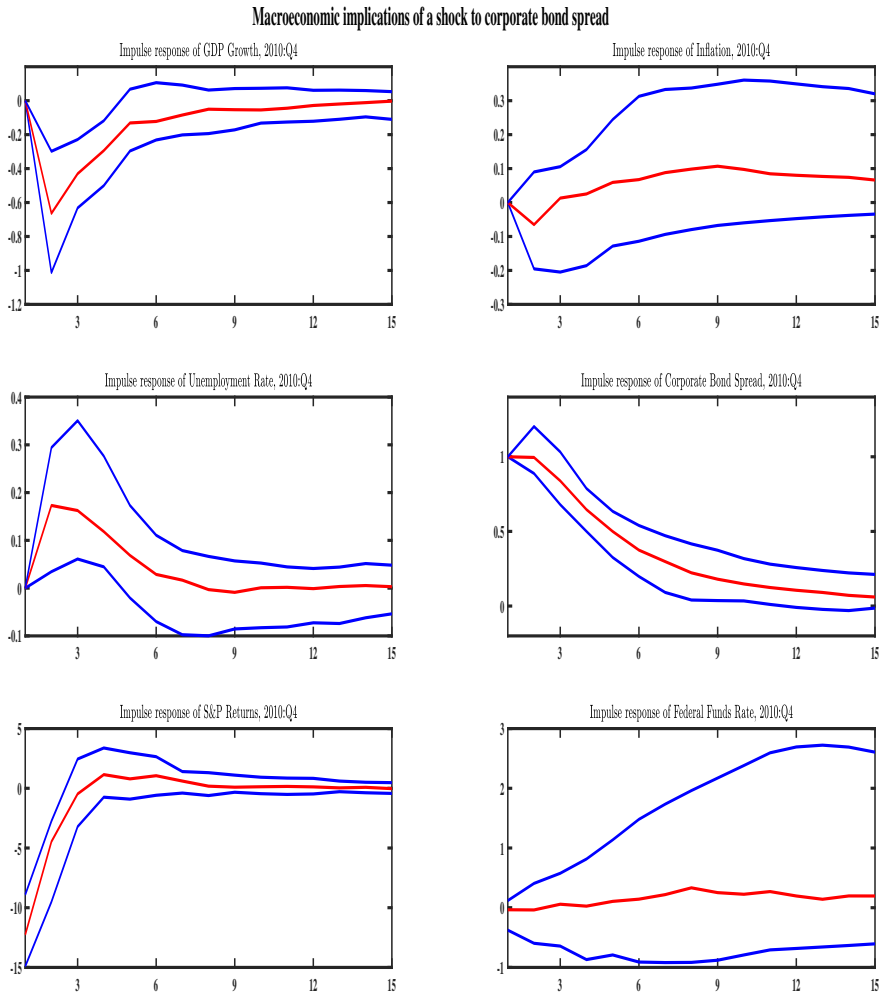

Figure 3.8: The figure depicts the impulse responses of GDP Growth, Inflation, Unemployment Rate, Corporate Bond Spread, S&P Returns, and Federal Funds Rate to the identified shock to the Corporate Bond Spread estimated from a six–variable TVP-VAR-SV model for a representative time 2010Q4. In this figure the posterior median is the red line and the blue lines are the 16th and 84th percentiles.

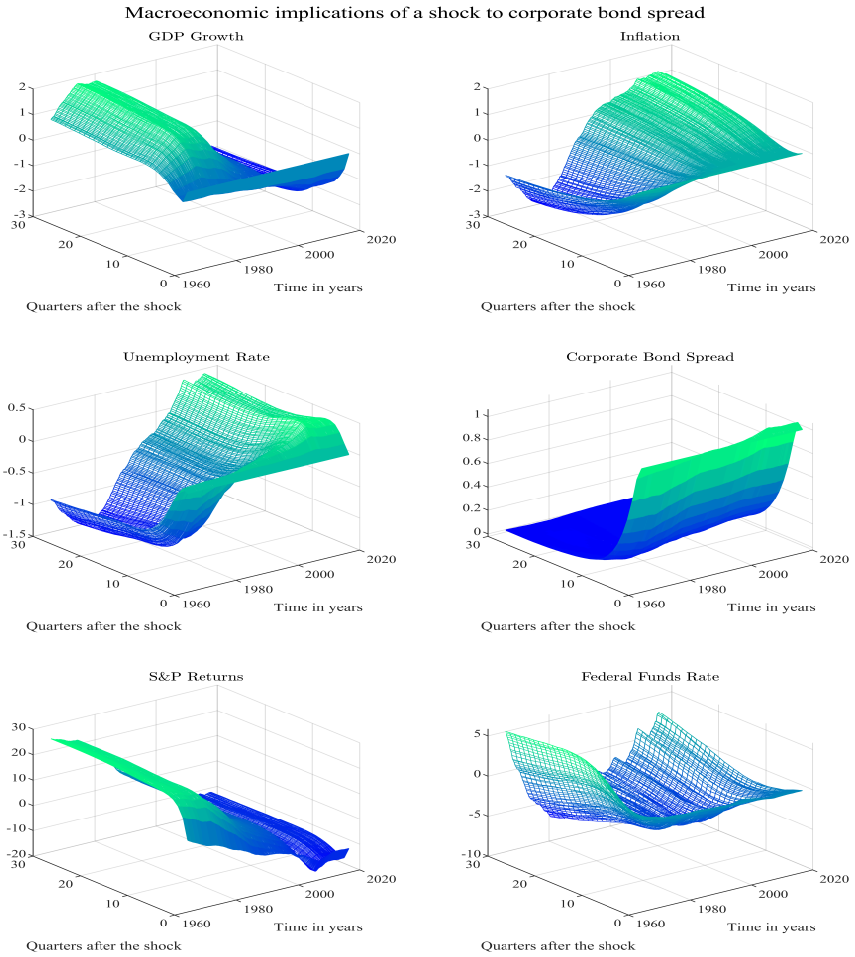

Figure 3.9: The figure depicts the impulse responses of GDP Growth, Inflation, Unemployment Rate, Corporate Bond Spread, S&P Returns, and Federal Funds Rate to the identified shock of the Corporate Bond Spread estimated from a six–variable TVP-VAR-SV model.

3.7.4 Data appendix

The quarterly time series variables used in the PCL-TVP-VAR-SV model are taken from the FRED database of the Federal Reserve Bank of St Louis spanning from 1959Q1 to 2018Q1. The columns of Table 3.1 denote the series numbers, Tcode denotes the data transformations based on McCracken and Ng (2020) and Gilchrist and Zakrajšek (2012), series denotes the FRED mnemonic except for CBS*, and description denotes a brief definition of the series.

The series CBS*, is a modified version of two series from FRED database, defined as the difference between Moody's Seasoned Baa Corporate Bond Yield and Moody's Seasoned Aaa Corporate Bond Yield. The modified Tcode, 1*, stands for no transformation of the series CBS*. The Effective Federal Funds Rate and the 10-Year Treasury Constant Maturity Rate are not transformed.

Time series used in the PCL-TVP-VAR-SV model with $N = 50$				
ID	Series	Tcode	Description	S / F
1	PCECC96	5	Real Personal Consumption Expenditures	Slow
2	GDPC1	5	Real Gross Private Domestic Investment, 3 decimal	Slow
3	GPDIC1	5	Real Gross Domestic Product, 3 Decimal	Slow
4	GDPCTPI	6	Gross Domestic Product: Chain-type Price Index	Slow
5	GCEC1	5	Real Government Consumption Expenditures & Gross Investment	Slow
6	EXPGSC1	5	Real Exports of Goods & Services, 3 Decimal	Slow
7	IMPGSC1	5	Real Imports of Goods & Services, 3 Decimal	Slow
8	DPIC96	5	Real Disposable Personal Income	Slow
9	OUTNFB	5	Nonfarm Business Sector: Real Output	Slow
10	OUTBS	5	Business Sector: Real Output	Slow
11	INDPRO	5	Industrial Production Index	Slow
12	IPDMAT	5	Industrial Production: Durable Materials	Slow
13	IPNMAT	5	Industrial Production: Nondurable Materials	Slow
14	IPDCONGD	5	Industrial Production: Durable Consumer Goods	Slow
15	IPNCONGD	5	Industrial Production: Nondurable Consumer Goods	Slow
16	CUMFNS	1	Capacity Utilization: Manufacturing (SIC)	Slow
17	PAYEMS	5	All Employees: Total nonfarm	Slow
18	USPRIV	5	All Employees: Total Private Industries	Slow
19	MANEMP	5	All Employees: Manufacturing	Slow
20	USEHS	5	All Employees: Education & Health Services	Slow
21	USFIRE	5	All Employees: Financial Activities	Slow
22	USINFO	5	All Employees: Information Services	Slow
23	CE16OV	5	Civilian Employment	Slow
24	CIVPART	2	Civilian Labor Force Participation Rate	Slow
25	UNRATE	2	Civilian Unemployment Rate	Slow
26	UNRATESTx	2	Unemployment Rate less than 27 weeks	Slow
27	UNRATELTx	2	Unemployment Rate for more than 27 weeks	Slow
28	UEMPLT5	5	Number of Civilians Unemployed - Less Than 5 Week	Slow
29	UEMP5TO14	5	Number of Civilians Unemployed for 5 to 14 Weeks	Slow
30	HOABS	5	Business Sector: Hours of All Persons	Slow
31	HOANBS	5	Manufacturing Sector: Hours of All Persons	Slow
32	CMRMTSPLx	5	Real Manufacturing and Trade Industries Sales	Slow
33	RSAFSx	5	Real Retail and Food Services Sales, deflated by Core PCE	Slow
34	PCECTPI	6	Personal Consumption Expenditures: Chain-type Price Index	Slow
35	CPIAUCSL	6	Consumer Price Index for All Urban Consumers: All Items	Slow
36	PPIACO	6	Producer Price Index for All Commodities	Slow
37	RCPHBS	5	Business Sector: Real Compensation Per Hour	Slow
38	OPHNFB	5	Nonfarm Business Sector: Real Output Per Hour of All Persons	Slow
39	OPHPBS	5	Business Sector: Real Output Per Hour of All Persons	Slow
40	CS*	1*	Baa-Aaa corporate credit spread	Fast
41	S&P 500	5	S&P's Common Stock Price Index: Composite	Fast
42	GS10	1*	10-Year Treasury Constant Maturity Rate	Fast
43	FEDFUNDS	1*	Effective Federal Funds Rate	Fast
44	TB3MS	2	3-Month Treasury Bill: Secondary Market Rate	Fast
45	TB6MS	2	6-Month Treasury Bill: Secondary Market Rate	Fast
46	GS1	2	1-Year Treasury Constant Maturity Rate	Fast
47	M1REAL	5	Real M1 Money Stock, deflated by CPI	Fast
48	M2REAL	5	Real M2 Money Stock, deflated by CPI	Fast
49	EXUSUKx	5	U.S. / U.K. Foreign Exchange Rate	Fast
50	EXCAUSx	5	Canada / U.S. Foreign Exchange Rate	Fast

Table 3.1: The quarterly time series variables used in the PCL-TVP-VAR-SV model

Chapter 4

Bayesian dynamic graphical models for high-dimensional vector autoregressions with time-varying parameters and volatility discounting

4.1 Introduction

Graphical models use pairwise conditional independence structures of a set of random variables to form sparsity features on the precision matrix of a multivariate Gaussian process, Lauritzen (1996) and Whittaker (2008). On the other hand, a graphical model can be constructed in a multiple regression model by assuming pairwise conditional independence structures formed with sparsity features on regression coefficients, Whittaker (2008). However, most graphical models either do not take into account time variation features on the coefficients and the covariance matrix of the model's disturbances or allow for a completely arbitrary mean. In this study, we propose a special case of graphical models, a Bayesian dynamic graphical model (BDGM) that characterises the relationships among variables by a directed graph, that is, the relationships among the variables are one directional. We consider a vector autoregressive model with time-varying parameters and volatility discounting (TVP-VAR-VD model) that allows a sparse

representation of dynamic state parameters and evolving covariance matrix elements within a BDGM framework.[1] The concept of studying the sparsity features on the state parameters of vector autoregressive models with time-varying parameters (TVP-VAR) has been explored in a number of papers, among others, see Korobilis (2013) and Koop and Korobilis (2013). Moreover, imposing zeros on the inverse of covariance matrices in VAR and other dynamic models using graphical models is quite common in statistics, Carvalho and West (2007), Nakajima and West (2015), Ahelegbey et al. (2016), Zhao et al. (2016), Gruber and West (2016), Gruber and West (2017), and for a discussion see West (2020). We combine those literature by studying sparsity features on both the state parameters and the inverse of covariance matrix using a BDGM approach.

We propose a BDGM approach that has an important property of splitting a complex and high-dimensional macroeconometric problem into locally sparse manageable components. The pairwise conditional independence structure is the key mechanism underlying the local model specifications and computations. Hence, feasible local computations can be performed rather than a joint high-dimensional one. In this regard, we consider splitting a joint multivariate dynamic linear model (multivariate DLM) into a componentwise multiple dynamic linear regression model (multiple regression DLM) and then combine those components to make inference from the joint model. More specifically, we split the multivariate TVP-VAR-VD model that falls within the multivariate DLM framework into a multiple regression on past and current values of predictors that falls within the multiple regression DLM framework. A key research question in building a joint TVP-VAR-VD model from component multiple regression DLMs using the BDGM approach concerns the selection of subset of predictors to include.

[1]BDGMs share many inferential tools with dynamic Bayesian networks that have been widely used in the literature on machine learning and artificial intelligence, including Koller and Friedman (2009), Murphy (2012) and Russell and Norvig (2010). In the dynamic Bayesian networks, the parameters are assumed to be constant, as noted by Murphy (2012). We prefer to use the term BDGM to distinguish our approach by relaxing this assumption and allowing for time-varying state parameters and volatilities.

The main contribution of this chapter is to resolve a high-dimensional sparse inference problem posed by parameter changes, many predictors and computational complexity with an extended version of graphical modelling on multiple regression models proposed by Whittaker (2008) and on multiregression dynamic models of Zhao et al. (2016). More precisely, we incorporate the conditional independence structure in the time-varying coefficient states to quantify the dynamics among autoregressive and cross-lagged values of the variables and the current values of the variables as an extension to the general ideas of graphical multiple regression models in Whittaker (2008), and in the time-varying covariance states to quantify the contemporaneous relationships among the variables inspired from Zhao et al. (2016) within a BDGM framework. Two sets of variables as predictors are considered, namely, dynamic autoregressive and cross-lagged values of the time series and current values of the time series. We achieve this framework by developing an efficient Bayesian graphical variable selection method that focuses on selecting from a set of dynamic and contemporaneous predictors to identify a set of locally good multiple regression DLMs in terms of posterior model probabilities. This approach can be applied recursively in parallel to find local pairwise independence structures among both the dynamic and the contemporaneous predictors using a Gray code algorithm. Then we perform a Bayesian model averaging on a range of good models with high posterior probabilities to forecast key macroeconomic and financial variables of interest.

We demonstrate the applicability of the BDGM approach on a real dataset comprising of ten quarterly U.S. macroeconomic and financial time series in an attempt to understand which subset of the dynamic and the contemporaneous predictors should be used in a compositional forecasting model. We compare results using the proposed Bayesian graphical variable selection on a set of $16,356$ distinct models in parallel. Our key findings are as follows.

First, the results of posterior model probabilities over the space of all com-

peting models show that there is a considerable model uncertainty within the best selected models for each series. As the number of the dynamic and the contemporaneous predictors increases in the recursive model, the model uncertainty increases too.

Second, we perform a sensitivity analysis by comparing median probability models and highest posterior probability models over the space of all competing models. The results show that the median probability model can be an important tool in assessing the effect of the dynamic and the contemporaneous predictors.

Third, out-of-sample forecast exercise shows that the joint model with Bayesian model averaging outperforms the joint model with the highest posterior probability for the variables consumption, investment, GDP growth, GDP deflator, industrial production, and unemployment rate over the majority of considered horizons. Other combinations of financial variables, such as corporate bond spread, S&P 500 stock returns, 10-year Treasury maturity rate, and Federal funds rate tend to improve predictions at longer forecast horizons.

The plan of this chapter is as follow. In Section 4.2 we define and motivate the BDGM framework to identify subsets of predictors and the number of graphical models. In Section 4.3 we show the component multiple regression DLMs, the joint TVP-VAR-VD model, and graph representation of the joint TVP-VAR-VD model. In Section 4.4, we summarize our Bayesian inferential tool based on Kalman filter algorithm. Section 4.5 reports the details of the Bayesian graphical variable selection approach. We report the empirical results for a range of data analysis in Section 4.6, such as exploratory data analysis, dynamic and contemporaneous pairwise dependence structures and out-of-sample forecast performance. A summary is present in Section 4.7. Finally, the Appendix provides the details of the Kalman filter algorithm, the joint model, more results, and data appendix.

4.2 Bayesian dynamic graphical Models

Let $\mathbf{y}_t$ denote an $N \times 1$ vector of time series and $y_{i,t}$ is a scalar with ith cross-sectional and tth time series element. For each cross-sectional variable $i = 1 : N$, $t = 1 : T$ denotes the time series observations. Let $\mathbf{y}_0$ represent all prior information set and $\mathbf{y}_{1:t} = \{\mathbf{y}_{1:t-1}, \mathbf{y}_t\}$ all available relevant information set at any time t. Given all the previous information at $t-1$, computing the joint one-step-ahead predictive density for the observed data $\mathbf{y}_t$ can be performed as a product of a set of conditional probability density functions

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = p(y_{1,t}|\mathbf{y}_{1:t-1}) \prod_{i=2}^{N} p(y_{i,t}|\mathbf{y}_{i-1:i,t-1:t-p}, \mathbf{y}_{1:i-1,t}, \mathbf{y}_{1:t-1}), \tag{4.1}$$

where p is the model order for $k = 1, \ldots, p$. In this setup, we have two types of dynamic lagged effects and one type of contemporaneous effects between pairs of variables, that is, dynamic autoregressive effects from $y_{i,t-k}$ to $y_{i,t}$, dynamic cross-lagged effects from $y_{j,t-k}$ to $y_{i,t}$ and contemporaneous effects from $y_{j,t}$ and $y_{i,t}$ with $i \neq j$. In terminology of graph theory, if $y_{j,t}$ is a predictor of $y_{i,t}$, $y_{j,t}$ is said to be parent of $y_{i,t}$. A cycle is defined as any directed path that starts and ends at the same variable. A graph which contains only directed path and does not have any cycles, is called a directed acyclic graph, Murphy (2012). The Bayesian dynamic graphical model is an example of a directed acyclic graph. Each $y_{i,t}$ has a conditional probability density function $p(y_{i,t}|\mathbf{F}_{pa(i),t}) = p(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1})$ that quantifies the effect of parents on the variable, where $\mathbf{F}_{pa(i),t} = (\mathbf{x}_{dp(i),t}, \mathbf{y}_{cp(i),t})'$ with a partition of $\mathbf{x}_{dp(i),t} = \{y_{i-1,t-1:t-p}, y_{i,t-1:t-p}\}$ and $\mathbf{y}_{cp(i),t} = \{y_{1,t}, \ldots, y_{i-1,t}\}$. We refer to two set of parents in Equation (4.1), dynamic parents with index $dp(i)$ and contemporaneous parents with index $cp(i)$. Hence, the joint parental set is partitioned $pa(i) = \{dp(i), cp(i)\}$ and Equation (4.1) can be equivalently

written as

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = p(y_{1,t}|\mathbf{F}_{pa(1),t}) \prod_{i=2}^{N} p(y_{i,t}|\mathbf{F}_{pa(i),t}), \tag{4.2}$$

where $dp(i)$ is a subset of $\{i-1, i\}$ and $cp(i)$ is a subset of $\{1, \ldots, i-1\}$, written as $dp(i) \subseteq \{i-1:i\}$ and $cp(i) \subseteq \{1:i-1\}$. We explain the reason behind this indexing in the next section.

4.2.1 The parental set

Suppose each univariate series $y_{i,t}$ is governed by a multiple regression DLM with a two-set of predictors, i.e., the autoregressive and the cross-lagged variables and the current values of the variables. It is just known that the order of the process does not exceed p and otherwise no prior knowledge of possible sparsity (absence of dynamic and contemporaneous effects) is available. To simplify the problem, we include only lower bidiagonal elements of the dynamic parents as predictors.[2] For each $i = 2 : N$, the lower bidiagonal elements of a vector autoregressive process of order one and a triangular system, even if we do not take into account the intercept terms, there exist $i-1:i$ bidiagonal dynamic and $i-1$ contemporaneous coefficients. Note that the set with $i-1:i$ distinct elements always consists of two elements except for the first variable, which contains one element (if the model order is one). Then, for simplicity, we may replace the elements of the set $\{i-1:i\}$ with 2. In this case, $\binom{2}{r}$ and $\binom{i-1}{r}$ subset with r elements can be chosen. Hence, there is a total of

$$2 + \sum_{i=2}^{N} \sum_{r=0}^{2} \binom{2}{r} = 2 + (N-1)2^2 \tag{4.3}$$

[2]If we consider the full set of dynamic parents, this requires performing graphical variable selection over 2^{Np} for each series. This property is infeasible to handle the system with many variables.

and

$$\sum_{i=1}^{N}\sum_{r=0}^{i-1}\binom{i-1}{r} = 2^N - 1, \tag{4.4}$$

subset models.

Our dynamic graphical model with dynamic and contemporaneous parents is a specification of the conditional density function $f(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1})$ for each series that incorporates a subset of the dynamic pairwise conditional independence structures $y_{i,t} \perp\!\!\!\perp x_{j,t}|(x_{dp(i),t}, y_{cp(i),t})$, where $dp(i) = \{i-1 : i\} \setminus \{j\}$ and $cp(i) = \{1, \dots i-1\} \setminus \{i\}$ with "\" meaning "excludes", and the contemporaneous pairwise conditional independence structures $y_{i,t} \perp\!\!\!\perp y_{j,t}|(\mathbf{x}_{dp(i),t}, \mathbf{y}_{cp(i),t})$, where $cp(i) = \{1, \dots, i-1\} \setminus \{i, j\}$.

Including the lag operator in the structure, this corresponds to perform graphical variable selection over a power set

$$\mathcal{P}(S) = 2^p + \sum_{i=2}^{N} 2^{2p+i-1} = 2^p + 2^{2p}(2^N - 2), \tag{4.5}$$

in parallel.[3] In this case, all possible $2^p + 2^{2p}(2^N - 2)$ model specifications can be evaluated using 2^{2p+i-1} subspace of models for each series in parallel and variable selection is equivalent to graphical model selection conducted on both the elements of a bidiagonal matrix and the elements of a strictly lower triangular covariance matrix.

[3]If we consider the full multivariate model, the total number of independence under consideration would be $\binom{N+Np}{2} = \binom{N}{2} + N^2p + \binom{Np}{2}$, including the examination of $\binom{Np}{2}$ independence among the dynamic vector autoregressive elements (which we ignore here). By definition, this equation holds true using the multinomial theorem for a lower factorial polynomial $(N)_2 = N(N-1)/2$ similar to umbral calculus as stated in Weisstein (2003), page 237. The total number of graphical models in this setup is $2^{\binom{N}{2}+N^2p}$. Inference in this multivariate model is obviously infeasible.

4.3 The Model

The model design consists of three steps. We show the component multiple regression DLMs, the joint TVP-VAR-VD model followed by the graph representation of the TVP-VAR-VD model.

4.3.1 Component multiple regression DLM

For each $i = 1 : N$, the dynamics of $y_{i,t}$ follow a univariate multiple regression DLM and elements of a triangular system that incorporate autoregressive effects, cross-lagged effects and contemporaneous effects,

$$\begin{aligned} y_{i,t} &= \mathbf{F}'_{pa(i),t}\boldsymbol{\theta}_{i,t} + \nu_{i,t}, \\ &= \mathbf{x}'_{dp(i),t}\boldsymbol{\beta}_{i,t} + \mathbf{y}'_{cp(i),t}\boldsymbol{\gamma}_{i,t} + \nu_{i,t}, \quad \nu_{i,t} \sim \mathcal{N}(0, 1/\lambda_{i,t}), \end{aligned} \tag{4.6}$$

with the state coefficients changing slowly over time according to random walk

$$\boldsymbol{\theta}_{i,t} = \boldsymbol{\theta}_{i,t-1} + \boldsymbol{\omega}_{i,t}, \quad \boldsymbol{\omega}_{i,t} \sim \mathcal{N}(0, \mathbf{W}_{i,t}), \tag{4.7}$$

where the state coefficients $\boldsymbol{\theta}_{i,t}$ are defined as column vectors, and the observational error terms $\nu_{i,t}$ and the evolution error terms $\boldsymbol{\omega}_{i,t}$ are assumed to be independent for all series i, independent over time such that $\nu_{i,t}$ and $\nu_{j,s}$ are independent, $\boldsymbol{\omega}_{i,t}$ and $\boldsymbol{\omega}_{j,s}$ are independent with $i \neq j$ and for all t and s.

We may partition the regression vectors and the state vectors as

$$\mathbf{F}_{pa(i),t} = \begin{pmatrix} \mathbf{x}_{dp(i),t} \\ \mathbf{y}_{cp(i),t} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\theta}_{i,t} = \begin{pmatrix} \boldsymbol{\beta}_{i,t} \\ \boldsymbol{\gamma}_{i,t} \end{pmatrix}, \tag{4.8}$$

where $\mathbf{x}_{dp(1),t} = (y_{1,t-1:t-p})'$ and for $i = 2 : N$, $\mathbf{x}_{dp(i),t} = (y_{i-1,t-1:t-p} : y_{i,t-1:t-p})'$ denote the dynamic parental series defined as column vectors. For $i = 2 : N$, $\mathbf{y}_{cp(i),t} = (y_{1,t}, \ldots, y_{i-1,t})'$ is the contemporaneous parental series with $\mathbf{y}_{cp(1),t} =$

$\{\emptyset\}$ being an empty set, $\mathbf{y}_{cp(2),t} = y_{1,t}$ and for $i = 3 : N$, $\mathbf{y}_{cp(i),t}$ is defined as column vectors, and $\mathbf{F}_{pa(1),t} = \mathbf{x}_{dp(1),t} = (y_{1,t-1:t-p})'$. The dimension of the time-varying coefficients $\boldsymbol{\gamma}_{i,t}$ is $P_{i,\gamma} = |cp(i)|$. The dimension of the time-varying coefficients $\boldsymbol{\beta}_{i,t}$ is $P_{i,\beta} = |dp(i)|$. The joint dimension of the coefficients is $P_{i,\beta,\gamma} = P_{i,\beta} + P_{i,\gamma}$, where for $i = 1$ it is p and for each series $i = 2 : N$ it is $2p + i - 1$.

Suppose that $\lambda_{i,t}$ is unknown and is subject to random changes over time. Following West and Harrison (1997), Prado and West (2010) and Zhao et al. (2016), the stochastic variation in the precision $\lambda_{i,t}$ can be modelled via a form of multiplicative Beta-Gamma random walk with volatility discounting

$$\lambda_{i,t} = \frac{\alpha_{i,t}\lambda_{i,t-1}}{\varphi_i}, \tag{4.9}$$

where the precision $\lambda_{i,t}$ have a Gamma distribution and innovations $\alpha_{i,t}$ follow a Beta distribution for some discount factors $\varphi_i \in (0, 1]$.

Furthermore, we assume that the component state evolution variance matrices $\mathbf{W}_{i,t}$ for $t = 1, \ldots, T$ are unknown. A reliable model specification requires the component evolution variance matrix $\mathbf{W}_{i,t}$ to be controlled due to its effects on the time-variation in $\boldsymbol{\theta}_{i,t}$. This can be achieved by a component discount factor $\delta_i \in (0, 1]$, which follows West and Harrison (1997).

4.3.2 Joint TVP-VAR-VD model

The model in Equation (4.6) through the partitioned form of the regression vectors and the state vectors can be written as a joint TVP-VAR-VD model as follows

$$(I - \boldsymbol{\Gamma}_t)\mathbf{y}_t = \sum_{k=1}^{p} \mathbf{B}_{k,t} y_{t-k} + \boldsymbol{\epsilon}_t, \tag{4.10}$$

where $\mathbf{B}_{k,t}$ is an $N \times N$ matrix of the time-varying parameters at time t and at lag k for $k = 1 : p$. The matrix $\mathbf{B}_{k,t}$ contains only the lower bidiagonal elements

of the time-varying parameters, which can be represented as

$$\mathbf{B}_{k,t} = \begin{pmatrix} b_{11,k,t} & 0 & \dots & 0 & 0 \\ b_{21,k,t} & b_{22,k,t} & 0 & \vdots & 0 \\ 0 & b_{32,k,t} & b_{33,k,t} & \vdots & 0 \\ \vdots & 0 & \ddots & b_{N-1N-1,k,t} & 0 \\ 0 & \dots & 0 & b_{NN-1,k,t} & b_{NN,k,t} \end{pmatrix}, \tag{4.11}$$

To establish a unique solution to the system, we further assume $\mathbf{\Gamma}_t$ is a strictly lower triangular matrix with elements $\boldsymbol{\gamma}_{i,t}$ below the diagonal and 0s along the principal diagonal, that is

$$\mathbf{\Gamma}_t = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \gamma_{21,t} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{N1,t} & \dots & \gamma_{NN-1,t} & 0 \end{pmatrix}, \quad \text{and} \tag{4.12}$$

$$\mathbf{\Lambda}_t = \begin{pmatrix} \lambda_{1,t} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_{N,t} \end{pmatrix}, \tag{4.13}$$

where the $\gamma_{i,j,t}$ elements are real-valued and the $\lambda_{i,t}$ are positive-valued. Recall that a model is recursive if the relationships among variables in $\mathbf{y}_t$ are one directional. A recursive model corresponds to an acyclic graph if $\mathbf{\Gamma}_t$ is defined as a strictly lower triangular matrix. In addition, defining $\mathbf{\Lambda}_t$ as a diagonal matrix is similar to a Gaussian directed graphical model as noted by Murphy (2012).

Furthermore, the model in Equation (4.10) can be written jointly in a compact form as

$$\mathbf{y}_t = \mathbf{B}_t\mathbf{x}_t + \mathbf{\Gamma}_t\mathbf{y}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{\Lambda}_t^{-1}), \tag{4.14}$$

where $\mathbf{B}_t$ is an $N \times Np$ matrix, $\mathbf{x}_t$ is an $Np \times 1$ vector, $\mathbf{\Gamma}_t$ is an $N \times N$ matrix, $\boldsymbol{\epsilon}_t$ is an $N \times 1$ vector, and $\mathbf{\Lambda}_t = \text{diag}(\lambda_{1,t}, \ldots, \lambda_{N,t})$. The matrices $\mathbf{B}_t$ and $\mathbf{\Gamma}_t$ contain the direct dynamic and the direct contemporaneous effect coefficients, respectively.

When working with graphical models, we may convert the joint structure to a reduced form. Solving for $\mathbf{y}_t$ in Equation (4.14), the reduced form for the model is

$$\mathbf{y}_t = (\mathbf{I} - \mathbf{\Gamma}_t)^{-1}\mathbf{B}_t\mathbf{x}_t + (\mathbf{I} - \mathbf{\Gamma}_t)^{-1}\boldsymbol{\epsilon}_t, \tag{4.15}$$

where $\mathbf{y}_t \sim \mathcal{N}((\mathbf{I} - \mathbf{\Gamma}_t)^{-1}\mathbf{B}_t\mathbf{x}_t, \mathbf{\Sigma}_t)$ with

$$\mathbf{\Omega}_t = \mathbf{\Sigma}_t^{-1} = (\mathbf{I} - \mathbf{\Gamma}_t)'\mathbf{\Lambda}_t(\mathbf{I} - \mathbf{\Gamma}_t). \tag{4.16}$$

4.3.3 Graph representation of the joint TVP-VAR-VD model

We may define the dynamic dependence structure by an indicator matrix $\mathbf{\Xi}_{B,k}$, such that $\xi_{\beta,ij,k} = 1$ if $j \in dp(i)$, otherwise $\xi_{\beta,ij,k} = 0$

$$\mathbf{\Xi}_{B,k} = \begin{pmatrix} \xi_{\beta,11,k} & 0 & \ldots & 0 & 0 \\ \xi_{\beta,21,k} & \xi_{\beta,22,k} & 0 & \vdots & 0 \\ 0 & \xi_{\beta,32,k} & \xi_{\beta,33,k} & \vdots & 0 \\ \vdots & 0 & \ddots & \xi_{\beta,N-1N-1,k} & 0 \\ 0 & \ldots & 0 & \xi_{\beta,NN-1,k} & \xi_{\beta,NN,k} \end{pmatrix},$$

and define the contemporaneous dependence structure by an indicator matrix $\mathbf{\Xi}_\gamma$, such that $\xi_{\gamma,ij} = 1$ if $j \in cp(i)$, otherwise $\xi_{\gamma,ij} = 0$

$$\mathbf{\Xi}_\gamma = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \xi_{\gamma,21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \xi_{\gamma,N1} & \dots & \xi_{\gamma,NN-1} & 0 \end{pmatrix},$$

then, we may write the graph representation of the model in Equation (4.14) as

$$\mathbf{y}_t = (\mathbf{\Xi}_B \circ \mathbf{B}_t)\mathbf{x}_t + (\mathbf{\Xi}_\gamma \circ \mathbf{\Gamma}_t)\mathbf{y}_t + \boldsymbol{\epsilon}_t, \tag{4.17}$$

and the reduced form model becomes

$$\mathbf{y}_t = (\mathbf{I} - \mathbf{\Xi}_\gamma \circ \mathbf{\Gamma}_t)^{-1}(\mathbf{\Xi}_B \circ \mathbf{B}_t)\mathbf{x}_t + (\mathbf{I} - \mathbf{\Xi}_\gamma \circ \mathbf{\Gamma}_t)^{-1}\boldsymbol{\epsilon}_t, \tag{4.18}$$

where the operator $\circ$ corresponds to the element by element multiplication.

To summarize, in a multivariate TVP-VAR-VD model, the contemporaneous conditional independence structures correspond to zeros in $\mathbf{\Sigma}_t^{-1}$ and the dynamic conditional independence structure corresponds to zeros in $\mathbf{\Sigma}_t^{-1}(\mathbf{I} - \mathbf{\Gamma}_t)^{-1}\mathbf{B}_t$. On the other hand, in the multiple regression DLM, the conditional independence (for each i) refers simply to zeros in the coefficient states, $\boldsymbol{\beta}_{i,t}$ and $\boldsymbol{\gamma}_{i,t}$. For details on DLM models in the context we use here, see Zhao et al. (2016), West and Harrison (1997), Chapters 4, 6 and 12 and Prado and West (2010), Chapter 4, and for details on graphical models in multivariate and multiple regression models see Whittaker (2008), Chapter 10.

4.4 Bayesian inference

Under Bayesian paradigm, our inferential problem can be solved by computing the component posterior distributions of the state parameters and the component predictive distributions of the observed data. Parallel computations follow a sequential updating and an online forecasting steps of the locally structured multiple regression DLMs. Inference is based on sequential updating of the component DLM structures by running N Kalman filter algorithm in parallel.

4.4.1 The Kalman filter algorithm

As a first step, the Kalman filter algorithm aims to sequentially compute the component posterior distribution $p(\boldsymbol{\theta}_{i,t}, \lambda_{i,t}|\mathbf{y}_{1:t})$ of the states $\boldsymbol{\theta}_{i,t}$ and $\lambda_{i,t}$ given observation $\mathbf{y}_{1:t}$ up to time t together with an observation relation $p(y_{i,t}|\mathbf{y}_{1:t-1})$ in $i = 1 : N$ parallel steps.

Let the posterior of the component state vector $\boldsymbol{\theta}_{i,t-1}$ for information up to time $t-1$ has a Normal distribution

$$(\boldsymbol{\theta}_{i,t-1}|\lambda_{i,t-1}, \mathbf{y}_{1:t-1}) \sim \mathcal{N}(\mathbf{m}_{i,t-1}, \mathbf{C}_{i,t-1}),$$

with parameters $\mathbf{m}_{i,t-1}$ and $\mathbf{C}_{i,t-1}$. The prior for the component state vector $\boldsymbol{\theta}_{i,t}$ has a Normal distribution

$$(\boldsymbol{\theta}_{i,t}|\lambda_{i,t}, \mathbf{y}_{1:t-1}) \sim \mathcal{N}(\mathbf{a}_{i,t}, \mathbf{R}_{i,t}),$$

with parameters $\mathbf{a}_{i,t}$ and $\mathbf{R}_{i,t}$ and the one-step-ahead predictive distribution of each observation

$$(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) \sim \mathcal{T}(f_{i,t}, q_{i,t}),$$

is a Student T distribution with parameters $f_{i,t}$ and $q_{i,t}$.

The filtering density of $\boldsymbol{\theta}_{i,t}$ given $\mathbf{y}_{1:t}$ is a Normal distribution

$$(\boldsymbol{\theta}_{i,t}|\lambda_{i,t}, \mathbf{y}_{1:t}) \sim \mathcal{N}(\mathbf{m}_{i,t}, \mathbf{C}_{i,t}),$$

with parameters $\mathbf{m}_{i,t}$ and $\mathbf{C}_{i,t}$.

Another key step is to decide on the stochastic evolution of the precision. A formal description of the variance discounting can be found in Uhlig (1994) and West and Harrison (1997). At time $t-1$, the precision $\lambda_{i,t-1}$ have Gamma distribution with posterior

$$(\lambda_{i,t-1}|\mathbf{y}_{1:t-1}) \sim \mathcal{G}(n_{i,t-1}/2, d_{i,t-1}/2),$$

and parameters $n_{i,t-1}/2$ and $d_{i,t-1}/2$, where $d_{i,t-1} = n_{i,t-1}s_{i,t-1}$, $n_{i,t-1}$ is the degrees of freedom and $s_{i,t-1}$ is the variance of the observations. Proceeding to time t, let's introduce the variance discount factor φ_i. Based on the posterior at the previous time $t-1$, suppose that $\lambda_{i,t}$ is obtained from $\lambda_{i,t-1}$ by some random walk model as in Equation (4.9). The implied prior at time t is a Gamma distribution

$$(\lambda_{i,t}|\mathbf{y}_{1:t-1}) \sim \mathcal{G}(\varphi_i n_{i,t-1}/2, \varphi_i d_{i,t-1}/2),$$

the prior for $\alpha_{i,t}$ have a Beta distribution

$$(\alpha_{i,t}|\mathbf{y}_{1:t-1}) \sim \mathcal{B}e(\varphi_i n_{i,t-1}/2, (1-\varphi_i)n_{i,t-1}/2),$$

and the filtering density of $\lambda_{i,t}$ given $\mathbf{y}_{1:t}$ is a Gamma distribution

$$(\lambda_{i,t}|\mathbf{y}_{1:t}) \sim \mathcal{G}(n_{i,t}/2, d_{i,t}/2).$$

The implication from the discount factor, $\varphi_{i,}$, is that as the value of φ_i gets larger, the random disturbance to the observational variance gets smaller at each

time, Prado and West (2010). Hence, the multiplicative random walk model in Equation (4.9) may be applied to model stochastic volatility in the observation precision over time, West and Harrison (1997) and Prado and West (2010).

Another unknown element of the DLM is the evolution variance matrix $W_{i,t}$. By definition, the posterior of the state vector has a variance matrix $\mathbf{P}_{i,t} = V(\boldsymbol{\theta}_{i,t-1}|\mathbf{y}_{1:t-1}) = \mathbf{C}_{i,t-1}$ at time $t-1$. Proceeding to time t, the prior of the state vector has a variance matrix $\mathbf{R}_{i,t} = V(\boldsymbol{\theta}_{i,t}|\mathbf{y}_{1:t-1}) = \mathbf{P}_{i,t} + \mathbf{W}_{i,t}$. This process describes the relationship between the state vectors $\boldsymbol{\theta}_{i,t-1}$ and $\boldsymbol{\theta}_{i,t}$ due to the increase in uncertainty in moving from $\mathbf{P}_{i,t}$ to $\mathbf{R}_{i,t} = \mathbf{P}_{i,t} + \mathbf{W}_{i,t}$. West and Harrison (1997), Section 6.3, recommend to set $\mathbf{R}_{i,t} = \mathbf{P}_{i,t}/\delta_i$ for a discount factor $\delta_i \in (0, 1]$. Hence, the evolution variance matrix for each i is

$$\mathbf{W}_{i,t} = \frac{1-\delta_i}{\delta_i}\mathbf{P}_{i,t},$$

where the discount factor δ_i associated with the series i component model controls the evolution variance matrix.

4.4.2 The compositional form of the joint model

To construct a joint model from the prediction and filtering (updating) steps of the Kalman filter algorithm, we compute the compositional form of the joint density function following West and Harrison (1997), Chapter 9 and Zhao et al. (2016). To predict y_t at time t, we assume that for each $i = 1 : N$, the dynamic and the contemporaneous parental predictors, $\mathbf{F}_{pa(i),t} = (\mathbf{x}_{dp(i),t}, \mathbf{y}_{cp(i),t})'$, are uncertain with a density function $p(\mathbf{F}_{pa(i),t}|\mathbf{y}_{1:t-1})$. Then, the compositional joint predictive density of $\mathbf{y}_t$ can be computed as

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \int \cdots \int \prod_{i=1}^{N} p(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) p(\mathbf{F}_{pa(i),t}|\mathbf{y}_{1:t-1}) d\mathbf{F}_{pa(i),t},$$

where $p(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1})$ is the one-step-ahead predictive Student T distribution. The joint predictive moments, the mean vector and the variance matrix

$$\mathbf{f}_t = E(\mathbf{y}_t|\mathbf{y}_{1:t-1}), \quad \text{and} \quad \mathbf{Q}_t = V(\mathbf{y}_t|\mathbf{y}_{1:t-1}),$$

are calculated in a compositional form for each $i = 1 : N$, where the elements of the vector $\mathbf{f}_t$ and the elements of the matrix $\mathbf{Q}_t$ are obtained by a plug-in rule following the methods developed by Zhao et al. (2016). The elements of the vector $\mathbf{f}_t$ and the elements of the matrix $\mathbf{Q}_t$ are plugged-in by updating the values of $f_{i,t}$ and $q_{i,t}$ to take into account the uncertainty of the dynamic and the contemporaneous parental predictors. Hence, the values of $f_{i,t}$ and $q_{i,t}$ that are obtained from the initial Kalman filter algorithm are not the same as the values that are used in the compositional form of the model. For all $i = 1 : N$, the Student T distribution has the form $(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) \sim \mathcal{T}_{n_{i,t}}(f_{i,t}, q_{i,t})$, on $n_{i,t}$ degrees of freedom, mean $f_{i,t}$ and scale $q_{i,t}$. If $n_{i,t} > 1$, $E(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) = f_{i,t}$. In addition, if $n_{i,t} > 2$, $V(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) = \frac{n_{i,t}}{n_{i,t}-2} q_{i,t}$. We show the functional forms of the updated values of $f_{i,t}$ and $q_{i,t}$ in the Appendix, Section 4.8.3.

The next step is to derive the h-step-ahead forecast density for the state vectors $p(\boldsymbol{\theta}_{i,t+h}, \lambda_{i,t+h}|\mathbf{y}_{1:t})$ and the h-step-ahead predictive density $p(y_{i,t+h}|\mathbf{y}_{1:t})$. We compute iterated forecasts for horizons between 1 and h. For the multi-step-ahead forecast, analytical solutions do not exist. Consequently, we evaluate the h-step-ahead of the posteriors of the states and the predictive density using Monte Carlo simulations.

The compositional form of the recursive model has the following patterns. We first sample the Normal-Gamma posterior $p(\boldsymbol{\theta}_{1,t}, \lambda_{1,t}|\mathbf{y}_{1:t})$ at time t to obtain samples from $p(\boldsymbol{\theta}_{1,t+1}, \lambda_{1,t+1}|\mathbf{y}_{1:t})$ and $p(y_{1,t+1}|\mathbf{y}_{1:t})$ at time $t+1$. In the next step, we move on series $i-1$ to sample from $p(\mathbf{y}_{1:i-1,t+1}|\mathbf{y}_{1:t})$. By this recursive process, we can generate full Monte Carlo samples from $p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})$ and can proceed

to generate full Monte Carlo samples from $p(\mathbf{y}_{t+1:t+h}|\mathbf{y}_{1:t})$. For full details and functional forms see Appendix, Section 4.8.1 through 4.8.3.

4.5 Bayesian graphical variable selection

In this section, we outline a practical approach to graphical variable search and selection, model selection and implied model averaging.

4.5.1 The model space

A large of number of papers has emphasized the challenging task of Bayesian variable selection when comparing a large number of models, key references include Madigan and Raftery (1994), Hoeting et al. (1999), Berger and Molina (2005) and Heaton and Scott (2010). We have already introduced a source of uncertainty in our modelling approach, i.e., the uncertainty of the unknown sparsity features of the dynamic and the contemporaneous parental predictors. We approach a practical way to specify a probability model $p(M)$ over the space of all models by denoting component model probabilities $p(M_i)$ over a subspace of models. Let's assume we have a set of $M = \{M_1, \ldots, M_N\}$ models over the space of all competing models, with $M_1 = M_{11:1K_1}$ and $K_1 = 2^p$ and for each $i = 2 : N$, $M_i = M_{i1:iK_i}$ and $K_i = 2^{2p+i-1}$ denote the subspace of models based on groupings of dynamic $dp(i)$ and contemporaneous parents $cp(i)$ treated as predictors. To implement a Bayesian graphical variable selection procedure for a recursively ordered variables described in Section 4.2 and 4.3, we propose an efficient variable search and selection algorithm for finding best and a range of good multiple regression DLMs. This procedure is able to compute posterior model probabilities over the space of all models by visiting the subspace of models in parallel.

We achieve this approach by ordering the graphical models in a Gray cyclic binary code order inspired from Cameron (1994) and Murphy (2012) and based

on an algorithm proposed by Boothroyd (1964). The Gray code algorithm shows a logical vector within the Gray cyclic binary code and has a beautiful interpretation in graph theory, Cameron (1994), Section 11.6. The algorithm starts from a null model and follows a path of vectors, which differ by dimension $1 \times p$ if $i = 1$ and $1 \times (2p + i - 1)$ if $i \geq 2$. The process changes one element of the vector to form the next by adding or removing a single variable. For example, if $i = 1$ and $p = 2$, the number of graphical models is $2^p = 4$ for a dimension of parental predictors $1 \times p$, the Gray cyclic binary code vector representation would be $\{0, 0\}$,$\{1, 0\}$,$\{1, 1\}$,$\{0, 1\}$.[4] The corresponding family of all models in the model subspace M_1 is

$$M_1 \equiv \{M_{1,(0,0)}, M_{1,(1,0)}, M_{1,(1,1)}, M_{1,(0,1)}\},$$

where $M_{1,(0,0)}$ refers to the null model. The component model $M_{1,(1,0)}$ is an alternative hypothesis that $dp(1) = y_{1,t-1}$ indicating a non-zero effect of the dynamic predictor $y_{1,t-1}$, $M_{1,(1,1)}$ is an alternative hypothesis that $dp(1) = (y_{1,t-1}, y_{1,t-2})$ defining non-zero effects of both dynamic predictors, and $M_{1,(0,1)}$ is an alternative hypothesis that $dp(1) = y_{1,t-2}$ pointing out a non-zero effect of the dynamic predictor $y_{1,t-2}$.[5]

We may analyse the models in the model subspaces for $i = 1 : N$. Assuming $N = 10$ and $p = 2$, the candidate models M_i are shown as a set of $|M_i|$ models over the subspace of models, and the cardinality of the parental sets are given in Table 4.1. The first row in Table 4.1 shows the order of models in the model subspaces. The second row evaluates the set of candidate models to be enumerated. Potential number of elements of the dynamic and contemporaneous parental sets.

[4]The Gray code representation $\{0, 0\}$,$\{1, 0\}$,$\{1, 1\}$,$\{0, 1\}$ is analogous to a Hamiltonian circuit in a hypercube graph representing the 2^p vector of binary variables for the number of elements p. A Hamiltonian circuit in a hypercube graph is a sequence of vector of binary variables that each vector varies in one position from the preceding vector, and the last vector varies from the first in one position, see Theorem 11.6.1 in Cameron (1994).

[5]We use a notation inspired from Barbieri and Berger (2004) and Müller et al. (2010) in defining the model space and subspace.

Model subspaces and parental sets										
Order	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
$\lvert M_i \rvert$	4	32	64	128	256	512	1024	2048	4096	8192
$\lvert dp(i) \rvert$	2	4	4	4	4	4	4	4	4	4
$\lvert cp(i) \rvert$		1	2	3	4	5	6	7	8	9

Table 4.1: The table shows order of models in the model subspaces, number of potential models to be enumerated, number of potential dynamic and contemporaneous parents.

4.5.2 A collection of priors for vector of binary indicator variables

A question arises, how to elicit the prior probabilities for the graphical models themselves? In high-dimensional problems, there is a consensus surrounding the use of "variable selection prior" with the elements of each of row vectors of $\boldsymbol{\Xi}_{B,k}$ and $\boldsymbol{\Xi}_{\gamma}$ are assumed to arise as a sequence of Bernoulli trials. Our aim is to discover models with high posterior probability. Thus, we focus on a nontraditional search algorithm that finds models with high posterior probability.[6] We list the models and their posterior model probabilities over the space of all competing models for each series i in parallel and compute the score for each one as the logarithm of the posterior probabilities.

In our Bayesian graphical variable selection, each row of $\boldsymbol{\Xi}_{B,k}$ and $\boldsymbol{\Xi}_{\gamma}$ are themselves defined as random vector of binary indicator variables. We define the prior for each row of the binary indicator matrix $\boldsymbol{\Xi}_{B,k}$ such that the columns of the $\mathbf{x}_{dp(i),t}$ are included in the multiple regression DLM if $\xi_{\beta,ij,k} = 1$ and excluded

[6] A nontraditional search algorithm refers to an approach that does not rely on using Markov chain Monte Carlo (MCMC) algorithms, such as the Gibbs Sampler or the Metropolis-Hastings algorithm. In a discrete state space, MCMC algorithms have proved to be inefficient. For a review see Heaton and Scott (2010) and for more details Murphy (2012).

if $\xi_{\beta,ij,k} = 0$. Then, for each $i = 2 : N$, we assume $y_{i,t-k}$ or $y_{j,t-k}$ is parent of $y_{i,t}$

$$\xi_{\beta,ij,k} = \begin{cases} 1, & \text{if } y_{i,t-k} \to y_{i,t} \quad \text{or} \quad y_{j,t-k} \to y_{i,t} \\ 0, & \text{otherwise,} \end{cases} \tag{4.19}$$

The number of dynamic parents that are included in the model, is given by

$$D_{\beta,i} = \sum_{j=i-1}^{i} \xi_{\beta,ij,k}, \tag{4.20}$$

where $D_{\beta,i} \leq |R_{dp(i)}|$.

As noted by Madigan and Raftery (1994), Hoeting et al. (1999), Cripps et al. (2005), and Zhao et al. (2016), a prior probability choice for the size of the graph on each dynamic parental series inclusion indictor may be a Bernoulli prior probability of the form

$$p(\xi_{\beta,i}) = \prod_{j=i-1}^{i} \mathcal{B}\text{er}(\xi_{\beta,ij,k}|\pi) = \pi^{D_{\beta,i}}(1-\pi)^{|R_{dp(i)}|-D_{\beta,i}}, \tag{4.21}$$

where we may interpret π as the probability that any two variables have a pairwise dependence structure with a maximum possible number of dependence $|R_{dp(i)}|$.

In a similar fashion, the prior for each row of the binary indicator matrix $\boldsymbol{\Xi}_\gamma$ such that the columns of the $\mathbf{y}_{cp(i),t}$ are included in the multiple regression DLM if $\xi_{\gamma,ij} = 1$ and excluded if $\xi_{\gamma,ij} = 0$ as

$$\xi_{\gamma,ij} = \begin{cases} 1, & \text{if } y_{j,t} \to y_{i,t} \\ 0, & \text{otherwise,} \end{cases} \tag{4.22}$$

For each $i = 2 : N$, the number of contemporaneous parental series that are

included in the model, is given by

$$D_{\gamma,i} = \sum_{j=1}^{i-1} \xi_{\gamma,ij}, \tag{4.23}$$

where $D_{\gamma,i} \leq i-1$.

The prior for the size of the graph on each contemporaneous parental series inclusion indictor is a Bernoulli prior of the form

$$p(\xi_{\gamma,i}) = \prod_{j=1}^{i-1} \mathcal{B}\text{er}(\xi_{\gamma,ij}|\pi) = \pi^{D_{\gamma,i}}(1-\pi)^{i-1-D_{\gamma,i}}. \tag{4.24}$$

Those two steps can be performed in one step for both the dynamic and contemporaneous parental series and for each p, if $i = 1$

$$p(\xi_{\beta,\gamma,i}) = \prod_{j=1}^{p} \mathcal{B}\text{er}(\xi_{\beta,\gamma,ij}|\pi) = \pi^{D_{\beta,\gamma,i}}(1-\pi)^{p-D_{\beta,\gamma,i}},$$

otherwise,

$$p(\xi_{\beta,\gamma,i}) = \prod_{j=1}^{2p+i-1} \mathcal{B}\text{er}(\xi_{\beta,\gamma,ij}|\pi) = \pi^{D_{\beta,\gamma,i}}(1-\pi)^{2p+i-1-D_{\beta,\gamma,i}}, \tag{4.25}$$

where if $i = 1$, $D_{\beta,\gamma,i} = \sum_{j=1}^{p} \xi_{\beta,\gamma,ij}$ otherwise, $D_{\beta,\gamma,i} = \sum_{j=1}^{2p+i-1} \xi_{\beta,\gamma,ij}$ and $D_{\beta,\gamma,i}$ is the number of dynamic and contemporaneous parents that are included in the model. In addition, if $i = 1$, $D_{\beta,\gamma,i} \leq p$ otherwise, $D_{\beta,\gamma,i} \leq 2p + i - 1$.

4.5.3 Priors and initial values for parameters

Analysing DLMs componentwise requires prior distributions to be specified for each individual component model by assuming independence between components as noted by West and Harrison (1997). At an initial time $t = 0$, the state vectors $\boldsymbol{\theta}_{i,0}$ and $\lambda_{i,0}$ are assumed to follow a Normal and a Gamma initial distributions,

respectively. The specified moments are

$$(\boldsymbol{\theta}_{i,0}|\mathbf{y}_0) \sim \mathcal{N}(\boldsymbol{\theta}_{i,0}|\mathbf{m}_{i,0}, \mathbf{C}_{i,0}), \tag{4.26}$$

$$(\lambda_{i,0}|\mathbf{y}_0) \sim \mathcal{G}(\lambda_{i,0}|n_{i,0}, d_{i,0}). \tag{4.27}$$

Following Koop and Korobilis (2013) and Zhao et al. (2016), for each $i = 1 : N$, we set the prior hyperparameters for $\mathbf{m}_{i,0} = (0, \ldots, 0)'$ and the initial values of $\mathbf{m}_{i,t}$ at time $t = 0$ are set to $\mathbf{m}_{i,0}$. The prior hyperparameters for the state covariance matrices $\mathbf{C}_{i,0} = \mathbf{I}$ and the initial values of $\mathbf{C}_{i,t}$ at time $t = 0$ is equal to $\mathbf{C}_{i,0}$. The prior hyperparameters of the degrees of freedom is set to $n_{i,0} = (1-\delta_1)^{-1} - p$ if $i = 1$, otherwise $n_{i,0} = (1-\delta_1)^{-1} - (2p+i-1)$, and similarly $d_{i,0} = (1-\delta_1)^{-1} - p$ if $i = 1$, otherwise $d_{i,0} = (1-\delta_1)^{-1} - (2p+i-1)$ leading to a prior hyperparameter value of the estimate of the observational variance $s_{i,0} = 1$.[7] And the implied sequence $\mathbf{W}_{i,t}$ for $t = 1, \ldots, T$ are identified given δ_i and $\mathbf{C}_{i,0}$. The discount factors for the residual volatilities φ_i and the discount factors for the variance of the states δ_i are evaluated over a grid of values that we describe in Section 4.6.

4.5.4 Predictive likelihood functions

The predictive likelihood function of the joint model defined in terms of the observed predictive density can be decomposed as the product of predictive likelihood functions of the M_i models at each point in time over $t = 1 : T$ as

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = p(y_{1,t}|\mathbf{F}_{pa(1),t}, \mathbf{y}_{1:t-1}) \prod_{i=2}^{N} p(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1})$$

[7] The value of the hyperparameter δ_1 is set to the value in Table 4.3.

where the predictive likelihood is defined as a univariate Student T distribution of the predictive density $p(y_{i,t}|\mathbf{F}_{pa(i)}, \mathbf{y}_{1:t-1})$

$$L_{i,t} = \frac{\Gamma[(n_{i,t-1}+1)/2]}{\Gamma[n_{i,t-1}/2]\sqrt{n_{i,t-1}q_{i,t}}}\left\{1 + \frac{(y_{i,t}-f_{i,t})^2}{n_{i,t-1}q_{i,t}}\right\}^{-(n_{i,t-1}+1)/2}, \tag{4.28}$$

and

$$(y_{i,t}|\mathbf{F}_{pa(i)}, \mathbf{y}_{1:t-1}) \sim \mathcal{T}_{n_{i,t-1}}(f_{i,t}, q_{i,t}). \tag{4.29}$$

for $t = 1, \ldots, T$.

4.5.5 Posterior over the model space

The posterior probability $p(M|\mathbf{y}_{1:t})$ over the space of all models M can be decomposed as a product of component posterior probabilities $p(M_i|y_{i,1:t})$ over the subspace of models M_i as

$$p(M|\mathbf{y}_{1:t}) \propto \prod_{i=1}^{N} p(M_i|y_{i,1:t}), \tag{4.30}$$

where for each $i = 1:N$ and $t = 1:T$, if $i = 1$

$$p(M_i|y_{i,1:t}) \propto p(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1})\pi^{D_{i,\beta}}(1-\pi)^{p-D_{i,\beta}},$$

otherwise

$$p(M_i|y_{i,1:t}) \propto p(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1})\pi^{D_{i,\beta,\gamma}}(1-\pi)^{2p+i-1-D_{i,\beta,\gamma}}. \tag{4.31}$$

4.5.6 Bayesian model selection

We have now identified the various elements of the Bayesian graphical variable selection problem, i.e. the full model space M and the subspace of models M_i, the two components of the prior probabilities, the priors for the vector of binary

indicator variables and the priors for the parameters, the predictive likelihood function of each model and the posteriors over the model space. In a procedure of identifying models in terms of posterior probabilities, we assume the model is another unknown parameter and obtain the posterior probabilities of the models under considerations. First, we define the marginal predictive density of the data as

$$p(y_{i,t}|\mathbf{y}_{1:t-1}) = \int p(y_{i,t}|\mathbf{F}_{pa(i)}, \boldsymbol{\theta}_{i,t}, \lambda_{i,t}) p(\boldsymbol{\theta}_{i,t}, \lambda_{i,t}|\mathbf{y}_{1:t-1}) d\boldsymbol{\theta}_{i,t} d\lambda_{i,t} \tag{4.32}$$

as the one-step-ahead predictive density for the observations.

Now let's write the number of models for each series explicitly and denote M_{ik} as any arbitrary choice of a sub-model of series i. Then, the relative posterior probability for model M_{ik} compared to all candidate models M_i is given by

$$p(M_{ik}|y_{i,1:t}) = \frac{p(y_{i,t}|\mathbf{F}_{pa(i)}, \mathbf{y}_{1:t-1}) p(M_{ik})}{\sum_{\ell=1}^{K_i} p(y_{i,t}|\mathbf{F}_{pa(i)} \cdot \mathbf{y}_{1:t-1}) p(M_{i\ell})}, \tag{4.33}$$

where $M_i \equiv \{1 : K_i\}$ is defined as the set of all models over a subspace of models in M_i with a maximum number of models K_i and M_{ik} is any arbitrary model in this set.

Next, to select the best models or a range of good models, we rank the candidate models based on their posterior model probabilities in Equation (4.33).

4.5.7 Bayesian model averaging

Evaluating a large number of multiple regression DLMs over the space of all models M from models for the subspaces M_i may not be practical. We integrate the variable search and selection components of previous section in a Bayesian model averaging approach. At this point, we truncate the space over the models and look at only the models that have a posterior model probability higher than

the threshold 0.005. Let the model M_{ij} be selected with a known probability

$$p_{ij} = \Pr[M_{ij}|\mathbf{y}_{1:t}],$$

where $j = 1 : J_i$ and $J_i < K_i$. The form of the density of $p(\boldsymbol{\theta}_{i,t}|\mathbf{y}_{1:t})$ is a mixture of J_i Normal distributions and $p(\lambda_{i,t}|\mathbf{y}_{1:t})$ is a mixture of J_i Gamma distributions given by following equations, respectively

$$p(\boldsymbol{\theta}_{i,t}|\mathbf{y}_{1:t}) = \sum_{j=1}^{J_i} f_N(\boldsymbol{\theta}_{ij,t}; \mathbf{m}_{ij,t}, \mathbf{C}_{ij,t}) p_{ij} \tag{4.34}$$

and

$$p(\lambda_{i,t}|\mathbf{y}_{1:t}) = \sum_{j=1}^{J_i} f_G(\lambda_{ij,t}; n_{ij,t}, d_{ij,t}) p_{ij} \tag{4.35}$$

with moments

$$\begin{aligned}
\mathbf{m}_{i,t} &= \sum_{j=1}^{J_i} \mathbf{m}_{ij,t} p_{ij}, \\
\mathbf{C}_{i,t} &= \sum_{j=1}^{J_i} [\mathbf{C}_{ij,t} + (\mathbf{m}_{i,t} - \mathbf{m}_{ij,t})(\mathbf{m}_{i,t} - \mathbf{m}_{ij,t})'] p_{ij}, \\
s_{i,t}^{-1} &= \sum_{j=1}^{J_i} s_{ij,t}^{-1} p_{ij},
\end{aligned} \tag{4.36}$$

where the number of components are fixed for all $t = 1 : T$.

The component densities have Student T distribution, which yields the following equation for a Student T finite mixture

$$p(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) = \sum_{j=1}^{J_i} f_{T_{n_t}}(y_{i,t}; n_{ij,t}, f_{ij,t}, q_{ij,t}) p_{ij}. \tag{4.37}$$

Given a nonnegative integer N, defining the cross-sectional dimension of the data, a nonnegative integer T, defining the time series dimension of the data, hyperparameter values reported in Table 4.2 and hyperparameter values of parameters, a vector of macroeconomic and financial variables $\mathbf{y}_t$, and a vector of lagged values the variables $(\mathbf{y}_{t-1}, \mathbf{y}_{t-2})$, a pseudocode for the Gray code algorithm and the Kalman filter algorithm that runs in parallel can be summarised as follows

4.5.8 A pseudocode to describe a parallel search algorithm

Start a loop that runs in parallel for $i = 1, \ldots, N$

1. Let $M_{i,0}$ denote the null sub-model, which contains no predictors.

2. **If** $i = 1$**, for** $j = 1 : K_1$

(a) Apply the Gray code algorithm and fit all 2^p models.

(b) Apply the Kalman filter algorithm over all model space as described in the Appendix, Section 4.8.2.

(c) Compute the score as the logarithm of Equation (4.31).

otherwise for $j = 1 : K_i$

(a) Apply the Gray code algorithm and fit all 2^{2p+i-1} models.

(b) Apply the Kalman filter algorithm over all model space.

(c) Compute the score as the logarithm of Equation (4.31).

3. Either select a single best model or perform a Bayesian model averaging by evaluating best models among those 2^p and 2^{2p+i-1} models. Here best is defined as having posterior probabilities as in Equation (4.33) exceeding the threshold 0.005.

4.6 Empirical Results

We assess the efficacy of the proposed graphical variable selection, the implied model selection and the model averaging approaches by measuring how well the resulting models predict future observations.

4.6.1 Data

Our data comprises a set of quarterly U.S. macroeconomic and financial time series data that are taken from the FRED database of the Federal Reserve Bank of St Louis spanning from 1959Q1 to 2022Q3. The variables are consumption, investment, GDP growth, GDP deflator, industrial production, unemployment rate, corporate bond spread, S&P 500 stock returns, 10-year Treasury rate, and Federal funds rate.

We work with ten variables and transform the majority of variables to stationarity using the benchmark transformation code of McCracken and Ng (2020). The Data appendix, Section 4.8.5, provides a complete listing of the variables.

4.6.2 Exploratory analysis

We perform exploratory analysis in order to discover the best approaches in selecting the discount factors, the model order p, and the graphical models. We try to answer three questions in order to proceed for our actual analysis. First, how to select the possible range of values for the discount factors? Second, how to choose the optimal values of the discount factors and the model order p for each series? Third, how to select a graphical model? To answer these questions, Table 4.2 provides some insight into various prior hyperparameter values to guide us in the model specification. The first two rows are commonly used possible grid of values for the discount factors φ_i and δ_i. The third quantity, π, is based on the value of the prior probability of graph structures, the prespecified threshold value, 0.005, is for evaluating the posterior model probabilities, and the model order p is considered over a lower bound and an upper bound values.

Range of values of series specific discount factors

Our first task is to choose the values of the discount factors δ_i and φ_i. As West and Harrison (1997) and Prado and West (2010) note, the variability of

Prior hyperparameters			
Hyperparameter	Value	Hyperparameter	Value
$\mathcal{D}_{\varphi_i}$	$\{0.974 : 0.004 : 0.998\}$	π	0.5
$\mathcal{D}_{\delta_i}$	$\{0.974 : 0.004 : 0.998\}$	Threshold	0.005
		$k = 1 : p$	$[1, 2]$

Table 4.2: The table shows a grid of values of prior hyperparameters of discount factors, value of the prior probability of graph structures, a threshold to evaluate high posterior probabilities, and the model order p.

the state coefficients can be controlled by the discount factors close to one. On the other hand, low values of the discount factors lead to high variability in the coefficient states. In practice, it is common to use $\delta_i, \varphi_i \geq .9$, Prado and West (2010). Thus, the range of values of the discount factors are evaluated over a grid of values as shown in Table 4.2. We evaluate every combination of $\mathcal{D}_{\varphi_i}$ and $\mathcal{D}_{\delta_i}$ for each series. This leads to a total combination of $\mathcal{D}_{\varphi_i,\delta_i} = 49$.

Optimal values of discount factors and model order

Second, we choose optimal values of the discount factors, δ_i and φ_i, and the model order p by maximizing the logarithm of the predictive likelihood functions. This is because we assume a simple noninformative uniform prior distribution for the hyperparameters of the model order p and discount factors, $p(\delta_i, \varphi_i, p) \propto 1$, which reduces the posterior inference problem to maximizing the logarithm of the predictive likelihood functions. We assume that the true model order p is unknown but a lower bound and an upper bound for the order is known. We set the lower bound to 1 and the upper bound to 2.[8] For each $i = 1 : N$, we choose the values of δ_i, φ_i, and p that maximize the function

[8]To ensure parsimonious models, the order of VAR(p) models with time-varying parameters and stochastic volatility is commonly fixed to 2, for instance, see Cogley and Sargent (2005), Primiceri (2005), and Koop and Korobilis (2010).

$$\begin{aligned} \log L(\delta_i, \varphi_i, p; \mathbf{y}) &\equiv \log[p(\mathbf{y}_{i,1:T}|\mathbf{y}_0, \delta_i, \varphi_i, p)], \\ &= \sum_{t=1}^{T} \log[p(y_{i,t}|\mathbf{y}_{1:t-1}, \delta_i, \varphi_i, p)], \end{aligned} \tag{4.38}$$

where $p(y_{i,t}|\mathbf{y}_{1:t-1})$ are the one-step-head univariate Student T predictive densities.

Table 4.3 displays the results for the optimal model order p, the optimal values of the discount factors and the logarithm of the predictive likelihoods for each series in the model space. In this experiment, for each series $i = 1 : N$, the algorithm explores 98 different combinations of δ_i, φ_i, and p over the subspace of models. Hence, in total $N \times 98$ models are explored. Let's first consider the inference about p. Models of the variables consumption and investment identify the order $p = 2$ as optimal. On the other hand, models of the variables GDP growth, GDP deflator, industrial production, unemployment rate, corporate bond spread, S&P 500 stock returns, 10-year Treasury rate, and Federal funds rate favour the order $p = 1$.

Obviously, the optimal values of φ is constant for all series. By contrast to φ_i, the optimal values of δ_i for the variables consumption and investment is 0.990 with an optimal lag order 2 and all other variables choose 0.986 as the optimal value for δ_i with the optimal lag order 1. This raises the question of whether the optimal value of the discount factors δ_i is sensitive to the model order.

To summarise, models of two orders are explored to identify the chosen order p. Similarly, the optimal values of the discount factors for the precision $\lambda_{i,t}$ and the state variance matrices $\mathbf{W}_{i,t}$, are chosen by exploring several values. Our actual analysis is based on the discount factors at the optimal values reported in Table 4.3. However, we fix the value of the model order to $p = 2$ for all series to explore the support of these results in the variable selection experiment.

Optimal model order p, discount factors and sum of L.P.L.									
Series	p	δ	φ	L.P.L.	Series	p	δ	φ	L.P.L.
PCECC96	2	0.990	0.974	-347.344	UNRATE	1	0.986	0.974	-209.895
GPDIC1	2	0.990	0.974	-318.572	CS	1	0.986	0.974	-184.588
GDPC1	1	0.986	0.974	-128.969	S&P 500	1	0.986	0.974	-332.989
GDPCTPI	1	0.986	0.974	-342.515	GS10	1	0.986	0.974	-352.916
INDPRO	1	0.986	0.974	-188.296	FEDFR	1	0.986	0.974	-310.682

Table 4.3: The table shows the optimal values of the model order p, the discount factors and the logarithms of predictive likelihoods (L.P.L.). The range of values of the discount factors are evaluated over a grid of values in $\mathcal{D}_\varphi = \{0.974 : 0.004 : 0.998\}$ and $\mathcal{D}_\delta = \{0.974 : 0.004 : 0.998\}$. Series descriptions are reported in the Data appendix.

Selecting graphical models

Third, applying a graphical variable selection approach with uncertainty over the number of parental series for each i, requires variable selection to be applied for each series as $2^p + \sum_{i=2}^{N} 2^{2p+i-1}$. We decide to apply the variable selection approach over the parental series only, i.e., the number of graphical models for each i would be $2^p + \sum_{i=2}^{N} 2^{2p+i-1}$ with the model order fixed to $p = 2$ and the discount factors would be preselected based on the values reported in Table 4.3. For a feasible computational approach in terms of time complexity of solving a problem, setting $N = 10$, leads to over $8,000$ possible models M_N and over 16,000 possible model $M_{1:N}$. A prior probability on the graph structure $\pi = 0.5$ is chosen and models with posterior probabilities less than a threshold, 0.005, are removed.

To illustrate the use of the variable selection of this section, we calculate the posterior model probabilities for each $i = 1 : N$. Table 4.4 through Table 4.13 present the series specific models that receive over .5% posterior model probability in descending order. The pattern of the dynamic and contemporaneous dependence structures takes values in the discrete space as 1's and the pattern of the independence structures are shown by 0's. The dimension of the vector of dynamic parents range from $1 \times p$ and $1 \times 2p$ and they refer to the dynamic parents in a lower bidiagonal order, that is, the first variable in the recursive or-

der has its first and second own-lags as dynamic parents and each other variable has a potential of first and second cross-lags of a preceding variable and first and second own-lags as parents. On the other hand, the dimension of the vector of the contemporaneous parents are presented for each $i = 2 : N$, as $1 \times (i - 1)$, representing the current values of variables in the same order that are reported in the Data appendix. For example, consumption = 1, investment = 2, GDP growth = 3, GDP deflator = 4, industrial production = 5, unemployment rate = 6, corporate bond spread = 7, S&P 500 stock returns = 8, 10-year Treasury rate = 9, and Federal funds rate = 10.

The preferred models for the variables consumption, investment and GDP growth are shown in Table 4.4. 3 out of 4 models are among the best models for consumption. It appears that first and second own-lags of consumption have much predictive power. Dynamic parents of the variable investment are first and second own-lags of investment and first and second cross-lags of the variable consumption. 7 out of 32 models are much confident that first cross-lag of consumption as a dynamic parent of the variable investment must be included in the model. Those models also support the inclusion of the current value of consumption as a contemporaneous parent of investment. The most likely 11 models for the variable GDP growth show that none of the variables from the dynamic parental set dominates the dependence structures. In addition, the model with the highest posterior probability represents 41% of the total posterior probability, indicating zero effects of the dynamic predictors. On the other hand, all models support the inclusion of the contemporaneous parents, consumption and investment, in the model featuring the variable GDP growth as a dependent variable. In total, 128 different models are visited for the variable GDP deflator as shown in Table 4.5. In this example, 29 models are chosen and all models support the inclusion of the first own-lag of the variable GDP deflator as a dynamic parent. Table 4.6 displays the results for the variable industrial production with a potential of 8 dynamic

and contemporaneous predictors. All chosen 12 models are confident that own first-lag of the variable industrial production and contemporaneous values of the variables consumption and investment must be included in the model. The results of the variable unemployment rate in Table 4.7 indicate considerable model uncertainty with the highest posterior model probability explaining only 15% of the total posterior probability. Those models are confident that the first-own lag of unemployment rate as a dynamic parent, and consumption and industrial production as contemporaneous parents must be included in the model.

For the variable corporate bond spread, Table 4.8 and Table 4.9 show the posterior model probabilities of 43 out of 1024 models. There is a considerable model uncertainty over the model space with the highest posterior model probability accounting for only 8% of the total posterior probability. We present the results for the variable S&P 500 stock returns in Table 4.10 with the search algorithm exploring all 2048 models. All models supports the inclusion of the first cross-lag of the variable corporate bond spread and the majority of models are confident that the first own-lag must be included in the model. The contemporaneous parent, corporate bond spread, has a predictive power on the variable S&P 500 stock returns, which is supported by all 38 models. Table 4.11 reports best models for the variable 10-year Treasury maturity rate with high model uncertainty across the model space and with few variables having explanatory power excluding the first own-lag as a dynamic parent. Finally, the results for variable Federal funds rate are reported in Table 4.12 and 4.13. 40 out of 8,192 models are supported by the method. The first own-lag of the variable Federal funds rate as a dynamic parent and the current value of the variable 10-year Treasury maturity rate have much predictive ability on the variable Federal funds rate.

To summarize, we have illustrated the inferential potential of a Bayesian variable selection problem. The results show that there is a considerable model uncertainty within the best selected models for each series. As the number of

predictors increases, the model uncertainty increases too. Inference based on a single best model with the highest posterior model probability would be ambiguous.

Range of good multiple regression DLMs exploring all possible $2^2 = 4$ models				
Consumption				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.857	$M_{1,(1,1)}$		3
2	0.112	$M_{1,(1,0)}$		2
3	0.031	$M_{1,(0,1)}$		4
Range of good multiple regression DLMs exploring all possible $2^5 = 32$ models				
Investment				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.763	$M_{2,(1,0,0,0)}$	$M_{2,(1)}$	31
2	0.090	$M_{2,(1,1,0,0)}$	$M_{2,(1)}$	30
3	0.064	$M_{2,(1,0,1,0)}$	$M_{2,(1)}$	26
4	0.057	$M_{2,(1,0,0,1)}$	$M_{2,(1)}$	18
5	0.010	$M_{2,(1,1,0,1)}$	$M_{2,(1)}$	19
6	0.009	$M_{2,(1,1,1,0)}$	$M_{2,(1)}$	27
7	0.006	$M_{2,(1,0,1,1)}$	$M_{2,(1)}$	23
Range of good multiple regression DLMs exploring all possible $2^6 = 64$ models				
GDP Growth				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.406	$M_{3,(0,0,0,0)}$	$M_{3,(1,1)}$	33
2	0.325	$M_{3,(1,0,0,0)}$	$M_{3,(1,1)}$	34
3	0.094	$M_{3,(1,1,0,0)}$	$M_{3,(1,1)}$	35
4	0.049	$M_{3,(0,0,1,0)}$	$M_{3,(1,1)}$	40
5	0.034	$M_{3,(0,1,0,0)}$	$M_{3,(1,1)}$	36
6	0.029	$M_{3,(1,0,1,0)}$	$M_{3,(1,1)}$	39
7	0.017	$M_{3,(0,0,0,1)}$	$M_{3,(1,1)}$	48
8	0.013	$M_{3,(1,0,0,1)}$	$M_{3,(1,1)}$	47
9	0.008	$M_{3,(1,1,1,0)}$	$M_{3,(1,1)}$	38
10	0.007	$M_{3,(0,0,1,1)}$	$M_{3,(1,1)}$	41
11	0.005	$M_{3,(1,1,0,1)}$	$M_{3,(1,1)}$	46

Table 4.4: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series consumption, investment and GDP growth, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^7 = 128$ models				
GDP Deflator				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.157	$M_{4,(0,1,0,1)}$	$M_{4,(1,1,0)}$	45
2	0.128	$M_{4,(0,1,0,1)}$	$M_{4,(0,1,0)}$	52
3	0.102	$M_{4,(0,1,0,1)}$	$M_{4,(1,1,1)}$	84
4	0.071	$M_{4,(0,1,0,0)}$	$M_{4,(1,1,0)}$	36
5	0.068	$M_{4,(0,1,0,0)}$	$M_{4,(0,1,0)}$	61
6	0.044	$M_{4,(0,1,0,1)}$	$M_{4,(0,1,1)}$	77
7	0.044	$M_{4,(0,1,0,0)}$	$M_{4,(1,1,1)}$	93
8	0.039	$M_{4,(0,1,0,1)}$	$M_{4,(0,0,1)}$	116
9	0.036	$M_{4,(0,1,0,1)}$	$M_{4,(1,0,0)}$	20
10	0.027	$M_{4,(0,1,0,1)}$	$M_{4,(1,0,1)}$	109
11	0.026	$M_{4,(0,1,0,0)}$	$M_{4,(0,1,1)}$	68
12	0.025	$M_{4,(0,1,1,1)}$	$M_{4,(1,1,0)}$	44
13	0.024	$M_{4,(0,1,1,1)}$	$M_{4,(1,1,1)}$	85
14	0.021	$M_{4,(1,1,0,1)}$	$M_{4,(1,1,0)}$	46
15	0.015	$M_{4,(0,1,1,1)}$	$M_{4,(0,1,0)}$	53
16	0.015	$M_{4,(1,1,0,1)}$	$M_{4,(1,1,1)}$	83
17	0.014	$M_{4,(1,1,0,1)}$	$M_{4,(0,1,0)}$	51
18	0.013	$M_{4,(0,1,0,0)}$	$M_{4,(0,0,1)}$	125
19	0.013	$M_{4,(1,1,0,1)}$	$M_{4,(1,0,0)}$	19
20	0.009	$M_{4,(0,1,0,0)}$	$M_{4,(1,0,0)}$	29
21	0.008	$M_{4,(0,1,0,0)}$	$M_{4,(1,0,1)}$	100
22	0.008	$M_{4,(0,1,1,1)}$	$M_{4,(1,0,0)}$	21
23	0.008	$M_{4,(0,1,1,1)}$	$M_{4,(0,1,1)}$	76
24	0.007	$M_{4,(1,1,0,0)}$	$M_{4,(1,1,0)}$	35
25	0.007	$M_{4,(1,1,0,1)}$	$M_{4,(0,1,1)}$	78
26	0.006	$M_{4,(1,1,0,1)}$	$M_{4,(0,0,1)}$	115
27	0.006	$M_{4,(1,1,0,1)}$	$M_{4,(1,0,1)}$	110
28	0.005	$M_{4,(0,1,1,1)}$	$M_{4,(0,0,1)}$	117
29	0.005	$M_{4,(1,1,0,0)}$	$M_{4,(0,1,0)}$	62

Table 4.5: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series GDP deflator, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^8 = 256$ models				
Industrial Production				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.394	$M_{5,(0,1,0,0)}$	$M_{5,(1,1,0,0)}$	36
2	0.295	$M_{5,(0,1,0,0)}$	$M_{5,(1,1,0,1)}$	221
3	0.051	$M_{5,(1,1,0,0)}$	$M_{5,(1,1,0,0)}$	35
4	0.047	$M_{5,(0,1,0,1)}$	$M_{5,(1,1,0,0)}$	45
5	0.042	$M_{5,(0,1,0,0)}$	$M_{5,(1,1,1,0)}$	93
6	0.032	$M_{5,(0,1,0,0)}$	$M_{5,(1,1,1,1)}$	164
7	0.030	$M_{5,(0,1,0,1)}$	$M_{5,(1,1,0,1)}$	212
8	0.027	$M_{5,(1,1,0,0)}$	$M_{5,(1,1,0,1)}$	222
9	0.021	$M_{5,(0,1,1,0)}$	$M_{5,(1,1,0,0)}$	37
10	0.015	$M_{5,(0,1,1,0)}$	$M_{5,(1,1,0,1)}$	220
11	0.010	$M_{5,(1,1,0,1)}$	$M_{5,(1,1,0,0)}$	46
12	0.005	$M_{5,(1,1,0,0)}$	$M_{5,(1,1,1,0)}$	94

Table 4.6: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series industrial production, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^9 = 512$ models				
Unemployment Rate				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.150	$M_{6,(1,1,0,1)}$	$M_{6,(1,1,0,0,1)}$	467
2	0.110	$M_{6,(1,1,0,1)}$	$M_{6,(1,0,0,0,1)}$	494
3	0.066	$M_{6,(0,1,1,0)}$	$M_{6,(1,0,0,0,1)}$	485
4	0.062	$M_{6,(1,1,1,0)}$	$M_{6,(1,0,0,0,1)}$	486
5	0.059	$M_{6,(1,1,1,0)}$	$M_{6,(1,1,0,0,1)}$	475
6	0.057	$M_{6,(0,1,0,1)}$	$M_{6,(1,1,0,0,1)}$	468
7	0.049	$M_{6,(0,1,0,1)}$	$M_{6,(1,0,0,0,1)}$	493
8	0.046	$M_{6,(0,1,1,0)}$	$M_{6,(1,1,0,0,1)}$	476
9	0.033	$M_{6,(0,1,0,0)}$	$M_{6,(1,0,0,0,1)}$	484
10	0.029	$M_{6,(1,1,0,1)}$	$M_{6,(1,1,1,0,1)}$	430
11	0.027	$M_{6,(1,1,0,1)}$	$M_{6,(1,0,1,0,1)}$	403
12	0.026	$M_{6,(1,1,0,0)}$	$M_{6,(1,0,0,0,1)}$	483
13	0.026	$M_{6,(0,1,0,0)}$	$M_{6,(1,1,0,0,1)}$	477
14	0.025	$M_{6,(1,1,0,0)}$	$M_{6,(1,1,0,0,1)}$	478
15	0.020	$M_{6,(1,1,1,1)}$	$M_{6,(1,1,0,0,1)}$	470
16	0.019	$M_{6,(0,1,1,1)}$	$M_{6,(1,1,0,0,1)}$	469
17	0.019	$M_{6,(0,1,1,1)}$	$M_{6,(1,0,0,0,1)}$	492
18	0.016	$M_{6,(1,1,1,1)}$	$M_{6,(1,0,0,0,1)}$	491
19	0.016	$M_{6,(0,1,1,0)}$	$M_{6,(1,0,1,0,1)}$	412
20	0.016	$M_{6,(1,1,1,0)}$	$M_{6,(1,0,1,0,1)}$	411
21	0.011	$M_{6,(0,1,0,1)}$	$M_{6,(1,0,1,0,1)}$	404
22	0.011	$M_{6,(0,1,0,1)}$	$M_{6,(1,1,1,0,1)}$	429
23	0.009	$M_{6,(1,1,1,0)}$	$M_{6,(1,1,1,0,1)}$	422
24	0.009	$M_{6,(0,1,0,0)}$	$M_{6,(1,0,1,0,1)}$	413
25	0.007	$M_{6,(1,1,0,0)}$	$M_{6,(1,0,1,0,1)}$	414
26	0.007	$M_{6,(0,1,1,0)}$	$M_{6,(1,1,1,0,1)}$	421

Table 4.7: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series unemployment rate, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^{10} = 1024$ models				
Corporate Bond Spread				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.077	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,0,0,1,0)}$	500
2	0.062	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,0,1,1,0)}$	269
3	0.048	$M_{7,(0,1,0,0)}$	$M_{7,(0,0,0,0,1,0)}$	509
4	0.047	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,0,0,1,0)}$	493
5	0.033	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,0,1,1,0)}$	276
6	0.032	$M_{7,(0,1,0,0)}$	$M_{7,(0,0,0,1,1,0)}$	260
7	0.030	$M_{7,(0,1,0,1)}$	$M_{7,(0,1,0,0,1,0)}$	461
8	0.026	$M_{7,(0,1,0,0)}$	$M_{7,(1,0,0,0,1,0)}$	484
9	0.026	$M_{7,(0,1,0,0)}$	$M_{7,(1,0,0,1,1,0)}$	285
10	0.026	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,0,0,1,1)}$	525
11	0.025	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,0,1,1,1)}$	756
12	0.025	$M_{7,(0,1,0,1)}$	$M_{7,(0,1,0,1,1,0)}$	308
13	0.024	$M_{7,(0,1,0,0)}$	$M_{7,(0,1,0,0,1,0)}$	452
14	0.021	$M_{7,(0,1,0,0)}$	$M_{7,(0,0,0,0,1,1)}$	516
15	0.019	$M_{7,(0,1,0,0)}$	$M_{7,(0,0,0,1,1,1)}$	765
16	0.019	$M_{7,(0,1,0,0)}$	$M_{7,(0,1,0,0,1,1)}$	573
17	0.018	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,0,0,1,1)}$	532
18	0.017	$M_{7,(0,1,0,1)}$	$M_{7,(0,1,0,1,1,1)}$	717
19	0.017	$M_{7,(0,1,0,0)}$	$M_{7,(0,1,0,1,1,1)}$	708
20	0.017	$M_{7,(0,1,0,1)}$	$M_{7,(0,1,0,0,1,1)}$	564
21	0.015	$M_{7,(0,1,0,0)}$	$M_{7,(0,1,0,1,1,0)}$	317

Table 4.8: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series corporate bond spread, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^{10} = 1024$ models				
Corporate Bons Spread (continued)				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
22	0.015	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,0,0,0,1)}$	1005
23	0.013	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,1,0,1,0)}$	404
24	0.012	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,0,1,1,1)}$	749
25	0.009	$M_{7,(0,1,0,1)}$	$M_{7,(1,1,0,0,1,0)}$	468
26	0.009	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,0,1,0,1)}$	788
27	0.009	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,1,0,1,0)}$	397
28	0.009	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,1,1,1,0)}$	365
29	0.009	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,1,0,0,1)}$	916
30	0.009	$M_{7,(0,1,0,0)}$	$M_{7,(1,0,0,0,1,1)}$	541
31	0.008	$M_{7,(1,1,0,0)}$	$M_{7,(0,0,0,1,1,1)}$	766
32	0.008	$M_{7,(0,1,0,0)}$	$M_{7,(1,0,0,1,1,1)}$	740
33	0.008	$M_{7,(0,1,0,0)}$	$M_{7,(1,0,1,0,1,0)}$	413
34	0.008	$M_{7,(0,1,0,1)}$	$M_{7,(1,0,1,0,1,1)}$	621
35	0.007	$M_{7,(0,1,0,0)}$	$M_{7,(1,0,1,1,1,0)}$	356
36	0.007	$M_{7,(0,1,0,1)}$	$M_{7,(1,1,0,1,1,0)}$	301
37	0.006	$M_{7,(1,1,0,0)}$	$M_{7,(0,1,0,1,1,1)}$	707
38	0.006	$M_{7,(0,1,0,0)}$	$M_{7,(1,1,0,0,1,0)}$	477
39	0.006	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,1,1,1,0)}$	372
40	0.006	$M_{7,(0,1,0,0)}$	$M_{7,(1,1,0,1,1,0)}$	292
41	0.006	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,1,0,1,1)}$	628
42	0.005	$M_{7,(0,1,0,1)}$	$M_{7,(1,1,0,0,1,1)}$	557
43	0.005	$M_{7,(0,1,0,1)}$	$M_{7,(0,0,1,0,0,1)}$	909

Table 4.9: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series corporate bond spread, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^{11} = 2048$ models				
S&P 500 Stock Returns				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.132	$M_{8,(1,1,0,0)}$	$M_{8,(0,1,1,0,1,0,1)}$	1603
2	0.092	$M_{8,(1,1,0,0)}$	$M_{8,(0,1,1,0,0,0,1)}$	1982
3	0.054	$M_{8,(1,1,0,0)}$	$M_{8,(0,0,1,0,1,0,1)}$	1662
4	0.052	$M_{8,(1,1,0,0)}$	$M_{8,(1,0,1,0,1,0,1)}$	1635
5	0.039	$M_{8,(1,1,1,0)}$	$M_{8,(0,1,1,0,1,0,1)}$	1606
6	0.037	$M_{8,(1,1,1,0)}$	$M_{8,(0,0,1,0,1,0,1)}$	1659
7	0.033	$M_{8,(1,1,0,0)}$	$M_{8,(0,1,1,0,0,1,1)}$	1091
8	0.031	$M_{8,(1,1,1,0)}$	$M_{8,(0,1,1,0,0,0,1)}$	1979
9	0.028	$M_{8,(1,1,1,0)}$	$M_{8,(1,0,0,0,0,0,1)}$	2022
10	0.027	$M_{8,(1,1,1,0)}$	$M_{8,(1,0,1,0,1,0,1)}$	1638
11	0.026	$M_{8,(1,1,0,0)}$	$M_{8,(1,0,0,0,1,0,1)}$	1566
12	0.023	$M_{8,(1,1,0,0)}$	$M_{8,(1,0,0,0,0,0,1)}$	2019
13	0.019	$M_{8,(1,1,0,0)}$	$M_{8,(1,1,1,0,1,0,1)}$	1630
14	0.016	$M_{8,(1,1,0,0)}$	$M_{8,(1,0,0,0,0,1,1)}$	1054
15	0.013	$M_{8,(1,1,1,0)}$	$M_{8,(1,0,0,0,1,0,1)}$	1563
16	0.013	$M_{8,(1,1,0,0)}$	$M_{8,(0,1,1,0,1,1,1)}$	1470
17	0.012	$M_{8,(1,1,0,0)}$	$M_{8,(1,1,1,0,0,0,1)}$	1955
18	0.011	$M_{8,(1,1,0,1)}$	$M_{8,(0,1,1,0,1,0,1)}$	1614
19	0.010	$M_{8,(1,1,1,0)}$	$M_{8,(0,0,1,0,0,0,1)}$	1926
20	0.010	$M_{8,(1,1,0,0)}$	$M_{8,(1,1,0,0,1,0,1)}$	1571
21	0.010	$M_{8,(1,1,0,1)}$	$M_{8,(0,1,1,0,0,0,1)}$	1971
22	0.009	$M_{8,(1,1,1,0)}$	$M_{8,(0,1,1,0,0,1,1)}$	1094
23	0.009	$M_{8,(1,1,0,0)}$	$M_{8,(1,0,1,0,0,1,1)}$	1123
24	0.009	$M_{8,(1,0,0,0)}$	$M_{8,(0,1,1,0,0,0,1)}$	1983
25	0.008	$M_{8,(1,1,0,0)}$	$M_{8,(0,0,1,0,0,1,1)}$	1150
26	0.008	$M_{8,(1,1,1,0)}$	$M_{8,(1,0,1,0,0,0,1)}$	1947
27	0.008	$M_{8,(1,1,1,0)}$	$M_{8,(1,0,0,0,0,1,1)}$	1051
28	0.007	$M_{8,(1,1,0,0)}$	$M_{8,(1,0,1,0,1,1,1)}$	1438
29	0.007	$M_{8,(1,1,0,0)}$	$M_{8,(0,0,1,0,1,1,1)}$	1411
30	0.007	$M_{8,(1,1,0,0)}$	$M_{8,(1,1,0,0,0,0,1)}$	2014
31	0.007	$M_{8,(1,1,0,0)}$	$M_{8,(0,0,0,0,0,0,1)}$	2046
32	0.006	$M_{8,(1,1,1,0)}$	$M_{8,(0,0,1,0,0,1,1)}$	1147
33	0.006	$M_{8,(1,1,1,0)}$	$M_{8,(1,1,1,0,1,0,1)}$	1627
34	0.006	$M_{8,(1,1,0,1)}$	$M_{8,(0,0,1,0,1,0,1)}$	1651
35	0.006	$M_{8,(1,1,0,0)}$	$M_{8,(1,0,1,0,0,0,1)}$	1950
36	0.006	$M_{8,(1,1,0,0)}$	$M_{8,(0,1,1,1,1,0,1)}$	1726
37	0.006	$M_{8,(1,1,0,0)}$	$M_{8,(0,0,1,0,0,0,1)}$	1923
38	0.005	$M_{8,(1,1,1,0)}$	$M_{8,(1,1,0,0,1,0,1)}$	1574

Table 4.10: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series S&P 500 stock returns, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^{12} = 4096$ models				
10-Year Treasury Maturity Rate				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.089	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,1,1,0,0,0)}$	261
2	0.065	$M_{9,(0,1,0,0)}$	$M_{9,(0,0,0,1,1,0,0,0)}$	260
3	0.050	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,0,1,0,0,0)}$	508
4	0.039	$M_{9,(0,1,0,0)}$	$M_{9,(0,0,0,0,1,0,0,0)}$	509
5	0.022	$M_{9,(0,1,0,0)}$	$M_{9,(0,0,0,0,1,0,0,1)}$	3588
6	0.021	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,0,1,0,0,1)}$	3589
7	0.018	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,1,1,0,0,1)}$	3836
8	0.018	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,1,0,1,0,0)}$	773
9	0.018	$M_{9,(1,1,1,0)}$	$M_{9,(0,0,0,1,1,0,0,0)}$	262
10	0.017	$M_{9,(0,1,0,0)}$	$M_{9,(0,0,0,1,1,0,0,1)}$	3837
11	0.015	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,1,1,1,0,0,0)}$	380
12	0.015	$M_{9,(0,1,1,0)}$	$M_{9,(0,1,0,1,1,0,0,0)}$	316
13	0.014	$M_{9,(0,1,1,0)}$	$M_{9,(0,1,0,1,0,0,0,0)}$	197
14	0.014	$M_{9,(0,1,0,0)}$	$M_{9,(0,1,0,1,0,0,0,0)}$	196
15	0.012	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,1,1,0,1,0)}$	1788
16	0.011	$M_{9,(0,1,0,0)}$	$M_{9,(0,0,0,1,0,1,0,0)}$	772
17	0.010	$M_{9,(0,1,0,0)}$	$M_{9,(0,1,0,1,1,0,0,0)}$	317
18	0.010	$M_{9,(0,1,1,0)}$	$M_{9,(0,1,0,0,1,0,0,0)}$	453
19	0.010	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,1,1,1,0,0)}$	764
20	0.010	$M_{9,(0,1,1,0)}$	$M_{9,(1,0,0,1,1,0,0,0)}$	284
21	0.009	$M_{9,(0,1,0,1)}$	$M_{9,(0,0,0,1,1,0,0,0)}$	269
22	0.009	$M_{9,(0,1,0,0)}$	$M_{9,(0,0,1,1,0,0,0,0)}$	132
23	0.009	$M_{9,(0,1,0,0)}$	$M_{9,(0,0,1,1,1,0,0,0)}$	381
24	0.009	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,0,0,1,0,1,0)}$	1541
25	0.008	$M_{9,(1,1,1,0)}$	$M_{9,(0,0,0,1,0,1,0,0)}$	774
26	0.008	$M_{9,(0,1,1,1)}$	$M_{9,(0,0,0,1,1,0,0,0)}$	268
27	0.008	$M_{9,(1,1,1,0)}$	$M_{9,(0,0,0,0,1,0,0,0)}$	507
28	0.007	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,1,0,1,0,0,0)}$	389
29	0.007	$M_{9,(0,1,0,1)}$	$M_{9,(0,0,0,0,1,0,0,0)}$	500
30	0.007	$M_{9,(0,1,0,0)}$	$M_{9,(0,1,0,0,1,0,0,0)}$	452
31	0.007	$M_{9,(0,1,1,0)}$	$M_{9,(0,0,1,1,0,0,0,0)}$	133
32	0.006	$M_{9,(0,1,1,1)}$	$M_{9,(0,0,0,0,1,0,0,0)}$	501
33	0.005	$M_{9,(0,1,0,0)}$	$M_{9,(1,0,0,1,1,0,0,0)}$	285

Table 4.11: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series 10-year Treasury maturity rate, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^{13} = 8192$ models				
Federal Funds Rate				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
1	0.114	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,0,0,1,0,0,0,1)}$	7684
2	0.059	$M_{10,(1,1,0,0)}$	$M_{10,(0,0,0,0,1,0,0,0,1)}$	7683
3	0.059	$M_{10,(0,1,1,0)}$	$M_{10,(0,0,0,0,1,0,0,0,1)}$	7685
4	0.051	$M_{10,(1,1,1,0)}$	$M_{10,(0,0,0,0,1,0,0,0,1)}$	7686
5	0.045	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,0,0,1,1,0,0,1)}$	7677
6	0.032	$M_{10,(1,1,0,0)}$	$M_{10,(0,0,0,0,1,1,0,0,1)}$	7678
7	0.026	$M_{10,(0,1,0,1)}$	$M_{10,(0,0,0,0,1,0,0,0,1)}$	7693
8	0.024	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,0,1,1,0,0,0,1)}$	7933
9	0.017	$M_{10,(1,1,1,0)}$	$M_{10,(0,0,0,0,1,1,0,0,1)}$	7675
10	0.016	$M_{10,(0,1,0,0)}$	$M_{10,(1,0,0,0,1,0,0,0,1)}$	7709
11	0.015	$M_{10,(0,1,0,0)}$	$M_{10,(0,1,0,0,1,0,0,0,1)}$	7741
12	0.014	$M_{10,(0,1,1,0)}$	$M_{10,(0,0,0,0,1,1,0,0,1)}$	7676
13	0.013	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,1,0,1,0,0,0,1)}$	7805
14	0.013	$M_{10,(0,1,0,1)}$	$M_{10,(0,0,0,0,1,1,0,0,1)}$	7668
15	0.012	$M_{10,(1,0,0,0)}$	$M_{10,(0,0,0,0,1,1,0,0,1)}$	7679
16	0.012	$M_{10,(1,1,1,0)}$	$M_{10,(0,1,0,0,1,0,0,0,1)}$	7739
17	0.012	$M_{10,(1,1,0,1)}$	$M_{10,(0,0,0,0,1,0,0,0,1)}$	7694
18	0.012	$M_{10,(0,1,1,0)}$	$M_{10,(0,1,0,0,1,0,0,0,1)}$	7740
19	0.011	$M_{10,(1,1,0,0)}$	$M_{10,(0,0,0,0,0,1,0,0,1)}$	7171
20	0.011	$M_{10,(0,1,1,0)}$	$M_{10,(0,0,0,1,1,0,0,0,1)}$	7932

Table 4.12: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series Federal funds rate, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

Range of good multiple regression DLMs exploring all possible $2^{13} = 8192$ models				
Federal Funds Rate (Continued)				
Choice	P.M.P.	Dynamic Parents	Contem. Parents	Model index
21	0.009	$M_{10,(0,1,1,0)}$	$M_{10,(0,0,1,0,1,0,0,0,1)}$	7804
22	0.009	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,0,0,0,1,0,0,1)}$	7172
23	0.008	$M_{10,(1,0,0,0)}$	$M_{10,(0,0,0,0,1,0,0,0,1)}$	7682
24	0.008	$M_{10,(0,1,1,0)}$	$M_{10,(1,0,0,0,1,0,0,0,1)}$	7708
25	0.008	$M_{10,(1,1,0,0)}$	$M_{10,(0,1,0,0,1,0,0,0,1)}$	7742
26	0.008	$M_{10,(1,1,0,1)}$	$M_{10,(0,0,0,0,1,1,0,0,1)}$	7667
27	0.007	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,0,1,1,1,0,0,1)}$	7428
28	0.007	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,0,0,1,0,0,1,1)}$	4605
29	0.007	$M_{10,(1,1,0,0)}$	$M_{10,(0,0,0,1,1,0,0,0,1)}$	7934
30	0.007	$M_{10,(0,1,0,0)}$	$M_{10,(0,1,0,0,1,1,0,0,1)}$	7620
31	0.006	$M_{10,(1,1,0,0)}$	$M_{10,(1,0,0,0,1,0,0,0,1)}$	7710
32	0.006	$M_{10,(1,1,1,0)}$	$M_{10,(0,0,1,0,1,0,0,0,1)}$	7803
33	0.006	$M_{10,(1,1,1,0)}$	$M_{10,(0,0,0,0,0,1,0,0,1)}$	7174
34	0.006	$M_{10,(1,1,0,0)}$	$M_{10,(0,1,0,0,0,1,0,0,1)}$	7230
35	0.006	$M_{10,(1,1,1,0)}$	$M_{10,(0,0,0,1,1,0,0,0,1)}$	7931
36	0.005	$M_{10,(1,1,1,0)}$	$M_{10,(0,1,0,0,1,1,0,0,1)}$	7622
37	0.005	$M_{10,(0,1,0,0)}$	$M_{10,(0,0,1,0,1,1,0,0,1)}$	7556
38	0.005	$M_{10,(1,1,0,0)}$	$M_{10,(0,0,1,0,1,0,0,0,1)}$	7806
39	0.005	$M_{10,(1,1,0,0)}$	$M_{10,(0,1,0,0,1,1,0,0,1)}$	7619
40	0.005	$M_{10,(1,1,1,0)}$	$M_{10,(1,0,0,0,1,0,0,0,1)}$	7707

Table 4.13: The table reports models with posterior model probabilities (P.M.P.) higher than a threshold of 0.005 for series Federal funds rate, excluded (included) dynamic and contemporaneous parents are shown by zero (one) independence (dependence) structures, and model indices.

4.6.3 Dynamic and contemporaneous pairwise dependence structures

The results of the previous section show that most of the models over the model space have small posterior probabilities. As detailed in Barbieri and Berger (2004) and Heaton and Scott (2010), a more useful summary of the posterior distributions is the median probability model. We define the median probability model as the model that includes the dynamic and the contemporaneous predictors having posterior inclusion probability of at least 0.5. In this case, we may perform a sensitivity analysis and compute the posterior inclusion probabilities of the individual dynamic and contemporaneous predictors. Those probabilities corresponds to search for features of the Bayesian dynamic graphical model, such as the presence or absence of an individual dynamic effect from the variable $y_{i,t-k}$ or $y_{j,t-k}$ at time $t-k$ to the variable $y_{i,t}$ at time t, and an individual contemporaneous effect from the variable $y_{i,t}$ to the variable $y_{j,t}$ with the same time index and $i \neq j$.

Figure 4.1 displays image plots of posterior inclusion probabilities $p(\xi_{\beta,ij,k} = 1|y_{1:t})$ for the dynamic parental series and posterior inclusion probabilities $p(\xi_{\gamma,ij} = 1|y_{1:t})$ for the contemporaneous parental series computed over the full model space. The scale moves from 0 (white) to 1 (black) with some intermediate shades of grey. In this setup, we have $\big(p + 2p(N-1)\big) = 38$ dynamic autoregressive and cross-lagged effects and $\binom{N}{2} = 45$ contemporaneous effects at each point in time t.[9]

There are a few interesting facts from the analysis of the median probability models displayed in Figure 4.1. First, the results show that only 16 out of 38 dynamic parents are included in the model. Likewise, only 18 out 45 contemporaneous parents are included in the model. Second, only one out of ten series

[9]Remember that unlike the state parameters, which are allowed to change in time, the graph structure is fixed in time.

has the median probability model that do not coincide with the highest posterior probability model. For instance, the median probability model of the variable unemployment rate, the sixth variable in the recursive order, does not coincide with highest posterior probability model displayed in Table 4.7.

Visiting the equation including the variable GDP growth as dependent variable, the median probability model and the highest posterior probability model displayed in Table 4.4 tend to agree with each other. The median probability model in the model subspace M_3, based on the groupings of the dynamic parents $dp(3)$ is $M_{3,(0,0,0,0)}$ and on the groupings of the contemporaneous parents $cp(3)$ is $M_{3,(1,1)}$. An interesting characteristic of this outcome is that the null model with zero subgroup dynamic effect corresponds to the highest posterior probability model. The variable corporate bond spread discovers the first and the second own-lags as good predictors. The corresponding highest posterior probability model and the median model based on the groupings of the dynamic parents is $M_{7,(0,1,0,1)}$. Likewise, the highest posterior probability model and the median model based on the groupings of the contemporaneous parents is $M_{7,(0,0,0,0,1,0)}$, which favours only the inclusion of the variable industrial production. Another interesting observation from Figure 4.1 is related to the variable S&P 500 stock returns with the first cross-lag of the variable corporate bond spread and the first own-lag of the S&P 500 stock returns having predictive power on the variable S&P 500 stock returns. This corresponds to the median probability model $M_{8,(1,1,0,0)}$, which favours the inclusion of two variables from the dynamic parental set $dp(8)$. The inclusion probabilities also show that the variables, investment, GDP growth, industrial production, and corporate bons spread, from the contemporaneous parental set should be included in the model $M_{8,(0,1,1,0,1,0,1)}$. The median probability model for the variable Federal funds rate, the last variable in the recursive order, based on the groupings of the dynamic parents $dp(10)$ is $M_{10,(0,1,0,0)}$ and on the groupings of the contemporaneous parents $cp(10)$ is

$M_{10,(0,0,0,0,1,0,0,0,1)}$. Clearly, the median probability model M_{10} coincides with the model, defined as that having highest posterior probability in Table 4.12.

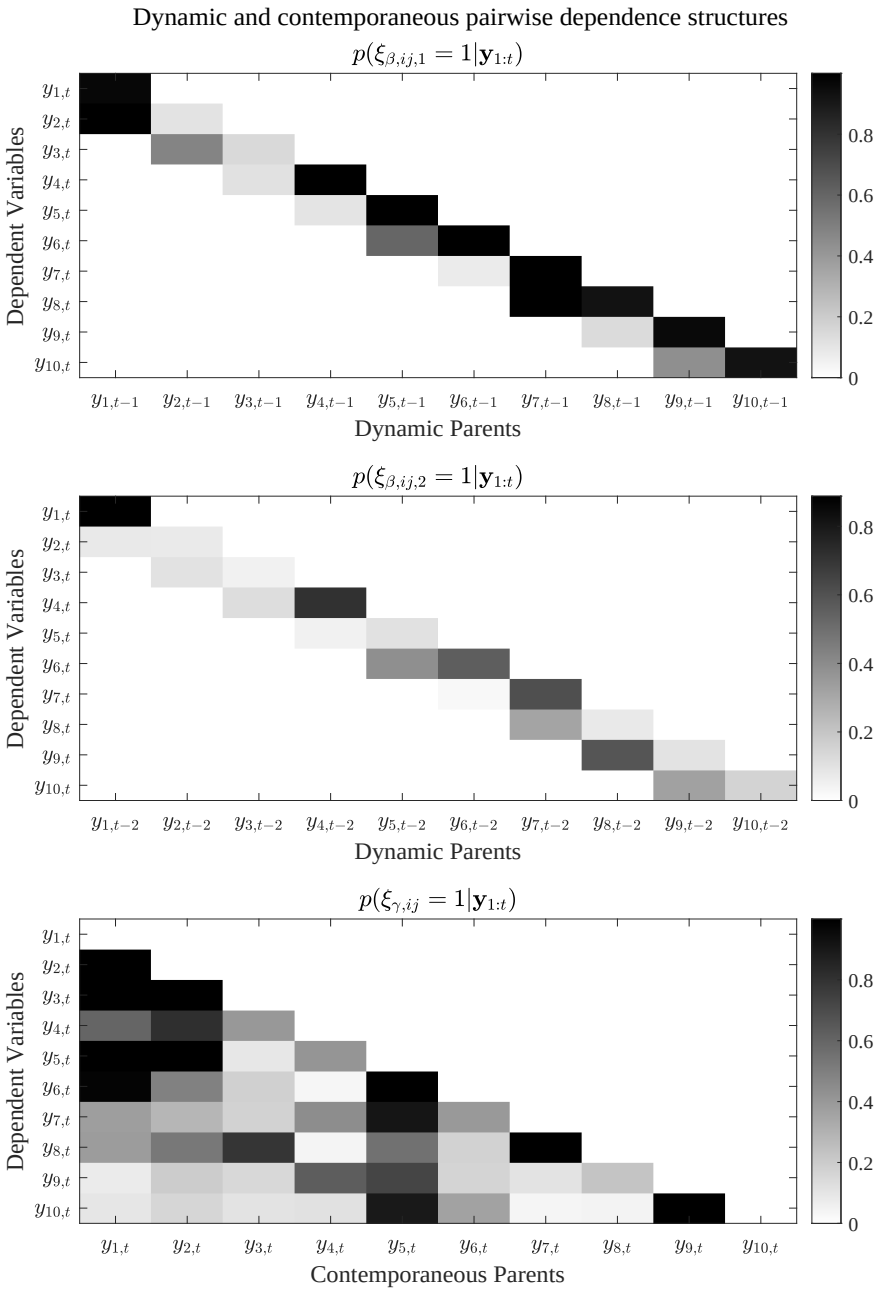

Figure 4.1: Image plots of dynamic and contemporaneous pairwise dependence structures. In each plot, the black (white) colour indicates strong (weak) evidence of dependence. The variables are $y_{1,t}$ = consumption, $y_{2,t}$ = investment, $y_{3,t}$ = GDP growth, $y_{4,t}$ = GDP deflator, $y_{5,t}$ = industrial production, $y_{6,t}$ = unemployment rate, $y_{7,t}$ = corporate bond spread, $y_{8,t}$ = S&P 500 stock returns, $y_{9,t}$ = 10-year Treasury rate and $y_{10,t}$ = Federal funds rate.

4.6.4 Out-of-sample forecast results

In this section, we evaluate the pseudo out-of-sample forecast performance of the joint model using Bayesian model averaging (joint model with BMA) versus the performance measured by basing predictive inference from a single best model (i.e. a model with highest posterior probability). At this point, we evaluate the range of good models M over a set of truncated candidate $|M|$ models and look at only the models that have a posterior model probability higher than the threshold 0.005, which can be partitioned as in Table 4.14

Truncated model subspace										
Order	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
Truncated $\|M_i\|$	3	7	11	29	12	26	43	38	33	40

Table 4.14: The table shows order of models and the truncated model subspace.

We do so by comparing the root mean squared forecast error (RMSFE) and the mean absolute forecast error (MAFE) statistics over the pseudo out-of-sample forecast period 1984:Q2-2022:Q3.[10] The former is defined as the size of the forecast error that is more sensitive to occasionally large errors when compared with the mean and the latter is a measure of the expected size of the forecast error. Hence, we can compare the RMSFE and MAFE to determine whether the forecast contains occasionally large errors. We present iterated forecasts for horizons between one and eight quarters. For the multi-step-ahead forecast, analytical solutions do not exist. Consequently, we evaluate the h-step-ahead predictive mean and variance using Monte Carlo simulations for a sample size $10,000$. The aim of this forecasting exercise is to assess the gains from using graphical variable selection with implied model averaging over the forecast performance resulting from a single best model with the highest posterior probability. As a measure of overall forecast performance, we use ten variables of interest, consumption,

[10]We calculate the RMSFE and MAFE as in Korobilis (2013).

investment, GDP growth, GDP deflator, industrial production, unemployment rate, corporate bond spread, S&P 500 stock returns, 10-year Treasury maturity rate and Federal funds rate.

Table 4.15 and Table 4.16 present RMSFEs and MAFEs for each of ten variables of interest, eight different forecast horizons and two different forecast metrics. The numbers in Table 4.15 are ratios of the RMSFE for the joint model with BMA divided by the RMSFE of the joint model with highest posterior probability. Ratio values smaller than one suggest better performance of the joint model with BMA relative to the joint model with the highest posterior probability.

In general, we find that the joint model with BMA to yield forecast performance improvements over the joint model having the highest posterior probability for the majority of forecast horizons and for all variables.

Looking at the relative RMSFEs for the variable consumption, we find forecast performance improvement of the joint model with BMA in terms of achieving lower RMSFE than the joint model with the highest posterior probability. In particular, the joint model with the highest posterior probability at eight-quarter horizon predicts much worse. The outcomes for the variables investment and GDP growth are slightly different than those for other variables. On one hand, the relative RMSFEs are very low at one-quarter and eight-quarter horizons. This may reflect the fact that the RMSFEs obtained from a single best model with highest posterior probability are too high. Investment and GDP growth are second and third in the recursive predictive path and is, as a consequence, more difficult to predict than the other variables. On the other hand, the difference in RMFSE between the joint model with BMA that averages over several good models and the joint model with highest posterior probability model, RMSFEs coming from the latter class of models remained higher as shown in Table 4.19. Interpreting the results for the variables GDP growth, GDP deflator, industrial production, unemployment rate, corporate bond spread, S&P 500 stock returns,

10-year Treasury maturity rate and Federal funds rate jointly, we see that is there a high instability in the RMSFEs obtained from the joint models with the highest posterior probability at eight-quarter horizon reported in Table 4.19 leading to very low values of the relative RMSFEs displayed in Table 4.15.

Comparing the RMSFE and the MAFE generated from both competing models can guide us on depicting the presence of any unusual forecast error size. As the difference between those quantities widens, the inconsistency of the forecast error size increases. The results displayed in Tables 4.17-4.18 and 4.19-4.20 do not show such large differences.

The dissatisfying results generated from the joint model with the highest posterior probability may have several reasons. First, for computational feasibility of the graphical variable selection approach, we imposed restrictions on the recursive models with only lower bidiagonal elements of the dynamic predictors to be included in or excluded from the model. The underlying model specification may be an issue. Second, the prior specification on the component parameter space and on the model space are crucial in this setup. The priors of the parameter space influence the predictive likelihoods and the posterior model probabilities. Similarly, the posterior model probabilities may be strongly influenced by the prior model probabilities. For instance, as a robustness test (not reported here), we tried alternative prior hyperparameter values for the degrees of freedom, $n_{i,0}$, and found that the RMSFE values are highly sensitive to different values of $n_{i,0}$. Likewise, alternative prior hyperparameter values for π generated unstable posterior weights. At this point, any certain conclusions on the unimpressive out-of-sample forecast performance of the joint single best model can be misleading.

RMSFE of joint model with BMA relative to highest posterior probability model								
F. Horizon	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
Variables	RMSFE							
PCECC96	1.0033	0.9997	1.0013	0.9858	1.0023	0.8315	0.9442	0.6260
GPDIC1	0.2481	0.7717	0.6976	0.7594	0.7559	0.7465	0.7299	0.7312
GDPC1	0.4506	0.9215	0.5705	0.8663	0.4818	0.6738	0.1937	0.3176
GDPCTPI	0.9547	0.9270	0.9365	0.6258	0.9789	0.2645	0.7562	0.0697
INDPRO	0.8901	0.9490	0.7148	0.4668	0.6222	0.1116	0.4891	0.0206
UNRATE	0.9925	0.9838	0.9292	0.7725	0.7224	0.3106	0.2136	0.0910
CS	1.0009	0.9974	1.0036	0.9688	0.9772	0.7368	0.4794	0.2024
S&P 500	1.0017	1.0001	1.0003	0.9928	0.9437	1.0009	0.7832	0.3207
GS10	0.9980	1.0015	0.9932	1.0013	0.9861	0.8697	1.0027	0.1417
FEDFUNDS	0.9999	1.0029	0.9956	0.9988	0.8768	0.7513	0.2962	0.1928

Table 4.15: The table reports out-of-sample forecast performance of the joint model with BMA relative to the model with the highest posterior probability, by computing the relative RMSFEs of ten macroeconomic and financial variables over the sample period 1984:Q2 - 2022:Q3.

MAFE of joint model with BMA relative to highest posterior probability model								
F. Horizon	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
Variables	MAFE							
PCECC96	1.0187	0.9938	1.0074	0.9220	1.0109	0.5107	0.7771	0.3488
GPDIC1	0.1741	0.6837	0.5928	0.6684	0.6623	0.6520	0.6303	0.6330
GDPC1	0.2281	0.7770	0.3097	0.6575	0.2486	0.3991	0.0933	0.1552
GDPCTPI	0.9162	0.8946	0.8929	0.5090	0.9558	0.1874	0.6579	0.0532
INDPRO	0.7931	0.7413	0.4997	0.2602	0.3951	0.0618	0.2809	0.0161
UNRATE	1.0233	0.8027	0.6650	0.3624	0.3199	0.1325	0.0710	0.0707
CS	1.0055	0.9864	1.0247	0.9024	1.0602	0.5872	0.3451	0.1375
S&P 500	1.0092	1.0003	1.0021	0.9748	0.8601	1.0055	0.6381	0.2392
GS10	0.9983	1.0012	0.9937	1.0011	0.9855	0.8876	1.0010	0.1165
FEDFUNDS	1.0000	0.9856	0.9937	0.9345	0.7592	0.5682	0.1969	0.1532

Table 4.16: The table reports out-of-sample forecast performance of the joint model with BMA relative to the model with the highest posterior probability, by computing the relative MAFEs of ten macroeconomic and financial variables over the sample period 1984:Q2 - 2022:Q3.

4.7 Summary

The methodology proposed in this chapter outlined a Bayesian dynamic graphical model approach for estimating a high-dimensional TVP-VAR-VD model.

This enabled the multivariate TVP-VAR-VD model to split into simpler multiple regression DLM components by allowing local computations equation-by-equation instead of a high-dimensional alternative. We proposed a nontraditional search algorithm that explored the model space to find high posterior probability graphs by using a Gray code algorithm. This corresponds to apply a graphical variable selection approach that followed by a Bayesian model selection and a Bayesian model averaging to explore complex and high-dimensional model spaces.

In specifying the TVP-VAR-VD model, we restricted the model parameters, defining $\boldsymbol{\beta}_t$ to be a bidiagonal matrix, specified the pattern of the coefficients in the matrix $\boldsymbol{\Gamma}_t$ as a strictly lower triangular matrix and $\boldsymbol{\Lambda}_t$ as a diagonal matrix. The recursive condition of the component multiple regression DLMs and the elements of the triangular system that incorporate autoregressive effects, cross-lagged effects and contemporaneous effects was sufficient for model identification implying the identification of all parameters.

Evaluating the performance of our modelling approach using ten quarterly U.S. macroeconomic and financial time series showed that there is a considerable model uncertainty within the range of good competing models for each series. The increase in the component model uncertainty was proportional to the number of the dynamic and the contemporaneous predictors. As a result, we compared the median probability models and the highest posterior probability models over the space of all competing models, which displayed almost identical results.

In addition, two competing models are used to obtain forecasts at one to eight quarter horizons for the variables consumption, investment, GDP growth, GDP deflator, industrial production, unemployment rate, corporate bond spread, S&P 500 stock returns, 10-year Treasury maturity rate and Federal funds rate. We can conclude that the joint model with the highest posterior probability do not perform significantly better than the joint model with BMA given that the joint model with BMA almost has the lowest RMSFE over the majority of forecast

horizons and all variables.

4.8 Appendix

4.8.1 Prediction and filtering steps for the component models

We apply the Kalman filter algorithm in a similar fashion as described in Zhao et al. (2016). Continuing from Section 4.4.1, the following statements hold for computing the predictive densities for the state parameters and the precision parameters in $i = 1 : N$ parallel steps.

i) The one-step-ahead predictive density of $\boldsymbol{\theta}_{i,t}$ given $\mathbf{y}_{1:t-1}$ is a Normal distribution and of $\lambda_{i,t}$ given $\mathbf{y}_{1:t-1}$ is a Gamma distribution, respectively

$$(\boldsymbol{\theta}_{i,t}|\lambda_{i,t}, \mathbf{y}_{1:t-1}) \sim \mathcal{N}(\mathbf{a}_{i,t}, \mathbf{R}_{i,t}),$$

$$(\lambda_{i,t}|\mathbf{y}_{1:t-1}) \sim \mathcal{G}(\varphi_i n_{i,t-1}/2, \varphi_i d_{i,t-1}/2),$$

with parameters

$$\begin{aligned} \mathbf{a}_{i,t} &= E(\boldsymbol{\theta}_{i,t}|\mathbf{y}_{1:t-1}) = \mathbf{m}_{i,t-1}, \\ \mathbf{R}_{i,t} &= V(\boldsymbol{\theta}_{i,t}|\mathbf{y}_{1:t-1}) = \mathbf{C}_{i,t-1}/\delta_i, \\ d_{i,t-1} &= n_{i,t-1} s_{i,t-1}, \\ s_{i,t-1} &= d_{i,t-1}/n_{i,t-1}, \end{aligned} \tag{4.39}$$

where $s_{i,t-1}$ is the point estimate of the observational variance $1/\lambda_{i,t}$.

ii) The one-step-ahead predictive density of $y_{i,t}$ given the parental set $\mathbf{F}_{pa(i),t}$ and $\mathbf{y}_{1:t-1}$ is a Student T distribution

$$(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) \sim \mathcal{T}(f_{i,t}, q_{i,t}),$$

with parameters

$$\begin{aligned} f_{i,t} &= E(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) = \mathbf{F}'_{pa(i),t}\mathbf{a}_{i,t}, \\ q_{i,t} &= V(y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}) = \mathbf{F}'_{pa(i),t}\mathbf{R}_{i,t}\mathbf{F}_{pa(i),t} + s_{i,t-1}. \end{aligned} \tag{4.40}$$

Let's partition $\mathbf{a}_{i,t}$ and $\mathbf{R}_{i,t}$ as follow

$$\mathbf{a}_{i,t} = \begin{pmatrix} \mathbf{a}_{i,\beta,t} \\ \mathbf{a}_{i,\gamma,t} \end{pmatrix},$$

and

$$\mathbf{R}_{i,t} = \begin{pmatrix} \mathbf{R}_{i,\beta,t} & \mathbf{R}_{i,\beta\gamma,t} \\ \mathbf{R}_{i,\gamma\beta,t} & \mathbf{R}_{i,\gamma,t} \end{pmatrix},$$

then, we obtain

$$\begin{aligned} f_{i,t} &= \mathbf{x}'_{dp(i),t}\boldsymbol{\beta}_{i,t} + \mathbf{y}'_{cp(i),t}\boldsymbol{\gamma}_{i,t}, \\ q_{i,t} &= \mathbf{y}'_{cp(i),t}\mathbf{R}_{i,\gamma,t}\mathbf{y}_{cp(i),t} + 2\mathbf{y}'_{cp(i),t}\mathbf{R}_{i,\beta\gamma,t}\mathbf{x}_{dp(i),t} + \mathbf{x}'_{dp(i),t}\mathbf{R}_{i,\beta,t}\mathbf{x}_{dp(i),t} + s_{i,t}. \end{aligned} \tag{4.41}$$

iii) The filtering densities of $\boldsymbol{\theta}_{i,t}$ given $\mathbf{y}_{1:t}$ is a Normal distribution and of $\lambda_{i,t}$ given $\mathbf{y}_{1:t}$ is a Gamma distribution

$$(\boldsymbol{\theta}_{i,t}|\lambda_{i,t}, \mathbf{y}_{1:t}) \sim \mathcal{N}(\mathbf{m}_{i,t}, \mathbf{C}_{i,t}),$$

$$(\lambda_{i,t}|\mathbf{y}_{1:t}) \sim \mathcal{G}(n_{i,t}/2, d_{i,t}/2),$$

with parameters

$$\begin{aligned}
&\mathbf{m}_{i,t} = E(\boldsymbol{\theta}_{i,t}|\mathbf{y}_{1:t}) = \mathbf{a}_{i,t} + \mathbf{K}_{i,t}e_{i,t},\\
&\mathbf{C}_{i,t} = V(\boldsymbol{\theta}_{i,t}|\mathbf{y}_{1:t}) = s_{i,t}/s_{i,t-1}\big(\mathbf{R}_{i,t} - \mathbf{K}_{i,t}\mathbf{K}'_{i,t}q_{i,t}\big),\\
&\mathbf{K}_{i,t} = \mathbf{R}_{i,t}\mathbf{F}_{pa(i),t}/q_{i,t},\\
&e_{i,t} = y_{i,t} - f_{i,t},\\
&n_{i,t} = \varphi_i n_{i,t-1} + 1,\\
&d_{i,t} = \varphi_i d_{i,t-1} + s_{i,t-1}e^2_{i,t}/q_{i,t},\\
&s_{i,t} = s_{i,t-1} + s_{i,t-1}/n_{i,t}\big(e^2_{i,,t}/q_{i,t} - 1\big),
\end{aligned} \tag{4.42}$$

where $\mathbf{K}_{i,t}$ is the Kalman gain, the state adaptive coefficient vector.

4.8.2 h-step ahead forecast

Distributions for future values of the state vectors and future observations are available at any time t for each of the N univariate series. Conditional on $\mathbf{y}_{1:t}$, the $h-$step-ahead forecast distribution for the state vectors and the corresponding $h-$step-ahead predictive distribution are

$$(\boldsymbol{\theta}_{i,t+h}, \lambda_{i,t+h}|\mathbf{y}_{1:t}) \sim \mathcal{NG}(\mathbf{a}_{i,t}(h), \mathbf{R}_{i,t}(h), n_{i,t}(h), d_{i,t}(h)), \tag{4.43}$$

$$(y_{i,t+h}|\mathbf{F}_{pa(i),t+h}, \mathbf{y}_{1:t}) \sim \mathcal{T}_{\varphi_i n_{i,t}}(f_{i,t+h}(\mathbf{F}_{pa(i),t+h}), q_{i,t+h}(\mathbf{F}_{pa(i),t+h})), \tag{4.44}$$

where the moments of the distributions of the state vectors are obtained sequentially from time t as

$$\begin{aligned}
\mathbf{a}_{i,t}(h) &= \mathbf{a}_{i,t}(h-1),\\
\mathbf{R}_{i,t}(h) &= \mathbf{R}_{i,t}(h-1) + \mathbf{W}_{i,t+h},\\
n_{i,t}(h) &= \varphi_i n_{i,t},
\end{aligned} \tag{4.45}$$

for $h = 1, 2, \ldots,$ with initial values $\mathbf{a}_{i,t}(0) = \mathbf{m}_{i,t}$, $\mathbf{W}_{i,t}(0) = \mathbf{C}_{i,t}/(1/\delta_i - 1)$, hence $\mathbf{R}_{i,t}(0) = \mathbf{C}_{i,t}/\delta_i$. Similarly, the moments of the $h-$step-ahead forecast predictive distribution are

$$\begin{aligned} f_{i,t}(h) &= \mathbf{F}'_{pa(i),t+h}\mathbf{a}_{i,t}(h), \\ q_{i,t}(h) &= \mathbf{F}'_{pa(i),t+h}\mathbf{R}_{i,t}(h)\mathbf{F}_{pa(i),t+h} + s_{i,t-1}. \end{aligned} \tag{4.46}$$

4.8.3 The joint model

Continuing from Section 4.4.2, let's assume that for each $i = 1 : N$, the mean and the variance matrix for $\mathbf{F}_{pa(i),t}$ exist and have the forms

$$\begin{aligned} \mathbf{h}_{pa(i),t} &= E[\mathbf{F}_{pa(i),t}|\mathbf{y}_{1:t-1}], \\ \mathbf{H}_{pa(i),t} &= V[\mathbf{F}_{pa(i),t}|\mathbf{y}_{1:t-1}], \end{aligned}$$

respectively. To compute the conditional Student T distribution of $y_{i,t}$, we set the degrees of freedom $n_{i,t} > 1$ and denote the mean

$$\begin{aligned} E[y_{i,t}|\mathbf{y}_{1:t-1}] &= E\{E[y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}]|\mathbf{y}_{1:t-1}\}, \\ &= E[\mathbf{F}'_{pa(i),t}\mathbf{a}_{i,t}|\mathbf{y}_{1:t-1}], \\ &= \mathbf{h}'_{pa(i),t}\mathbf{a}_{i,t}. \end{aligned} \tag{4.47}$$

Setting $n_{i,t} > 2$, the conditional variance of $y_{i,t}$ is

$$\begin{aligned}
V[y_{i,t}|\mathbf{y}_{1:t-1}] &= E\{V[y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}]|\mathbf{y}_{1:t-1}\} \\
&+ V\{E[y_{i,t}|\mathbf{F}_{pa(i),t}, \mathbf{y}_{1:t-1}]|\mathbf{y}_{1:t-1}\} \\
&= E[\frac{n_{i,t}}{n_{i,t}-2} q_{i,t}|\mathbf{y}_{1:t-1}] + V[f_{i,t}|\mathbf{y}_{1:t-1}], \\
&= \frac{n_{i,t}}{n_{i,t}-2}\{s_{i,t} + E[\mathbf{F}'_{pa(i),t}\mathbf{R}_{i,t}\mathbf{F}_{pa(i),t}|\mathbf{y}_{1:t-1}]\} \\
&+ V[\mathbf{F}'_{pa(i),t}\mathbf{a}_{i,t}|\mathbf{y}_{1:t-1}], \\
&= \frac{n_{i,t}}{n_{i,t}-2}[s_{i,t} + \mathbf{h}'_{pa(i),t}\mathbf{R}_{i,t}\mathbf{h}_{pa(i),t}] \\
&+ \text{trace}\{\mathbf{R}_{i,t}\mathbf{H}_{pa(i),t}\}] + \mathbf{a}'_{i,t}\mathbf{H}_{pa(i),t}a_{i,t},
\end{aligned} \tag{4.48}$$

where the two terms in the last equation in the squared brackets are from the property of the expected value of a quadratic form

$$\begin{aligned}
E[\mathbf{F}'_{pa(i),t}\mathbf{R}_{i,t}\mathbf{F}_{pa(i),t}|\mathbf{y}_{1:t-1}] &= E[\text{trace}\mathbf{F}_{pa(i),t}\mathbf{R}_{i,t}\mathbf{F}'_{pa(i),t}|\mathbf{y}_{1:t-1}], \\
&= \text{trace}E[\mathbf{F}_{pa(i),t}\mathbf{R}_{i,t}\mathbf{F}'_{pa(i),t}|\mathbf{y}_{1:t-1}], \\
&= \text{trace}\mathbf{R}_{i,t}E[\mathbf{F}_{pa(i),t}\mathbf{F}'_{pa(i),t}|\mathbf{y}_{1:t-1}], \\
&= \text{trace}\mathbf{R}_{i,t}[V(\mathbf{F}_{pa(i),t}) + E(\mathbf{F}_{pa(i),t})E(\mathbf{F}_{pa(i),t})'], \\
&= \text{trace}\mathbf{R}_{i,t}V(\mathbf{F}_{pa(i),t}) + \text{trace}\mathbf{R}_{i,t}E(\mathbf{F}_{pa(i),t})E(\mathbf{F}_{pa(i),t})', \\
&= \text{trace}\mathbf{R}_{i,t}V(\mathbf{F}_{pa(i),t}) + \text{trace}E(\mathbf{F}_{pa(i),t})'\mathbf{R}_{i,t}E(\mathbf{F}_{pa(i),t}), \\
&= \text{trace}\mathbf{R}_{i,t}V(\mathbf{F}_{pa(i),t}) + E(\mathbf{F}_{pa(i),t})'\mathbf{R}_{i,t}E(\mathbf{F}_{pa(i),t}), \\
&= \text{trace}\{\mathbf{R}_{i,t}\mathbf{H}_{pa(i),t}\} + \mathbf{h}'_{pa(i),t}\mathbf{R}_{i,t}\mathbf{h}_{pa(i),t}.
\end{aligned} \tag{4.49}$$

Using the partitioned vector $\mathbf{a}_{i,t}$ and matrix $\mathbf{R}_{i,t}$, we get the following results For i = 1:N

if $i = 1$

$$
\begin{aligned}
f_{i,t} &= \mathbf{x}'_{dp(i),t}\mathbf{a}_{i,\beta,t} \\
q_{i,t} &= \frac{n_{i,t}}{n_{i,t}-2}\big(s_{i,t-1} + \mathbf{x}'_{dp(i),t}\mathbf{R}_{i,\beta,t}\mathbf{x}_{dp(i),t}\big),
\end{aligned}
$$

otherwise

$$
\begin{aligned}
f_{i,t} &= \mathbf{x}'_{dp(i),t}\mathbf{a}_{i,\beta,t} + \mathbf{h}'_{cp(i),t}\mathbf{a}_{i,\gamma,t}, \\
q_{i,t} &= \frac{n_{i,t}}{n_{i,t}-2}\big(s_{i,t-1} + \mathbf{h}'_{cp(i),t}\mathbf{R}_{i,\gamma,t}\mathbf{h}_{cp(i),t} \\
&\quad + \operatorname{trace}(\mathbf{R}_{i,\gamma,t}\mathbf{H}_{cp(i),t}) + 2\mathbf{x}'_{dp(i),t}\mathbf{R}_{i,\beta\gamma,t}\mathbf{h}_{cp(i),t} \\
&\quad + \mathbf{x}'_{dp(i),t}\mathbf{R}_{i,\beta,t}\mathbf{x}_{dp(i),t}\big) + \mathbf{h}'_{cp(i),t}\mathbf{R}_{i,\gamma,t}\mathbf{h}_{cp(i),t}.
\end{aligned}
\tag{4.50}
$$

The off-diagonal elements of the covariance matrix $C(y_{i,t}, \mathbf{y}_{1:i-1,}|\mathbf{y}_{1:t-1})$ may be calculated as

$$
(y_{i,t}, \mathbf{y}_{1:i-1,t}|\mathbf{y}_{1:t-1}) = \mathbf{Q}_{1:i-1,t}\mathbf{a}_{1:i-1,\gamma,t}, \tag{4.51}
$$

and we end the process.

Finally, we plug-in the values obtained from Equation (4.50) and (4.51) into the joint mean $\mathbf{f}_t$ and $\mathbf{Q}_t$ as appropriate. Those steps may be repeated for the h-step-ahead forecast exercise. For brevity, only the steps using the one-step-ahead forecast model are reported (see Zhao et al. (2016) for details).

4.8.4 More Results

RMSFE of joint model with BMA								
F. Horizon	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
Variables	RMSFE							
PCECC96	1.1413	1.1437	1.1480	1.1505	1.1520	1.1532	1.1559	1.1945
GPDIC1	0.8345	0.8345	0.8372	0.8383	0.8381	0.8409	0.8432	0.8460
GDPC1	1.0327	1.0339	1.0389	1.0410	1.0462	1.0488	1.0594	1.0584
GDPCTPI	0.9156	0.9178	0.9208	0.9241	0.9255	0.9367	0.9335	1.1521
INDPRO	0.9428	0.9450	0.9498	0.9485	0.9552	0.9534	0.9655	1.1820
UNRATE	1.2256	1.2300	1.2338	1.2399	1.2424	1.2677	1.2575	1.6854
CS	0.8415	0.8411	0.8392	0.8410	0.8389	0.8410	0.8349	0.8433
S&P 500	1.0248	1.0277	1.0229	1.0218	1.0248	1.0291	1.0185	1.0268
GS10	0.8742	0.8711	0.8473	0.8328	0.8346	0.8359	0.8405	0.8161
FEDFUNDS	0.5135	0.5147	0.5158	0.5099	0.5074	0.5087	0.5106	0.5516

Table 4.17: The table reports out-of-sample forecast performance of the joint model with BMA, by computing the RMSFEs of ten macroeconomic and financial variables over the sample period 1984:Q2 - 2022:Q3.

MAFE of joint model with BMA								
F. Horizon	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
Variables	MAFE							
PCECC96	0.4904	0.4886	0.4924	0.4908	0.4879	0.4839	0.4826	0.5717
GPDIC1	0.5684	0.5664	0.5688	0.5682	0.5661	0.5689	0.5701	0.5726
GDPC1	0.4807	0.4814	0.4851	0.4853	0.4893	0.4902	0.5021	0.4978
GDPCTPI	0.6309	0.6306	0.6354	0.6351	0.6375	0.6424	0.6466	0.8776
INDPRO	0.4869	0.4911	0.4930	0.4965	0.4943	0.5247	0.4961	0.9270
UNRATE	0.4103	0.4171	0.4111	0.4363	0.4076	0.5312	0.4088	1.3068
CS	0.5732	0.5717	0.5669	0.5685	0.5591	0.5662	0.5275	0.5638
S&P 500	0.7161	0.7183	0.7113	0.7091	0.7106	0.7181	0.7010	0.7257
GS10	0.6905	0.6867	0.6737	0.6640	0.6647	0.6679	0.6725	0.6642
FEDFUNDS	0.3270	0.3276	0.3277	0.3234	0.3195	0.3343	0.3242	0.4314

Table 4.18: The table reports out-of-sample forecast performance of the joint model BMA, by computing the MAFEs of ten macroeconomic and financial variables over the sample period 1984:Q2 - 2022:Q3.

RMSFE of joint model with highest posterior probability								
F. Horizon	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
Variables	RMSFE							
PCECC96	1.1376	1.1441	1.1465	1.1671	1.1494	1.3870	1.2243	1.9081
GPDIC1	3.3630	1.0814	1.2002	1.1039	1.1087	1.1265	1.1552	1.1569
GDPC1	2.2920	1.1220	1.8210	1.2016	2.1716	1.5565	5.4695	3.3323
GDPCTPI	0.9591	0.9900	0.9833	1.4766	0.9455	3.5415	1.2344	16.5360
INDPRO	1.0593	0.9958	1.3287	2.0320	1.5353	8.5434	1.9741	57.5167
UNRATE	1.2349	1.2503	1.3277	1.6050	1.7197	4.0817	5.8884	18.5261
CS	0.8407	0.8433	0.8361	0.8681	0.8584	1.1414	1.7416	4.1661
S&P 500	1.0231	1.0276	1.0225	1.0292	1.0859	1.0282	1.3004	3.2013
GS10	0.8759	0.8698	0.8531	0.8316	0.8464	0.9610	0.8382	5.7599
FEDFUNDS	0.5135	0.5132	0.5181	0.5105	0.5787	0.6771	1.7236	2.8615

Table 4.19: The table reports out-of-sample forecast performance of the joint model with the highest posterior probability, by computing the RMSFEs of ten macroeconomic and financial variables over the sample period 1984:Q2 - 2022:Q3.

MAFE of joint model with highest posterior probability								
F. Horizon	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
Variables	MAFE							
PCECC96	0.4814	0.4917	0.4888	0.5323	0.4827	0.9474	0.6211	1.6389
GPDIC1	3.2641	0.8285	0.9596	0.8501	0.8547	0.8726	0.9044	0.9046
GDPC1	2.1075	0.6195	1.5663	0.7381	1.9683	1.2282	5.3817	3.2075
GDPCTPI	0.6886	0.7049	0.7116	1.2477	0.6670	3.4279	0.9829	16.5096
INDPRO	0.6139	0.6624	0.9866	1.9077	1.2512	8.4901	1.7659	57.5087
UNRATE	0.4009	0.5196	0.6181	1.2039	1.2740	4.0094	5.7541	18.4836
CS	0.5701	0.5796	0.5532	0.6299	0.5274	0.9642	1.5286	4.1001
S&P 500	0.7096	0.7181	0.7098	0.7274	0.8262	0.7142	1.0986	3.0341
GS10	0.6917	0.6859	0.6779	0.6633	0.6745	0.7524	0.6718	5.7022
FEDFUNDS	0.3270	0.3324	0.3298	0.3460	0.4208	0.5883	1.6466	2.8157

Table 4.20: The table reports out-of-sample forecast performance of the joint model with the highest posterior probability, by computing the MAFEs of ten macroeconomic and financial variables over the sample period 1984:Q2 - 2022:Q3.

4.8.5 Data appendix

The quarterly time series variables used in the TVP-VAR-VD models are taken from the FRED database of the Federal Reserve Bank of St Louis spanning from 1959Q1 to 2022Q3. The columns of Table 4.21, denote the series numbers, Tcode denotes the data transformations based on McCracken and Ng (2020), series denotes the FRED mnemonic, and description denotes a brief definition of the series.

Corporate bond spread is defined as Moody's Baa corporate bond yield minus Moody's Aaa corporate bond yield. The modified Tcode, 1^*, stands for no transformation of the series.

Time series used in the TVP-VAR-VD model			
ID	Series	Tcode	Description
1	PCECC96	5	Real Personal Consumption Expenditures
2	GPDIC1	5	Real Gross Private Domestic Investment
3	GDPC1	5	Real Gross Domestic Product
4	GDPCTPI	6	Gross Domestic Product: Chain-type Price Index
5	INDPRO	5	Industrial Production Index
6	UNRATE	2	Civilian Unemployment Rate
7	CS	1^*	Moody's Seasoned Baa-Aaa Corporate Bond Spread
8	S&P 500	5	S&P's Common Stock Price Index: Composite
9	GS10	2	10-Year Treasury Constant Maturity Rate
10	FEDFUNDS	2	Effective Federal Funds Rate

Table 4.21: The quarterly time series variables used in the TVP-VAR-VD models.

Chapter 5

Discussion

We have drawn on ideas of high-dimensional inference problems from pairwise composite likelihood and high-dimensional sparse inference problems from dynamic graphical models on addressing the challenges of modern empirical macroeconomics. We offer some suggestions about future directions that seem promising for further research in the identification of financial shocks, theory, application and computation of composite likelihood methods and dynamic graphical models within the Bayesian inferential paradigm.

Chapter 2 The first objective was to understand the real effects of credit market disruption through a measure of financial distress, the financial external premium, which allowed to vary over time. Although our focus was on the transmission of the financial shocks into the real economy in one directional setup, identification of financial shocks is a challenging task due to simultaneity problem, as noted by Gertler and Gilchrist (2019). Extending the analysis to consider aspects of the simultaneity problem using a TVP-VAR-SV model is an important direction for future research.

Chapter 3 The second objective was to address the high-dimensional inference problem of the TVP-VAR-SV model through a novel Bayesian pairwise composite

likelihood approach. This method becomes particularly relevant when estimating the full multivariate TVP-VAR-SV model is computationally infeasible. A sufficient condition for using the Bayesian pairwise composite likelihood approach is that parameter estimation for each bivariate TVP-VAR-SV model remains computationally tractable. However, implementing the FFBS algorithm across $\binom{N}{2}$ parallel steps for $N > 50$ posed severe memory challenges, as it required storing every element of the stacked parameters $\boldsymbol{\beta}_{c,t}$, $\mathbf{h}_{c,t}$, and $\boldsymbol{\alpha}_{c,t}$ for $i = 1, \ldots, \binom{N}{2}$ models, over $r = 1, \ldots, R$ MCMC iterations (after burn-in), and for all $t = 1, \ldots, T$ time periods. Developing algorithms that reduce this space complexity represents a promising direction for future research in Bayesian estimation using composite likelihoods.

Another avenue for improvement concerns the weighting of pairwise likelihood components. Future work will explore optimal weighting schemes to better capture the contribution of each pairwise component to model assessment. As noted by Verbeke and Molenberghs (2005), when random (non-constant) weights are used in the pairwise score functions, it is unclear whether the expected value of the weighted pairwise score remains zero, raising important theoretical considerations.

Additionally, we observed an issue related to the magnitude of composite impulse responses. Although the responses are economically plausible, further investigation is needed to address this magnitude discrepancy.

Finally, the Direct Averaging Method introduced in this thesis was interpreted as a computational approximation to a MH algorithm, where parameter draws from each bivariate model are implicitly accepted with probability one. While this interpretation is conceptually appealing, deriving the formal MH acceptance ratio in the second step remains an open problem, and future research will aim to establish these theoretical conditions.

Chapter 4 The third objective was to establish a feasible computational algorithm for the problem of variable selection using the Bayesian dynamic graphical model approach. Although the proposed Bayesian dynamic graphical model approach enables the multivariate TVP-VAR-VD model to split into simpler multiple regression DLM components, the problem of selecting the good models from among the $2^p + 2^{2p}(2^N - 2)$ possibilities is not trivial. We have applied a non-traditional search approach in parallel to sets of component univariate DLMs. Nevertheless, the implied model selection and model averaging that proceeded the variable selection step cannot be applied with a very large number of variables.

Another potential concern is that we have ordered the variables as though the macroeconomic variables comes before the financial variables, which may reflect economic reasoning and theory, as have been emphasized in previous studies Bańbura et al. (2010) and Gilchrist and Zakrajšek (2012). In a forecasting exercise, as noted by West (2020), ordering of variables may be redundant because the results are affected from the precision matrices and the multiple regression DLM components. A future research may investigate ordering free algorithms.

Furthermore, the posterior probabilities over the model space have shown appreciable sensitivity to two components of the prior probabilities, that is, the prior for the graphical models and the prior for the dynamic parameters in the component multiple regression DLMs. We may investigate Bayesian variable selection approaches that encourage posterior probabilities over the model space less sensitive to the prior specification.

Bibliography

Ahelegbey, D. F., M. Billio, and R. Casarin (2016), "Bayesian graphical models for structural vector autoregressive processes." *Journal of Applied Econometrics*, 31, 357–386.

Andrew, A., M. Piazzesi, and M. Wei (2006), "What does the yield curve tell us about GDP growth?" *Journal of econometrics*, 131, 359–403.

Azzalini, A. (1983), "Maximum likelihood estimation of order m for stationary stochastic processes." *Biometrika*, 70, 381–387.

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

Barbieri, M. M. and J. O. Berger (2004), "Optimal predictive model selection." *The Annals of Statistics*, 32, 870—897.

Berger, J. O. and G. Molina (2005), "Posterior model probabilities via path-based pairwise priors." *Statistica Neerlandica*, 59, 3–15.

Bernanke, B. and M. Gertler (1989), "Agency costs, net worth, and business fluctuations." *American Economic Review*, 79, 14–31.

Bernanke, B. S. (2003), "An unwelcome fall in inflation? speech, the economics roundtable, University of California–San Diego, July 23, 2003."

Bernanke, B. S. (2018), "The real effects of disrupted credit: Evidence from the global financial crisis." *Brookings Papers on Economic Activity*, 251–342.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005), "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach." *The Quarterly journal of economics*, 120, 387–422.

Bernanke, B. S., M. Gertler, and S. Gilchrist (1999), "The financial accelerator in a quantitative business cycle framework." *Handbook of macroeconomics*, 1, 1341–1393.

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems." *Journal of the Royal Statistical Society*, 36, 192–236.

Blake, A. P. and H. Mumtaz (2012), "Applied Bayesian econometrics for central bankers." *Technical Books*.

Boivin, J., M. P. Giannoni, and D. Stevanovic (2020), "Dynamic effects of credit shocks in a data-rich environment." *Journal of Business & Economic Statistics*, 38, 272–284.

Boothroyd, J. (1964), "Algorithm 246: Graycode." *Communications of the ACM*, 7, 701.

Cameron, P. J. (1994), *Combinatorics: topics, techniques, algorithms*. Cambridge University Press.

Canova, F. and C. Matthes (2021), "A composite likelihood approach for dynamic structural models." *The Economic Journal*, 131, 2447–2477.

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

Carvalho, C. M. and M. West (2007), "Dynamic matrix-variate graphical models." *Bayesian Analysis*, 2, 69—-98.

Chan, J. C. and E. Eisenstat (2018a), "Bayesian model comparison for time-varying parameter VARs with stochastic volatility." *Journal of applied econometrics*, 33, 509–532.

Chan, J. C. C. and E. Eisenstat (2018b), "Comparing hybrid time-varying parameter VARs." *Economics Letters*, 171, 1–5.

Chan, J. C. C., E. Eisenstat, C. Hou, and G. Koop (2020), "Composite likelihood methods for large Bayesian VAR's with stochastic volatility." *Journal of Applied Econometrics*, 35, 692–711.

Cogley, T. and T. J. Sargent (2002), "Evolving post-world War II US inflation dynamics." *NBER macroeconomics annual*, 16, 331–373.

Cogley, T. and T. J. Sargent (2005), "Drifts and volatilities: monetary policies and outcomes in the post wwii us." *Review of Economic dynamics*, 8, 262–302.

Cox, D. R. (1975), "Partial likelihood." *Biometrika*, 62, 269–276.

Cox, D. R. and R. Reid (2004), "A note on pseudo likelihood constructed from marginal densities." *Biometrika*, 91, 729–737.

Cripps, E., C. Carter, and R. Kohn (2005), "Variable selection and covariance selection in multivariate regression models." *Handbook of Statistics*, 25, 519–552.

DasGupta, A. (2008), *Asymptotic theory of statistics and probability*. Springer.

Fieuws, S. and G. Verbeke (2006), "Pairwise fitting of mixed models for the joint modelling of multivariate longitudinal profiles." *Biometrics*, 62, 424–431.

Friel, N. (2012), "Bayesian inference for gibbs random fields using composite likelihoods." *Proceedings of the Winter Simulation Conference*, 28, 1–8.

Frühwirth-Schnatter, S. (1994), "Data augmentation and dynamic linear models." *Journal of time series analysis*, 15, 183–202.

Galí, J. and L. Gambetti (2015), "The effects of monetary policy on stock market bubbles: Some evidence." *American Economic Journal: Macroeconomics*, 7, 233–57.

Gambetti, L. and A. Musso (2017), "Loan supply shocks and the business cycle." *Journal of Applied Econometrics*, 32, 764–782.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013), *Bayesian Data Analysis*, 3rd edition. Chapman and Hall/CRC.

Gertler, M. and S. Gilchrist (2018), "What happened: Financial factors in the great recession." *Journal of Economic Perspectives*, 32, 3–30.

Gertler, M. and S. Gilchrist (2019), "The channels of financial distress during the great recession: Some evidence on the aggregate effects."

Gilchrist, S., J. W. Sim, and E. Zakrajšek (2014), "Uncertainty, financial frictions, and investment dynamics." Technical report, National Bureau of Economic Research.

Gilchrist, S., V. Yankov, and E. Zakrajšek (2009), "Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets." *Journal of monetary Economics*, 56, 471–493.

Gilchrist, S. and E. Zakrajšek (2012), "Credit spreads and business cycle fluctuations." *American Economic Review*, 102, 1692–1720.

Godambe, V. P. (1960), "An optimum property of regular maximum likelihood estimation." *The Annals of Mathematical Statistics*, 31, 1208–1211.

Gruber, L. F. and M. West (2016), "GPU-accelerated Bayesian learning and forecasting in simultaneous graphical dynamic linear models." *Bayesian Analysis*, 11, 125–149.

Gruber, L. F. and M. West (2017), "Bayesian online variable selection and scalable multivariate volatility forecasting in simultaneous graphical dynamic linear models." *Econometrics and Statistics*, 3, 3–22.

Hamilton, J. D. (1994), *Time series analysis*. Princeton university press Princeton, NJ.

Hamilton, J. D. and D. H. Kim (2002), "A re-examination of the predictability of the yield spread for real economic activity." *Journal of Money, Credit, and Banking*, 34, 340–360.

Heaton, M. J. and J. G. Scott (2010), "Bayesian computation and the linear model." In *Frontiers of Statistical Decision Making and Bayesian Analysis* (C. Ming-Hui, K. D. Dipak, P. Müller, S. Dongchu, and Keying Y., eds.), 527–545, Springer.

Helbling, T., R. Huidrom, M. A. Kose, and C. Otrok (2011), "Do credit shocks matter? A global perspective." *European Economic Review*, 55, 340–353.

Hjort, N. L. and C. Varin (2008), "Ml, pl, ql in markov chain models." *Scandinavian Journal of Statistics*, 35, 64–82.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), "Bayesian model averaging: A tutorial." *Statistical science*, 14, 382–417.

Karlsson, S. and P. Österholm (2023), "Is the US Phillips curve stable? evidencefrom Bayesian vector autoregressions." *The Scandinavian Journal of Economics*, 125, 287–314.

Kent, J. (1982), "Robust properties of likelihood ratio tests." *Biometrika*, 69, 19–27.

Kilian, L. and H. Lütkepohl (2017), *Structural vector autoregressive analysis*. Cambridge University Press.

Kim, C. J. and C. R. Nelson (1999), *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. The MIT press.

Kim, S., N. Shephard, and S. Chib (1998), "Stochastic volatility: likelihood inference and comparison with arch models." *The review of economic studies*, 65, 361–393.

Kiyotaki, N. and J. Moore (1997), "Credit cycles." *Journal of political economy*, 105, 211–248.

Koller, D. and N. Friedman (2009), *Probabilistic graphical models: principles and techniques*. MIT press.

Koop, G. (2003), *Bayesian econometrics*. Wiley.

Koop, G. (2013), "Using VARs and TVP-VARs with many macroeconomic variables." *Central European Journal of Economic Modelling and Econometrics*, 4, 143–167.

Koop, G. (2017), "Bayesian methods for empirical macroeconomics with big data." *Review of Economic Analysis*, 9, 33–56.

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

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

Koop, G., R. Leon–Gonzalez, and R. Strachan (2009), "On the evolution of the monetary policy transmission mechanism,." *Journal of Economic Dynamics and Control*, 33, 997–1017.

Korobilis, D. (2013), "VAR forecasting using Bayesian variable selection." *Journal of Applied Econometrics*, 28, 204–230.

Kroese, D. P. and J. C. C. Chan (2014), *Statistical modelling and computation*. Springer.

Lauritzen, S. L. (1996), *Graphical models*. Clarendon Press.

Lindsay, B. G. (1988), "Composite likelihood methods." *Contemporary Mathematics*, 80, 221–239.

Madigan, D. and A. E. Raftery (1994), "Model selection and accounting for model uncertainty in graphical models using occam's window." *Journal of the American Statistical Association*, 89, 1535–1546.

McCracken, M. and S. Ng (2020), "FRED-QD: A quarterly database for macroeconomic research." *Federal Reserve Bank of St. Louis Working Paper 2020-005*.

Müller, P., S. Sivaganesan, and W. L. Purushottam (2010), "A Bayes rule for subgroup reporting." In *Frontiers of Statistical Decision Making and Bayesian Analysis* (C. Ming-Hui, K. D. Dipak, P. Müller, S. Dongchu, and Keying Y., eds.), 277–284, Springer.

Murphy, K. P. (2012), *Machine learning: a probabilistic perspective*. MIT press: London.

Nakajima, J. and M. West (2015), "Dynamic network signal processing using latent threshold models." *Digital Signal Processing*, 47, 5–16.

Pakel, C., N. Shephard, K. Sheppard, and R. Engle (2014), "Fitting vast dimensional time-varying covariance models." *Woking Paper*.

Pauli, F., W. Racugno, and L. Ventura (2011), "Bayesian composite marginal likelihoods." *Statistica Sinica*, 21, 149–164.

Prado, R. and M. West (2010), *Time series: Modelling, computation and inference*. Chapman and Hall/CRC Press.

Prieto, E., S. Eickmeier, and M. Marcelino (2016), "Time variation in macro–financial linkages." *Journal of Applied Econometrics*, 31, 1215–1233.

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

Ribatet, M., D. Cooley, and A. C. Davison (2012), "Bayesian inference from composite likelihoods with an application to spatial extremes." *Statistica Sinica*, 22, 813–845.

Roche, A. (2016), "Composite Bayesian inference." *arXiv preprint arXiv:1512.07678v3*.

Russell, S. and P. Norvig (2010), *Artificial Intelligence A Modern Approach*. Pearson Education, Inc.

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

Stock, J. H. and M. W. Watson (2001), "Vector autoregressions." *Journal of Economic perspectives*, 15, 101–115.

Stock, J. H. and M. W. Watson (2003), "Forecasting output and inflation: The role of asset prices." *Journal of Economic Literature*, 16, 788–829.

Stock, J. H. and M. W. Watson (2012), "Disentangling the channels of the 2007-2009 recession." Technical report, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (2016), "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in

macroeconomics." In *Handbook of macroeconomics*, volume 2, 415–525, Elsevier.

Uhlig, H. (1994), "On singular wishart and singular multivariate beta distributions." *The Annals of Statistics*, 395–405.

Vandekerkhove, P. (2005), "Consistent and asymptotically normal parameter estimates for hidden markov mixtures of markov models." *Bernoulli*, 11, 103–129.

Varin, C. (2008), "On composite marginal likelihoods." *ASta advances in statistical analysis*, 92, 1–28.

Varin, C., N. Reid, and D. Firth (2011), "An overview of composite likelihood methods." *Statistica Sinica*, 21, 5–24.

Verbeke, G. and G. Molenberghs (2005), *Models for discrete longitudinal data*. New York: Springer.

Wakefield, J. (2013), *Bayesian and frequentist regression methods*. Springer.

Weisstein, E. W. (2003), *CRC concise encyclopedia of mathematics*, 2nd edition. CRC press.

West, M. (2020), "Bayesian forecasting of multivariate time series: scalability, structure uncertainty and decisions." *Annals of the Institute of Statistical Mathematics*, 72, 1–31.

West, M. and J. Harrison (1997), *Bayesian forecasting and dynamic models*, 2nd edition. Springer: New York.

White, H. (1994), *Estimation, inference and specification analysis*. Cambridge University Press.

Whittaker, J. (2008), *Graphical models in applied multivariate statistics*. Wiley Publishing.

Zhao, Z. Y., M. Xie, and M. West (2016), "Dynamic dependence networks: Financial time series forecasting and portfolio decisions." *Applied Stochastic Models in Business and Industry*, 32, 311–332.