Bayesian Mixture Models with Applications in Macroeconomics

Chenghan Hou

A thesis submitted for the degree of
Doctor of Philosophy at
The Australian National University

June 2017

Except where otherwise indicated, this thesis is my own original work.

Chenghan Hou
20 June 2017

I dedicate this dissertation to my mother.

Acknowledgements

I would like to express my deepest gratitude to my principal supervisor Dr. Joshua Chan for his constant patience, guidance and encouragement for the last few years, and also for his valuable time and advice and corrections to this thesis. Without his endless supports, this thesis would not have been possible. It is a great honour to have his supervision during my Ph.D. study.

I would like to extend my sincere appreciation to my supervisory panel members Dr. John Stachurski and Dr. Tue Gorgens for their constructive comments and suggestions to this thesis. In particular, I thank Dr. John Stachurki for giving me the great opportunity to take part in the Quantitative Economics project, where I learned programming techniques for economic modeling. His guidance in learning the rigourous probability theory tremendously benefitted me in understanding many econometric theories often employed in empirical economic studies. I would like to thank Dr. Tue Gorgens for his valuable time in discussing many research questions inside or outside the scope of my research, which broadened my mind and enlightened me. He always encouraged me to think about economic questions from different angles which were very helpful in my Ph.D. study.

In addition, I would like to thank Dr. Thomas Yang for his supports and time to discuss my thesis chapter. I thank Dr. Yong Song for his helpful comments and suggestions to the fourth chapter of this thesis.

I am grateful to my friends and colleagues Minhee Chae, Jamie Cross, Jim Hancock, Anpeng Li, Sanghyeok Li, Zhuo Li, Qingyin Ma, Arm Nakornthab, Michinao Okachi, Aubrey Poon, Sen Xue, Luis Uzeda-Garcia, Guanlong Ren, and Jilu Zhang for their challenges and encouragement. Especially, I thank my partner Jiawei Cai for her patience and seemingly reasonable suggestions to this thesis.

I am deeply indebted to Suya Cheng, Qinxuan Xiang and Ling Liu for their persistent support and care for all these years which made it possible to finish my study.

Abstract

A vast empirical literature has documented the widespread nature of structural instability in many macroeconomic time series. In order to accommodate such a feature, there has been an increasing interest in models that allow time-variation in the parameters. One important issue for modeling this time-variation is to decide which type of time-varying processes is more suitable in applications. For instance, one might want to choose between a model where the parameters are gradually evolving over time or one in which there are a small number of abrupt change-points. The objective of this thesis is to investigate the performance of Bayesian mixture models in modeling such changes in macroeconomic time series.

First, we examine the performance of two basic types of mixture models, a scale mixture of Gaussian models and a finite Gaussian mixture model, in forecasting inflation rates of G7 countries. Since it is well-known that many heavy-tailed distributions can be represented as a scale mixture of Gaussian distributions, we build upon the frequently employed stochastic volatility (SV) models and allow the error terms to have different distributional assumptions, such as the t distribution and double exponential (or Laplace) distribution. The results suggest that allowing for heavy-tailed distributed error terms is as important as allowing stochastic volatility in improving point and density forecast accuracy.

Next, we propose a Gaussian mixture innovation model with time-varying mixture probabilities to detect the in-sample breaks in the relationship between inflation and inflation uncertainty. By allowing the time-variation in the mixture probabilities, we find that the proposed model produces more robust estimates and better in-sample fit. Our empirical study provides strong evidence of the existence of breaks in the relationship between inflation and inflation uncertainty in the last few decades.

Finally, we develop a class of vector autoregressive (VAR) models with infinite hidden Markov structures. We first improve the computational efficiency by developing a new Markov chain Monte Carlo method built upon the precision-based algorithms. We then investigate the performance of these infinite hidden Markov models with various dynamics to predict the US inflation, GDP growth and interest rate. The results show that it is better to model separately the time variation in the conditional

mean coefficients and that in the variance process.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A voluminous literature has highlighted the empirical importance of allowing time-variation in model parameters for modeling financial and macroeconomic variables. Bayesian mixture models provide a flexible framework for modeling such a feature. This thesis contributes to this literature by proposing novel Bayesian mixture models and investigating their properties. It consists of three main chapters (Chapter 2 - Chapter 4). Each chapter studies the performances of various types of mixture models in different macroeconomic applications.

In Chapter 2, we compare the forecasting performance of autoregressive (AR) models with a variety of error distributional assumptions and stochastic volatility specifications in terms of both point and density forecasts. Conventional priors like uniform and exponential priors for the degree of freedom parameter have been pointed out to be inappropriate for estimating the t-distributed error regression model. Therefore, we compare the forecast performance of models with t-distributed errors under different prior assumptions. In the context of inflation forecast, we provide empirical evidence to support the claim that some of the main findings in the US inflation forecast literature seem to be unique but not common to the other G7 countries. Two main conclusions can be drawn from our forecasting results. First, the prior of the degree of freedom parameter has little influence on the forecasting performance for models with t-distribution errors. Second, the heavy-tailed distributed error assumption and stochastic volatility component more often improve rather than harm the forecast accuracy in both point forecast and density forecast.

Chapter 3 investigates whether the relationship between inflation and inflation uncertainty has changed and whether the change in this relationship has been gradual or abrupt. We extend the time-varying parameter with stochastic volatility in mean model (TVP-SVM) to include a mixture innovation disturbance in the time-varying parameter process. The proposed model produces more reliable estimates and al-

lows us to investigate the occurrence of breaks in the gradually evolving process of the time-varying coefficients. Using data from the US, Germany, Canada and New Zealand, we find that: 1) the relationship between inflation and inflation uncertainty substantially varies over time over the last few decades; 2) there is strong support for the existence of abrupt changes in the inflation-inflation uncertainty relationship; 3) our empirical results for Canada and New Zealand show that the correlation between inflation and inflation uncertainty has been much weaker since the early 1990s, which coincides with the timing of the implementation of inflation targeting.

Chapter 4 develops a class of vector autoregressive models with infinite hidden Markov structures. This is motivated by the recent empirical success of hierarchical Dirichlet process mixture models in financial and macroeconomic applications. We first develop a new Markov chain Monte Carlo (MCMC) method built upon the precision-based algorithms to improve computational efficiency. We then investigate the forecasting performance of these infinite hidden Markov switching models. Our forecasting results suggest that 1) models with separate infinite hidden Markov processes for the VAR coefficients and volatilities in general forecast better than other specifications of infinite hidden Markov switching models; 2) using a single infinite hidden Markov process to govern all model parameters tends to forecast poorly; 3) most of the gains in forecasting GDP inflation and GDP growth seem to come from allowing for time-variation in volatilities rather than conditional mean coefficients. In contrast, allowing time-variation in all model parameters is important in forecasting the short-term interest rate.

Chapter 2

Stochastic Volatility Models with Alternative Heavy-tailed Distributions

2.1 Introduction

Many US macroeconomic series have been less volatile since the early 1980s. This phenomenon is first termed as the Great Moderation by Stock and Watson [2003]. Some views about the causes of this reduction in volatility include stabilized economic structure, improvement in monetary policy and small shock. Despite the fact that studies on the causes of the Great Moderation have not come to a consensus over the last few decades, most recent empirical studies turn out to be in line with a critical finding: allowing for time-varying volatility plays an essential role in macroeconomic forecast [Stock and Watson, 2007; Clark and Doh, 2011; Chan, 2013; Clark and Ravazzolo, 2014] and policy making [Cogley and Sargent, 2005; Primiceri, 2005; Eisenstat and Strachan, 2015]. Recently, Clark [2009] finds a significant increase in the volatility in the late 2000s. This finding also highlights the empirical importance of allowing for time-varying volatility in the recent macroeconomic study.

Two popular families of models are often used to capture the time-variation in volatility: the autoregressive conditional heteroskedasticity models (GARCH) [Engle, 1982; Bollerslev, 1986] and the stochastic volatility models (SV) [Taylor, 1994; Jacquier et al., 2002; Kim et al., 1998]. With the prodigious advance in computational capacity, the simulation-based econometric inference has been broadly adopted as an alternative approach to the traditional likelihood-based inference. In the Bayesian econometric framework, estimation of SV models based on Markov chain Monte Carlo (MCMC) methods are computationally tractable. This led to extension and application based on the SV models have emerged in recent macroeconomic studies [Clark and Doh,

2011; Clark and Ravazzolo, 2014; Chan, 2015b; Chan and Eisenstat, 2015].

Several empirical studies have shown that the fluctuation of the US inflation is more likely to be explained by the changes of the variance rather than the conditional mean coefficients [Primiceri, 2005]. In addition, the finding that allowing for time-varying volatility significantly improves the forecast performance of a model has been repeatedly verified in many recent macroeconomic studies [Clark, 2011; Clark and Ravazzolo, 2014; Chan, 2013, 2015b]. However this finding is based on the US data, there are few studies focus on the forecast performance of the SV models for other countries' data. One contribution of chapter is to fill up this gap. This chapter investigates the inflation forecast performance of various specifications of stochastic volatility model using data from G7 countries. Our empirical results suggest that allowing for time-varying volatility does not always guarantee an improvement in forecast accuracy.

One prominent feature common in many macroeconomic time series is the presence of outliers. This feature makes the conventional assumption of a Gaussian distributed error inappropriate in many macroeconomic studies. One popular approach is to assume the error term to follow a t-distribution [Chib et al., 2002; Jacquier et al., 2004; Nakajima and Omori, 2012; Chan and Hsiao, 2013]. It has been shown that allowing for a t-distributed error reduces the influence of outliers and produces more robust estimates [Maronna and Yohai, 1998; West, 1984; Lange et al., 1989]. In Bayesian analysis, two priors are often employed in the estimation of the degree of freedom parameter in t-distributed linear regression models: the uniform prior and the exponential prior [Jacquier et al., 2004; Chan and Hsiao, 2013; Clark and Ravazzolo, 2014; Geweke, 1993]. However Fonseca et al. [2008] indicate that using these two conventional priors might lead to unreliable posterior inference. To overcome these issues, they propose a Jeffreys prior for the degree of freedom parameter for the t-distributed linear regression model. In this chapter, we compare the forecast performance of t-distributed error models with alternative prior specifications for the degree of freedom parameter. It is evident from our results that the prior specification for the degree of freedom parameter has little influence on the forecast performance.

Another contribution of this chapter is to investigate the forecast performance of a model under various error distributional assumptions. In the current literature, the feature of the heavy-tailedness is often captured by assuming a t-distributed error term [Clark and Ravazzolo, 2014]. One reason for this is perhaps that the t-

distribution has a few worthy properties such as being symmetric, bell-shaped and most importantly it is a limiting case of a Gaussian distribution. As t-distribution is only one member in the family of heavy-tailed distributions, it is important to investigate the performance of other types of heavy-tailed distribution in macroeconomic forecast. However, current literature has paid little attention to this issue.

In this chapter, we investigate G7 inflation forecast performance of autoregressive models with various error structures such as Gaussian distribution, t-distribution, double exponential distribution and mixture of Gaussian distributions. Our empirical study enables us to draw a more general conclusion: a model involving a stochastic volatility component and heavy-tailed error does not always guarantee an improvement in forecast accuracy, however including these two components seems to more often improve rather than harm the forecast accuracy.

The rest of the chapter is organized as follows. Section 2.2 discusses a few properties of various error distributional assumptions considered in this chapter. Section 2.3 presents all models in our recursive out-of-sample forecasting exercise. Section 2.4 discusses the data and the priors. Section 2.5 briefly discusses the algorithms for model estimation. Section 2.6 presents forecast metrics and results. Section 2.7 concludes.

2.2 Error Distributions and Motivations

One of the main objectives of this chapter is to investigate the forecast performance of a model with different error distributional assumptions. We consider four families of distributions for the error term: 1) the Gaussian distribution, 2) the student-t distribution, 3) the double exponential distribution and 4) the mixture of Gaussian distributions. To illustrate the features of these four classes of distribution, we plot one probability density function from each of these classes, as shown in Figure 2.1. To facilitate comparison, the density functions plotted are standardized such that they all have the identical means 0 and variances 2. The density function of the mixture normal plotted in Figure 2.1 is an equally weighted of a standard Gaussian distribution and a Gaussian distribution with mean 0 and variance 3.

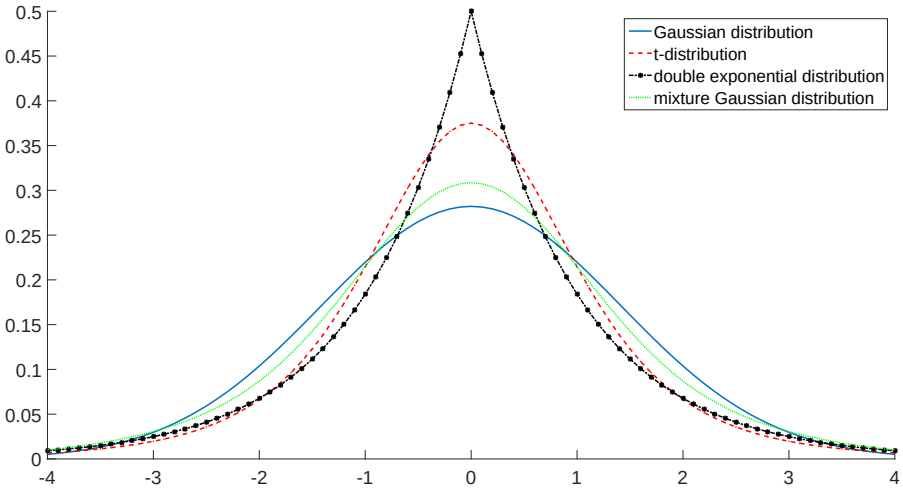

Figure 2.1: Error distributions (with mean 0 and variance 2).

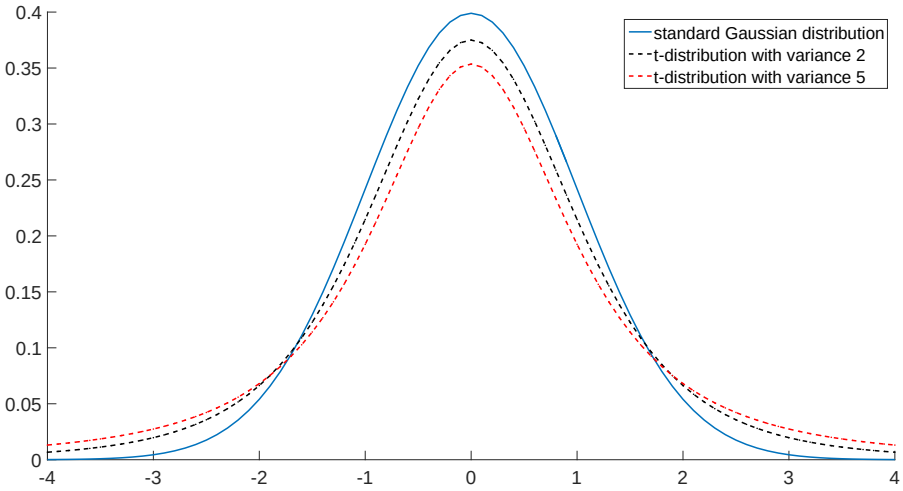

Figure 2.2: *t*-distributions VS. Gaussian distribution.

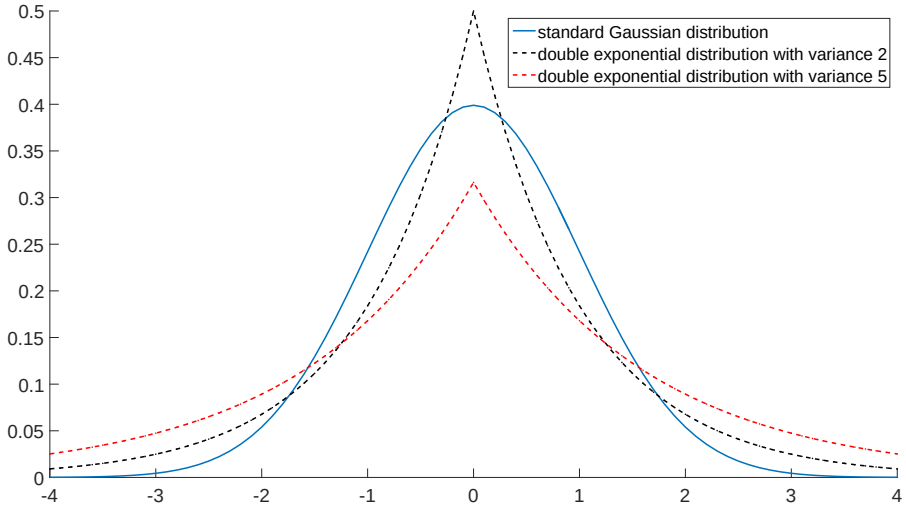

Figure 2.3: Double exponential distribution VS. Gaussian distribution.

As shown in Figure 2.1 - Figure 2.3, greater probability density is concentrated at the tail of the t-distributions compared with the Gaussian distributions. Similar heavy-tailed features are also shown in the double exponential probability density functions. To be more specific, the density functions from different distribution families have tails with different decay rates. The tail of the Gaussian density functions has decay rate that is proportional to $e^{-c_1 x^2}$ for some $c_1 > 0$. In contrast, the decay rates of the tails for t-distributions and double exponential distributions are proportional to x^{-c_2} and $e^{-c_3 x}$ respectively, for some $c_2 > 0$ and $c_3 > 0$. This implies that the Gaussian density function approaches to zero at the highest rate compared with the density function of double exponential and t-distribution as x approaches to infinity. As the tail of a t-distribution decays polynomially and the tail of a double exponential distribution decays exponentially, the t-distribution has the slowest tail decaying rate which results in the heaviest tailed density among the three families of distributions.

Another observation is that the density functions of Gaussian distribution and t-distribution are bell shaped. Instead, the density function of double exponential distribution is not smooth but kinked at its mean. In addition, the density function of double exponential distribution has a constant exponential decay rate not only at the tailed part of the distribution, but over the whole support. This leads to a higher central tendency within a small region around its mean comparing to both Gaussian distribution and t-distribution. It is well known that the t-distribution converges to the Gaussian distribution as the degree of freedom parameter approaches to infinity.

As such, the Gaussian distribution is a limiting case of the t-distribution, which is illustrated in Figure 2.2. Similar graph for the double exponential distributions is plotted in Figure 2.3 for a comparison with standard Gaussian distribution.

Due to the various properties described above, modeling the error term by using different distributional assumptions allows a model to accommodate the data in a different way which may effect on the forecast performance of the model. In this chapter, we conduct a recursive out-of-sample forecast exercise to provide empirical evidence for justifying this question. Before that, we first discuss the models and data we used in the following sections.

2.3 Models

In this section, we present the models considered in this chapter. To be specific, we consider a class of autoregressive (AR) models with different error distributions. We follow [Clark and Ravazzolo, 2014] by setting the orders of the autoregressive process as 4, i.e. AR(4). In the rest of this section, we focus on discussion about various error distributional assumptions.

2.3.1 Models with Constant Variance

First we consider the standard AR(4) model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + \epsilon_t^y, \quad \epsilon_t^y \sim \mathcal{N}(0, \sigma^2),$$

where $\mathcal{N}(\cdot, \cdot)$ denotes the Gaussian distribution.

Next, we extend the standard AR(4) model by allowing a t-distributed error term. In particular, we write the t-distribution as a scale mixture of Gaussian distributions[Geweke, 1993],

$$\begin{aligned} &y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + \epsilon_t^y, \\ &\epsilon_t^y | \lambda_t \sim \mathcal{N}(0, \lambda_t \sigma^2), \quad \lambda_t \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \end{aligned}$$

where $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse gamma distribution. The scale-mixing variable λ_t are assumed to be independent with each other. It can be shown that ϵ_t^y follows a t-distribution with scale parameter σ^2 and degree of freedom ν when λ_t is marginalized out.

Since ν is treated as a parameter to be estimated, we denote the prior distribution for ν as $p(\nu)$. As Fonseca et al. [2008] point out that the prior distribution for ν may lead to improper posterior analysis. To investigate the influence of the prior distributions of ν on the forecast performance for a model with t-distributed error term, we consider three different priors of ν: 1) the exponential prior, 2) the uniform prior and 3) the Jeffreys prior, where the Jeffreys prior is developed in Fonseca et al. [2008]. We proceed our discussion on other error structures and postpone the detailed discussion of different prior specification in Section 2.4.2.

The third distribution we consider is the double exponential distribution which has the density function

$$f(\epsilon_t^y|\sigma^2) = \frac{1}{2\sigma} e^{-\frac{|\epsilon_t^y|}{\sigma}}.$$

It has a latent variable representation:

$$\epsilon_t^y \sim \mathcal{N}(0, \lambda_t \sigma^2), \quad \lambda_t \sim \mathcal{E}(\frac{1}{2}),$$

where $\mathcal{E}(\theta)$ denotes the exponential distribution with mean θ and variance θ^2.

Lastly, we consider a mixture of two Gaussian distributions. To avoid the identification issue, we fix one component of this mixture distribution as the standard Gaussian distribution,

$$\epsilon_t^y \sim p\mathcal{N}(0,1) + (1-p)\mathcal{N}(0,\sigma^2),$$

where $0 \leq p \leq 1$ is the mixture probability. This implies that the error term ϵ_t^y follows $\mathcal{N}(0,1)$ with probability p and $\mathcal{N}(0,\sigma^2)$ with probability $1-p$. To facilitate estimation, we rewrite this mixture Gaussian distribution as

$$\begin{aligned} \epsilon_t^y|\tau_t &= \tau_t \mathcal{N}(0,1) + (1-\tau_t)\mathcal{N}(0,\sigma^2), \\ \Pr(\tau_t = 1) &= p, \quad \Pr(\tau_t = 0) = 1-p, \end{aligned}$$

which means that the mixture component indicator τ_t follows a Bernoulli distribution with parameter p. It can be easily shown that the presence of the τ_t does not affect the joint posterior distribution of other parameters of interest. More details for the estimation are referred to the Appendix.

2.3.2 Models with Stochastic Volatility

A voluminous literature has highlighted the importance of allowing a time-varying volatility in modeling and forecasting macroeconomic variables [Cogley and Sargent, 2005; Stock and Watson, 2003; Chan, 2013; Clark and Ravazzolo, 2014]. In the context of SV models, the time-variation in the volatility is often incorporated into a model through modeling the log-volatilities. More specifically, we assume $\epsilon_t^y \sim \mathcal{N}(0, e^{h_t})$. Two popular SV specifications are often used to model the evolving process of the log-volatility h_t. One approach is to assume h_t to evolve according to a random walk process

$$h_t = h_{t-1} + \epsilon_t^h, \quad \epsilon_t^h \sim \mathcal{N}(0, \sigma_h^2),$$

in which the states are initialized with $h_1 \sim \mathcal{N}(0, V_h)$. Another alternative is to assume the log-volatility h_t to follow a stationary AR(1) process,

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \epsilon_t^h, \quad \epsilon_t^h \sim \mathcal{N}(0, \sigma_h^2),$$

with $|\phi_h| < 1$, and the states are initialized with the stationary distribution $h_1 \sim \mathcal{N}(\mu_h, \sigma_h^2/(1-\phi_h^2))$.

The constant volatility models discussed in Section 2.3.1 can be extended by using these two SV specifications. In particular, we consider a general framework for the error term

$$\epsilon_y^2 \sim \mathcal{N}(0, \eta_t e^{h_t}).$$

Modeling the suitable process of the scale-mixing variable η_t allows us to generalize the conventional stochastic volatility models, such as:

1. model with stochastic volatility: $\eta_t = 1$.

2. model with stochastic volatility and t-distributed error: $\eta_t \sim \mathcal{IG}(\nu/2, \nu/2)$.

3. model with stochastic volatility and double exponential distributed error: $\eta_t \sim \mathcal{E}(1/2)$.

4. model with stochastic volatility and mixture Gaussian distributed error: $\eta_t|\tau_t = \tau_t + (1-\tau_t)\sigma^2$ that is equivalent to

$$\tau_t \mathcal{N}(0, e^{h_t}) + (1-\tau_t)\mathcal{N}(0, \sigma^2 e^{h_t})$$

We combine two SV specifications with various error distributional assumptions and denote the random walk stochastic volatility as RWSV and the stationary AR(1) stochastic volatility as ARSV. We also consider the t-distributed error models with various priors for the degree of freedom parameter. We summarize all model specifications in Table 2.1.

Table 2.1: A list of models.

Model	Description
AR	Gaussian distributed error
AR-t-EXP	t-distributed error with exponential prior
AR-t-U	t-distributed error with uniform prior
AR-t-J	t-distributed error with Jeffreys prior
AR-de	double exponential distributed error
AR-mixn	mixture Gaussian distributed error
AR-RWSV	random walk stochastic volatility
AR-t-RWSV-EXP	RWSV with t-distributed error with exponential prior
AR-t-RWSV-U	RWSV with t-distributed error with uniform prior
AR-t-RWSV-J	RWSV with t-distributed error with Jeffreys prior
AR-de-RWSV	RWSV with double exponential distributed error
AR-mixn-RWSV	RWSV with mixture Gaussian distributed error
AR-ARSV	stationary AR(1) stochastic volatility
AR-t-ARSV-EXP	ARSV with t-distributed error with exponential prior
AR-t-ARSV-U	ARSV with t-distributed error with uniform prior
AR-t-ARSV-J	ARSV with t-distributed error with Jeffreys prior
AR-de-ARSV	ARSV with double exponential distributed error
AR-mixn-ARSV	ARSV with mixture Gaussian distributed error

2.4 Data and Priors

2.4.1 Data

The data we use are the quarterly inflation rates of the G7 countries over the period 1955Q1-2014Q1. More specifically, given the quarterly price index at time t, PI_t, the inflation rate at time t is computed as $400\log(PI_t/PI_{t-1})$. The price indices (CPI) for the UK, Germany, Italy and France are obtained from the OECD statistics database. The price indices for Canada are obtained from the Statistics Canada. For the price indices of Japan, we first obtain the monthly consumer price index from Statistics Bureau of Japan and then compute the quarterly price indices by averaging the monthly indices in each quarter. The US price indices are obtained from the Federal Reserve Bank of Philadelphia's Real Time Dataset for Macroeconomists.

Figure 2.4 - Figure 2.10 plots the inflation rates for the G7 countries from 1955Q1

- 2014Q1. There are apparent drops in volatility of the inflation in the late 1980s for the US, the UK, Japan, France and Italy. However, it is hard to detect a similar reduction for Canada and Germany. In contrast, the inflation for Canada and Germany has been fluctuating more steadily within a certain range − around -5% to 15% for Canada and -4% to 10% for Germany − for the last four decades. It seems that the volatility of the inflation rates for Canada and Germany have not substantially changed over time. For forecasting Canada and Germany inflation rates, including a stochastic volatility component and heavy-tailed error term might not be useful for improving forecast accuracy. Our results presented in Section 2.6.2 provide evidence to support this observation. One of our main findings in this chapter is that including a stochastic volatility component and heavy-tailed error term into a model provides little improvement in the forecast accuracy for low volatility data series, but significantly improve the forecast accuracy for high volatility data series.

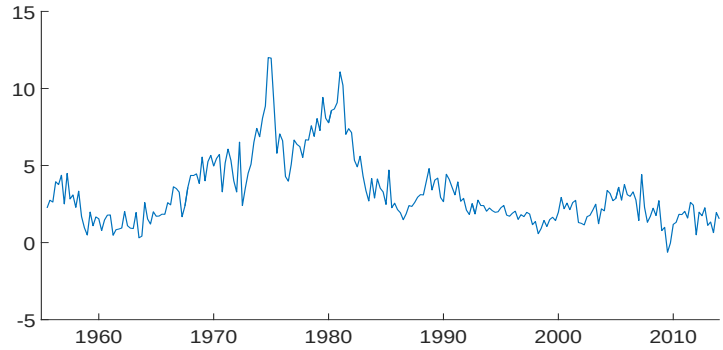

Figure 2.4: The US quarterly inflation from 1955Q1 to 2014Q1.

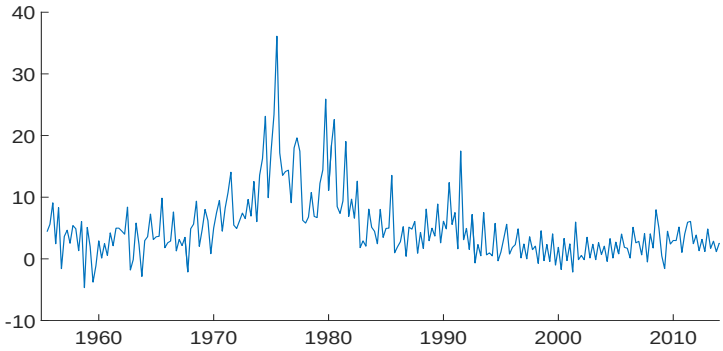

Figure 2.5: The UK quarterly inflation from 1955Q1 to 2014Q1.

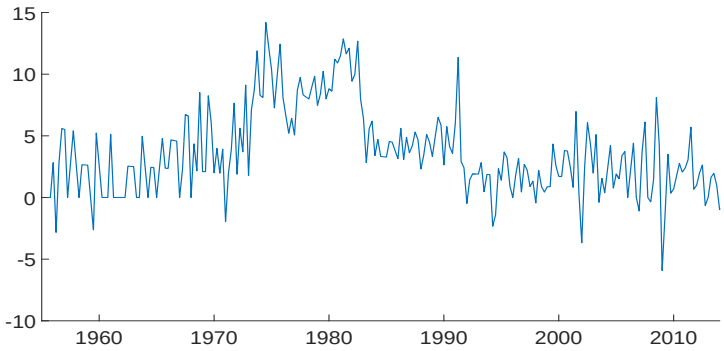

Figure 2.6: Canada quarterly inflation from 1955Q1 to 2014Q1.

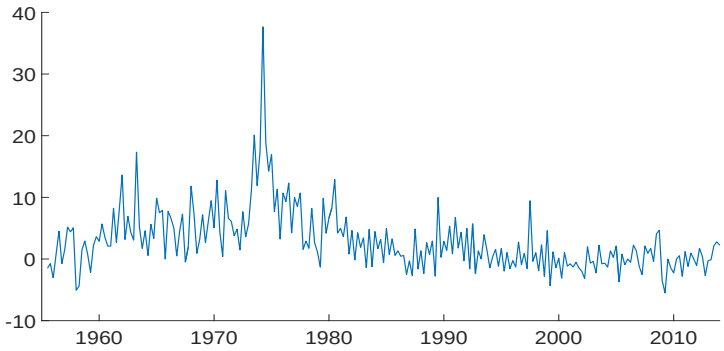

Figure 2.7: Japan quarterly inflation from 1955Q1 to 2014Q1.

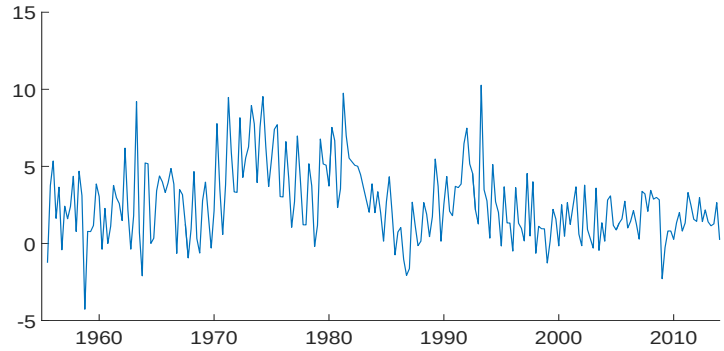

Figure 2.8: Germany quarterly inflation from 1955Q1 to 2014Q1.

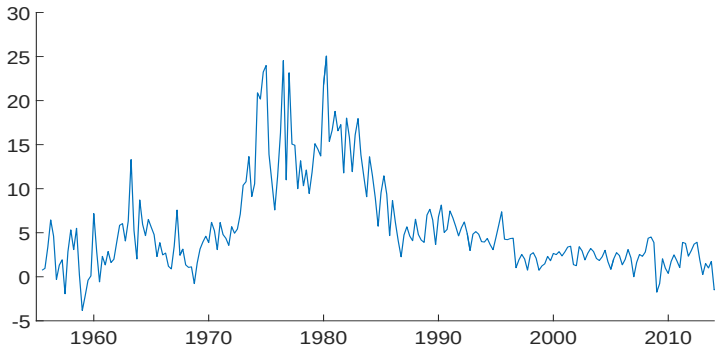

Figure 2.9: Italy quarterly inflation from 1955Q1 to 2014Q1.

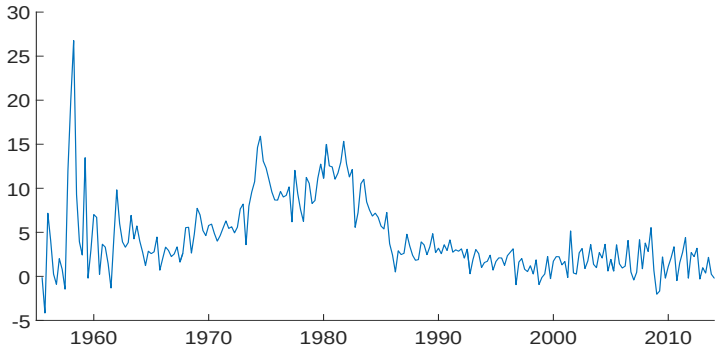

Figure 2.10: France quarterly inflation from 1955Q1 to 2014Q1.

2.4.2 Priors

We use the same priors for the common model parameters and also assume independent prior distributions for all model parameters. In particular, for the models with constant volatility, the priors for the AR coefficients and the variance are

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}^0, V_\beta)\mathbb{1}(\boldsymbol{\beta} \in S), \tag{2.1}$$

$$\sigma_y^2 \sim \mathcal{IG}(\nu_y, S_y), \tag{2.2}$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_4)'$ and $\mathbb{1}(\boldsymbol{\beta} \in S)$ denotes the indicator function such that $\mathbb{1}(\boldsymbol{\beta} \in S) = 1$ if $\boldsymbol{\beta} \in S$ and 0 otherwise. S is the stationarity region where all roots of the characteristic equation associated with β lie outside the unit circle. We set $\nu_y = 10$ and $S_y = 9$, which implies that the prior mean $\mathbb{E}(\sigma^2) = 1$. For the truncated Gaussian prior, we set $\boldsymbol{\beta}^0 = 0$ and $V_\beta = 10$.

To investigate the influence of the prior of the ν on the forecast performance, we

consider three priors for ν: 1) the exponential prior, 2) the uniform prior, and 3) the Jeffreys prior. To be specific, the probability density functions of these three priors are

1. Exponential prior:

$$p_{\mathcal{E}}(\nu) = \frac{1}{8}\exp(-\frac{\nu+2}{8}), \quad \nu > 2. \tag{2.3}$$

2. Uniform prior:

$$p_{\mathcal{U}}(\nu) = \frac{1}{16}, \quad 2 < \nu < 18. \tag{2.4}$$

3. Jeffreys prior:

$$p_{\mathcal{J}}(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left(\psi'(\frac{\nu}{2}) - \psi'(\frac{\nu+1}{2}) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right)^{\frac{1}{2}} \mathbb{1}(\nu > 2), \tag{2.5}$$

where $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ and $\psi'(x) = \frac{d}{dx}\psi(x)$ are the digamma and trigamma functions. The restriction $\nu > 2$ guarantees the existence of first and second moments of the t-distribution. It is easy to check that under the uniform prior and exponential prior, we have $\mathbb{E}(\nu) = 10$.

Before we move on, we first have a brief discussion about these three priors for the degree of freedom parameter ν. The uniform prior and exponential priors are often used in many financial and macroeconomic studies [Geweke, 1993; Jacquier et al., 2004; Chan and Hsiao, 2013; Clark and Ravazzolo, 2014]. However Fonseca et al. [2008] have shown that these two priors might result in unreliable inference. They point out that the uniform prior is inappropriate since the estimate of ν is highly sensitive to where the uniform prior is truncated. For the exponential prior, they also provide empirical evidence that the posterior estimate of ν is prior dependent. To overcome these issues, Fonseca et al. [2008] derive the Jeffreys prior for ν. In general, Jeffreys prior is defined such that it is proportional to the square root of the determinant of the Fisher information matrix. It is well known that Jeffreys prior possesses two nice properties: 1) it is invariant under reparameterization, 2) it is noninformative. More details about the Jeffreys prior can be found in Jeffreys [1946]; Kass and Wasserman [1996].

For models with a mixture of Gaussian distributed error, we assume

$$p \sim \mathcal{B}(p_1, p_2),$$

where $\mathcal{B}$ denotes beta distribution. We set $p_1 = 5.75$ and $p_2 = 5.75$. This implies $\mathbb{E}(p) = 0.5$. For models with the stationary AR(1) stochastic volatility, we use exactly the same priors for the common model parameters and assume that

$$\phi_h \sim \mathcal{N}(\phi_{h_0}, V_{\phi_h})\mathbb{1}(|\phi_h| < 1), \quad \mu_h \sim \mathcal{N}(\mu_{h_0}, V_{\mu_h}), \quad \sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h).$$

We let $\nu_h = 10$, $S_h = 0.09$ which implies $\mathbb{E}(\sigma_h^2) = 0.01$. For the prior of the intercept, we set $\mu_{h_0} = 0$ and $V_{\mu_h} = 0.25$. For the AR(1) coefficient, we let $\phi_{h_0} = 0.8$ and $V_{\phi_h} = 0.04$. The random walk SV specification is a special case of the stationary AR(1) stochastic volatility with $\phi_h = 1$ and $\mu_h = 0$. For the initial state of the random walk SV, we let $h_1 \sim \mathcal{N}(0, V_h)$ and set $V_h = 10$. This setting is comparable to those used in previous studies, such as Clark and Ravazzolo [2014]; Cogley and Sargent [2005].

2.5 Algorithms

All models are estimated using Markov chain Monte Carlo (MCMC) methods. Models with a stochastic volatility component are estimated using the approach of auxiliary mixture sampler of Kim et al. [1998]. Most algorithms of models are standard and we refer the readers to the discussion in Clark and Ravazzolo [2014]; Chan and Hsiao [2013]; Koop et al. [2007]. The MCMC algorithms for the AR-t-ARSV-J model and AR-mixn-ARSV model are provided in the Appendix. It is easy to verify that most models considered in this chapter are a special case of AR-t-ARSV-J. For example, the AR models with constant volatility are equivalent to the AR-t-ARSV-J model if we set $h_t = 0$ for all t. Models without a heavy-tailed error is a special case of the AR-t-ARSV-J model if we set the scale-mixing variable $\lambda_t = 1$ for all t. Finally, to estimate the t-distributed error models with the degree of freedom ν following the uniform prior or exponential priors, it can easily be achieved by replacing the corresponding block in the Gibbs sampler. We refer the reader to Chan and Hsiao [2013]; Geweke [1993] for more details.

2.6 Forecast Metrics and Results

This section first discusses the forecast metrics and then presents the results of the recursive out-of-sample forecasting exercise.

2.6.1 Forecast Metrics

We use $\mathbf{y_{1:t}}$ to denote the data up to time t. The posterior mean $\mathbb{E}(y_{t+k}|\mathbf{y_{1:t}})$ is used as the k-step-ahead point forecast. To compute an estimate for $\mathbb{E}(y_{t+k}|\mathbf{y_{1:t}})$, we first use the posterior draw obtained from each MCMC iteration after a burn-in period to simulate the future states from $t+1, \cdots, t+k$ using the relevant transition equations. Next, we exploit the autoregressive structure of the model to iterate forward along with the simulated error until a draw y_{t+k} is obtained. We then compute the average over all posterior draws to produce an estimate for $\mathbb{E}(y_{t+k}|\mathbf{y_{1:t}})$.

Let y_t^o be the observed value of y_t, the metric used to evaluate the point forecasts is the root mean squared forecast error(RMSFE) which is defined as

$$\text{RMSFE} = \sqrt{\frac{\sum_{t=t_0}^{T-k} \left(y_{t+k}^o - \mathbb{E}(y_{t+k}|\mathbf{y_{1:t}})\right)^2}{T-k-t_0+1}},$$

We use two metrics, the average log-predictive likelihoods and the average continuous ranked probability score, to evaluate the density forecast accuracy. The average log-predictive likelihoods (ALPL) is defined as

$$\text{ALPL} = \frac{1}{T-k-t_0+1} \sum_{t=t_0}^{T-k} \log p_{t+k}(y_{t+k} = y_{t+k}^o|\mathbf{y_{1:t}}),$$

where p_{t+k} denotes the k-step-ahead forecast density. There is a close connection between the predictive likelihood and the marginal likelihood. More detailed discussion of the log-predictive likelihood is provided in Geweke and Amisano [2010].

Another metric used to evaluate the density forecast is the average continuous rank probability score (ACRPS),

$$\text{ACRPS} = \frac{1}{T-k-t_0+1} \sum_{t=t_0}^{T-k} \text{CRPS}_t, \tag{2.6}$$

where $\text{CRPS}_t = \int_{-\infty}^{\infty} \left(F_{t+k}(z) - \mathbb{1}(y_{t+k}^o < z)\right)^2 dz = \mathbb{E}_{p_{t+k}}|y_{t+k} - y_{t+k}^o| - 0.5\mathbb{E}_{p_{t+h}}|y_{t+h} - y'_{t+h}|$. F_{t+k} is the cumulative distribution corresponding to the predictive density p_{t+k}. The $\mathbb{E}_f(x)$ denotes the expected value of x with respect to the density f. y_{t+k} and y'_{t+k} are two independent draws from the predictive distribution p_{t+k}. Some authors consider the ACRPS as a better scoring rule for measuring the accuracy of density forecast than the ALPL. The main advantage is that it is defined directly in terms of the predictive cumulative distribution which provides a consistent measure to com-

pare between deterministic forecast and probability forecast and it is also found to be less sensitive to outliers [Gneiting and Ranjan, 2012].

2.6.2 Forecasting Results

This section conducts a recursive out-of-sample forecasting exercise to compare the G7 inflation forecast performance of those models listed in Table 2.1. We report the forecasting results at various forecast horizons $k = 1$ (one quarter), $k = 2$ (two quarters), $k = 4$ (one year) and $k = 8$ (two years). Our forecast evaluation period is from 1975Q1 - 2014Q1. All forecast estimates are computed using 50000 posterior draws retained after a burn-in period of 20000. For easy comparison, we report the relative scores to the AR-RWSV benchmark. For the AR-RWSV, we report the actual forecast estimates. The relative RMSFEs of a given model are the ratios of RMSFEs to those of AR-RWSV. Hence values less than 1 indicate better forecast performance compared with the AR-RWSV. For the density forecasts, we report the difference of ALPLs of a given model to those of AR-RWSV. Thus positive values indicate better forecast performance. The relative scores for the ACRPSs are the ratio of ACRPSs of a given model to those of AR-RWSV, then the values less than 1 indicate better forecast performance.

Table 2.2 - Table 2.8 present the forecasting results for each G7 country. In each table, panel a) presents the point forecast results, i.e. the RMSFEs and panel b) and c) present the density forecast results, i.e. the ALPL and ACRPS respectively. To facilitate comparison, the results are grouped into three panels: the upper panel contains results for models with constant volatility, the middle panel contains results for models with random walk stochastic volatility and the lowest panel contains results for models with stationary AR(1) stochastic volatility.

Table 2.2: Forecasting results for the US.

	a) RMSFEs				b) ALPLs				c) ACRPSs			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
AR-RWSV	**0.92**	**1.11**	**1.33**	**1.69**	**-1.35**	**-1.45**	**-1.60**	**-1.80**	**0.50**	**0.59**	**0.72**	**0.93**
AR	1.02	1.03	1.03	1.04	-0.19	-0.21	-0.21	-0.21	1.04	1.07	1.08	1.09
AR-t-EXP	1.01	1.03	1.03	1.05	-0.13	-0.18	-0.20	-0.23	1.03	1.06	1.07	1.10
AR-t-U	1.01	1.03	1.03	1.05	-0.14	-0.18	-0.19	-0.22	1.03	1.06	1.07	1.10
AR-t-J	1.01	1.03	1.04	1.05	-0.13	-0.18	-0.20	-0.23	1.03	1.06	1.08	1.10
AR-de	1.02	1.03	1.04	1.05	-0.10	-0.17	-0.22	-0.27	1.03	1.06	1.08	1.10
AR-mixn	1.02	1.03	1.03	1.04	-0.17	-0.20	-0.20	-0.21	1.04	1.06	1.08	1.09
AR-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-t-RWSV-EXP	1.00	1.00	1.01	1.01	0.01	0.00	-0.01	-0.02	1.00	1.01	1.01	1.01
AR-t-RWSV-U	1.00	1.00	1.00	1.01	0.01	0.00	-0.01	-0.02	1.00	1.00	1.01	1.01
AR-t-RWSV-J	1.00	1.00	1.01	1.01	0.01	0.00	-0.01	-0.02	1.00	1.00	1.01	1.01
AR-de-RWSV	1.01	1.02	1.02	1.03	0.01	-0.03	-0.07	-0.10	1.01	1.02	1.03	1.04
AR-mixn-RWSV	1.01	1.03	1.02	1.03	-0.02	-0.04	-0.05	-0.06	1.01	1.02	1.03	1.04
AR-ARSV	1.00	1.01	1.01	1.01	-0.01	-0.02	-0.02	-0.03	1.00	1.01	1.01	1.02
AR-t-ARSV-EXP	1.01	1.02	1.02	1.03	-0.01	-0.03	-0.05	-0.06	1.01	1.02	1.03	1.03
AR-t-ARSV-U	1.01	1.02	1.02	1.02	-0.01	-0.03	-0.05	-0.06	1.00	1.02	1.02	1.03
AR-t-ARSV-J	1.01	1.02	1.02	1.03	-0.01	-0.03	-0.04	-0.06	1.00	1.02	1.02	1.03
AR-de-ARSV	1.01	1.03	1.03	1.04	-0.03	-0.08	-0.12	-0.16	1.02	1.04	1.05	1.06
AR-mixn-ARSV	1.00	1.01	1.01	1.02	-0.01	-0.02	-0.03	-0.04	1.00	1.01	1.02	1.02

Table 2.3: Forecasting results the UK.

	a) RMSFEs				b) ALPLs				c) ACRPSs			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
AR-RWSV	**3.72**	**3.93**	**3.69**	**4.01**	**-2.54**	**-2.57**	**-2.56**	**-2.68**	**1.85**	**1.92**	**1.84**	**2.11**
AR	1.03	1.04	1.05	1.08	-0.34	-0.36	-0.34	-0.33	1.11	1.12	1.13	1.15
AR-t-EXP	1.02	1.02	0.98	1.00	-0.15	-0.16	-0.12	-0.13	1.06	1.07	1.03	1.02
AR-t-U	1.02	1.02	0.98	1.00	-0.15	-0.16	-0.12	-0.13	1.06	1.07	1.03	1.02
AR-t-J	1.02	1.02	0.98	1.00	-0.14	-0.16	-0.12	-0.13	1.06	1.06	1.02	1.02
AR-de	1.02	1.02	0.99	1.01	-0.14	-0.16	-0.13	-0.14	1.06	1.07	1.03	1.03
AR-mixn	1.03	1.03	1.02	1.03	-0.26	-0.28	-0.25	-0.25	1.07	1.08	1.07	1.07
AR-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-t-RWSV-EXP	1.00	1.00	0.98	0.98	0.01	0.01	0.02	0.01	1.01	1.00	0.99	0.98
AR-t-RWSV-U	1.00	1.00	0.98	0.98	0.01	0.01	0.01	0.01	1.00	1.00	0.99	0.99
AR-t-RWSV-J	1.01	1.00	0.98	0.98	0.01	0.01	0.01	0.02	1.01	1.00	0.99	0.98
AR-de-RWSV	1.02	1.00	0.97	0.98	0.01	0.01	0.02	0.01	1.01	1.01	0.98	0.98
AR-mixn-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-ARSV	1.00	1.00	1.00	1.00	0.02	0.02	0.03	0.03	1.00	1.00	1.00	0.99
AR-t-ARSV-EXP	1.01	1.00	0.98	0.98	0.02	0.02	0.04	0.04	1.01	1.01	0.99	0.97
AR-t-ARSV-U	1.00	1.00	0.98	0.98	0.03	0.03	0.04	0.04	1.00	1.00	0.99	0.98
AR-t-ARSV-J	1.01	1.00	0.98	0.98	0.02	0.02	0.04	0.04	1.01	1.01	0.99	0.97
AR-de-ARSV	1.02	1.01	0.97	0.98	0.02	0.02	0.04	0.04	1.02	1.01	0.98	0.97
AR-mixn-ARSV	1.00	1.00	1.00	1.00	0.02	0.02	0.03	0.03	1.00	1.00	0.99	0.99

Table 2.4: Forecasting results for Canada.

	a) RMSFEs				b) ALPLs				c) ACRPSs			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
AR-RWSV	**2.20**	**2.39**	**2.44**	**2.87**	**-2.31**	**-2.37**	**-2.41**	**-2.56**	**1.25**	**1.36**	**1.40**	**1.70**
AR	1.00	1.00	1.00	1.00	0.00	0.01	0.02	0.04	0.99	1.00	0.99	0.97
AR-t-EXP	1.00	1.00	1.00	1.00	0.02	0.02	0.03	0.04	0.99	0.99	0.99	0.97
AR-t-U	1.00	1.00	1.00	1.00	0.02	0.02	0.04	0.04	0.99	0.99	0.99	0.97
AR-t-J	1.00	1.00	1.00	1.00	0.02	0.02	0.03	0.04	0.99	0.99	0.99	0.97
AR-de	1.01	1.01	1.01	1.02	0.03	0.02	0.02	0.01	1.00	1.00	1.00	1.00
AR-mixn	1.00	1.00	1.01	1.00	0.02	0.02	0.04	0.05	0.99	0.99	0.99	0.98
AR-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-t-RWSV-EXP	1.00	1.00	1.00	1.00	0.02	0.01	0.02	0.01	1.00	1.00	1.00	1.00
AR-t-RWSV-U	1.00	1.00	1.00	1.00	0.02	0.02	0.02	0.01	1.00	1.00	1.00	1.00
AR-t-RWSV-J	1.00	1.00	1.00	1.00	0.01	0.01	0.01	0.01	1.00	1.00	1.00	1.00
AR-de-RWSV	1.01	1.01	1.01	1.02	0.02	0.01	0.01	-0.02	1.01	1.01	1.01	1.02
AR-mixn-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-ARSV	1.00	1.00	1.00	1.00	0.02	0.02	0.03	0.04	0.99	0.99	0.99	0.99
AR-t-ARSV-EXP	1.00	1.00	1.00	1.00	0.03	0.03	0.04	0.04	0.99	0.99	0.99	0.98
AR-t-ARSV-U	1.00	1.00	1.00	1.00	0.03	0.03	0.04	0.04	0.99	0.99	0.99	0.98
AR-t-ARSV-J	1.00	1.00	1.00	1.00	0.03	0.03	0.04	0.04	0.99	0.99	0.99	0.98
AR-de-ARSV	1.01	1.01	1.01	1.02	0.03	0.02	0.02	0.00	1.00	1.00	1.00	1.00
AR-mixn-ARSV	1.00	1.00	1.00	1.00	0.02	0.02	0.02	0.02	1.00	1.00	1.00	1.00

Table 2.5: Forecasting results for Japan.

	a) RMSFEs				b) ALPLs				c) ACRPSs			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
AR-RWSV	**2.88**	**2.69**	**2.63**	**2.98**	**-2.54**	**-2.54**	**-2.56**	**-2.65**	**1.61**	**1.55**	**1.57**	**1.81**
AR	1.10	1.08	1.12	1.19	-0.29	-0.29	-0.30	-0.30	1.14	1.12	1.15	1.21
AR-t-EXP	0.94	0.97	1.00	1.00	-0.01	0.00	-0.02	-0.05	0.96	0.97	0.97	0.97
AR-t-U	0.94	0.97	1.00	1.00	-0.01	0.00	-0.02	-0.05	0.96	0.97	0.98	0.98
AR-t-J	0.94	0.97	1.00	0.99	-0.01	0.00	-0.01	-0.05	0.96	0.96	0.97	0.98
AR-de	0.94	0.97	1.01	1.01	-0.04	-0.03	-0.04	-0.07	0.96	0.97	0.98	0.99
AR-mixn	0.99	1.01	1.04	1.08	-0.18	-0.18	-0.19	-0.21	1.02	1.02	1.04	1.07
AR-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-t-RWSV-EXP	0.95	0.97	0.99	0.96	0.07	0.07	0.07	0.06	0.96	0.96	0.96	0.94
AR-t-RWSV-U	0.95	0.97	0.99	0.97	0.06	0.07	0.06	0.06	0.96	0.97	0.96	0.94
AR-t-RWSV-J	0.95	0.97	0.99	0.96	0.07	0.07	0.07	0.06	0.95	0.96	0.96	0.94
AR-de-RWSV	0.92	0.95	0.98	0.95	0.07	0.08	0.08	0.07	0.94	0.95	0.95	0.93
AR-mixn-RWSV	0.99	1.00	1.00	0.99	0.01	0.01	0.01	0.02	0.99	0.99	0.99	0.98
AR-ARSV	1.00	1.00	1.00	1.00	0.02	0.02	0.03	0.03	0.99	0.99	0.99	0.99
AR-t-ARSV-EXP	0.95	0.97	0.99	0.96	0.09	0.09	0.09	0.09	0.95	0.96	0.95	0.93
AR-t-ARSV-U	0.95	0.97	0.99	0.96	0.08	0.09	0.09	0.09	0.95	0.96	0.95	0.93
AR-t-ARSV-J	0.94	0.97	0.99	0.96	0.09	0.09	0.09	0.09	0.95	0.96	0.95	0.93
AR-de-ARSV	0.92	0.95	0.98	0.95	0.09	0.10	0.10	0.09	0.93	0.95	0.94	0.92
AR-mixn-ARSV	0.99	0.99	0.99	0.98	0.04	0.04	0.04	0.04	0.98	0.98	0.98	0.97

Table 2.6: Forecasting results for Germany.

	a) RMSFEs				b) ALPLs				c) ACRPSs			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
AR-RWSV	**1.72**	**1.78**	**1.79**	**2.09**	**-2.05**	**-2.10**	**-2.11**	**-2.25**	**0.98**	**1.03**	**1.03**	**1.23**
AR	0.99	0.99	0.99	0.99	-0.07	-0.05	-0.04	-0.02	0.99	0.99	0.98	0.98
AR-t-EXP	0.99	0.99	0.99	0.99	-0.03	-0.01	-0.01	-0.01	0.99	0.98	0.98	0.97
AR-t-U	0.99	0.99	0.99	0.99	-0.03	-0.01	-0.01	-0.00	1.00	0.98	0.98	0.97
AR-t-J	0.99	0.99	0.99	0.99	-0.02	-0.01	-0.01	-0.01	0.99	0.98	0.98	0.97
AR-de	1.00	1.00	0.99	0.97	-0.01	0.00	0.01	0.00	1.00	0.99	0.97	0.96
AR-mixn	0.99	0.99	0.99	0.98	-0.04	-0.02	-0.01	0.01	0.99	0.98	0.97	0.96
AR-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-t-RWSV-EXP	1.00	1.00	1.00	0.99	0.01	0.01	0.01	0.01	1.00	1.00	0.99	0.99
AR-t-RWSV-U	1.00	1.00	1.00	0.99	0.02	0.02	0.02	0.01	1.00	1.00	0.99	0.99
AR-t-RWSV-J	1.00	1.00	1.00	0.99	0.01	0.01	0.01	0.01	1.00	1.00	1.00	0.99
AR-de-RWSV	1.01	1.00	0.99	0.98	0.00	0.01	0.02	0.00	1.01	1.01	0.99	0.98
AR-mixn-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-ARSV	1.00	1.00	1.00	1.00	-0.01	0.00	0.00	0.00	1.00	0.99	0.99	0.99
AR-t-ARSV-EXP	0.99	0.99	0.99	0.99	-0.01	0.00	0.00	0.00	0.99	0.99	0.98	0.98
AR-t-ARSV-U	0.99	0.99	0.99	0.99	-0.01	0.00	0.00	0.00	0.99	0.99	0.98	0.98
AR-t-ARSV-J	0.99	0.99	0.99	0.99	-0.01	0.00	0.00	0.00	0.99	0.99	0.98	0.98
AR-de-ARSV	1.00	1.00	0.99	0.97	-0.02	-0.01	0.01	-0.01	1.00	0.99	0.98	0.96
AR-mixn-ARSV	1.00	1.00	1.00	1.00	0.02	0.02	0.02	0.02	1.00	1.00	1.00	0.99

Table 2.7: Forecasting results for Italy.

	a) RMSFEs				b) ALPLs				c) ACRPSs			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
AR-RWSV	**2.80**	**2.99**	**3.25**	**3.39**	**-2.12**	**-2.27**	**-2.32**	**-2.42**	**1.30**	**1.45**	**1.59**	**1.81**
AR	1.01	1.03	1.02	1.00	-0.47	-0.40	-0.42	-0.43	1.14	1.12	1.13	1.16
AR-t-EXP	1.02	1.04	1.06	1.09	-0.25	-0.20	-0.25	-0.33	1.10	1.07	1.10	1.15
AR-t-U	1.02	1.04	1.06	1.09	-0.25	-0.20	-0.25	-0.33	1.10	1.07	1.10	1.15
AR-t-J	1.02	1.05	1.07	1.10	-0.25	-0.19	-0.25	-0.33	1.10	1.07	1.11	1.16
AR-de	1.02	1.04	1.07	1.07	-0.24	-0.19	-0.25	-0.32	1.10	1.06	1.09	1.12
AR-mixn	1.01	1.03	1.03	1.01	-0.36	-0.30	-0.34	-0.37	1.10	1.07	1.09	1.12
AR-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-t-RWSV-EXP	1.00	1.01	1.03	1.04	0.02	0.02	0.00	-0.01	1.00	1.01	1.02	1.02
AR-t-RWSV-U	1.00	1.01	1.02	1.04	0.02	0.02	0.00	-0.01	1.00	1.01	1.01	1.02
AR-t-RWSV-J	1.00	1.01	1.03	1.04	0.02	0.02	0.00	-0.01	1.01	1.01	1.02	1.03
AR-de-RWSV	1.01	1.04	1.09	1.12	0.01	0.02	-0.02	-0.06	1.02	1.03	1.06	1.08
AR-mixn-RWSV	1.00	1.00	1.00	1.01	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-ARSV	1.00	1.00	1.00	1.00	0.01	0.02	0.02	0.02	1.00	1.00	0.99	0.99
AR-t-ARSV-EXP	1.00	1.02	1.04	1.04	0.03	0.03	0.01	0.01	1.01	1.01	1.02	1.02
AR-t-ARSV-U	1.00	1.02	1.03	1.04	0.03	0.03	0.02	0.01	1.01	1.01	1.01	1.01
AR-t-ARSV-J	1.00	1.02	1.04	1.05	0.02	0.03	0.01	0.00	1.01	1.01	1.02	1.03
AR-de-ARSV	1.01	1.04	1.08	1.11	0.01	0.03	-0.01	-0.04	1.03	1.02	1.05	1.07
AR-mixn-ARSV	1.00	1.00	1.01	1.00	0.02	0.02	0.02	0.02	1.00	1.00	0.99	0.99

Table 2.8: Forecasting results for France.

	a) RMSFEs				b) ALPLs				c) ACRPSs			
	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8	k=1	k=2	k=4	k=8
AR-RWSV	**1.71**	**1.86**	**1.96**	**2.45**	**-2.04**	**-2.10**	**-2.15**	**-2.32**	**0.97**	**1.06**	**1.12**	**1.44**
AR	1.08	1.15	1.19	1.24	-0.38	-0.43	-0.43	-0.40	1.15	1.25	1.29	1.30
AR-t-EXP	1.00	1.01	1.00	1.00	-0.02	-0.07	-0.08	-0.13	0.99	1.00	1.00	1.00
AR-t-U	1.00	1.01	1.00	1.00	-0.02	-0.07	-0.08	-0.13	1.00	1.00	1.00	1.00
AR-t-J	1.00	1.00	1.00	1.00	-0.02	-0.07	-0.08	-0.14	0.99	1.00	1.00	1.01
AR-de	1.01	1.02	1.02	1.03	-0.08	-0.14	-0.14	-0.17	1.01	1.04	1.04	1.03
AR-mixn	1.02	1.02	1.02	1.01	-0.17	-0.23	-0.25	-0.26	1.03	1.05	1.07	1.05
AR-RWSV	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
AR-t-RWSV-EXP	1.00	1.00	1.00	1.00	0.04	0.02	0.02	0.00	0.99	1.00	0.99	0.98
AR-t-RWSV-U	1.00	1.00	1.00	1.00	0.04	0.02	0.03	0.01	0.99	1.00	0.99	0.99
AR-t-RWSV-J	1.00	1.01	1.00	0.99	0.04	0.02	0.02	0.00	0.99	1.00	0.99	0.98
AR-de-RWSV	1.00	1.01	1.00	0.99	0.04	0.02	0.03	0.02	0.99	1.00	0.98	0.97
AR-mixn-RWSV	1.00	1.00	1.00	1.00	0.01	0.00	0.01	0.01	1.00	1.00	1.00	1.00
AR-ARSV	1.00	1.00	1.00	1.00	0.01	0.01	0.01	0.01	1.00	1.00	1.00	1.00
AR-t-ARSV-EXP	1.00	1.00	1.00	1.00	0.05	0.02	0.02	0.00	0.99	0.99	0.99	0.98
AR-t-ARSV-U	1.00	1.00	1.00	1.00	0.04	0.02	0.02	0.00	0.99	0.99	0.99	0.98
AR-t-ARSV-J	1.00	1.00	1.00	1.00	0.05	0.02	0.02	-0.01	0.99	1.00	0.99	0.98
AR-de-ARSV	1.00	1.01	1.00	0.99	0.05	0.02	0.03	0.01	0.99	0.99	0.98	0.97
AR-mixn-ARSV	1.00	1.00	1.00	1.00	0.02	0.02	0.02	0.02	1.00	1.00	0.99	0.99

2.6.2.1 Priors for ν

First, we investigate the forecast performance of the t-distributed error models under various prior specifications for the degree of freedom parameter. It is evident from our forecasting results that t-distributed models with various prior specifications for the degree of freedom parameter, either with the stochastic volatility or without the stochastic volatility component, perform similarly for both point and density forecasts. It suggests that the prior for the degree of freedom has little effect on the forecast accuracy. According to this finding, we precede our discussion by simply denoting all models with t-distributed error as AR-t. Similarly, for models with the random walk or stationary AR(1) stochastic volatility, we denote them as AR-t-RWSV and AR-t-ARSV respectively.

2.6.2.2 Point Forecasts

The US inflation forecasting results presented in Table 2.2a) show that none of the AR models consistently outperform the AR-RWSV benchmark. Our results are also in line with the previous studies of US inflation forecast, that models with stochastic volatility perform better than their counterparts with only constant variance. These results are broadly consistent with the findings in [Chan, 2015b; Clark, 2011; Clark and Ravazzolo, 2014]. However, these findings do not seem to hold for the other

G7 countries. For instance, the results for Japan show that both AR-t-RWSV and AR-t-ARSV strictly outperform the AR-RWSV at all forecast horizons. Furthermore, the results for Canada and Germany indicate that models with stochastic volatility do not improve the point forecast performance over their constant volatility counterparts. For example, all models produce almost the same point forecast estimates at all forecast horizons for the Canada inflation forecast.

It is evident that, in general, allowing for a t-distributed error term improves the forecast performance. For example, the results for Japan show that the AR-t model strictly dominates the AR model. In particular, the RMSFEs evaluated at all forecast horizons for the AR-t model are at least 8% lower than the values for the AR model. Among models with a heavy-tailed distributed error term, it is hard to make a conclusion about which specification of the heavy-tailed distribution, t-distribution or double exponential distribution, has a better forecast performance. Yet the model with a double exponential distributed error term performs less stable than the model with a t-distributed error term. For example, the AR-de model performs the worst in forecasting the inflation of Italy, whereas performs the best in forecasting the inflation of Japan.

In general the AR-t-RWSV and AR-t-ARSV models perform well in the G7 inflation forecast. None of the models we considered produce sizable and consistent improvements over the AR-t-RWSV and AR-t-ARSV models. However, it is unclear of which specification of the stochastic volatility, random walk or stationary AR(1), forecasts better. Although there are cases for which AR-mixn, AR-mixn-RWSV and AR-mixn-ARSV perform better than the AR-RWSV, these improvements are more likely to be occasional and unremarkable.

2.6.2.3 Density Forecasts

The density forecast results are presented in the panel b) and c) of Table 2.2 - Table 2.8. The patterns are broadly similar to the point forecast results. Models including the stochastic volatility component substantially improve the performance for forecasting the US inflation. The best model for forecasting the US inflation is the AR-RWSV. This is again consistent with the finding in Clark and Ravazzolo [2014]. Although both AR-t-RWSV and AR-t-ARSV models are slightly less accurate than the AR-RWSV, these losses are relatively small and negligible.

Similar to the point forecast results, adding the stochastic volatility component does

not seem to improve the density forecast accuracy in forecasting the inflation of Germany. It is evident that, in general, both AR-t-RWSV and AR-t-ARSV forecast better than the AR-RWSV. For example, the AR-t-RWSV or AR-t-ARSV perform the best in forecasting the inflation of the UK, Canada, Japan and France whereas for the US, Italy and Germany, the AR-t-RWSV or AR-t-ARSV models perform quite well. Our density forecast results also suggest that none of the models we considered produce notable improvements over the AR-t-RWSV and AR-t-ARSV.

A few conclusions can be drawn from our forecasting results. First, adding stochastic volatility and heavy-tailed distributed error term tend to improve both point and density forecast performances. Second, none of the models considered in this chapter yield significant and consistent improvements over the AR-t-RWSV and AR-t-ARSV. Third allowing for both stochastic volatility and heavy-tailed distributed error term does not always guarantee an improvement in the forecast accuracy, but it is more likely to improve rather than harm the forecast accuracy.

2.6.3 Heavy-tailedness and Stochastic Volatility

It is difficult to conclude which specification of stochastic volatility performs better in forecasting the G7 inflation. In forecasting the US inflation, models with the random walk stochastic volatility outperform their counterparts with the stationary AR(1) stochastic volatility. However, this finding is unique and not common to the other G7 countries. It is evident that many G7 countries seem to be in favor of the stationary AR(1) specification. For example, the density forecast results for Japan show that models with stationary AR(1) stochastic volatility strictly dominate their counterparts with the random walk stochastic volatility.

One possible reason that the t-distributed error model tends to outperform the normally distributed error model is that the degree of freedom parameter is treated as a parameter to be estimated in our out-of-sample forecasting exercise. As outlined in Section 2.2, the degree of freedom parameter has a close connection to both the tailed density of the t-distribution and the Gaussian distribution. In particular, a small value for the degree of freedom parameter implies a heavier tailed t-distribution. In contrast, a large value of the degree of freedom parameter makes the corresponding t-distribution a good approximation to Gaussian distribution. Allowing the degree of freedom parameter to be estimated along with the other model parameters generalize the Gaussian distributed error model to some extent which is more effective in capturing outliers in the data.

The main purpose of allowing a model to include the heavy-tailed error structure and stochastic volatility component is for reducing the influence of the outliers and capturing the time-varying volatility exhibited in data series. One nature question to ask is whether a model including these two components performs well if outliers and time-varying volatility do not seem to be present in the data series. Our forecasting results show that the heavy-tailed error term and the stochastic volatility component can be treated as the latent components which help to reduce the influence of the outliers and to capture the time-varying volatility only when these properties are exhibited in the data series. For example, as demonstrated in Section 2.4.1, the inflation data for Canada and Germany are relatively more stable compared with those for the other countries. Our forecasting results for Canada and Germany indicate that there is no significant improvement by adding the heavy-tailed error term and stochastic volatility component.

2.7 Conclusion

This chapter compares the G7 inflation forecast performance of autoregressive models with different error distributional assumptions and stochastic volatility specifications. Two main conclusions can be made from our results. First, the prior for the degree of freedom parameter has little influence on the forecast performance for models with a t-distribution error term. Second, both the heavy-tailed distributed error term and stochastic volatility component are important in point and density forecasts. Although a model including these two components does not always guarantee an improvement in the forecast accuracy, they tend to more often improve rather than harm the forecast accuracy. This implies that the heavy-tailed error term and the stochastic volatility component can be treated as the latent components which help to reduce the influence of the outliers and to capture the time-varying volatility only when these properties are exhibited in the data series.

2.8 Appendix

In this Appendix we provide the MCMC algorithms for AR-t-ARSV-J and AR-mixn-ARSV models.

2.8.1 AR-t-ARSV-J

First we consider the AR-t-ARSV-J model:

$$
\begin{aligned}
&y_t = \beta_0 + \beta_1 y_{t-1} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + \epsilon_t^y, \\
&h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \epsilon_t^h, \\
&\epsilon_t^y \sim \mathcal{N}(0, \lambda_t e^{h_t}), \quad \lambda_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2}), \\
&\epsilon_t^h \sim \mathcal{N}(0, \sigma_h^2).
\end{aligned}
$$

This can be written in matrix form,

$$
\begin{aligned}
&\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}^y, && \boldsymbol{\epsilon}^y \sim \mathcal{N}(0, \boldsymbol{\Sigma}_y), \\
&\mathbf{H}_{\phi_h}\overline{\mathbf{h}} = \boldsymbol{\epsilon}^h, && \boldsymbol{\epsilon}^h \sim \mathcal{N}(0, \boldsymbol{\Sigma}_h),
\end{aligned}
$$

where $\mathbf{y} = (y_1, \ldots, y_T)'$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_4)'$, $\boldsymbol{\epsilon}^y = (\epsilon_1^y, \ldots, \epsilon_T^y)'$, $\boldsymbol{\epsilon}^h = (\epsilon_1^h, \ldots, \epsilon_T^h)'$, $\overline{\mathbf{h}} = (\bar{h}_1, \ldots, \bar{h}_T)'$ with $\bar{h}_t = h_t - \mu_h$ for $t = 1, 2, \cdots, T$, $\boldsymbol{\Sigma}_y = \text{diag}(\lambda_1 e^{h_1}, \cdots, \lambda_T e^{h_T})$, $\boldsymbol{\Sigma}_h = \text{diag}(\sigma_h^2/(1-\phi_h^2), \cdots, \sigma_h^2)$ and

$$
\mathbf{X} = \begin{pmatrix}
1 & y_0 & y_{-1} & y_{-2} & y_{-3} \\
1 & y_1 & y_0 & y_{-1} & y_{-2} \\
1 & y_2 & y_1 & y_0 & y_{-1} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
1 & y_{T-1} & y_{T-2} & y_{T-3} & y_{T-4}
\end{pmatrix}, \quad
\mathbf{H}_{\phi_h} = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
-\phi_h & 1 & 0 & \cdots & 0 \\
0 & -\phi_h & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & -\phi_h & 1
\end{pmatrix}.
$$

For notational convenience, we denote $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_T)'$. We assume $|\phi_h| < 1$ and the states are initialized with $h_1 \sim \mathcal{N}(\mu_h, \sigma_h^2/(1-\phi_h^2))$. We also assume all parameters follow independent priors with

$$
\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}^0, V_\beta)\mathbb{1}(\boldsymbol{\beta} \in S), \quad \phi_h \sim \mathcal{N}(\phi_{h0}, V_{\phi_h})\mathbb{1}(|\phi_h| < 1), \quad \mu_h \sim \mathcal{N}(\mu_{\mathbf{h}_0}, V_{\mu_h}), \quad \sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h).
$$

The Jefferys prior developed by Fonseca et al. [2008] is given as

$$p_{\mathcal{J}}(\nu) \propto (\frac{\nu}{\nu+3})^{\frac{1}{2}} \left(\psi'(\frac{\nu}{2}) - \psi(\frac{\nu+1}{2}) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right)^{\frac{1}{2}},$$

where $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ and $\psi'(x) = \frac{d}{dx}\psi(x)$ are the digamma and trigamma functions. The posterior draws of model parameters can be obtained by sequentially sampling from:

1. $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{h}, \sigma_h^2, \phi_h, \mu_h, \boldsymbol{\lambda}, \nu) = p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\lambda}, \mathbf{h})$.

2. $p(\mathbf{h}|\mathbf{y}, \boldsymbol{\beta}, \sigma_h^2, \phi_h, \mu_h, \boldsymbol{\lambda}, \nu) = p(\mathbf{h}|\mathbf{y}, \sigma_h^2, \phi_h, \mu_h, \boldsymbol{\lambda})$.

3. $p(\sigma_h^2|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \phi_h, \mu_h, \boldsymbol{\lambda}, \nu) = p(\sigma_h^2|\mathbf{h}, \phi_h, \mu_h)$.

4. $p(\phi_h|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \sigma_h^2, \mu_h, \boldsymbol{\lambda}, \nu) = p(\phi_h|\mathbf{h}, \sigma_h^2, \mu_h)$.

5. $p(\mu_h|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \sigma_h^2, \phi_h, \boldsymbol{\lambda}, \nu) = p(\mu_h|\mathbf{h}, \sigma_h^2, \phi_h)$.

6. $p(\boldsymbol{\lambda}|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \sigma_h^2, \phi_h, \mu_h, \nu) = p(\boldsymbol{\lambda}|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \nu)$.

7. $p(\nu|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \sigma_h^2, \phi_h, \mu_h, \boldsymbol{\lambda}) = p(\nu|\boldsymbol{\lambda})$.

Step 1. The conditional posterior probability density

$$\begin{aligned} p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\lambda}, \mathbf{h}) &\propto p(\mathbf{y}|\mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\beta}) p(\boldsymbol{\beta}) \\ &\propto \exp\left(-\frac{1}{2} \left(\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{\Sigma}_y^{-1}\mathbf{X} + V_\beta^{-1})\boldsymbol{\beta} - 2\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{\Sigma}_y^{-1}\mathbf{y} + V_\beta^{-1}\boldsymbol{\beta}^0) \right) \right) \mathbb{1}(\boldsymbol{\beta} \in S), \end{aligned}$$

which implies

$$\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\lambda}, \mathbf{h} \sim \mathcal{N}(\overline{\boldsymbol{\beta}}, D_\beta)\mathbb{1}(\boldsymbol{\beta} \in S),$$

where $D_\beta = (\mathbf{X}'\boldsymbol{\Sigma}_y^{-1}\mathbf{X} + V_\beta^{-1})^{-1}$, $\overline{\boldsymbol{\beta}} = D_\beta(\mathbf{X}'\boldsymbol{\Sigma}_y^{-1}\mathbf{y} + V_\beta^{-1}\boldsymbol{\beta}^0)$.

Step 2. applies the auxiliary mixture sampler of Kim et al. [1998] to obtain draws of $\mathbf{h}$.

In Step 3., the posterior probability density is given by

$$\begin{aligned} p(\sigma_h^2|\mathbf{h},\mu_h,\phi_h) &\propto p(\mathbf{h}|\mu_h,\phi_h,\sigma_h^2)p(\sigma_h^2) \\ &\propto (\sigma_h^2)^{-\frac{T}{2}-\nu_h-1}\exp\big(-\frac{1}{2}\overline{\mathbf{h}}'\mathbf{H}'_{\phi_h}\mathbf{\Sigma}_h^{-1}\mathbf{H}_{\phi_h}\overline{\mathbf{h}}-\frac{S_h}{\sigma_h^2}\big) \\ &\propto (\sigma_h^2)^{-(\frac{T}{2}+\nu_h)-1}\exp\left(-\big(\frac{(1-\phi_h^2)\bar{h}_1^2+\sum_{t=2}^T(\bar{h}_t-\phi_h\bar{h}_{t-1})^2}{2}+S_h\big)/\sigma_h^2\right), \end{aligned}$$

we have

$$\sigma_h^2 \sim \mathcal{IG}(\frac{T}{2}+\nu_h, \tilde{S}_h),$$

where $\tilde{S}_h = \frac{1}{2}\left((1-\phi_h^2)\bar{h}_1^2+\sum_{t=2}^T(\bar{h}_t-\phi_h\bar{h}_{t-1})^2\right)+S_h$.

For Step 4.

$$\begin{aligned} &p(\phi_h|\mathbf{h},\mu_h,\phi_h^2) \\ \propto &p(\mathbf{h}|\phi_h,\mu_h,\sigma_h^2)p(\phi_h) \\ \propto &c(1-\phi_h^2)^{\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\phi_h^2\left(\frac{\sum_{t=2}^{T-1}\bar{h}_t}{\sigma_h^2}+V_{\phi_h}^{-1}\right)-2\phi_h\left(\frac{\sum_{t=2}^T\bar{h}_t\bar{h}_{t-1}}{\sigma_h^2}+\frac{\phi_{h_0}}{V_{\phi_h}}\right)\right)\right)\mathbb{1}(|\phi_h|<1), \end{aligned}$$

where $c = \exp(-\frac{\sum_{t=2}^T\bar{h}_{t-1}^2}{2\sigma_h^2}-\frac{\phi_{h_0}^2}{2V_{\phi_h}})$. The conditional distribution of ϕ_h is a truncated Gaussian distribution. We implement an independence-chain Metropolis-Hastings sampler with a proposal distribution $\mathcal{N}(\hat{\phi_h}, D_{\phi_h})\mathbb{1}(|\phi_h|<1)$. We denote its density as $q(x)$. So the probability for accepting a proposal draw ϕ_h^* given a current draw ϕ_h is

$$\alpha(\phi_h,\phi_h^*) = \min\left(1, \frac{p(\phi_h^*|\mathbf{h},\mu_h,\phi_h^2)q(\phi_h)}{p(\phi_h|\mathbf{h},\mu_h,\phi_h^2)q(\phi_h^*)}\right) = \min\left(1, \frac{(1-\phi_h^{*2})^{\frac{1}{2}}\exp(-\frac{1}{2\sigma_h^2}(1-\phi_h^{*2})(h_1-\mu_h)^2)}{(1-\phi_h^2)^{\frac{1}{2}}\exp(-\frac{1}{2\sigma_h^2}(1-\phi_h^2)(h_1-\mu_h)^2)}\right).$$

Step 5. The conditional density of μ_h is given by

$$\begin{aligned} p(\mu_h|\mathbf{h},\sigma_h^2,\phi_h^2) &\propto p(\mathbf{h}|\mu_h,\sigma_h^2,\phi_h^2)p(\mu_h) \\ &\propto \exp\left(-\frac{1}{2}\big((\iota'\mathbf{H}'_{\phi_h}\mathbf{\Sigma}_h^{-1}\mathbf{H}_{\phi_h}\iota+V_{\mu_h}^{-1})\mu_h^2-2\mu_h(\iota'\mathbf{H}'_{\phi_h}\mathbf{\Sigma}_h^{-1}\mathbf{H}_{\phi_h}\mathbf{h}+\mu_{h_0}/V_{\mu_h})\big)\right). \end{aligned}$$

In particular,

$$\mu_h|\mathbf{h},\phi_h,\sigma_h^2 \sim \mathcal{N}(\bar{\mu}, D_{\mu_h}),$$

where $D_{\mu_h} = (\iota' \mathbf{H}'_{\phi_h} \mathbf{\Sigma}_h^{-1} \mathbf{H}_{\phi_h} \iota + V_{\mu_h}^{-1})^{-1}$ and $\bar{\mu} = D_{\mu_h}(\iota' \mathbf{H}'_{\phi_h} \mathbf{\Sigma}_h^{-1} \mathbf{H}_{\phi_h} \boldsymbol{h} + \mu_{h_0}/V_{\mu_h})$.

For Step 6., since λ_t are assumed to be independent, then sampling $\boldsymbol{\lambda}$ can be achieved by sampling each of them sequentially. To be specific we have

$$p(\boldsymbol{\lambda}|\mathbf{y}, \mathbf{h}, \boldsymbol{\beta}, \nu) = \prod_{t=1}^{T} p(\lambda_t|y_t, h_t, \boldsymbol{\beta}, \nu).$$

Let y_t^d be the demeaned series of y_t, i.e. $y_t^d = y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \beta_3 y_{t-3} - \beta_p y_{t-4}$, then

$$\begin{aligned} p(\lambda_t|y_t, h_t, \boldsymbol{\beta}, \nu) &\propto p(y_t|h_t, \boldsymbol{\beta}, \lambda_t, \nu) p(\lambda_t) \\ &\propto \lambda_t^{-\frac{\nu+1}{2}-1} \exp\Big(-\left(\frac{(y_t^d)^2}{2e^{h_t}} + \frac{\nu}{2}\right)/\lambda_t\Big), \end{aligned}$$

we have

$$\lambda_t|y_t, h_t, \boldsymbol{\beta}, \nu \sim \mathcal{IG}\big(\frac{\nu+2}{2}, \frac{(y_t^d)^2}{2e^{h_t}} + \frac{\nu}{2}\big).$$

Step 7., to obtain samples from the conditional distribution of ν

$$p(\nu|\lambda) \propto p(\lambda|\nu) p_{\mathcal{J}}(\nu),$$

we first derive the log-density $\log p(\nu|\lambda) = \log p(\lambda|\nu) + \log p_{\mathcal{J}}(\nu)$,

$$\begin{aligned} \log p(\lambda|\nu) &= \frac{T\nu}{2} \log\big(\frac{\nu}{2}\big) - T \log \Gamma\big(\frac{\nu}{2}\big) - \big(\frac{\nu}{2} + 1\big) \sum_{t=1}^{T} \log \lambda_t - \frac{\nu}{2} \sum_{t=1}^{T} \lambda_t^{-1} + k \\ \log p_{\mathcal{J}}(\nu) &= \frac{1}{2} \left(\log \nu - \log(\nu+3) + \log g(\nu)\right) + \bar{k} \end{aligned}$$

where k and $\bar{k}$ are normalizing constants and

$$g(\nu) = \psi'\big(\frac{\nu}{2}\big) - \psi'\big(\frac{\nu+1}{2}\big) - \frac{2(\nu+3)}{\nu(\nu+1)^2}.$$

The first and second derivatives of $\log p(\lambda|\nu)$ with respect to ν can easily be derived

$$\frac{d\log p(\lambda|\nu)}{d\nu} = \frac{T}{2}\log(\frac{\nu}{2}) + \frac{T}{2} - \frac{T}{2}\Psi(\frac{\nu}{2}) - \frac{1}{2}\sum_{t=1}^{T}\log\lambda_t - \frac{1}{2}\sum_{t=1}^{T}\lambda_t^{-1},$$
$$\frac{d^2\log p(\lambda|\nu)}{d\nu^2} = \frac{T}{2\nu} - \frac{T}{4}\psi'(\frac{\nu}{2}).$$

The first and second derivatives of the $\log p_{\mathcal{J}}(\nu)$ are given

$$\frac{d\log p_{\mathcal{J}}(\nu)}{d\nu} = \frac{1}{2}\left(\frac{1}{\nu} - \frac{1}{\nu+3} + \frac{g'(\nu)}{g(\nu)}\right),$$
$$\frac{d^2\log p_{\mathcal{J}}(\nu)}{d\nu^2} = \frac{1}{2}\left(\frac{1}{(\nu+3)^2} - \frac{1}{\nu^2} + \frac{g''(\nu)g(\nu) - g'(\nu)^2}{g(\nu)^2}\right),$$

where

$$g'(\nu) = \frac{1}{2}\left(\psi''(\frac{\nu}{2}) - \psi''(\frac{\nu+1}{2})\right) - \frac{2}{\nu(\nu+1)^2} + \frac{2(\nu+3)}{\nu^2(\nu+1)^2} + \frac{4(\nu+3)}{\nu(\nu+1)^3},$$
$$g''(\nu) = \frac{1}{4}\left(\psi'''(\frac{\nu}{2}) - \psi'''(\frac{\nu+1}{2})\right) + \frac{4}{\nu^2(\nu+1)^2} + \frac{8}{\nu(\nu+1)^3} - \frac{4(\nu+3)}{\nu^3(\nu+1)^2} - \frac{8(\nu+3)}{\nu^2(\nu+1)^3} - \frac{12(\nu+3)}{\nu(\nu+1)^4},$$

hence we have

$$\log p(\nu|\lambda) \propto \log p(\lambda|\nu) + \log p_{\mathcal{J}}(\nu),$$
$$\frac{d\log p(\nu|\lambda)}{d\nu} = \frac{d\log p(\lambda|\nu)}{d\nu} + \frac{d\log p_{\mathcal{J}}(\nu)}{d\nu},$$
$$\frac{d^2\log p(\nu|\lambda)}{d\nu^2} = \frac{d^2\log p(\lambda|\nu)}{d\nu^2} + \frac{d^2\log p_{\mathcal{J}}(\nu)}{d\nu^2}.$$

We implement an independence-chain Metropolis-Hastings with proposal density $N(\hat{v}, K_v^{-1})$, where $\hat{v}$ is the maximizer of log-density function $\log p(\nu|\lambda)$ and K_v is the negative Hessian evaluated at the $\hat{v}$.

2.8.2 AR-mixn-ARSV

For convenience, we reproduce the model AR-mixn-ARSV below

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + \epsilon_t^y,$$
$$h_t = \phi_h h_{t-1} + \epsilon_t^h,$$
$$\epsilon_t^y \sim p\mathcal{N}(0, e^{h_t}) + (1-p)\mathcal{N}(0, \sigma_y^2 e^{h_t}),$$
$$\epsilon_t^h \sim \mathcal{N}(0, \sigma_h^2).$$

Common model parameters follow the same priors as mentioned in the last section.

To complete the model specification, we assume

$$\sigma_y^2 \sim \mathcal{IG}(\nu_y, S_y), \quad p \sim \mathcal{B}(p_1, p_2), \tag{2.7}$$

where $\mathcal{B}$ denotes beta distribution. Equivalently we can rewrite the expression of ϵ_t^y by introducing a mixture component indicator $\tau_t \in \{1, 0\}$ in term of conditional linear Gaussian distribution

$$\begin{aligned} \epsilon_t^y | \tau_t &= \tau_t \mathcal{N}(0, e^{h_t}) + (1 - \tau_t)\mathcal{N}(0, \sigma_y^2 e^{ht}), \\ &= \left(\tau_t + (1 - \tau_t)\sigma_y\right) \mathcal{N}(0, e^{h_t}), \end{aligned}$$

where $\Pr(\tau_t = 1) = p$ and $\Pr(\tau_t = 0) = 1 - p$. For simplicity, we denote $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_T)'$. The posterior samples for the model parameters can be obtained by sequentially sampling from:

1. $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{h}, \sigma_h^2, \phi_h, \mu_h, p, \boldsymbol{\tau}, \sigma_y^2) = p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{h}, \boldsymbol{\tau}, \sigma_y^2)$.
2. $p(\mathbf{h}|\mathbf{y}, \boldsymbol{\beta}, \sigma_h^2, \phi_h, \mu_h, p, \boldsymbol{\tau}, \sigma_y^2)$
3. $p(\sigma_h^2|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \phi_h, \mu_h, p, \boldsymbol{\tau}, \sigma_y^2) = p(\sigma_h^2|\mathbf{h}, \phi_h, \mu_h)$.
4. $p(\phi_h|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \sigma_h^2 \mu_h, p, \boldsymbol{\tau}, \sigma_y^2) = p(\phi_h|\mathbf{h}, \sigma_h^2, \mu_h)$.
5. $p(p|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \sigma_h^2, \phi_h, \mu_h, \boldsymbol{\tau}, \sigma_y^2) = p(p|\boldsymbol{\tau})$.
6. $p(\boldsymbol{\tau}|\mathbf{y}, \beta, \mathbf{h}, \sigma_h^2, \phi_h, \mu_h, p, \sigma_y^2)$.
7. $p(\sigma_y^2|\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \sigma_h^2, \phi_h, \mu_h, p, \boldsymbol{\tau})$

In Step 1. the conditional distribution is standard. For Step 2, we first transform

$$y_t^* = (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \beta_3 y_{t-3} - \beta_4 y_{t-4}) \,/\, \left(\tau_t + (1 - \tau_t)\sigma_y\right),$$

then $y_t^* | \boldsymbol{\beta}, \mathbf{h}, \boldsymbol{\tau}, \sigma_y^2 \sim \mathcal{N}(0, e^{h_t})$. Hence Step 1 - Step 4 can be achieved similarly to Section 2.8.1. Note that for identification issue we assume the stationary AR(1) stochastic volatility process with zero intercept. To accomplish Step 5 - Step 7, we first make another transform

$$y_t^{**} = (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \beta_3 y_{t-3} - \beta_4 y_{t-4}) \,/ e^{h_t/2},$$

then $y_t^{**} | \beta, \tau_t, \sigma_y^2 = \tau_t \mathcal{N}(0, 1) + (1 - \tau_t)\mathcal{N}(0, \sigma_y^2)$, which implies that y^{**} follows a two component mixture Gaussian distribution.

For Step 6. we have

$$\begin{aligned} p(p|\boldsymbol{\tau}) &\propto p(\boldsymbol{\tau}|p)p(p) \\ &\propto p^{\sum_{t=1}^{T}\tau_t+p_1-1}(1-p)^{\sum_{t=1}^{T}(1-\tau_t)+p_2-1}, \end{aligned}$$

which implies that

$$p|\boldsymbol{\tau} \sim \mathcal{B}\left(\sum_{t=1}^{T}\tau_t + p_1, \sum_{t=1}^{T}(1-\tau_t) + p_2\right).$$

To achieve Step 7, we observe that $p(\tau_t|y,\beta,h,\sigma_h^2,\phi_h,\mu_h,p,\sigma_y^2) \propto p(y_t^{**}|\tau_t,\sigma_y^2)p(\tau_t|p)$. This implies that

$$\begin{aligned} p(\tau_t = 1|y,\beta,h,\sigma_h^2,\phi_h,\mu_h,p,\sigma_y^2) &= p\psi(y_t^{**};0,1)/\left(p\psi(y_t^{**};0,1) + (1-p)\psi(y_t^{**};0,\sigma_y^2)\right), \\ p(\tau_t = 0|y,\beta,h,\sigma_h^2,\phi_h,\mu_h,p,\sigma_y^2) &= (1-p)\psi(y_t^{**};0,\sigma_y^2)/\left(p\psi(y_t^{**};0,1) + (1-p)\psi(y_t^{**};0,\sigma_y^2)\right), \end{aligned}$$

where $\phi(x;a,b)$ denotes the value obtained at x for Gaussian density with mean a and variance b. Hence the conditional distribution of τ_t follows Bernoulli distribution.

For Step 8, we first let $\mathbf{y}^{**} = (y_1^{**},\cdots,y_t^{**})$ then

$$\begin{aligned} &p(\sigma_y^2|\boldsymbol{\tau},\mathbf{y},\boldsymbol{\beta},\mathbf{h},\sigma_h^2,\phi_h,\mu_h,p) \propto p(\mathbf{y}^{**}|\boldsymbol{\tau},\sigma_y^2)p(\sigma_y^2) \\ &\propto (\sigma_y^2)^{-(\frac{\sum_{t=1}^{T}(1-\tau_t)}{2}+\nu_y)-1}\exp\left(\left(\sum_{t=1}^{T}(1-\tau_t)y_t^{**2}/2 + S_y\right)/\sigma_y^2\right), \end{aligned}$$

therefore

$$\sigma_y^2|\boldsymbol{\tau},\mathbf{y},\boldsymbol{\beta},\mathbf{h},\sigma_h^2,\phi_h,\mu_h,p \sim \mathcal{IG}\left(\frac{\sum_{t=1}^{T}(1-\tau_t)}{2} + \nu_y, \sum_{t=1}^{T}(1-\tau_t)y_t^{**2}/2 + S_y\right).$$

Chapter 3

Time-varying Relationship between Inflation and Inflation Uncertainty

3.1 Introduction

The vast literature on the relationship between inflation and inflation uncertainty can be traced back to Friedman [1977]. On the theoretical side, Ball [1992] formalizes Friedman's idea through a repeated game between the public and the monetary authority. Ball [1992] suggests that policymakers with different reactions to high inflation who alternate power stochastically would generate future inflation uncertainty, which predicts that higher inflation results in higher inflation uncertainty. On the other hand, Cukierman and Meltzer [1986] build a game-theoretic model in which the central bank has an incentive to create higher inflation uncertainty so as to exploit the trade-off between inflation and unemployment for stimulating output growth. Thus, their model predicts that higher inflation uncertainty causes higher inflation.

Empirical studies on the relationship between inflation and inflation uncertainty have often reached different conclusions in the direction of higher inflation uncertainty causing higher inflation. On the one hand, Holland [1995] uses a survey-based proxy as a measure of inflation uncertainty and finds a negative relationship between inflation and inflation uncertainty. He suggests that policymakers who view inflation uncertainty as costly are likely to have an incentive to control future inflation to a low level when current inflation is high. On the other hand, Grier and Perry [1998] conduct a cross countries analysis by using one-step ahead conditional variance from a GARCH model as a proxy for inflation uncertainty. Their results show that a negative relationship exists in countries with more independent central banks, whereas a positive relationship is found in countries with less independent central banks. Subsequent research has pointed out that the dispersion of professional eco-

nomic forecast which is used as a proxy for inflation uncertainty by Holland [1995] does not reflect the uncertainty of individual forecasters. Hence, this measure may not be appropriated to be treated as a proxy for inflation uncertainty. Furthermore using the conditional variance estimated by GARCH as a proxy for inflation uncertainty, as in Grier and Perry [1998], is also problematic since it does not account for the structural instability.

To avoid these problems, Grier and Perry [2000] adopt a GARCH in mean (GARCH-M) model in which the time varying volatility and its impact on inflation are jointly estimated and find no evidence that higher inflation uncertainty raises the inflation rate. In contrast, Berument et al. [2009] investigate the inflation-inflation uncertainty relationship using stochastic volatility in mean (SVM) model and find evidence supporting the view of Cukierman and Meltzer [1986] that higher inflation uncertainty increases inflation.

As many recent studies have pointed out the widespread nature of structural instability in many macroeconomic time series [Stock and Watson, 1996, 2007; Cogley and Sargent, 2002; Kim et al., 2004], Chan [2015b] extends the SVM model of Koopman and Hol Uspensky [2002] to allow for time-varying parameters and re-examine the relationship between inflation and inflation uncertainty. By allowing the coefficient associated with the volatility in the conditional mean to change over time, Chan [2015b] finds substantial time-variation in this relationship using data from the US, the United Kingdom and Germany. He suggests that models with constant coefficients are likely to average out the potential time-varying effects which give misleading results. However, the TVP-SVM model proposed by Chan [2015b] shares the common drawback of most conventional time-varying parameter models – the changes in the parameters are restricted in magnitude and do not accommodate the occurrence of drastic changes immediately [Koop and Potter, 2007]. Instead, a large level shift in the parameters can only be gradually captured by a sequence of small changes after the change-point, which may lead to bias estimates and result in unreliable inference.

The main goal of this chapter is to investigate the time-varying relationship between inflation and inflation uncertainty. To this end, we extend the TVP-SVM model of Chan [2015b] so as to allow the gradually evolving parameters to have potential drastic changes. For this purpose, we draw on the mixture innovation approach [McCulloch and Tsay, 1993; Gerlach et al., 2000; Giordani and Kohn, 2008] and focus on modeling the time-varying coefficients associated with the volatility

in the conditional mean. To be specific, the time-varying parameters are assumed to evolve as a random walk with a two-component Gaussian mixture innovation. Conventional mixture innovation model usually assumes the mixture probabilities to be time-invariant. However we find that this setting might not be appropriate in identifying in-sample breaks. This problem can be surmounted by assuming the mixture probabilities to be time-varying. We conduct a small simulation study to shed light on this issue in Section 3.3. In our empirical study, the sensitivity analysis also shows that the proposed model produces more reliable estimates for detecting in-sample breaks. More details will be discussed in Section 3.5.3.

We use the proposed model to investigate the inflation-inflation uncertainty relationship for the US, Germany, Canada and New Zealand. Our empirical results show that the relationship between inflation and inflation uncertainty substantially changes over time, which is consistent with the findings in Chan [2015b]. Another important finding is that it is unlikely the inflation-inflation uncertainty relationship has been gradually evolving over the last few decades. Specifically, we find a large downward shift, from positive to negative, in this relationship in the US and Germany during the Global Financial Crisis in the late 2000s. Our empirical results strongly suggest that ignoring the potential drastic changes in this relationship may lead to biased estimates. Lastly, we find that the correlation between inflation and inflation uncertainty has been much weaker since late 1990s for Canada and New Zealand which coincides the timing of the implementation of inflation targeting.

The rest of the chapter is organized as follows. Section 3.2 presents the new model for investigating the inflation-inflation uncertainty relationship. Section 3.3 conducts a simulation study. Section 3.4 discusses the posterior sampler for fitting the proposed model. Section 3.5 first discusses the data, priors, empirical results and then conducts a sensitivity analysis and model comparison exercise.

3.2 TVP-SVM with Time-varying Mixture Innovation Model

In this section we discuss the model used to investigate the relationship between inflation and inflation uncertainty. We extend the time-varying parameter stochastic volatility in mean model (TVP-SVM) of Chan [2015b] by allowing the time varying coefficients associated with the volatility in the conditional mean to have a Gaussian

mixture innovation (TVP-SVM-TVMI). Specifically, we consider

$$y_t = \tau_t + \alpha_t e^{h_t} + \epsilon_t^y, \qquad \epsilon_t^y \sim \mathcal{N}(0, e^{h_t}), \tag{3.1}$$
$$\tau_t = \tau_{t-1} + \epsilon_t^\tau, \qquad \epsilon_t^\tau \sim \mathcal{N}(0, \sigma_\tau^2), \tag{3.2}$$
$$h_t = h_{t-1} + \epsilon_t^h, \qquad \epsilon_t^h \sim \mathcal{N}(0, \sigma_h^2), \tag{3.3}$$
$$\alpha_t = \alpha_{t-1} + \epsilon_t^\alpha, \qquad \epsilon_t^\alpha \sim \mathcal{N}(0, \sigma_{K_t}^2), \tag{3.4}$$
$$\Pr(K_t = 1|\pi_t) = \pi_t, \tag{3.5}$$

where $\mathcal{N}(\cdot, \cdot)$ denotes the Gaussian distribution and $K_t \in \{0, 1\}$. The disturbances ϵ_t^y, ϵ_t^τ, ϵ_t^h and ϵ_t^α are mutually and serially uncorrelated. The time-varying volatility, i.e. e^{h_t}, is included in the conditional mean as a covariate as shown in equation (3.1). The log-volatility h_t and the state τ_t are assumed to follow random walk processes. The variances of the two-component Gaussian disturbance ϵ_t^α are restricted as $\sigma_0^2 < \sigma_1^2$.

In contrast to the TVP-SVM model of Chan [2015b], we assume α_t follows a random walk with a mixture Gaussian innovation. We also impose the restriction $\sigma_0^2 < \sigma_1^2$ that is not only for identification purpose but also for detecting drastic changes in the process of α_t. To illustrate how a drastic change can be modeled by a two-component Gaussian mixture innovation, suppose that we have a sample $(z_1, \ldots, z_T)$ with $z_t = 0$ for $t < t_b$ and $z_t = s$ for $t \geq t_b$ — i.e. a break occurs at t_b with break size s. Suppose a simple random walk process $z_t = z_{t-1} + \epsilon_t^z$ with $\epsilon_t^z \sim \mathcal{N}(0, \sigma_z^2)$ is used to model the data. Given all observed data up to time $t_b - 1$ $(z_1, \ldots, z_{t_b-1})$, the conditional 90% credible interval of z_{t_b-1+k} is given as $(-1.645k\sigma_z, 1.645k\sigma_z)$. If the level shift is large, says $s = 5$, and we assume $\sigma_z^2 = 1$, then the smallest k for the 90% credible interval to contain the changed level is 4 (because $s \in (-1.645 \times 4, 1.645 \times 4) = (-6.58, 6.58)$). Instead if $\sigma_z^2 = 10$, then this drastic level shift will be located in the 90% credible interval for $k = 1$. It is easy to see that the 90% credible interval for $k = 1$ is $(-1.654\sqrt{10}, 1.654\sqrt{10}) = (-5.20, 5.20)$ which contain s. However if the break size is small, say $s = 1$, then a Gaussian innovation with small variance, $\sigma^2 = 1$, will be better in explaining this change than that with a large variance $\sigma_z^2 = 10$. It is because the Gaussian likelihood evaluated at $s = 1$ with $\sigma_z^2 = 1$ which is 0.24 is larger than the likelihood with $\sigma_z^2 = 10$ which is 0.12.

A few implications can be drawn from this example. First, a random walk process with innovations that have a small variance is better for modeling gradually evolving changes. And second, innovations that have a large variance are better for capturing drastic changes. More importantly, assuming innovations with a small variance, as in

most TVP models, one cannot capture a drastic change immediately. Instead, such a drastic change can only be captured by a sequence of small changes after the change-point. Our previous example shows that when $s = 5$ and $\sigma_z^2 = 1$, the random walk process will take at least 4 steps to allow the 90% credible interval to contain this change. However, if we assume $\sigma_z^2 = 10$, this drastic change can be captured immediately by only 1 step. In Chan [2015b], the time-varying parameters α_t are assumed to be driven by a single Gaussian innovation with a small variance. Such a setting is suitable for modeling a gradual evolution in the inflation-inflation uncertainty relationship, but it may be inappropriate if drastic changes exist in this relationship. Our empirical study also provides strong evidence that ignoring the potential drastic changes in the inflation-inflation uncertainty relationship would typically result in biased estimates.

To allow our proposed model to accommodate both gradually evolving and drastic changes, we adopt the mixture innovation approach [McCulloch and Tsay, 1993; Gerlach et al., 2000]. The key element to combine the two component Gaussian innovations is the mixture probability π_t. Previous studies in modeling structure change often model the occurrence of breaks, K_t, as an i.i.d Bernoulli process [Koop et al., 2009; Liu and Morley, 2014], i.e. $\pi_t = \pi$ for all t. In contrast, our proposed model assumes that the mixture probability is varying over time as shown in equation (3.5). The proposed setting is important in identifying in-sample breaks. As we will show in Section 3.3, our proposed approach performs better in detecting the in-sample breaks and produces more reliable estimates.

3.3 Simulation Study

In this section, we conduct a simulation study to investigate the performance of two types of Gaussian mixture innovation models in identifying small numbers of in-sample breaks with small break size. Our simulation study shows that compared to models with an invariant mixture probability, models with time-varying probabilities perform much better in detecting in-sample breaks and are less sensitive to the sample size.

First, we consider the traditional Gaussian mixture innovation model

$$\begin{aligned} y_t &\sim \mathcal{N}(0, \sigma_{K_t}^2), \\ \Pr(K_t = 1|\pi) &= \pi, \end{aligned}$$

and refer to this model as M_1 where $K_t \in \{0, 1\}$ are independent and identically distributed Bernoulli variables with parameter π.

Second, we consider another Gaussian mixture innovation model where K_t are independent but non-identical Bernoulli random variables. To be specific we consider

$$\begin{aligned} y_t &\sim \mathcal{N}(0, \sigma^2_{K_t}), \\ \Pr(K_t = 1|\pi_t) &= \pi_t \end{aligned}$$

and refer to this model as M_2. For both M_1 and M_2, we assume $\sigma_0^2 < \sigma_1^2$.

To facilitate direct comparison, we assume the same prior for parameters common across models. For (σ_0^2, σ_1^2), we assume

$$(\sigma_0^2, \sigma_1^2) \sim (\mathcal{IG}(\nu_0, S_0) \times \mathcal{IG}(\nu_1, S_1))\, \mathbb{1}(\sigma_0^2 < \sigma_1^2),$$

where $\mathcal{IG}(\cdot, \cdot)$ denotes the Inverse-Gamma distribution and $\mathbb{1}(A)$ is the indicator function that is equal to one if statement A is true and zero otherwise. In other words, the prior density of (σ_0^2, σ_1^2) is a product of two Inverse-Gamma distributions but truncated so as to satisfy the restriction $\sigma_0^2 < \sigma_1^2$. We set the shape parameter $\nu_0 = \nu_1 = 10$ and scale parameter $S_0 = 0.09$ and $S_1 = 90$. For the mixture probabilities π and π_t, we assume

$$\pi \sim \mathcal{B}(p_1, p_2), \quad \pi_t \sim \mathcal{B}(p_{1t}, p_{2t})$$

where $\mathcal{B}(\cdot, \cdot)$ indicates the Beta distribution. For simplicity, we set the shape parameters $p_1 = p_{1t} = 1$ and $p_2 = p_{2t} = 99$ for all $t = 1, \cdots, T$. This implies that $\mathbb{E}(\pi) = \mathbb{E}(\pi_t) = 0.01$ for all $t = 1, \ldots, T$.

To investigate the performance of these two models in identifying in-sample breaks, we simulate a time series according to

$$y_t \sim \mathcal{N}(0, \sigma^2) \quad \text{for} = 1, \cdots, t_b - 1, t_b + 1, \cdots, T, \tag{3.6}$$

$$y_{t_b} \sim \mathcal{N}(s, \sigma^2), \tag{3.7}$$

where the break date t_b is set at $\lceil T/2 \rceil$ for sample size T with break size $s = 0.5$. We set the variance of the disturbance $\sigma^2 = 0.01$. $\lceil z \rceil$ denotes the smallest integer larger than z. For example, $\lceil 3.5 \rceil = 4$. Two models discussed before are considered for detecting the break data t_b with break size s by using the posterior mean of K_{t_b}.

Our simulation study is conducted as follows, for a given sample size T,

S1: we first generate 500 data sets according to equation (3.6) and equation (3.7);

S2: then for each data set $i = 1, \ldots, 500$, we compute the posterior mean of $K^i_{t_b}$ and denote it as $P^i_{t_b}$;

S3: lastly, we define $\bar{K}^T_{t_b} = \frac{1}{500} \sum_{i=1}^{500} P^i_{t_b}$ as the estimate for sample size T;

We repeat S1 - S3 for sample size $T = 500, 1000, \cdots, 4500, 5000$ and evaluate $\bar{K}^T_{t_b}$ accordingly. We use the estimate $\bar{K}^T_{t_b}$ for each model, M_1 and M_2, to measure the performance of identifying the in-sample breaks. As the time t_b is the known break date in the simulated data series, we expect $\bar{K}^T_{t_b}$ to be close to 1.

Figure 3.1 plots $\bar{K}^T_{t_b}$ against sample size T. A few conclusions can be drawn from this simulation study. First, M_1 consistently produces a lower $\bar{K}^T_{t_b}$ compared with M_2 at any given sample size T. In other words, at the known break date t_b, M_2 performs better than M_1 in identifying the in-sample break. Second, the estimate $\bar{K}^T_{t_b}$ for M_1 is dependent on the sample size T. In Figure 3.1, the estimates $\bar{K}^T_{t_b}$ from M_1 are more likely to be small for a large sample size compared with those for a small sample size. For example $\bar{K}^{500}_{t_b}$ is about 0.7 and $\bar{K}^{5000}_{t_b}$ is about 0.55. Third, it is worth stressing that matching the prior mean of π with the exact ratio of break numbers to sample size cannot improve the performance of M_2, instead it would perform even worse. For example, for $T = 500$ the actual ratio of break numbers to sample size is $\frac{1}{500} = 0.002$ which is much less than the prior mean of $\mathbb{E}(\pi) = 0.01$. For a larger sample size, this ratio becomes even smaller. This implies that the prior belief of the occurrence of an in-sample break has already been larger than the actual ratio of the break numbers to the sample size.

We compare the performance of two types of Gaussian mixture innovation models in identifying in-sample breaks in this section Our results show that modeling the mixture probability to be invariant over time, as in most mixture innovation models, is not appropriate for identifying in-sample breaks. One drawback of this type of models is that their performance is dependent on the sample size. We find that our proposed approach, in which the mixture probabilities are assumed to be time-varying, does not suffer from this problem and produces more reliable estimates in identifying in-sample breaks than does the conventional approach.

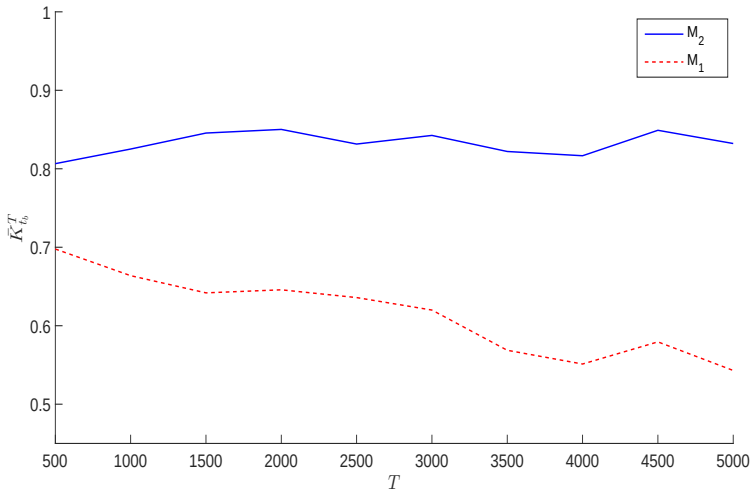

Figure 3.1: Results of Simulation Study: M_1 assumes K_t following independent and **identical** Bernoulli process; M_2 assumes K_t following independent but **non-identical** Bernoulli process.

3.4 Bayesian Estimation

In this section, we briefly discuss the posterior sampler for estimating the TVP-SVM-TVMI model in (3.1) - (3.5). For notational convenience, we denote $\mathbf{y} = (y_1, \ldots, y_T)'$, $\mathbf{h} = (h_1, \ldots, h_T)'$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_T)'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)'$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_T)'$ and $\mathbf{K} = (K_1, \ldots, K_T)'$. Posterior draws can be obtained by sequentially sampling from:

1. $p(\mathbf{h}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_\tau^2, \sigma_h^2, \mathbf{K}, \boldsymbol{\pi}, \sigma_0^2, \sigma_1^2, \mathbf{y}) = p(\mathbf{h}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_h^2, \mathbf{y})$;
2. $p(\boldsymbol{\tau}, \boldsymbol{\alpha}|\mathbf{h}, \sigma_\tau^2, \sigma_h^2, \mathbf{K}, \boldsymbol{\pi}, \sigma_0^2, \sigma_1^2, \mathbf{y}) = p(\boldsymbol{\tau}, \boldsymbol{\alpha}|\mathbf{h}, \sigma_\tau^2, \mathbf{K}, \sigma_0^2, \sigma_1^2, \mathbf{y})$;
3. $p(\sigma_\tau^2|\mathbf{h}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_h^2, \mathbf{K}, \boldsymbol{\pi}, \sigma_0^2, \sigma_1^2, \mathbf{y}) = p(\sigma_\tau^2|\boldsymbol{\tau})$;
4. $p(\sigma_h^2|\mathbf{h}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_\tau^2, \mathbf{K}, \boldsymbol{\pi}, \sigma_0^2, \sigma_1^2, \mathbf{y}) = p(\sigma_h^2|\mathbf{h})$;
5. $p(\mathbf{K}|\mathbf{h}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_\tau^2, \sigma_h^2, \boldsymbol{\pi}, \sigma_0^2, \sigma_1^2, \mathbf{y}) = p(\mathbf{K}|\boldsymbol{\alpha}, \boldsymbol{\pi}) = \prod_{t=1}^T p(K_t|\boldsymbol{\alpha}, \pi_t)$;
6. $p(\boldsymbol{\pi}|\mathbf{h}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_\tau^2, \sigma_h^2, \mathbf{K}, \sigma_0^2, \sigma_1^2, \mathbf{y}) = p(\boldsymbol{\pi}|\mathbf{K}) = \prod_{t=1}^T p(\pi_t|K_t)$;
7. $p(\sigma_0^2, \sigma_1^2|\mathbf{h}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_\tau^2, \sigma_h^2, \mathbf{K}, \boldsymbol{\pi}, \mathbf{y}) = p(\sigma_0^2, \sigma_1^2|\boldsymbol{\alpha}, \mathbf{K})$.

In Step 1, we follow the approach of Chan [2015b] and Chan and Strachan [2014] by exploiting the fact that the Hessian of $\log p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}, \sigma_h^2)$ is a band matrix. A Gaussian approximation can be obtained and used as a proposal density in an acceptance-rejection Metropolis-Hastings algorithm. In addition, the Hessian of this Gaussian proposal density is a band matrix, and therefore draws from the proposal density can

be obtained efficiently using the precision sampler of Chan and Jeliazkov [2009]. The conditional distribution in Step 2. is Gaussian. To obtain draws from this step we again adopt the more efficient precision sampler [Chan and Jeliazkov, 2009] instead of the conventional Kalman filter-based algorithms. Step 3. - Step 6. are standard, and it is worth noting that the probability π_t of occurrence of a break is time-varying. In Step 7., since the support of the conditional distribution is truncated as $\sigma_0^2 < \sigma_1^2$, a Metropolis-Hastings method is implemented. Full details of the estimation algorithm are provided in the Appendix.

3.5 Application

In this section, we use the proposed model to analyze the relationship between inflation and inflation uncertainty for the US, Germany, Canada and New Zealand. We first discuss data and priors used in our empirical study, then present the full sample results. At the end, we compare the estimates of our proposed model with those produced by the traditional mixture innovation model in the sensitivity analysis and conduct a model comparison exercise.

3.5.1 Data and Priors

The data we used are quarterly CPI inflation of the US, Germany, Canada and New Zealand. More specifically, given the quarterly CPI index x_t, we use $y_t = 400 \log(x_t / x_{t-1})$ as the CPI inflation. We use all available data for each country in the post-World War II period. To be specific, the CPI inflation data for the US, Canada, New Zealand are from 1947Q1 - 2014Q4. The CPI inflation data for Germany are from 1955Q2-2014Q4. All countries selected are either explicit inflation targeting countries like Canada and New Zealand, or behave as inflation targeting countries like the US and Germany. Figure 3.2 displays the quarterly CPI inflation data.

We assume independent priors for (σ_0^2, σ_1^2), σ_τ^2, σ_h^2 and $(\pi_1, \ldots, \pi_T)$. The priors for (σ_0^2, σ_1^2) and $(\pi_1, \ldots, \pi_T)$ are set to be the same as those discussed in Section 3.3. For σ_τ^2 and σ_h^2, we assume

$$\sigma_\tau^2 \sim \mathcal{IG}(\nu_\tau, S_\tau), \quad \sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h).$$

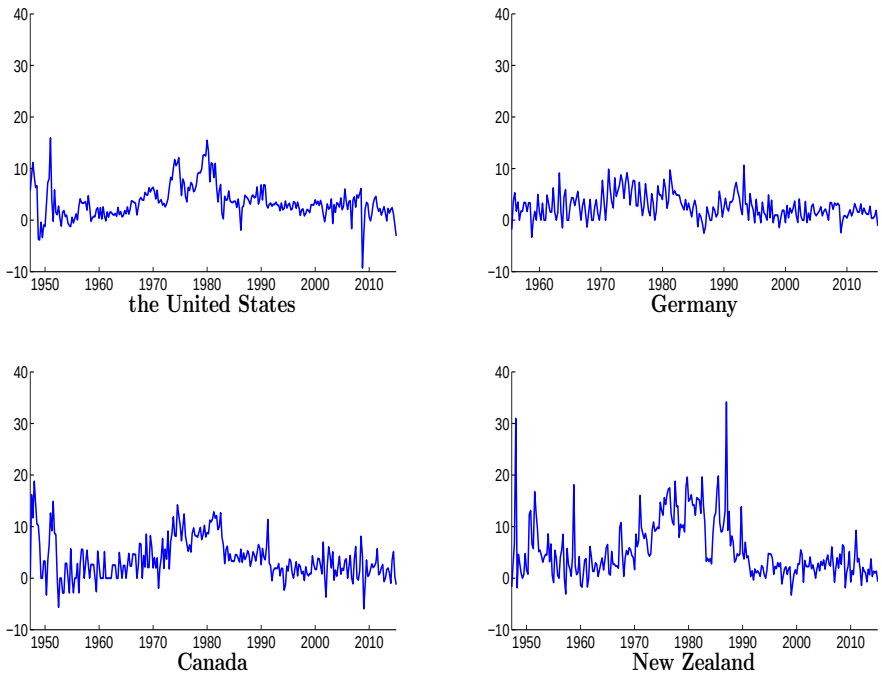

Figure 3.2: Quarterly CPI inflation for the US, Germany, Canada and New Zealand.

All state variables are initialized as

$$\tau_0 \sim \mathcal{N}(0, V_\tau), \quad h_0 \sim \mathcal{N}(0, V_h), \quad \alpha_0 \sim \mathcal{N}(0, V_\alpha).$$

For the hyperparameters of the prior distributions, we set a large variance for the prior Gaussian distribution for all initial state variable, $V_\tau = V_h = V_\alpha = 5$ and choose a relatively noninformative priors by setting $\nu_\tau = \nu_h = \nu_\alpha = 10$ and $S_\tau = 0.5625$, $S_h = 0.36$ and $S_\alpha = 0.09$. These imply that $\mathbb{E}\sigma_h^2 = 0.2^2$, $\mathbb{E}\sigma_\alpha^2 = 0.1^2$ and $\mathbb{E}\sigma_\tau^2 = 0.25^2$. These priors are comparable to those used in Chan [2015b].

3.5.2 Full Sample Results

Models are estimated using Markov chain Monte Carlo (MCMC) methods. All results are based on 50000 posterior draws after a burnin period of 5000. For a comparison purpose, we also replicate the estimation results of TVP-SVM of Chan [2015b]. In Figure 3.3 - Figure 3.6, we plot the estimated posterior means and the 90% credible intervals of α_t in panel (a). In panel (b), we plot the estimated posterior means and the 90% credible intervals of α_t (left axis) alongside with the posterior probabilities of occurrence of a break $P(K_t = 1|\mathbf{y})$ (right axis). Since the main objective of this chapter is to investigate the relationship between inflation and inflation uncertainty,

for the sake of brevity, other estimation results are presented in the Appendix.

The estimation results for both TVP-SVM and TVP-SVM-TVMI suggest that the relationship between inflation and inflation uncertainty varies substantially over the last few decades. For example, the estimates of α_t for the US, Canada and New Zealand are positive during the 1970s and then drop almost to 0 in the late 1990s. This suggests that models that assume time-invariant relationships between inflation and inflation uncertainty might not be appropriate.

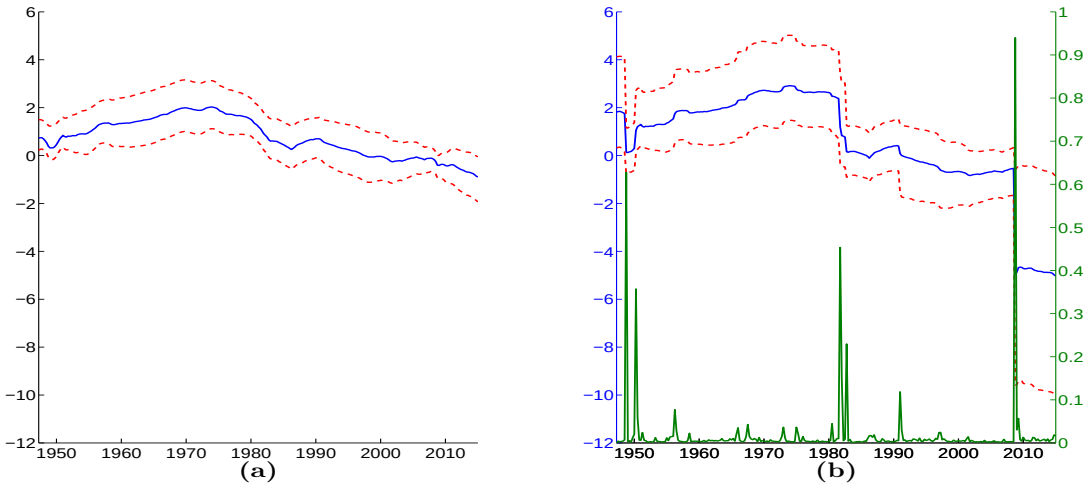

Figure 3.3: Estimation Results of the US: a) Posterior means and 90% credible intervals of α_t from TVP-SVM; and b) Posterior means and 90% credible intervals of α_t (left axis) and probabilities of occurrence of break $P(K_t = 1|\mathbf{y})$ (right axis) from TVP-SVM-TVMI.

Furthermore, we find that this relationship not always gradually evolves over time. The posterior probabilities of break occurrence, $P(K_t = 1|\mathbf{y})$, provide clear evidence supporting the occurrence of breaks in the relationship between inflation and inflation uncertainty for the US, Germany and Canada. Ignoring the existence of breaks might lead to unreliable inference. In general, the posterior mean of α_t produced by the TVP-SVM is more moderate compared with those produced by the TVP-SVM-TVMI. One possible explanation for this is that drastic changes in parameters cannot be captured immediately in traditional TVP models and the effect of such changes is likely to be averaged out within the estimates during periods where breaks occur. Evidence of this can be clearly found by taking a close look at the results of the US. In the early 1980s, a considerable large downward shift is detected by our proposed model TVP-SVM-TVMI. Before this period, the estimates of the TVP-SVM-TVMI are mostly around 2, and they decrease instantaneously to almost 0 after the break. In contrast, the estimates produced by the TVP-SVM are much smaller in the early

1980s, mostly between 1 and 2. They start declining slowly in the early 1980s and almost reach 0 in the late 1980s. Without taking account of the effects of the break in the early 1980s, the estimates for α_t are more likely to be underestimated before this period and overestimated after this period.

Table 3.1 reports the five breaks identified with the highest posterior probabilities by the TVP-SVM-TVMI. For example, the posterior probability that a break occurs at 2008Q4 for Germany is 0.54, i.e. $P(K_{2008Q4} = 1) = 0.54$. A few breaks are detected by our proposed model for the US. They are mainly in the late 1950s, early 1980s and late 2000s. A few previous studies have found that most macroeconomic series of the US experienced a change in dynamics around the 1950s [Romer and Romer, 2002] and we also find evidence that such an instability exists in the relationship between inflation and inflation uncertainty. Moreover, in the three decades between 1960 and 1980, inflation and inflation uncertainty are highly positively correlated — all of the 90% credible intervals exclude 0 throughout this period, which is consistent with recent studies by Berument et al. [2009] and Chan [2015b]. This positive correlation is maintained until the early 1980s before a significant break has been detected at 1981Q3 (with probability 0.45), which results in a steep downward drop in the correlation between inflation and inflation uncertainty to almost 0. One possible cause for this drastic change is the contractionary monetary policy in the early 1980s as the Volcker Federal Reserve focuses more on targeting the inflation from the 1980s than before. Strong support for the existence of a break can also be found in the late 2000s during the Global Financial Crisis. Interestingly, after the occurrence of this break, the correlation between inflation and inflation uncertainty experiences a drastic drop from positive to negative. One possible explanation for this result is that right after the Global Financial Crisis, the Federal Reserve may have perceived inflation uncertainty as more costly than pre-Crisis periods. This hypothesis is consistent with the view of Holland [1995]. Although the 90% credible intervals of α_t become much wider after the Global Financial Crisis period, the intervals all exclude zero, which provide further evidence supporting a negative relationship between inflation and inflation uncertainty.

Table 3.1: Five breaks identified with the highest posterior probabilities by TVP-SVM-TVMI.

the US	Germany	Canada	New Zealand
2008Q3 (0.94)	2008Q4 (0.54)	1991Q2 (0.78)	1983Q1 (0.33)
1948Q3 (0.63)	2014Q3 (0.04)	1973Q3 (0.14)	1987Q4 (0.23)
1981Q3 (0.45)	1985Q3 (0.03)	1975Q1 (0.11)	1970Q4 (0.23)
1950Q1 (0.36)	2008Q3 (0.02)	1991Q1 (0.09)	1970Q1 (0.21)
1982Q3 (0.23)	1955Q3 (0.02)	1952Q1 (0.07)	1986Q4 (0.10)

Germany has a well-known reputation for maintaining a low level of inflation over the past decades. This can be observed in Figure 4.1. Of particular interest to this study is the fact that with the exception of the late 2000s, the estimates produced by the TVP-SVM-TVMI are almost identical to those produced by TVP-SVM. Specifically, the inflation-inflation uncertainty relationship is mostly between 0 and 1. Nevertheless, similar to the US, an abrupt fall in the estimated posterior mean of α_t has been found and the 90% credible intervals become much wider after the Global Financial Crisis. However, unlike the US, it is hard to conclude a positive, or negative, relationship between inflation and inflation uncertainty as the 90% credible intervals always include 0 throughout the whole sample period. This finding is consistent with the results in Grier and Perry [2000].

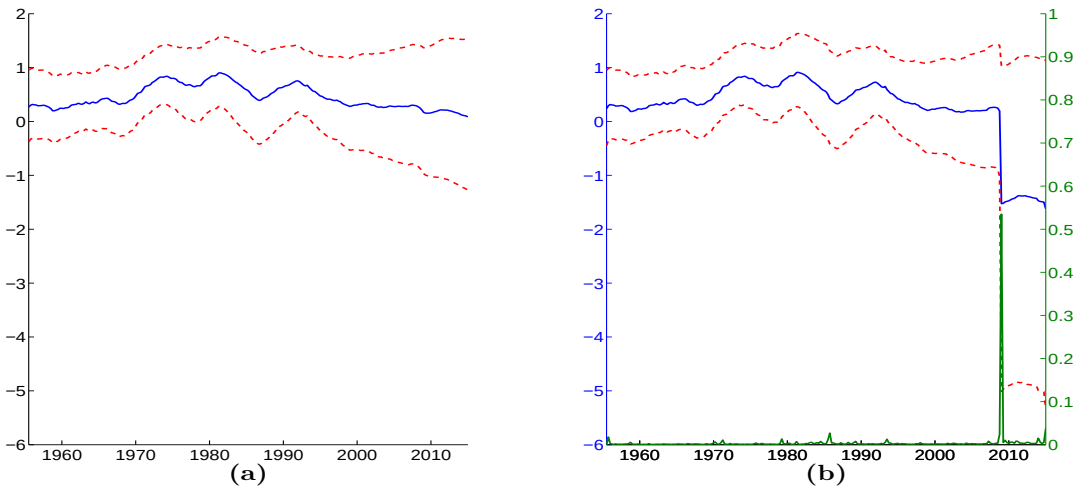

Figure 3.4: Estimation Results of Germany: a) Posterior means and 90% credible intervals of α_t from TVP-SVM; and b) Posterior means and 90% credible intervals of α_t (left axis) and probabilities of occurrence of break $P(K_t = 1|\mathbf{y})$ (right axis) from TVP-SVM-TVMI.

New Zealand experienced high inflation in the 1970s-1980s, see Figure 4.1. To control the high inflation, a few attempts were made by the Reserve Bank of New Zealand in the 1980s. They announced the adoption of inflation targeting in the late 1980s

and the inflation rate has been successfully controlled at under 5% in the 1990s, see Sherwin [1997] for more details. We find that the inflation and inflation uncertainty are highly positive correlated during the 1970s and 1980s. The 90% credible intervals of α_t produced by both TVP-SVM and TVP-SVM-TVMI exclude 0 during this period, which provides strong statistical evidence supporting a positive relationship between inflation and inflation uncertainty. Although a few breaks have been detected by the TVP-SVM-TVMI, their corresponding posterior probabilities are all smaller than 0.4, which provides weak evidence supporting the occurrence of breaks during the whole sample period. We conduct a formal model comparison exercise in Section 3.5.4 for further investigation of whether modeling the breaks in 1970s-1980s fits the data better. The α_t estimates start decreasing in the early 1990s and remain under 1 in 2000s and their corresponding 90% credible intervals all include 0 between the 1990s and 2000s. From 1970 to 1980, we find evidence to support the view of Cukierman and Meltzer [1986] that higher inflation uncertainty has a positive impact on inflation. However, this highly positive relationship has been weakened since the late 1980s and drops to nearly zero during the early 1990s, which coincides with the timing of the implementation of the inflation targeting by the Reserve Bank of New Zealand.

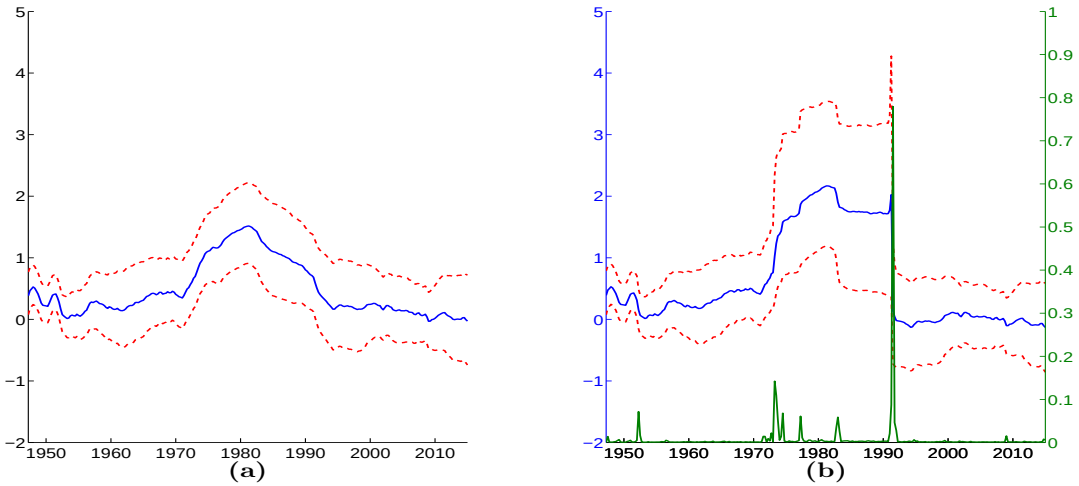

Figure 3.5: Estimation Results of Canada: a) Posterior means and 90% credible interval of α_t from TVP-SVM; and b) Posterior means and 90% credible interval of α_t (left axis) and probabilities of occurrence of break $P(K_t = 1|\mathbf{y})$ (right axis) from TVP-SVM-TVMI.

The estimation results for Canada provide strong evidence for the occurrence of a break in the early 1990s, similar to the result of New Zealand. This also matches the timing of the adoption of the inflation targeting for Canada[Melino, 2012]. The correlation between inflation and inflation uncertainty experiences a drastic drop at 1991Q2 — the probability of occurrence of a break $P(K_{1991Q2} = 1|\mathbf{y}) = 0.78$ and all

the 90% credible intervals of α_t exclude zero before 1991Q2 and include zero after 1991Q2. One possible explanation for the drop of the correlation between inflation and inflation uncertainty for Canada and New Zealand in the early 1990s is that the inflation targeting has been effective in weakening the relationship between inflation and inflation uncertainty. However, the results of Canada show that the drop of the α_t estimates is abrupt and permanent. In contrast, the results of New Zealand presented in Figure 3.6 show that the drop of the α_t estimates is more incremental over time. Similar to New Zealand, the estimation results for Canada also provide strong evidence of a highly positive correlation between inflation and inflation uncertainty in the 1970s and 1980s — as all the 90% credible intervals of α_t exclude zero during this period. This supports the view of Cukierman and Meltzer [1986] that higher inflation uncertainty raises higher inflation.

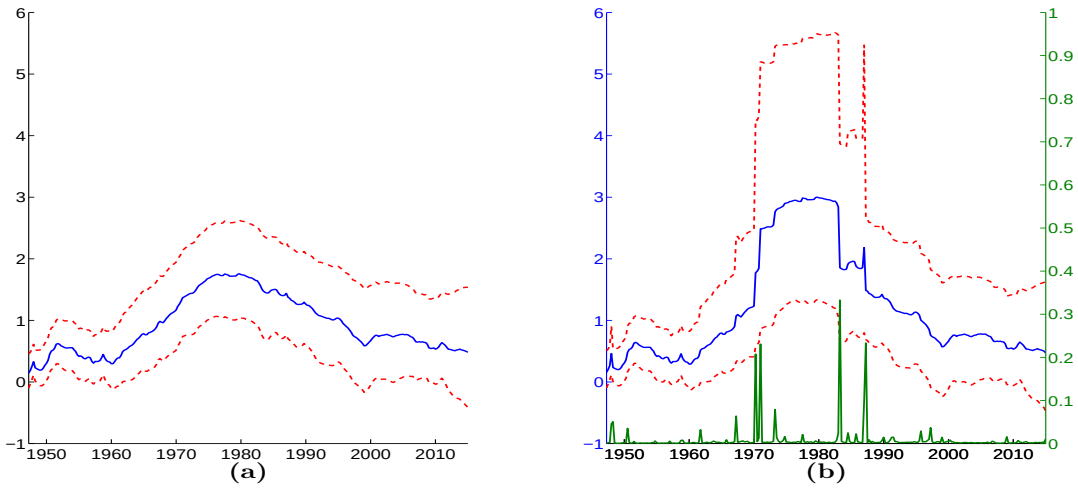

Figure 3.6: Estimation Results of New Zealand: a) Posterior means and 90% credible intervals of α_t from TVP-SVM; and b) Posterior means and 90% credible intervals of α_t (left axis) and probabilities of occurrence of break $P(K_t = 1|\mathbf{y})$ (right axis) from TVP-SVM-TVMI.

3.5.3 Sensitivity Analysis

We compare our proposed model with the traditional mixture Gaussian innovation model in which the latent variables K_t are assumed to be i.i.d. Bernoulli random

variables with the same success probability π. To be specific, we consider

$$y_t = \tau_t + \alpha_t e^{h_t} + \epsilon_t^y, \qquad \epsilon_t^y \sim \mathcal{N}(0, e^{h_t}), \tag{3.8}$$

$$\tau_t = \tau_{t-1} + \epsilon_t^\tau, \qquad \epsilon_t^\tau \sim \mathcal{N}(0, \sigma_\tau^2), \tag{3.9}$$

$$h_t = h_{t-1} + \epsilon_t^h, \qquad \epsilon_t^h \sim \mathcal{N}(0, \sigma_h^2), \tag{3.10}$$

$$\alpha_t = \alpha_{t-1} + \epsilon_t^\alpha, \qquad \epsilon_t^\alpha \sim \mathcal{N}(0, \sigma_{K_t}^2), \tag{3.11}$$

$$\Pr(K_t = 1|\pi) = \pi. \tag{3.12}$$

For simplicity, we refer this model as TVP-SVM-MI. We re-estimate models TVP-SVM-TVMI and TVP-SVM-MI based on a subsample of the US covering the period 1960Q1 - 2005Q4, and then compare with the full sample estimation results. We choose this subsample because we want to eliminate the volatile periods in the early 1950s and late 2000s. The estimates of the posterior probabilities of the occurrence of a break $P(K_t = 1|\mathbf{y})$ based on the full sample are plotted in Figure 3.7 and the subsample estimation results are presented in Figure 3.8. Considering the period of 1960Q1-2005Q4 only, we find that the TVP-SVM-MI model in which K_t are assumed to follow an identical and independent Bernoulli random variables detects more breaks in the full samplecase but fewer breaks in the subsample case. Furthermore, the posterior probabilities $P(K_t = 1|\mathbf{y})$ produced by the TVP-SVM-MI are more sensitive to different sample periods. For example, the posterior probability produced by the TVP-SVM-MI using the full sample is $P(K_{1990Q4} = 1|\mathbf{y}) = 0.27$, which is almost four times higher than the subsample estimation result $P(K_{1990Q4} = 1|\mathbf{y}) = 0.07$. This finding suggests that the conventional specification is sensitive to the sample period. In contrast, our proposed model TVP-SVM-TVMI is less sensitive to different sample periods. For example, $P(K_{1990Q4} = 1|\mathbf{y}) = 0.12$ for full sample result and $P(K_{1990Q4} = 1|\mathbf{y}) = 0.05$ for subsample result — the drop of the posterior probability $P(K_t = 1|\mathbf{y})$ is smaller comparing that of TVP-SVM-MI. This is because assuming that K_t as i.i.d Bernoulli random variables allows the information of occurrence of breaks in-sample to be shared among each observation which may not be a desirable property. The results from the sensitivity analysis help us to conclude that the proposed model TVP-SVM-TVMI produces more reliable and less sensitive estimates than those of the traditional approach for identifying in-sample structure breaks.

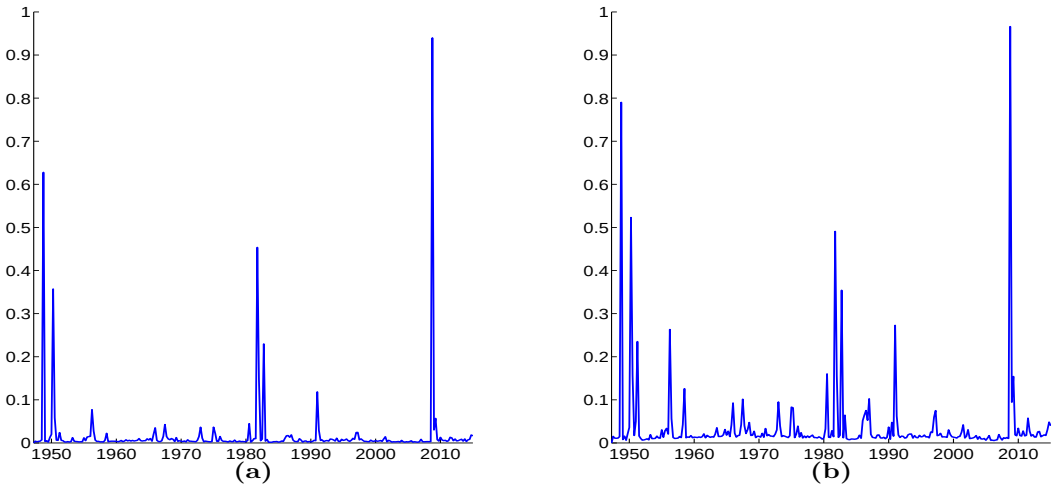

Figure 3.7: Posterior probabilities $P(K_t = 1|y)$ estimated by a) TVP-SVM-TVMI; and b) TVP-SVM-MI based on sample 1947Q1 - 2014Q4.

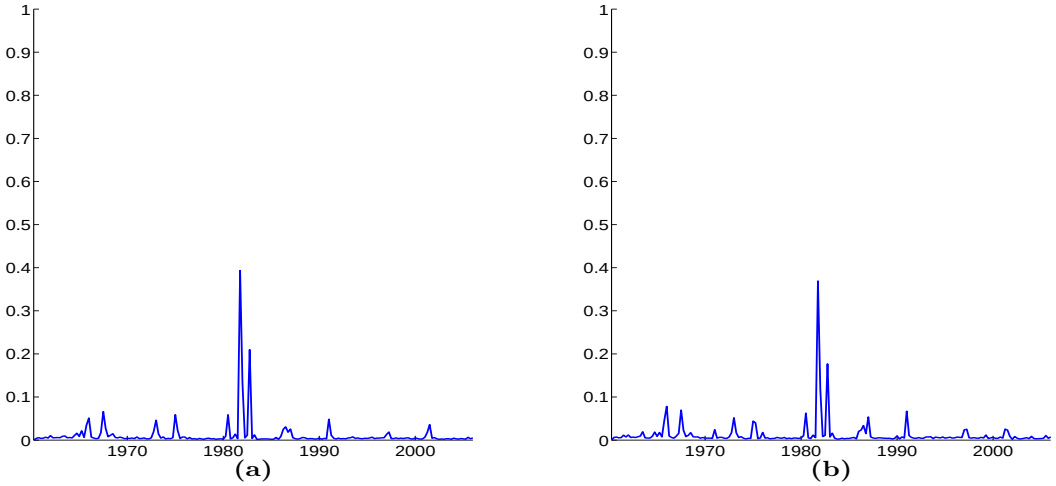

Figure 3.8: Posterior probabilities $P(K_t = 1|y)$ estimated by a) TVP-SVM-TVMI; and b) TVP-SVM-MI based on sample 1960Q1 - 2005Q4.

3.5.4 Model Comparison

In this section, we conduct a model comparison exercise. We use the marginal likelihood as a criterion to compare different models. To be specific, given model M_i, the marginal likelihood is defined as

$$p(\mathbf{y^o}|M_i) = \int p(\mathbf{y^o}|\theta_i, M_i)p(\theta_i|M_i)d\theta_i,$$

where $\mathbf{y^o} = (y_1^o, \ldots, y_T^o)$ is the observed data with sample size T and θ_i is a vector

of the parameters of model M_i. In addition, the marginal likelihood of model M_i can be rewritten as a product of one-step-ahead predictive likelihoods evaluated at the observed data. To be specific, $p(\mathbf{y^o}|M_i) = p(y_1^o|M_i)\prod_{t=2}^{T} p(y_t^o|y_1^o,\ldots,y_{t-1}^o,M_i)$. We will use this expression to compute the marginal likelihood. More discussions about the marginal likelihood and predictive likelihood can be found in Geweke and Amisano [2011].

Kass and Raftery [1995] suggest that the evidence in favor of model M_i against M_j can be interpreted by the relative log-marginal likelihood $\log \frac{p(\mathbf{y^o}|M_i)}{p(\mathbf{y^o}|M_j)}$ as: not worth more than a bare mention if $0 \leq \log \frac{p(\mathbf{y^o}|M_i)}{p(\mathbf{y^o}|M_j)} < 1$; positive if $1 \leq \log \frac{p(\mathbf{y^o}|M_i)}{p(\mathbf{y^o}|M_j)} < 3$; strong if $3 \leq \log \frac{p(\mathbf{y^o}|M_i)}{p(\mathbf{y^o}|M_j)} < 5$; very strong if $5 \leq \log \frac{p(\mathbf{y^o}|M_i)}{p(\mathbf{y^o}|M_j)}$.

Table 3.2: log marginal likelihood estimates of competing models.

	US	Germany	Canada	New Zealand
TVP-SVM	−555.6	−521.6	−655.2	−707.7
TVP-SVM-MI	−550.5	−522.3	−655.4	−707.7
TVP-SVM-TVMI	−551.1	−522.5	−655.2	−708.0

Table 3.2 reports the log-marginal likelihood of each model. The log-marginal likelihoods of the TVP-SVM-MI and TVP-SVM-TVMI are very close to each other. As we have shown in Section 3.5.3, the estimation results of model TVP-SVM-TVMI are more reliable. Hence, the following discussion will focus on comparing the TVP-SVM and the TVP-SVM-TVMI.

The log-marginal likelihoods are very similar for Germany, Canada and New Zealand. This suggests all three models fit the data equally well. This finding is not surprising. As we can see in Figure 3.4 - Figure 3.6, the estimation results of α_t for the TVP-SVM and TVP-SVM-TVMI are very similar. For example, the estimated 90% credible intervals of α_t for Canada and New Zealand exclude zeros in 1970s-1980s and include zeros in 1990s-2000s. Although the posterior means of α_t incur an abrupt downward shift in the results of Germany in late 2000s, there is no strong evidence that supports these values are significantly different from zero. However, for the results of the US, the difference of the log-marginal likelihoods between TVP-SVM and TVP-SVM-TVMI is 4.5. Based on the criterion suggested by Kass and Raftery [1995], there is strong evidence in favor of the TVP-SVM-TVMI compared to TVP-SVM. The results of the log-marginal likelihood further emphasize the empirical importance of the proposed model.

3.6 Conclusion

This chapter re-examines the relationship between inflation and inflation uncertainty for the US, Germany, Canada and New Zealand. We extend the stochastic volatility in mean model with time-varying parameter by allowing the time-varying coefficient associated with the volatility to include a mixture of Gaussian distributed errors. This allows us to investigate whether the changes in the inflation-inflation uncertainty relationship have been gradual or abrupt over time. There are three key findings. First, we find substantial time-variation in the relationship between inflation and inflation uncertainty over the last few decades for the US, Germany, Canada and New Zealand. Second, we find strong evidence of the existence of breaks in this relationship. A large change has been found in the US during the Global Financial Crisis. Interestingly, the correlation between inflation and inflation uncertainty drops from positive to negative after this break. Third, from the results of Canada and New Zealand, we find that the correlation between inflation and inflation uncertainty has been weakened since the early 1990s which coincide with the timing of the implementation of inflation targeting. A sensitivity analysis is conducted based on a subsample of the US and we find that our proposed model produces more reliable estimates in identifying in-sample breaks than the traditional mixture innovation model. Lastly, the results of the model comparison exercise provide evidence which highlight the empirical importance of our proposed model.

3.7 Appendix

3.7.1 MCMC Estimation Algorithm

In this Appendix we provide the details of the estimation procedure for the TVP-SVM-TVMI model given by equation (3.8) - (3.12). The posterior draws can be obtained by sequentially sampling from Step 1. - Step 8. in section 3.4. To obtain sample from Step 1. we follow the approach of Chan [2015b] and Chan and Strachan [2014] by exploiting the feature of the band matrix of the Hessian of $\log p(\mathbf{h}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_h^2, \mathbf{y})$. Then a Gaussian proposal can be obtained efficiently for an acceptance-rejection Metropolis-Hastings algorithm. To be specific, first note that by Bayes' rule, we have

$$p(\mathbf{h}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \sigma_h^2, \mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \mathbf{h}) p(\mathbf{h}|\sigma_h^2),$$

where the log-likelihood $\log p(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\tau}) = \sum_{t=1}^{T} \log p(y_t|\alpha_t, h_t, \tau_t)$. Then a second-order Taylor expansion around an arbitrary point $\tilde{\mathbf{h}} = (\tilde{h}_1, \ldots, \tilde{h}_T)'$ is given by

$$\begin{aligned}\log p(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\tau}) &\approx \log p(\mathbf{y}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \tilde{\mathbf{h}}) + (\mathbf{h} - \tilde{\mathbf{h}})'\mathbf{f} - \frac{1}{2}(\mathbf{h} - \tilde{\mathbf{h}})'\mathbf{G}(\mathbf{h} - \tilde{\mathbf{h}}) \\ &= -\frac{1}{2}\left(\mathbf{h}'\mathbf{G}\mathbf{h} - 2\mathbf{h}'(\mathbf{f} + \mathbf{G}\tilde{\mathbf{h}})\right) + c_1,\end{aligned}$$

where c_1 is a normalizing constant which is independent of $\mathbf{h}$, $\mathbf{f} = (f_1, \cdots, f_T)'$ and $\mathbf{G} = \text{diag}(G_1, \cdots, G_T)$ with

$$f_t = \frac{\partial}{\partial h_t} \log p(y_t|\alpha_t, h_t, \tau_t)|_{h_t = \tilde{h}_t}, \quad G_t = -\frac{\partial^2}{\partial h_t^2} \log p(y_t|\alpha_t, h_t, \tau_t)|_{h_t = \tilde{h}_t}.$$

It is worth stressing that the diagonal matrix $\mathbf{G}$ is the negative Hessian of the log-likelihood evaluated at $\tilde{\mathbf{h}}$, which is the key to obtain efficient Gaussian proposal for the acceptance-rejection Metropolis-Hastings step. The log-likelihood is given by

$$\begin{aligned}\log p(y_t|\alpha_t, h_t, \tau_t) &= -\frac{1}{2}\log 2\pi - \frac{1}{2}h_t - \frac{1}{2}e^{-h_t}(y_t - \tau_t - \alpha_t e^{h_t})^2 \\ &= -\frac{1}{2}\log 2\pi - \frac{1}{2}h_t - \frac{1}{2}\left(\alpha_t^2 e^{h_t} + e^{-h_t}(y_t - \tau_t)^2 - 2\alpha_t(y_t - \tau_t)\right),\end{aligned}$$

and the first and second derivatives of the log-likelihood can be obtained by straightforward calculations as follows:

$$\begin{aligned}\frac{\partial}{\partial h_t} \log p(y_t|\alpha_t, h_t, \tau_t) &= -\frac{1}{2} - \frac{1}{2}\alpha_t^2 e^{h_t} + \frac{1}{2}e^{-h_t}(y_t - \tau_t)^2, \\ \frac{\partial^2}{\partial h_t^2} \log p(y_t|\alpha_t, h_t, \tau_t) &= -\frac{1}{2}\alpha_t^2 e^{h_t} - \frac{1}{2}e^{-h_t}(y_t - \tau_t)^2.\end{aligned}$$

The derivation of the prior density $p(\mathbf{h}|\sigma_h^2)$ is standard. Let

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix},$$

which is the first order difference matrix. Then rewrite the state equation of h_t in matrix form,

$$\mathbf{Hh} = \boldsymbol{\epsilon}^h, \quad \boldsymbol{\epsilon}^h \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\mathbf{h}),$$

where $\boldsymbol{\Omega}_\mathbf{h} = \text{diag}(V_h, \sigma_h^2, \cdots, \sigma_h^2)$. That is, $(\mathbf{h}|\sigma_h^2) \sim \mathcal{N}(\mathbf{0}, (\mathbf{H}'\boldsymbol{\Omega}_\mathbf{h}^{-1}\mathbf{H})^{-1})$ with log-density

$$\log p(\mathbf{h}|\sigma_h^2) = -\frac{1}{2}(\mathbf{h}'\mathbf{H}'\boldsymbol{\Omega}_\mathbf{h}^{-1}\mathbf{Hh}) + c_2,$$

where c_2 is a normalizing constant independent of $\mathbf{h}$. Finally we obtain

$$\begin{aligned} \log p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\tau}) &= \log p(\mathbf{y}|\boldsymbol{\tau}, \boldsymbol{\alpha}, \mathbf{h}) + \log p(\mathbf{h}|\sigma_h^2) + c_3 \\ &\approx -\frac{1}{2}(\mathbf{h}'\mathbf{K}_\mathbf{h}\mathbf{h} - 2\mathbf{h}'\mathbf{k}_\mathbf{h}) + c_4, \end{aligned}$$

where c_3 and c_4 are constants independent of $\mathbf{h}$, $\mathbf{K}_\mathbf{h} = \mathbf{H}'\boldsymbol{\Omega}_\mathbf{h}^{-1}\mathbf{H} + \mathbf{G}$ and $\mathbf{k}_\mathbf{h} = \mathbf{f} + \mathbf{G}\tilde{\mathbf{h}}$. It can be shown that this is the log-kernel of the $\mathcal{N}(\hat{\mathbf{h}}, \mathbf{K}_\mathbf{h}^{-1})$ density, where $\hat{\mathbf{h}} = \mathbf{K}_\mathbf{h}^{-1}\mathbf{k}_\mathbf{h}$. In other words, $p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}, \mathbf{h})$ can be approximated by the Gaussian density with mean vector $\hat{\mathbf{h}}$ and precision matrix $\mathbf{K}_\mathbf{h}$. The Gaussian approximation is then used as the proposal density in the acceptance-rejection Metropolis-Hastings step. We choose $\tilde{\mathbf{h}}$ to be the mode of $p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\tau})$. To quickly locate $\tilde{\mathbf{h}}$, we apply the Newton-Raphson method. To be specific, $\mathbf{K}_\mathbf{h}$ is the negative Hessian and the score vector of $\log p(\mathbf{h}|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\tau})$ evaluate at $\mathbf{h} = \tilde{\mathbf{h}}$ is $-\mathbf{K}_\mathbf{h}\tilde{\mathbf{h}} + \mathbf{k}_\mathbf{h}$. Given a initial value $\mathbf{h_0}$, for $t = 0, 1, 2, \cdots$, evaluate $\mathbf{K}_\mathbf{h}$ and $\mathbf{k}_\mathbf{h}$ at $\tilde{\mathbf{h}} = \mathbf{h}^{(t)}$ by iterating

$$\mathbf{h}^{(t+1)} = \mathbf{h}^{(t)} + \mathbf{K}_\mathbf{h}^{-1}(-\mathbf{K}_\mathbf{h}\mathbf{h}^{(t)} + \mathbf{k}_\mathbf{h}) = \mathbf{K}_\mathbf{h}^{-1}\mathbf{k}_\mathbf{h},$$

until some convergence criterion is reached, e.g. $||\mathbf{h}^{(t+1)} - \mathbf{h}^{(t)}|| < \epsilon$ for some pre-fixed tolerance level ϵ.

To obtain draws from Step 2., we denote $\gamma_t = (\tau_t, \alpha_t)'$, $\boldsymbol{\gamma} = (\gamma_1' \ldots, \gamma_T')'$ and $\mathbf{z_t} = (1, e^{h_t})'$. Then the measurement and the state equations can be rewritten in matrix

form

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}^{\mathbf{y}}, \quad \boldsymbol{\epsilon}^{\mathbf{y}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{y}})$$
$$\mathbf{H}_{\gamma}\boldsymbol{\gamma} = \boldsymbol{\epsilon}^{\gamma}, \quad \boldsymbol{\epsilon}^{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\gamma})$$

where $\boldsymbol{\Omega}_{\mathbf{y}} = \text{diag}(e^{h_1}, \cdots, e^{h_T})$ and $\boldsymbol{\Omega}_{\gamma} = \text{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T)$,

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}_T' \end{pmatrix}, \quad \mathbf{H}_{\gamma} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_2 & \mathbf{I}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_2 & \mathbf{I}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{I}_2 & \mathbf{I}_2 \end{pmatrix},$$

and

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} V_{\tau} & 0 \\ 0 & V_{\alpha} \end{pmatrix}, \quad \boldsymbol{\Sigma}_t = \begin{pmatrix} \sigma_{\tau}^2 & 0 \\ 0 & \sigma_{K_t}^2 \end{pmatrix},$$

thus $(\boldsymbol{\gamma}|\mathbf{h}, \sigma_{\tau}^2, \mathbf{K}, \sigma_0^2, \sigma_1^2, \mathbf{y}) \sim \mathcal{N}(\hat{\boldsymbol{\gamma}}, \mathbf{D}^{-1})$, where $\mathbf{D} = \mathbf{Z}'\boldsymbol{\Omega}_{\mathbf{y}}^{-1}\mathbf{Z} + \mathbf{H}'_{\gamma}\boldsymbol{\Omega}_{\gamma}^{-1}\mathbf{H}_{\gamma}$ and $\hat{\boldsymbol{\gamma}} = \mathbf{D}^{-1}\mathbf{Z}'\boldsymbol{\Omega}_{\mathbf{y}}^{-1}\mathbf{y}$. Exact draws from this full conditional Gaussian distribution can be obtained by implementing the precision sampler of Chan and Jeliazkov [2009].

Step 3. and Step 4. are implemented by drawing from the Inverse-Gamma distributions

$$\sigma_{\tau}^2|\boldsymbol{\tau} \sim \mathcal{IG}(\hat{\nu}_{\tau}, \hat{S}_{\tau}),$$
$$\sigma_h^2|\mathbf{h} \sim \mathcal{IG}(\hat{\nu}_h, \hat{S}_h),$$

where $\hat{\nu}_{\tau} = (T-1)/2 + \nu_{\tau}$, $\hat{S}_{\tau} = S_{\tau} + \frac{1}{2}\sum_{t=2}^{t=T}(\tau_t - \tau_{t-1})^2$ and $\hat{\nu}_h = (T-1)/2 + \nu_h$, $\hat{S}_h = S_h + \frac{1}{2}\sum_{t=2}^{t=T}(h_t - h_{t-1})^2$.

For Step 5, the conditional distribution of the latent variable $K_t \in \{0, 1\}$ is given by

$$p(K_t = 1|\boldsymbol{\alpha}, \pi_t) \propto p(\boldsymbol{\alpha}|K_t)p(K_t = 1|\pi_t) \propto \pi_t \phi(\alpha_t - \alpha_{t-1}; 0, \sigma_1^2),$$
$$p(K_t = 0|\boldsymbol{\alpha}, \pi_t) \propto p(\boldsymbol{\alpha}|K_t)p(K_t = 0|\pi_t) \propto (1 - \pi_t)\phi(\alpha_t - \alpha_{t-1}; 0, \sigma_0^2),$$

where $\phi(x; \mu, \sigma^2)$ denotes the Gaussian density with mean μ and variance σ^2 evaluated at x.

In Step 6, since $(\pi_1, \ldots, \pi_T)$ are conditionally independent given $(K_1, \ldots, K_T)$, we can sample each of them sequentially. By Bayes' rule, we have

$$\begin{aligned} p(\pi_t|K_t) &\propto p(K_t|\pi_t)p(\pi_t) \\ &= (\pi_t)^{K_t+\underline{p}_1-1}(1-\pi_t)^{1-K_t+\underline{p}_2-1}, \end{aligned}$$

which then implies that $\pi_t|K_t \sim \mathcal{B}(K_t + \underline{p}_1, 1 - K_t + \underline{p}_2)$, where $\mathcal{B}(\cdot,\cdot)$ denotes the Beta distribution.

The conditional distribution of (σ_0^2, σ_1^2) in Step 7. has a truncated support and is a non-standard distribution. To be specific, the conditional density is

$$\begin{aligned} &p(\sigma_0^2, \sigma_1^2|\boldsymbol{\alpha}, \mathbf{K}) \propto p(\boldsymbol{\alpha}|\sigma_0^2, \sigma_1^2, K)p(\sigma_0^2, \sigma_1^2) \\ &\propto {\sigma_0^2}^{-(\frac{n_0}{2}+\nu_0)-1}{\sigma_1^2}^{-(\frac{n_1}{2}+\nu_1)-1}e^{-\frac{1}{2\sigma_0^2}\sum_{t=2}^{T}K_t(\alpha_t-\alpha_{t-1})^2-\frac{1}{2\sigma_1^2}\sum_{t=2}^{T}(1-K_t)(\alpha_t-\alpha_{t-1})^2-\frac{S_0}{\sigma_0^2}-\frac{S_1}{\sigma_1^2}}\mathbb{1}(\sigma_0^2<\sigma_1^2). \end{aligned}$$

We implement an independence-chain Metropolis-Hastings step with a proposal density constructed by the product of two Inverse-Gamma distributions. To be specific, a draws from the proposal can be obtained by independently drawing ${\sigma_0^2}^* \sim \mathcal{IG}(\tilde{\nu}_0, \tilde{S}_0)$ and ${\sigma_1^2}^* \sim \mathcal{IG}(\tilde{\nu}_1, \tilde{S}_1)$, where $\tilde{\nu}_0 = \frac{n_0}{2} + \nu_0$, $\nu_1 = \frac{n_1}{2} + \nu_1$, $\tilde{S}_0 = S_0 + \frac{1}{2}\sum_{t=2}^{T}K_t(\alpha_t - \alpha_{t-1})^2$ and $\tilde{S}_1 = S_1 + \frac{1}{2}\sum_{t=2}^{T}(1-K_t)(\alpha_t - \alpha_{t-1})$. Given the current draw (σ_0^2, σ_1^2), the acceptance probability for the draw $({\sigma_0^2}^*, {\sigma_1^2}^*)$ is $\min\{1, \mathbb{1}\left({\sigma_0^2}^* < {\sigma_1^2}^*\right)\}$.

3.7.2 Estimation Results

In this section we provide the estimation results for model TVP-SVM, TVP-SVM-MI and TVP-SVM-TVMI. The posterior means and standard deviations of model parameters for each country are reported in Table 3.3 - Table 3.6. The results obtained are similar to those in previous studies. The acceptance rate of the acceptance-rejection Metropolis-Hastings step to sample $\mathbf{h}$ for each model is higher than 90%, which indicates that the Gaussian proposal approximates the full conditional distribution of $\mathbf{h}$ well in this estimation step. The acceptance rate of the Metropolis-Hastings step to sample (σ_0^2, σ_1^2) of each model is close to 100%. These results are expected because of the tight priors imposed on (σ_0^2, σ_1^2) and π_t.

Table 3.3: Estimation results for the US.

	TVP-SVM		TVP-SVM-MI		TVP-SVM-TVMI	
Parameters	Mean	Std.Dev	Mean	Std.Dev.	Mean	Std.Dev
σ_τ^2	0.071	0.026	0.073	0.026	0.079	0.029
σ_h^2	0.069	0.021	0.062	0.019	0.067	0.019
σ_0^2			0.012	0.005	0.012	0.006
σ_1^2			9.264	3.068	9.899	3.361
σ_α^2	0.017	0.008				
	Acceptance Rate for (A-R)M-H step					
AR-**h**	93%		96%		95%	
AR-(σ_0^2, σ_1^2)			99%		99%	

Table 3.4: Estimation results for Germany.

	TVP-SVM		TVP-SVM-MI		TVP-SVM-TVMI	
Parameters	Mean	Std.Dev	Mean	Std.Dev.	Mean	Std.Dev
σ_τ^2	0.058	0.019	0.059	0.019	0.058	0.018
σ_h^2	0.029	0.008	0.029	0.009	0.029	0.008
σ_0^2			0.009	0.003	0.009	0.003
σ_1^2			9.874	3.445	9.742	3.376
σ_α^2	0.009	0.003				
	Acceptance Rate for (A-R)M-H step					
AR-**h**	93%		94%		94%	
AR-(σ_0^2, σ_1^2)			99%		99%	

Table 3.5: Estimation results for Canada.

	TVP-SVM		TVP-SVM-MI		TVP-SVM-TVMI	
Parameters	Mean	Std.Dev	Mean	Std.Dev.	Mean	Std.Dev
σ_τ^2	0.061	0.020	0.057	0.019	0.055	0.018
σ_h^2	0.037	0.010	0.036	0.010	0.036	0.009
σ_0^2			0.010	0.003	0.009	0.003
σ_1^2			9.547	3.305	9.324	3.163
σ_α^2	0.010	0.003				
	Acceptance Rate for (A-R)M-H step					
AR-**h**	93%		93%		93%	
AR-(σ_0^2, σ_1^2)			99%		99%	

Table 3.6: Estimation results for New Zealand.

	TVP-SVM		TVP-SVM-MI		TVP-SVM-TVMI	
Parameters	Mean	Std.Dev	Mean	Std.Dev.	Mean	Std.Dev
σ_τ^2	0.058	0.019	0.058	0.019	0.059	0.019
σ_h^2	0.069	0.021	0.068	0.021	0.069	0.021
σ_0^2			0.010	0.003	0.010	0.003
σ_1^2			9.604	3.293	9.478	3.210
σ_α^2	0.010	0.004				
		Acceptance Rate for (A-R)M-H step				
AR-**h**	97%		97%		97%	
AR-(σ_0^2, σ_1^2)			99%		99%	

Chapter 4

Infinite Hidden Markov Switching VARs with Application to Macroeconomic Forecast

4.1 Introduction

A voluminous literature has highlighted the widespread nature of structural instability in macroeconomic time series [Stock and Watson, 1996; Cogley and Sargent, 2002; Kim et al., 2004; Koop and Potter, 2007]. In order to accommodate such a feature, there has been an increasing interest in models that allow for time-variation in both conditional mean coefficients and volatilities. There are two popular families of models often used in this line of study: time-varying parameters models with stochastic volatility (TVP-SV) [Cogley and Sargent, 2005; Primiceri, 2005; Liu and Morley, 2014; Koop et al., 2009; Chan and Eisenstat, 2015] and Markov switching (MS) models [Sims and Zha, 2006; Geweke and Amisano, 2011; Hubrich and Tetlow, 2015]. Many recent studies have compared alternative specifications of TVP-SV models and found that allowing for time-variation in conditional mean coefficients and volatilities is important in improving forecast accuracy [D'Agostino et al., 2013; Clark, 2011; Clark and Ravazzolo, 2014]. However, there are few papers that focus on the forecast performance of MS models, especially in the multivariate setting. This chapter contributes to the current literature by developing vector autoregressive models with infinite hidden Markov structures. To improve computational efficiency, we develop a new MCMC sampling method built upon precision-based algorithms. We then investigate the point and density forecast performances of these infinite hidden Markov switching (IHM) models based on US quarterly GDP inflation, GDP growth and short-term interest rate.

Most recent macroeconomic studies of structural break models and Markov switch-

ing models are restricted to univariate models[1]. Koop and Potter [2007] extend the structural break model of Chib [1998] by allowing the number of structural breaks to be unknown and estimated from data. On the other hand, Giordani and Kohn [2012] model the random number of structural breaks within the framework of Gaussian state space model through mixture distributions in the state innovations. Bauwens et al. [2014] conduct a large forecasting exercise to compare various types of structural break models and highlight the importance of structural breaks in macroeconomic forecast. As pointed out by Song [2014], structural break models may incur a loss of estimation precision, because pre-break data are not directly used in the estimation of parameters of the new born regimes. To this end, he proposes an IHM model with a second hierarchical structure on the model parameters. To allow richer dynamics, Bauwens et al. [2015] extend the IHM model by imposing two independent infinite hidden Markov processes which are respectively governing the conditional mean coefficients and the volatilities. However, they have not investigated the forecast performance of IHM models with different types of dynamics. In addition, all these studies are limited to univariate models and have paid little attention to the key practical problem of multivariate macroeconomic forecasting under the structural instability. We make two main contributions to the current literature. First, we extend the IHM models into the framework of vector autoregressive models. Second, we evaluate the forecast performance of IHM models with various dynamics in both univariate and multivariate setting.

The IHM model is a nonparametric model built upon the hierarchical Dirichlet process of Teh et al. [2006], which extends the conventional MS models to allow for a possibly infinite number of regimes. One of the main advantages of IHM models over MS models is that its number of regimes for an IHM model needs not be predetermined before estimation. For a traditional MS model, selecting the number of regimes is essential for both in-sample inference and out-of-sample forecast performance. This is often done by comparing or averaging models with different numbers of regimes using a model comparison criterion [Hubrich and Tetlow, 2015; Jochmann and Koop, 2015]. However, conducting such a comparison in general is tedious and computationally demanding. More importantly, it may not be feasible for high dimensional multivariate models. Motivated by the recent empirical success of the IHM models in many empirical finance and macroeconomic studies [Jensen and Maheu, 2014; Song, 2014; Jochmann, 2015; Maheu and Yang, 2015], we adopt the IHM model approach and extend it into the multivariate setting. To capture the state

[1]We refer a structure break model as the model that do not restrict the magnitude of changes in its parameters when a break occurs.

persistence we consider the sticky version of IHM model in this chapter. We provide a brief discussion about this model in section 4.2 and refer the readers to Fox et al. [2011] for more details.

First, we contribute to this emerging research by developing vector autoregressive models with infinite hidden Markov structures. A vast literature has highlighted the empirical importance of allowing for time-variation in both conditional mean coefficients and volatilities, especially in macroeconomic forecast [Stock and Watson, 2007; D'Agostino et al., 2013; Clark, 2011; Clark and Ravazzolo, 2014; Chan, 2013, 2015a]. However, most of these studies are based on the TVP-SV models. For MS models, as a competitive counterpart of TVP-SV models, there are few studies in the literature that focus on the comparison of MS model with various dynamics. The main purpose of this chapter is to fill this gap. To be specific, this chapter compares the forecast performance of alternative IHM models within the AR and VAR specification and try to address the following three main issues: 1) whether the IHM model improves the forecast accuracy upon the time invariant AR or VAR models; 2) whether the findings in the literature of TVP-SV models are consistent to those of IHM models; 3) which specification of IHM models performs better comparing with other specifications.

The second contribution of this chapter is to propose an algorithm to improve the computational efficiency of estimating IHM models. In order to simulate the regime parameters, the conventional sampling approach requires regrouping date according to different regimes. This approach can be slow since the number of regimes for an IHM model is treated as random and it is possible that a large number of (active) regimes might be realized in some specific MCMC iterations. Our proposed algorithm allows us to simulate the parameters in distinct regime efficiently. The main idea is from the literature of nonparametric additive regression [Jeliazkov, 2008]. To be specific, selection matrices are constructed on the fly during the estimation process which helps quickly reordering the data corresponding to distinct regimes. By exploiting the selection matrices, the precision-based algorithm of Chan and Jeliazkov [2009] can be used to simulate parameters in distinct regime efficiently.

Third, our recursive out-of-sample forecast exercise shows that time-variation in model parameters is important for both point and density forecasts. Most of the gains in forecasting GDP inflation and GDP growth seem to come from allowing for time-variation in volatilities rather than conditional mean coefficients. In contrast, allowing for time-variation in conditional mean coefficients and volatilities are both important in improving forecast accuracy for short-term interest rate. In addition,

models where breaks of all model parameters occur at the same time in general have poor forecast performance. Furthermore, none of the alternative specifications of IHM models consistently outperform the model with two independent infinite hidden Markov process: one governs the changes of the conditional mean coefficients and one governs the changes of the volatilities.

The rest of the chapter is organized as follows. Section 4.2 briefly discusses the Dirichlet process and its associated mixture models. Section 4.3 presents various specifications of IHM models. Section 4.4 develops an efficient algorithm for estimation. Section 4.5 first conducts a posterior analysis, followed by a recursive out-of-sample forecasting exercise to evaluate the forecast performance of various specifications of IHM models in both univariate and multivariate settings. Section 4.6 concludes.

4.2 Hierarchical Dirichlet Process Mixture Model

In this section, we briefly discuss the Dirichlet process (DP) which is the main building block of infinite hidden Markov switching models. The Dirichlet process, denoted by $\text{DP}(c, G_0)$, is first introduced by Ferguson [1973] and defined as a distribution on distributions. It is parameterized by a base distribution G_0 over a sample space $\mathbf{\Theta}$, and a positive concentration parameter α. For illustrative purpose, we leave the base distribution G_0 unspecified at this stage. Suppose $G \sim \text{DP}(c, G_0)$. It can be shown that G has the following stick-breaking representation [Sethuraman, 1994]:

$$G = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}, \tag{4.1}$$

where δ_k is the degenerate probability measure at k and $\theta_i \sim G_0$ for $i = 1, 2, \ldots$. The probability weights $\mathbf{p} = (p_1, p_2, \ldots)$ are obtained from a stick-breaking process

$$V_i \sim \mathcal{B}(1, c), \quad p_i = V_i \prod_{j=1}^{i-1} (1 - V_j), \tag{4.2}$$

where $\mathcal{B}(\cdot, \cdot)$ represents the Beta distribution. We denote the process in (4.2) as $\mathbf{p} \sim \mathbf{SBP}(c)$. It is worth noting that a draw $\mathbf{p} \sim \mathbf{SBP}(c)$ satisfies $\sum_{i=1}^{\infty} p_i = 1$ and can be interpreted as a distribution over natural number. Intuitively, the process in (4.2) can be thought of as breaking a stick of unit length infinitely many times. At the *ith* time of breaking the stick, a draw $V_i \sim \mathcal{B}(1, c)$ will determine the proportion of the remaining stick to be broken.

The discrete nature of the DP makes it well suited for a prior in nonparametric mixture modeling. Let $(y_1, \ldots, y_T)$ be the variables of interest. The generic form of a Dirichlet process mixture model can be written as

$$y_t | s_t, \{\theta_i\}_{i=1}^{\infty} \sim F(\theta_{s_t}), \tag{4.3}$$

$$\theta_i \sim G_0, \tag{4.4}$$

$$s_t \sim \mathbf{p}, \tag{4.5}$$

$$\mathbf{p} \sim \mathbf{SBP}(c), \tag{4.6}$$

where s_t is the state variable taking values over natural number and F is a distribution parameterized by θ_i. We leave F to be unspecified for now and will discuss more about it in section 4.3. The Dirichlet process mixture model described above can be thought of as an extension of the finite mixture model to an infinite number of mixture components. However, the lack of state dependency makes the Dirichlet process mixture model less suitable for time series analysis. To this end, Teh et al. [2006] introduce a hierarchical DP prior to a set of Dirichlet process mixture models. This allows those Dirichlet process mixture models to be related and share information to each other through the hierarchical structure. The infinite hidden Markov switching model is one important variant of the hierarchical DP mixture model. Its generic form can be represented as

$$y_t | s_t, \{\theta_i\}_{i=1}^{\infty} \sim F(\theta_{s_t}), \tag{4.7}$$

$$\theta_i \sim G_0, \tag{4.8}$$

$$s_t | s_{t-1}, \{\mathbf{p}_i\}_{i=1}^{\infty} \sim \mathbf{p}_{s_{t-1}}, \tag{4.9}$$

$$\mathbf{p}_i | c, \boldsymbol{\pi} \sim \mathrm{DP}(c, \boldsymbol{\pi}), \tag{4.10}$$

$$\boldsymbol{\pi} | \gamma \sim \mathbf{SBP}(\gamma), \tag{4.11}$$

From equation (4.9), we can see that the state variable s_t follows a first order Markov process. As each $\mathbf{p}_i$ is an infinite dimensional vector, s_t can be thought of as following a Markov process which is governed by an infinite dimensional transitional matrix. In addition, all $\mathbf{p}_i$ are drawn from the common distribution $\mathrm{DP}(c, \boldsymbol{\pi})$, this allows them to learn and share information with each other.

As many macroeconomic time series are evolving with high persistence, the sticky version of the Dirichlet process prior is more suitable in modeling such a feature [Fox et al., 2011]. To be specific, the sticky version of an infinite hidden Markov switching

model is defined as the same as (4.7) - (4.11) but with equation (4.10) replaced by

$$\mathbf{p}_i|c, \boldsymbol{\pi} \sim \text{DP}(c, (1-\rho)\boldsymbol{\pi} + \rho\delta_i), \tag{4.12}$$

where $0 < \rho < 1$ is referred as the sticky parameter which reinforces the state self-transition probability. As different values of parameters (c, γ, ρ) reflect different belief of the hidden Markov process, instead of setting them to specific values, we treat them as parameters to be estimated. A brief overview of the MCMC estimation is presented in section 4.4. More details are provided in the Appendix. We refer readers to Teh et al. [2006]; Jochmann [2015] for more discussions about the DP and its associated mixture models.

4.3 Infinite Hidden Markov Switching VAR

In this section, we discuss three specifications of infinite hidden Markov switching vector autoregressive model. To set the stage, let $(\mathbf{y}_1, \ldots, \mathbf{y}_T)$ be the T observed variables of interest and each $\mathbf{y}_t$ is a $n \times 1$ vector. We consider a VAR(q) model:

$$\mathbf{y}_t = \mathbf{c}_{s_t} + \mathbf{A}_{1,s_t}\mathbf{y}_{t-1} + \cdots + \mathbf{A}_{q,s_t}\mathbf{y}_{t-q} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{s_t}), \tag{4.13}$$

where $\mathcal{N}(\cdot,\cdot)$ denotes the Gaussian distribution, $\mathbf{c}_{s_t}$ is an $n \times 1$ vector of intercepts, $\mathbf{A}_{1,s_t}, \ldots, \mathbf{A}_{q,s_t}$ are $n \times n$ coefficient matrices and $\boldsymbol{\Sigma}_{s_t}$ is the $n \times n$ covariance matrix. We rewrite equation (4.13) as a linear regression model

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta}_{s_t} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{s_t}), \tag{4.14}$$

where $\boldsymbol{\beta}_{s_t} = \text{vec}\left((\mathbf{c}_{s_t}, \mathbf{A}_{1,s_t}, \ldots, \mathbf{A}_{q,s_t})'\right)$ is $k_\beta \times 1$ with $k_\beta = n(nq+1)$ and $\mathbf{X}_t = \mathbf{I}_n \otimes (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-q})$. The time-variation of the VAR coefficients $\boldsymbol{\beta}_{s_t}$ and covariance $\boldsymbol{\Sigma}_{s_t}$ are determined by the regime indicator variable $s_t \in \{1, 2, \ldots\}$ which is following a infinite hidden Markov process:

$$s_t|s_{t-1}, \{\mathbf{p}_i\}_{i=1}^{\infty} \sim \mathbf{p}_{s_{t-1}}, \tag{4.15}$$

$$\mathbf{p}_i|c, \rho, \boldsymbol{\pi} \sim \text{DP}(c, (1-\rho)\boldsymbol{\pi} + \rho\delta_i), \tag{4.16}$$

$$\boldsymbol{\pi}|\gamma \sim \mathbf{SBP}(\gamma), \tag{4.17}$$

$$(\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i) \sim G_{(\boldsymbol{\beta},\boldsymbol{\Sigma})}, \tag{4.18}$$

where $G_{(\boldsymbol{\beta}_i,\boldsymbol{\Sigma}_i)}$ denotes the base distribution which generates the regime parameters. We refer the model (4.14) - (4.18) as IHM-VAR(q). The IHM-VAR(q) assumes that both $\boldsymbol{\beta}_{s_t}$ and $\boldsymbol{\Sigma}_{s_t}$ are governed by a single infinite hidden Markov process. This im-

plies that the breaks in the VAR coefficients and the volatilities have to occur at the same time.

To allow for richer dynamics, we also consider a version that incorporates two independent infinite hidden Markov processes: one governs the changes of the VAR coefficients and the other governs the changes of the volatilities. To be specific, we consider the model:

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_{s_t} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{z_t}), \tag{4.19}$$

$$s_t | s_{t-1}, \{\mathbf{p}_i^s\}_{i=1}^{\infty} \sim \mathbf{p}_{s_{t-1}}^s, \tag{4.20}$$

$$\mathbf{p}_i^s | c^s, \rho^s, \boldsymbol{\pi}^s \sim \mathrm{DP}(c^s, (1-\rho^s)\boldsymbol{\pi}^s + \rho^s \delta_i), \tag{4.21}$$

$$\boldsymbol{\pi}^s | \gamma^s \sim \mathbf{SBP}(\gamma^s), \tag{4.22}$$

$$z_t | z_{t-1}, \{\mathbf{p}_i^z\}_{i=1}^{\infty} \sim \mathbf{p}_{z_{t-1}}^z, \tag{4.23}$$

$$\mathbf{p}_i^z | c^z, \rho^z, \boldsymbol{\pi}^z \sim \mathrm{DP}(c^z, (1-\rho^z)\boldsymbol{\pi}^z + \rho^z \delta_i), \tag{4.24}$$

$$\boldsymbol{\pi}^z | \gamma^z \sim \mathbf{SBP}(\gamma^z), \tag{4.25}$$

$$\boldsymbol{\beta}_i \sim G_{\boldsymbol{\beta}}, \quad \boldsymbol{\Sigma}_i \sim G_{\boldsymbol{\Sigma}}. \tag{4.26}$$

We refer to this VAR model which consists of double infinite hidden Markov processes as DIHM-VAR(q). The main difference of DIHM-VAR(q) from IHM-VAR(q) is that the variations of $\boldsymbol{\beta}_{s_t}$ and $\boldsymbol{\Sigma}_{z_t}$ depend respectively on s_t and z_t, which are independent of each other. This can also be seen by comparing the base distributions in equation (4.18) and (4.26). In IHM-VAR(q), all model parameters are generated from a single base distribution $G_{(\boldsymbol{\beta},\boldsymbol{\Sigma})}$. In contrast two base distributions, $G_{\boldsymbol{\beta}}$ and $G_{\boldsymbol{\Sigma}}$, are incorporated in DIHM-VAR(q) which generates $\boldsymbol{\beta}_{s_t}$ and $\boldsymbol{\Sigma}_{z_t}$ independently.

Many studies have found that the conditional mean coefficients of a VAR model are less likely to be varying over time [Primiceri, 2005; Koop et al., 2009; Chan and Eisenstat, 2015]. Based on this reason, we consider a VAR model with constant conditional mean coefficients:

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{z_t}), \tag{4.27}$$

$$z_t | z_{t-1}, \{\mathbf{p}_i^z\}_{i=1}^{\infty} \sim \mathbf{p}_{z_{t-1}}^z, \tag{4.28}$$

$$\mathbf{p}_i^z | c^z, \rho^z, \boldsymbol{\pi}^z \sim \mathrm{DP}(c^z, (1-\rho^z)\boldsymbol{\pi}^z + \rho^z \delta_i), \tag{4.29}$$

$$\boldsymbol{\pi}^z | \gamma^z \sim \mathbf{SBP}(\gamma^z), \tag{4.30}$$

$$\boldsymbol{\Sigma}_i \sim G_{\boldsymbol{\Sigma}}, \tag{4.31}$$

and we refer to this model as C-VAR(q)-IHM.

To complete model specification, we specify our base distribution of DIHM-VAR(q) and IHM-VAR(q) as,

$$\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_0), \quad \boldsymbol{\Sigma}_i^{-1} \sim \mathcal{W}(\boldsymbol{\Sigma}_0^{-1}, \nu_0). \tag{4.32}$$

For C-VAR(q)-IHM, we assume the priors for the VAR coefficients and base distribution of the covariance matrix as

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_c, \mathbf{V}_c), \quad \boldsymbol{\Sigma}_i^{-1} \sim \mathcal{W}(\boldsymbol{\Sigma}_0^{-1}, \nu_0). \tag{4.33}$$

As proposed by Song [2014], imposing a second hierarchical structure on the base distribution allows for regime parameters in the new born regime to learn from the existing regimes, which helps improving the forecast accuracy. We follow his approach and set the second hierarchical priors as

$$\boldsymbol{\beta}_0 \sim \mathcal{N}(\boldsymbol{\beta}_{00}, \mathbf{B}_{00}), \quad \mathbf{V}_0^{-1} \sim \mathcal{W}(\mathbf{A}_{00}, a_{00}), \quad \boldsymbol{\Sigma}_0 \sim \mathcal{W}(\mathbf{Q}_{00}, b_{00}), \quad \nu_0 \sim \mathcal{E}(\lambda_{00})\mathbb{1}(\nu_0 > n), \tag{4.34}$$

where $\mathcal{W}(\mathbf{S}, \nu)$ denotes the Wishart distribution with scale matrix $\mathbf{S}$ and degrees of freedom ν. $\mathcal{E}(\lambda)$ denotes the exponential distribution with mean λ. The indicator function $\mathbb{1}(\nu_0 > n)$ implies that the support of the prior for ν_0 is restricted to be greater than n. Lastly, we assume the hyperparameters $(c_s, c_z, \gamma_s, \gamma_z, \rho_s, \rho_z)$ to follow independent priors given as

$$c^s \sim \mathcal{G}(w_s, \theta_s), \quad \gamma^s \sim \mathcal{G}(h_s, \eta_s), \quad \rho^s \sim \mathcal{B}(f_s, g_s), \tag{4.35}$$

$$c^z \sim \mathcal{G}(w_z, \theta_z), \quad \gamma^z \sim \mathcal{G}(h_z, \eta_z), \quad \rho^z \sim \mathcal{B}(f_z, g_z), \tag{4.36}$$

where $\mathcal{G}(\kappa_1, \kappa_2)$ denotes the Gamma distribution with mean κ_1/κ_2 and variance κ_1/κ_2^2.

4.4 Bayesian Estimation

This section provides an overview of the Markov chain Monte Carlo (MCMC) posterior sampler for model estimation. To be concise, we focus on the estimation procedure of the DIHM-VAR(q) model. Only minor modifications are needed for the estimation of the other models. For notational convenience, let $\mathbf{s} = (s_1, \ldots, s_T)$, $\mathbf{z} = (z_1, \ldots, z_T)$, $\mathbf{H}^s = (c^s, \gamma^s, \rho^s)$, $\mathbf{H}^z = (c^z, \gamma^z, \rho^z)$, $\boldsymbol{\Phi} = (\boldsymbol{\beta}_0, \mathbf{V}_0, \boldsymbol{\Sigma}_0, \nu_0)$ and $\mathbf{P}^s = (\mathbf{p}_1^{s\prime}, \mathbf{p}_2^{s\prime}, \ldots)'$, $\mathbf{P}^z = (\mathbf{p}_1^{z\prime}, \mathbf{p}_2^{z\prime}, \ldots)'$. We also use $\boldsymbol{\Theta}$ to denote the collection of all

regime parameters, i.e. $\boldsymbol{\Theta} = \left(\{\boldsymbol{\beta}_i\}_{i=1}^{\infty}, \{\boldsymbol{\Sigma}_j\}_{j=1}^{\infty}\right)$. Since IHM models allow for a possibly infinite number of states, the algorithm of Chib [1996] that handles a finite number of states cannot be applied. One approach is to work only with a large finite number of states to approximate the IHM model [Fox et al., 2011; Jochmann, 2015]. In contrast, we follow the approach of beam sampler proposed by Van Gael et al. [2008] for obtaining samples from the exact posterior distribution.

To apply the beam sampler, auxiliary variables $\mathbf{u}^s = (u_1^s, \ldots, u_T^s)$ and $\mathbf{u}^z = (u_1^z, \ldots, u_T^z)$ are introduced such that they are sampled alongside with the other model parameters, but they do not change the marginal distributions of the other model parameters. To be specific, for each t, u_t^s is introduced with conditional density

$$p(u_t^s | s_{t-1}, s_t) = \frac{p(u_t^s, s_t | s_{t-1})}{p(s_t | s_{t-1})} = \frac{\mathbb{1}(0 < u_t^s < p^s_{s_{t-1}s_t})}{p^s_{s_{t-1}s_t}}. \tag{4.37}$$

We suppress the rest of the conditioning variables for notational simplicity. This implies that the joint density of (u_t^s, s_t) given s_{t-1} is $p(u_t^s, s_t | s_{t-1}) = \mathbb{1}(0 < u_t^s < p^s_{s_{t-1}s_t})$. Apparently, marginalizing out u_t^s returns the original model. Hence sampling $\mathbf{u}^s$ alongside with the other model parameters does not change the marginal distributions of the other model parameters. In addition, the conditional density of s_t given (u_t^s, s_{t-1}) is

$$p(s_t | u_t^s, s_{t-1}) = \frac{\mathbb{1}(0 < u_t^s < p^s_{s_{t-1}s_t})}{\sum_i \mathbb{1}(0 < u_t^s < p^s_{s_{t-1}i})}. \tag{4.38}$$

As $\sum_i p^s_{s_{t-1}i} = 1$, it is not hard to see that the set $\{i : 0 < u_t^s < p^s_{s_{t-1}i}\}$ is finite. This implies that, given $\mathbf{u}^s$, there is only a finite number of state trajectory needed to be considered. Thus the algorithm of Chib [1996] can be applied to obtain samples of the state variables $\mathbf{s}$. The samples for the state variables $\mathbf{z}$ can also be obtained using a similar approach. More details about the beam sampler can be found in Van Gael et al. [2008] and the Appendix.

The posterior draws for the DIHM-VAR(q) model can be obtained by sequentially sampling from:

1. $p(\boldsymbol{\Theta} | \mathbf{s}, \mathbf{z}, \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{P}^s, \mathbf{P}^z, \mathbf{u}^s, \mathbf{u}^z, \mathbf{H}^s, \mathbf{H}^z, \boldsymbol{\Phi}, \mathbf{y}_{1:T}) = p(\boldsymbol{\Theta} | \boldsymbol{\Phi}, \mathbf{s}, \mathbf{z})$;
2. $p(\boldsymbol{\Phi} | \mathbf{s}, \mathbf{z}, \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{P}^s, \mathbf{P}^z, \mathbf{H}^s, \mathbf{H}^z, \mathbf{u}^s, \mathbf{u}^z, \boldsymbol{\Theta}, \mathbf{y}_{1:T}) = p(\boldsymbol{\Phi} | \boldsymbol{\Theta}, \mathbf{s}, \mathbf{z})$;
3. $p(\mathbf{u}^s, \mathbf{u}^z | \mathbf{s}, \mathbf{z}, \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{P}^s, \mathbf{P}^z, \mathbf{H}^s, \mathbf{H}^z, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathbf{y}_{1:T}) = p(\mathbf{u}^s, \mathbf{u}^z | \mathbf{s}, \mathbf{z}, \mathbf{P}^s, \mathbf{P}^z)$
4. $p(\mathbf{s}, \mathbf{z} | \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{P}^s, \mathbf{P}^z, \mathbf{H}^s, \mathbf{H}^z, \mathbf{u}^s, \mathbf{u}^z, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathbf{y}_{1:T}) = p(\mathbf{s}, \mathbf{z} | \mathbf{P}^s, \mathbf{P}^z, \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{u}^s, \mathbf{u}^z, \boldsymbol{\Theta}, \mathbf{y}_{1:T})$;

5. $p(\boldsymbol{\pi}^s, \boldsymbol{\pi}^z|\mathbf{s}, \mathbf{z}, \mathbf{P}^s, \mathbf{P}^z, \mathbf{H}^s, \mathbf{H}^z, \mathbf{u}^s, \mathbf{u}^z, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathbf{y}_{1:T}) = p(\boldsymbol{\pi}^s, \boldsymbol{\pi}^z|\mathbf{s}, \mathbf{z}, \mathbf{H}^s, \mathbf{H}^z)$;
6. $p(\mathbf{P}^s, \mathbf{P}^z|\mathbf{s}, \mathbf{z}, \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{u}^s, \mathbf{u}^z, \mathbf{H}^s, \mathbf{H}^z, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathbf{y}_{1:T}) = p(\mathbf{P}^s, \mathbf{P}^z|\mathbf{s}, \mathbf{z}, \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{H}^s, \mathbf{H}^z)$;
7. $p(\mathbf{H}^s, \mathbf{H}^z|\mathbf{s}, \mathbf{z}, \boldsymbol{\pi}^s, \boldsymbol{\pi}^z, \mathbf{P}^s, \mathbf{P}^z, \mathbf{u}^s, \mathbf{u}^z, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathbf{y}_{1:T}) = p(\mathbf{H}^s, \mathbf{H}^z|\mathbf{s}, \mathbf{z}, \mathbf{P}^s, \mathbf{P}^z)$.

A regime is said to be active if at least one observation belongs to it. It is worth noting that conditional on the auxiliary variables $\mathbf{u}^s$ and $\mathbf{u}^z$, there is only a finite number of active regimes. We denote L_s and L_z as the numbers of active regimes implied by $\mathbf{s}$ and $\mathbf{z}$ respectively. To improve computational efficiency, we propose an efficient way to obtain samples from $p(\Theta|\Phi, \mathbf{s}, \mathbf{z})$ in step 1. The idea is from the literature of nonparametric additive regression [Jeliazkov, 2008]. In our proposed approach, selection matrices are constructed on the fly in each MCMC iteration to avoid explicitly regrouping the data into distinct regimes. To be specific, we first stack all T observations and rewrite equation (4.19) as

$$\mathbf{Y} = \mathbf{X}\widetilde{\mathbf{B}} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{4.39}$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_T \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma}_{z_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{z_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{z_T} \end{pmatrix},$$

and $\mathbf{Y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_T)'$ and $\widetilde{\mathbf{B}} = (\boldsymbol{\beta}'_{s_1}, \ldots, \boldsymbol{\beta}'_{s_T})'$.

It is important to note that only the VAR coefficients for distinct active regimes, $\mathbf{B} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_{L_s})'$, are needed to be sampled in each MCMC iteration. To improve efficiency, we define a $T \times L_s$ selection matrix $\mathbf{D_s}$ which has entry $\mathbf{D_s}(i, j) = 1$ if $s_i = j$ and 0 otherwise. It is easy to check that $\widetilde{\mathbf{B}} = \left(\mathbf{D_s} \otimes \mathbf{I}_{k_\beta}\right) \mathbf{B}$. Hence equation (4.39) can be written as

$$\mathbf{Y} = \mathbf{X_s}\mathbf{B} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{4.40}$$

where $\mathbf{X_s} = \mathbf{X}\left(\mathbf{D_s} \otimes \mathbf{I}_{k_\beta}\right)$. Equation (4.40) is in the form of linear regression model, a standard result can be applied to obtain draws of $\mathbf{B}$ from

$$\mathbf{B} \sim \mathcal{N}\left(\widehat{\mathbf{B}}, \mathbf{V_B}\right), \tag{4.41}$$

where $\mathbf{V_B}^{-1} = \left(\mathbf{X_s}'\boldsymbol{\Omega}^{-1}\mathbf{X_s} + \underline{\mathbf{V}}_{L_s}^{-1}\right)$, $\widehat{\mathbf{B}} = \mathbf{V_B}\left(\mathbf{X_s}'\boldsymbol{\Omega}^{-1}\mathbf{Y} + \underline{\mathbf{V}}_{L_s}^{-1}\underline{\boldsymbol{\beta}}_{L_s}\right)$, $\underline{\mathbf{V}}_{L_s} = \mathbf{I}_{L_s} \otimes \mathbf{V}_0$,

$\underline{\boldsymbol{\beta}}_{L_s} = \mathbf{1}_{L_s} \otimes \boldsymbol{\beta}_0$. $\mathbf{1}_{L_s}$ is a $L_s \times 1$ column vector with all one at each entry.

To obtain draws of $\boldsymbol{\Sigma}_i$ for $i = 1, \ldots, L_z$, we apply a similar trick to sort out the data by using a selection matrix. To be specific, given draws of $\mathbf{B}$, $\mathbf{s}$ and $\mathbf{z}$ we first construct

$$\mathbf{Y^d} = \begin{pmatrix} \left(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_{s_1}\right)\left(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_{s_1}\right)' \\ \left(\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_{s_2}\right)\left(\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_{s_2}\right)' \\ \vdots \\ \left(\mathbf{y}_T - \mathbf{X}_T\boldsymbol{\beta}_{s_T}\right)\left(\mathbf{y}_T - \mathbf{X}_T\boldsymbol{\beta}_{s_T}\right)' \end{pmatrix}, \tag{4.42}$$

and a $L_z \times T$ selection matrix $\mathbf{D_z}$ that has $\mathbf{D_z}(i,j) = 1$ if $z_j = i$ and 0 otherwise. It is straightforward to check that

$$\left(\mathbf{D_z} \otimes \mathbf{I}_n\right)\mathbf{Y^d} = \begin{pmatrix} \sum_{t:z_t=1}\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)' \\ \sum_{t:z_t=2}\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)' \\ \vdots \\ \sum_{t:z_t=L_z}\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)' \end{pmatrix}, \tag{4.43}$$

then for $i = 1, \ldots, L_z$ sample $\boldsymbol{\Sigma}_i \sim \mathcal{IW}\left(\sum_{t:z_t=i}\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{s_t}\right)' + \boldsymbol{\Sigma}_0, \nu_0 + T_i\right)$, where $T_i = \sum_j \mathbf{D_z}(i,j)$.

We make two remarks on the computation. First, it is important to observe that the $k_\beta L_s \times k_\beta L_s$ matrix $\mathbf{V_B^{-1}}$ is block-banded i.e., all non-zero elements are concentrated around the main diagonal. This structure can be exploited to efficiently compute $\widehat{\mathbf{B}}$ without calculating the inverse of $\mathbf{V_B^{-1}}$. In addition, the precision-based sampler of Chan and Jeliazkov [2009] can be applied to quickly obtain draws from $\mathcal{N}(\widehat{\mathbf{B}}, \mathbf{V_B})$.

Second, in principle the number of active regimes in the framework of IHM model can be as large as the sample size T. In the traditional approach, parameters are sampled sequentially after regrouping the data into distinct regimes in each MCMC iteration. The computational efficiency is low especially when a large number of active regimes is realized at some specific MCMC iterations. Our proposed approach vectorizes all the operations through constructing selection matrices on the fly during estimation, which is very fast in programming environments such as MATLAB, GAUSS and R.

Table 4.1 reports the time taken to obtain 10000 posterior draws for various types of IHM models with 2 lags. The computational times for two univariate IHM models, DIHM-AR and IHM-AR, are also reported[2]. More specifically, the DIHM-AR (IHM-AR) is the DIHM-VAR (IHM-VAR) with $n = 1$. We use U.S. quarterly GDP inflation rate, GDP growth and short-term interest rate from 1954Q3 - 2015Q1 with a total of $T = 241$ observations. We will discuss more about the data in the next section. All the algorithms are implemented using MATLAB on a desktop with an Intel Core i7-2600 @3.40GHz processor.

Table 4.1: Time taken (in minutes) to obtain 10000 posterior draws for various types of IHM models with 2 lags.

Model	DIHM-VAR	IHM-VAR	DIHM-AR	IHM-AR
Time	3.7	2.5	2.3	1.4

As shown in the Table 4.1, it takes around $2-4$ minutes for the multivariate models and $1-3$ minutes for the univariate models to obtain 10000 posterior draws.

Step 2 - step 11 of the posterior sampler are following Song [2014] which are provided in the Appendix.

4.5 Application

In this section, we first discuss the data and priors used in the empirical study of this chapter. We then report some full sample posterior estimates of two IHM models, DIHM-VAR(2) and IHM-VAR(2), to illustrate a few properties of IHM models with different dynamics. Lastly, a recursive out-of-sample forecasting exercise is conducted to evaluate the forecast performance of alternative specifications of IHM models.

4.5.1 Data and Priors

The data are U.S. quarterly inflation rate computed from the GDP deflator, GDP growth and short-term interest rate from 1954Q3 - 2015Q1. The GDP deflator and real GDP are transformed to annualized growth rates. For example, given GDP deflator (real GDP) x_t we use $y_t = 400\log(x_t/x_{t-1})$ as the GDP inflation (GDP growth). The

[2]The computational times for the univariate models, DIHM-AR and DIHM-AR, are the average estimation times for GDP inflation rate, GDP growth and short-term interest rate.

short-term interest rate is the effective federal fund rate. These series are sourced from the Federal Reserve Bank of St. Louis. We plot the data in Figure 4.1.

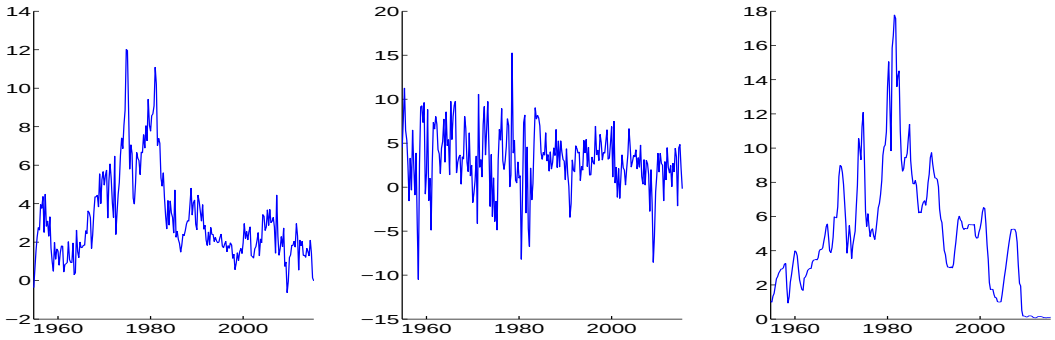

Figure 4.1: The U.S. quarterly GDP inflation (left), GDP growth (middle), and effective federal fund rate (right) from 1954Q3 to 2015Q1.

For easy comparison, we assume the same prior for parameters common across models. We set informative priors for those hyperparameters of the base distribution. In particular we set the hyperparameters $\boldsymbol{\beta}_c = \boldsymbol{\beta}_{00} = \mathbf{0}$, $\mathbf{V}_c = \mathbf{B}_{00} = \mathbf{I}_{k_\beta}$, $a_{00} = k_\beta + 10$, $\mathbf{A}_{00} = 2a_{00}\mathbf{I}_{k_\beta}$, $\mathbf{Q}_{00} = b_{00}^{-1}\mathbf{I}_n$, $b_{00} = 5$, $\lambda_{00} = n + 2$. For the priors of those hidden Markov parameters, we set $w_s = w_z = 5$, $\theta_s = \theta_z = 1$, $h_s = h_z = \eta_s = \eta_z = 1$. For the sticky parameters, we set $f_s = f_z = 10$ and $g_s = g_z = 1$ which implies a relatively high self-transition probability. As univariate models often forecast well, we also compare the forecast performance of different types of AR(q) models with various IHM structures. For each VAR(q) or AR(q) model, we consider $q = 1, 2$ for the lag order.

We report the full sample estimates in section 4.5.2 and the out-of-sample forecasting results in section 4.5.3. All estimates are based on 50000 posterior draws after a burn-in period of 5000. One main identification issue for mixture models is label switching. Ignoring this may result in misleading inference. However, Geweke [2007] shows that inference on the label invariant statistics are valid and can be conducted by as usual. Hence, all estimates reported in the following sections are label invariant.

4.5.2 Posterior Analysis

This section presents some full sample estimation results for the DIHM-VAR(2) and IHM-VAR(2). The estimation results show substantial time-variation in both the VAR coefficients and volatilities over the sample period. The main difference between the DIHM-VAR(2) and IHM-VAR(2) appear in the estimated posterior volatilities but not

VAR coefficients. More importantly, it is also evident that breaks of the parameters are not likely to occur at the same time.

Figure 4.2 - Figure 4.5 plot the estimated posterior means (solid lines) for the VAR coefficients and covariance matrices with the corresponding 68% credible intervals (dashed lines) for the DIHM-VAR(2) and IHM-VAR(2). In Figure 4.2 and Figure 4.3, each row plots the coefficients for one equation and each column reports the coefficients for one variables. For example, the third panel in the first row plots the coefficients of the first lagged GDP growth in the equation of inflation.

Overall, the estimates of the VAR coefficients from both models are mostly the same over the sample period. There are cases that the estimated 68% credible intervals of some coefficients for DIHM-VAR(2) and IHM-VAR(2) almost exclude each other, but the difference of their posterior means are small. For example, in the equation of interest rate, the 68% credible intervals of the coefficients of the first lag interest rate for DIHM-VAR(2) and IHM-VAR(2) barely exclude each other in most periods, however, the difference of their posterior means is only around 0.2.

Figure 4.4 and Figure 4.5 plots the estimates of the covariance matrices for the DIHM-VAR(2) and IHM-VAR(2) models. We use $1, 2, 3$ to label the equation of inflation, GDP growth and interest rate accordingly. Thus, σ_{ij} denotes the covariance of innovations in equation i and equation j. For example, σ_{11} is the variance of the innovation in the equation of inflation. The estimated covariance matrices for both models indicate that the volatilities of the innovations are typically high in the 1970s, followed by a decline in the early 1980s. This finding is generally consistent with the studies of the Great Moderation [Chan and Eisenstat, 2015; Primiceri, 2005; Sims and Zha, 2006]. However, the estimates for the DIHM-VAR(2) detect a drastic rise in volatilities during the Global Financial Crisis in the late 2000s. In contrast, there are no notable changes have been found in the volatilities during the whole 2000s from the estimates of the IHM-VAR(2). Furthermore, the estimated covariance among variables for the IHM-VAR(2) are much larger comparing with those for the DIHM-VAR(2) in this high volatile period.

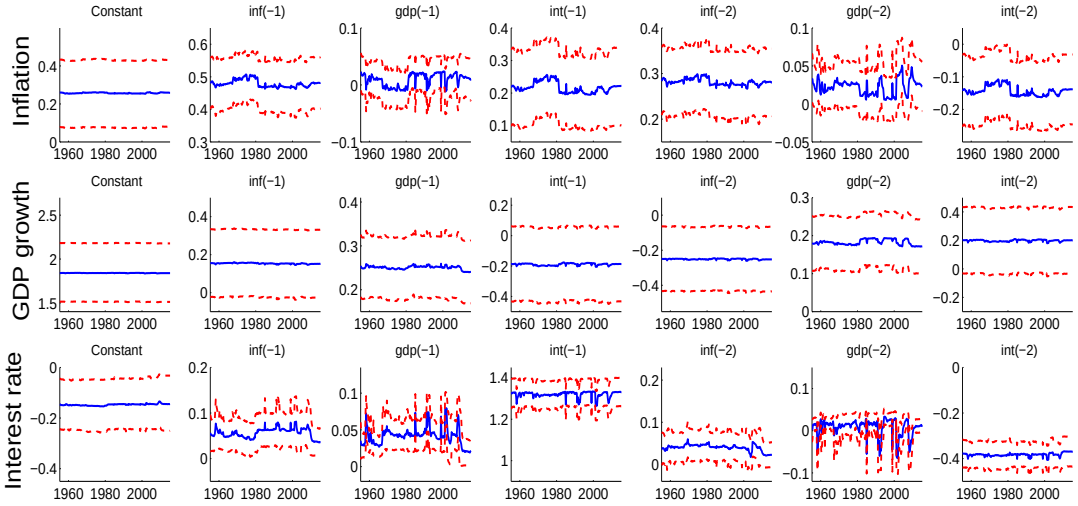

Figure 4.2: The estimated VAR coefficients for the DIHM-VAR(2): Each row reports estimates for the corresponding equation. Each column reports estimates of the lag variables.

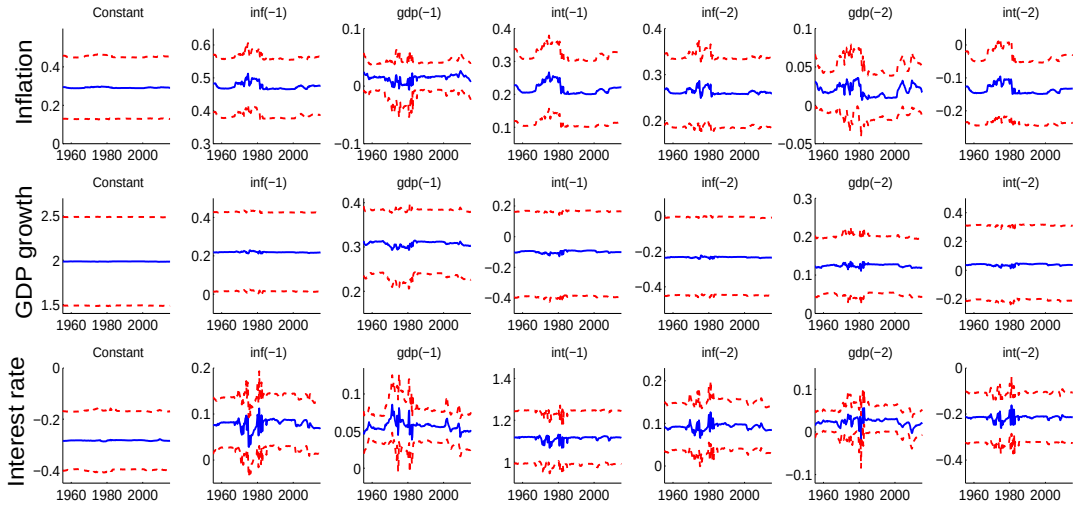

Figure 4.3: The estimated VAR coefficients for the IHM-VAR(2): Each row reports estimates for the corresponding equation. Each column reports estimates of the lag variables.

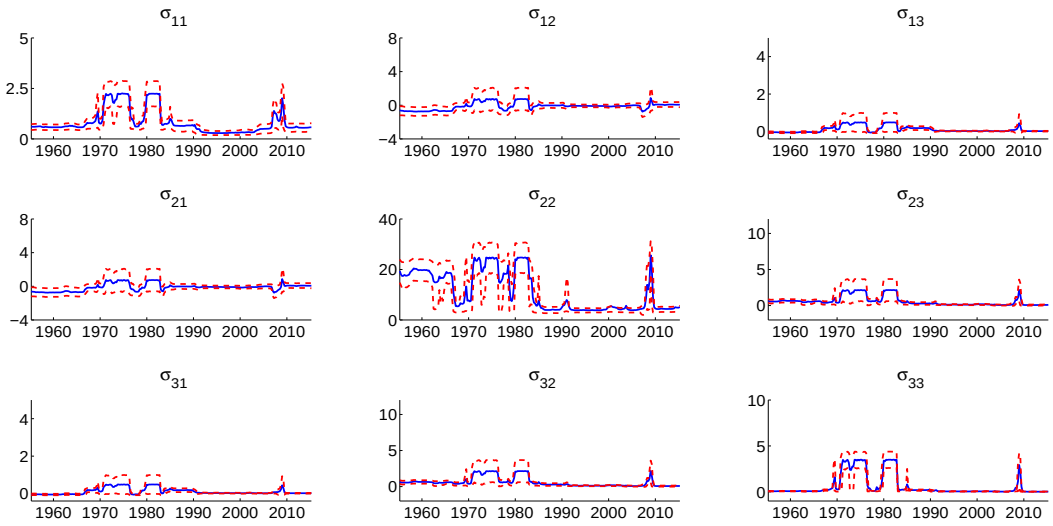

Figure 4.4: The estimated covariance matrix for the DIHM-VAR(2).

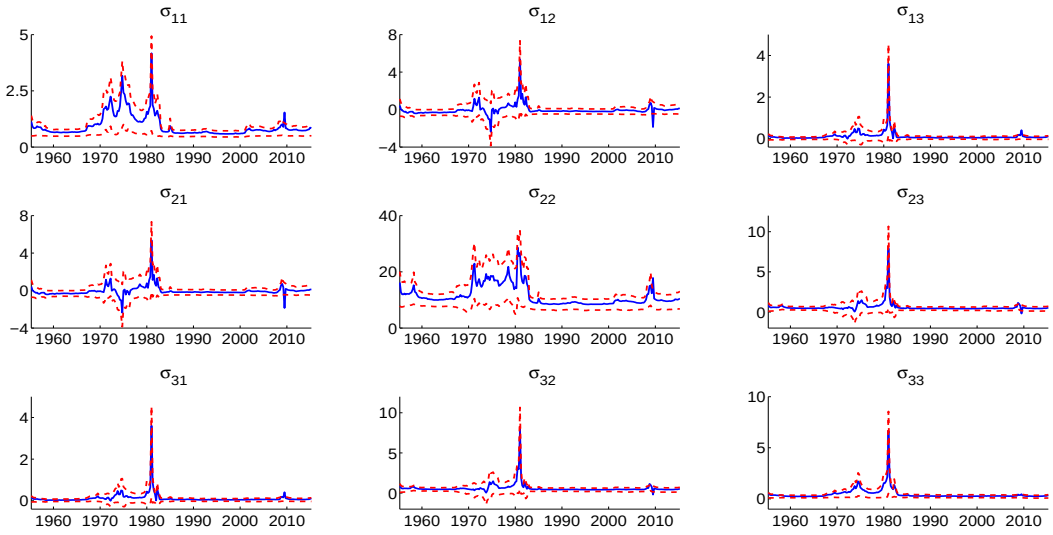

Figure 4.5: The estimated covariance matrix for the IHM-VAR(2).

Figure 4.6 and Figure 4.7 display the estimated posterior distributions of the numbers of the active regimes for the DIHM-VAR(2) and IHM-VAR(2). These figures show that the regime uncertainty has been taken into account by both models. Moreover, it is obvious that the shapes of these posterior distributions are very different from each other. As there is only one infinite hidden Markov process for the IHM-VAR(2) to drive the time-variation of all model parameters, it is not surprising that the IHM-VAR(2) tends to use more regimes to explain the uncertainty of the model

parameters.

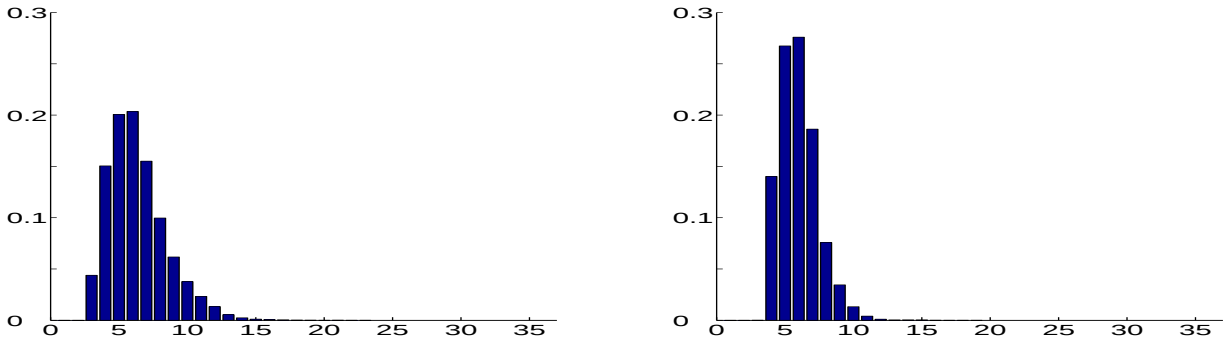

Figure 4.6: Posterior distribution of the numbers of the active regimes of the VAR coefficients (left) and the covariance matrix (right) for the DIHM-VAR(2).

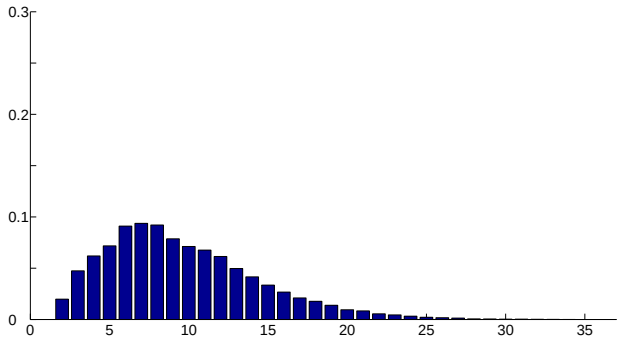

Figure 4.7: Posterior distribution of the numbers of the active regimes for the IHM-VAR(2).

To further investigate on how the data are categorized into different regimes, we plot the $T \times T$ heat maps in Figure 4.8 and Figure 4.9 (see, e.g., Geweke and Jiang [2011]; Song [2014]; Maheu and Yang [2015]). A heat map displays the posterior probabilities that the regime indicators of different time periods have the same value, i.e. $\Pr(s_i = s_j|\mathbf{y}_{1:T})$ (or $\Pr(z_i = z_j|\mathbf{y}_{1:T})$) for $i = 1, \ldots, T$, $j = 1, \ldots, T$. To be specific, the lighter (darker) the colour is in a ceil (i, j) of a heat map, the higher (lower) the probability that the time i and time j are clustered into the same regime. From the heat maps of the DIHM-VAR(2), it is evident that the clustering of regimes for the VAR coefficients and the volatilities are quite different over time. For instance, the right panel of Figure 4.8 shows clearly that the regimes for the volatilities are not likely to be recurrent over time. However, on the left panel of Figure 4.8, there is no visible distinct regime can be found for the VAR coefficients.

To shed light on whether the breaks of the VAR coefficients and volatilities occur at the same time, we plot a weighted heat map for the DIHM-VAR(2), i.e.

$\frac{1}{2}\Pr(s_i = s_j|\mathbf{y}_{1:T}) + \frac{1}{2}\Pr(z_i = z_j|\mathbf{y}_{1:T})$ along side with the heat map associated with the IHM-VAR(2) in Figure 4.9. Again, these two heat maps show very different patterns from each other. It provides strong evidence supporting that the break points of the VAR coefficients and volatilities occur at different time periods. More interestingly, although the estimates for both models indicate a high volatile period in the 1970s, their heat maps tell a very different story. For the DIHM-VAR(2), this highly volatile period is likely to be clustered into one or two regimes. However, for the IHM-VAR(2), this period is categorized into many distinct regimes.

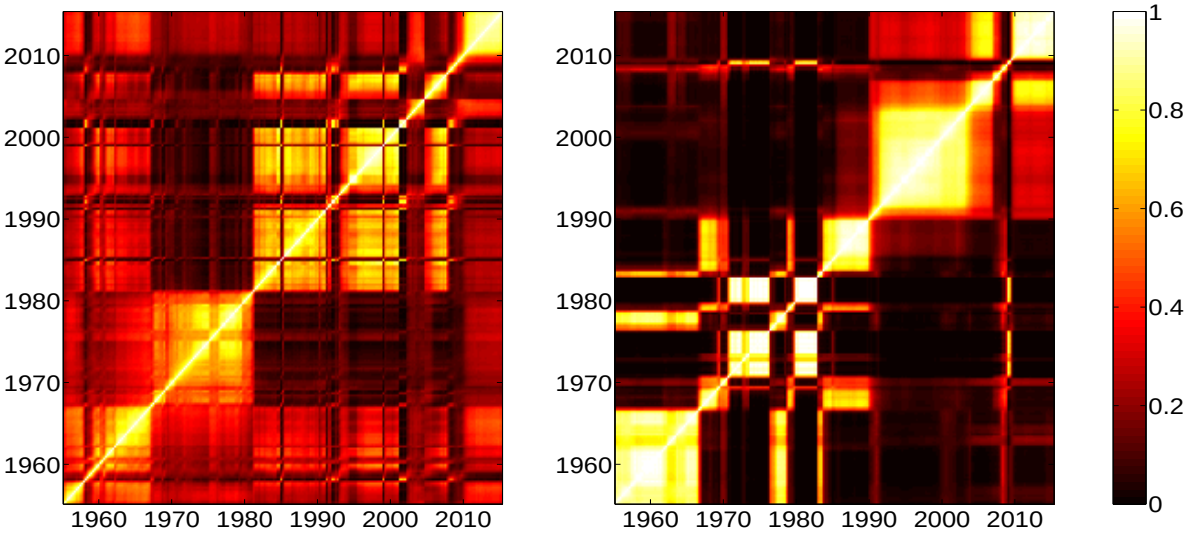

Figure 4.8: The estimated heat maps of the VAR coefficients (left) and the covariance matrix (right) for the DIHM-VAR(2).

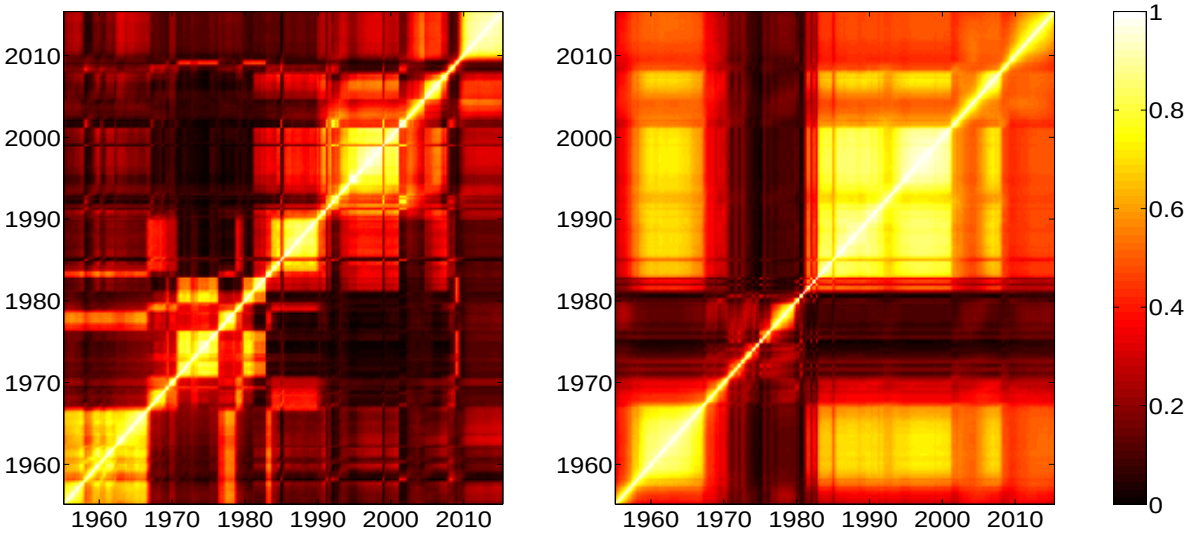

Figure 4.9: The estimated weighted heat map for the DIHM-VAR(2) (left) and the estimated heat map for the IHM-VAR(2) (right).

A few implications can be drawn from the results reported in this section. The estimation results show substantial time-variation in the parameters of both models. The main difference between the DIHM-VAR(2) and IHM-VAR(2) appears in the estimates of volatilities rather than the VAR coefficients. More importantly, it is evident that the breaks for the VAR coefficients and volatilities are not likely to occur at the same time. In addition, it is worth noting that even though some of the estimated parameters for both models seem to share some similarities, but the latent dynamics, as shown in Figure 4.6 - Figure 4.9, are quite different. As various types of IHM model are likely to possess different features, it is important to investigate which type of these models has better forecast performance. To this end, we proceed by conducting a formal recursive out-of-sample forecasting exercise in the next section.

4.5.3 Forecasting Results

This section performs a recursive out-of-sample forecasting exercise to evaluate the performance of the models discussed in section 4.3. We evaluate the iterated h-step-ahead forecasts of each model with $h = 1$ (one quarter), $h = 2$ (two quarters), $h = 4$ (one year), $h = 8$ (two years). Our forecast evaluation period is from 1970Q1 - 2015Q1.

4.5.3.1 Forecast Metrics

Let $\mathbf{y}_{t+h}^o$ be the observed value of $\mathbf{y}_{t+h}$ and $\mathbf{y}_{1:t}$ denotes the data up to time t. We compute the h-step-ahead predictive posterior median $\hat{\mathbf{y}}_{t+h}$ as the point forecast for a given model. The mean absolute forecast error (MAFE) is used to measure the accuracy of the point forecast of a model, which is defined as

$$\text{MAFE} = \frac{1}{T-h-t_0+1} \sum_{t=t_0}^{T-h} \left| \mathbf{y}_{t+h}^o - \hat{\mathbf{y}}_{t+h} \right|.$$

For evaluating the performance of density forecasts, we use the average log-predictive likelihoods (ALPL) which is defined as

$$\text{ALPL} = \frac{1}{T-h-t_0+1} \sum_{t=t_0}^{T-h} \log p_{t+h}(\mathbf{y}_{t+h} = \mathbf{y}_{t+h}^o | \mathbf{y}_{1:t}),$$

where p_{t+h} denotes h-step-ahead predictive density function. $\mathbf{y}_{1:t}$ denotes the given the data up to time t. The log-predictive likelihood is often used to compare forecast performance of models in Bayesian framework since there is a close connection between the predictive likelihood and the marginal likelihood. More discussions about the log-predictive likelihoods can be found in Geweke and Amisano [2011].

To facilitate comparison, the relative scores to the benchmark models, AR(1) for the univariate models and VAR(1) for the multivariate models, are reported. More specifically, we report the ratios of MAFEs of a given model to those of the benchmark. Hence values less than unity indicate better point forecast performance than the benchmark. For density forecasts, we report the difference of ALPLs of a given models to those of the benchmark. Thus positive values indicate better density forecast performance than the benchmark.

4.5.3.2 Point Forecasts

Table 4.2 - Table 4.5 report the relative MAFEs and ALPLs. For each table, the forecasting results for GDP inflation, GDP growth and short-term interest rate are presented in panel a), panel b) and panel c) respectively. Our point forecast results, shown in table 4.2, suggest that allowing for time-variation in model parameters is important in improving the accuracy of point forecasts. None of the models we consider produce sizable and consistent improvements over DIHM-VAR(2) for forecasting GDP inflation, GDP growth and short-term interest rate. Although the IHM-VAR(1) and IHM-VAR(2) tend to outperform the DIHM-VAR(2) on forecasting the

short term GDP inflation, these improvements are not likely to be shared for forecasting the other variables. For example, the IHM-VAR(2) performs relatively worse at forecasting the short-run interest rate. It is worth noting that most of the gains in forecast accuracy appear to have come from allowing for time-variation in the VAR coefficients. This can be shown by comparing the performance of the DIHM-VAR(2) with C-VAR(2)-IHM. For example, both DIHM-VAR(2) and C-VAR(2)-IHM produce almost the same forecast estimates for all variables at various forecast horizons.

For AR models, the patterns of the point forecast results shown in Table 4.3 are broadly similar to those of VARs. The DIHM-AR(2) is always among the top forecasting models. Likewise, most of the forecasting gains appear to have come from allowing for time-variation in volatilities. Even through allowing the AR coefficients to be time-varying tends to improve the point forecast performance, such improvements are likely to be small. These results are in general consistent with those findings in the literature of TVP-SV models [Clark and Ravazzolo, 2014; Chan, 2015a].

Table 4.2: Relative MAFEs; GDP Inflation (panel a)), GDP growth (panel b)), short-term interest rate (panel c)).

	a) GDP inflation				b) GDP growth				c) short-term interest rate			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
VAR(1)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
VAR(2)	0.99	0.93	0.86	0.88	0.99	0.98	1.02	1.01	0.96	1.03	1.01	1.00
DIHM-VAR(1)	0.98	0.93	0.87	0.84	0.99	0.97	0.97	0.98	0.95	0.95	0.96	0.97
DIHM-VAR(2)	0.99	0.89	0.81	0.76	1.00	0.96	0.98	1.00	0.86	0.93	0.91	0.91
IHM-VAR(1)	0.96	0.93	0.87	0.86	0.99	0.97	0.98	1.01	1.01	1.04	0.99	1.01
IHM-VAR(2)	0.95	0.88	0.80	0.79	1.00	0.97	1.02	1.01	0.97	1.07	1.01	1.00
C-VAR(1)-IHM	0.99	0.97	0.95	0.93	0.99	0.98	0.98	0.99	0.98	0.96	0.97	0.94
C-VAR(2)-IHM	0.98	0.89	0.81	0.77	0.99	0.97	0.98	1.00	0.86	0.92	0.91	0.90

Table 4.3: Relative MAFEs; GDP Inflation (panel a)), GDP growth (panel b)), short-term interest rate (panel c)).

	a) GDP inflation				b) GDP growth				c) short-term interest rate			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
AR(1)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR(2)	0.98	0.95	0.95	0.96	1.00	1.00	0.99	1.00	0.90	1.03	1.00	0.99
DIHM-AR(1)	0.99	0.96	0.92	0.89	1.01	1.00	1.00	1.01	0.87	0.93	0.96	0.99
DIHM-AR(2)	0.97	0.92	0.90	0.88	1.00	1.00	0.99	1.01	0.82	0.95	0.95	0.96
IHM-AR(1)	0.98	0.95	0.92	0.86	1.01	1.00	1.00	1.01	0.88	0.95	0.93	0.95
IHM-AR(2)	0.97	0.92	0.90	0.87	1.00	0.97	1.02	1.01	0.83	0.96	0.93	0.91
C-AR(1)-IHM	0.99	0.97	0.94	0.92	1.01	1.00	1.00	1.01	0.96	0.97	0.99	0.99
C-AR(2)-IHM	0.97	0.92	0.92	0.90	1.00	1.00	0.99	1.01	0.82	0.94	0.95	0.97

4.5.3.3 Density Forecasts

We report the results of density forecasts in Table 4.4 and Table 4.5. For both VAR and AR models, it is apparent that models allowing for time-variation in parameters, either in the VAR coefficients or volatilities, substantially outperform the benchmarks at all forecast horizons. Among the VAR models with infinite hidden Markov component(s), the DIHM-VAR(2) performs the best at all horizons in forecasting GDP inflation, GDP growth and short-term interest rate. With AR models, the patterns are mostly the same as those of VAR models. Except in the case of forecasting short-term interest rate, the DIHM-AR(2) and IHM-AR(2) perform quite similarly at short horizons, however, the IHM-AR(2) tends to be forecasting better than the DIHM-VAR(2) at longer horizons.

Table 4.4: Relative ALPLs; GDP Inflation (panel a)), GDP growth (panel b)), short-term interest rate (panel c)).

	a) GDP inflation				b) GDP growth				c) short-term interest rate			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
VAR(1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VAR(2)	0.02	0.05	0.12	0.17	-0.01	0.01	-0.02	-0.01	0.01	0.00	0.05	0.06
DIHM-VAR(1)	0.11	0.23	0.30	0.40	0.08	0.11	0.10	0.07	0.58	0.46	0.24	0.09
DIHM-VAR(2)	0.13	0.26	0.33	0.40	0.09	0.13	0.11	0.07	0.69	0.49	0.26	0.15
IHM-VAR(1)	0.12	0.20	0.26	0.35	0.04	0.05	0.04	0.03	0.39	0.32	0.16	0.11
IHM-VAR(2)	0.12	0.22	0.29	0.37	0.04	0.06	0.02	0.03	0.44	0.32	0.17	0.10
C-VAR(1)-IHM	0.10	0.18	0.19	0.24	0.08	0.11	0.10	0.06	0.47	0.35	0.16	0.09
C-VAR(2)-IHM	0.13	0.26	0.31	0.37	0.08	0.12	0.11	0.07	0.67	0.47	0.24	0.14

Table 4.5: Relative ALPLs; GDP Inflation (panel a)), GDP growth (panel b)), short-term interest rate (panel c)).

	a) GDP inflation				b) GDP growth				c) short-term interest rate			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
AR(1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AR(2)	-0.01	-0.10	0.04	-0.09	0.01	0.00	0.01	0.00	-0.11	-0.05	1.24	1.56
DIHM-AR(1)	0.46	0.47	0.52	0.93	0.21	0.20	0.20	0.18	1.50	1.77	2.31	2.65
DIHM-AR(2)	0.49	0.51	0.59	0.98	0.23	0.21	0.20	0.17	1.55	1.76	2.31	2.66
IHM-AR(1)	0.47	0.47	0.53	0.97	0.20	0.18	0.16	0.15	1.50	1.77	2.35	2.72
IHM-AR(2)	0.48	0.48	0.56	0.95	0.21	0.18	0.18	0.16	1.54	1.77	2.36	2.74
C-AR(1)-IHM	0.48	0.49	0.55	0.93	0.21	0.20	0.20	0.18	1.38	1.65	2.17	2.55
C-AR(2)-IHM	0.48	0.49	0.55	0.93	0.23	0.21	0.20	0.17	1.51	1.71	2.25	2.61

To investigate the forecast performance of IHM models at different period of time, we plot the one-quarter-ahead cumulative sums of log-predictive likelihoods in Figure 4.10 - Figure 4.15. In each of these figures, we plot the relative cumulative log-predictive likelihoods for the DIHM-VAR(2), IHM-VAR(2) and C-VAR(2)-IHM to the

VAR(1) benchmark on the left panel. Furthermore, to investigate whether allowing for time-variation in the conditional mean coefficients helps improving forecast accuracy, we also plot the relative cumulative likelihoods for the DIHM-VAR(2) to the C-VAR(2)-IHM on the right panel of each figure. Similar plots for the AR models are shown in Figure 4.13 - Figure 4.12.

Overall, models incorporating any infinite hidden Markov components consistently dominate the benchmarks over the whole evaluation period in forecasting all variables. Except for forecasting GDP growh, the DIHM-VAR(2) and IHM-VAR(2) only outperform the benchmark after the early 1990s. In addition, a few broad conclusions can be drawn from the results. First, models restricting the breaks of the VAR coefficients and volatilities to occur at the same time, i.e. IHM-VAR(2) and IHM-AR(2), forecast poorer than the other variants of IHM models for most periods (except for forecasting GDP growth — the DIHM-VAR(2) and C-VAR(2)-IHM only outperform the IHM-VAR(2) after the early 1990s). Second, models allowing for time-variation in the VAR coefficients tend to improve rather than harm forecast accuracy. This can be seen from the right panel of Figure 4.10 - Figure 4.15. For example, the DIHM-VAR(2) and IHM-VAR(2) produce very similar one-quarter-ahead cumulative sums of log-predictive likelihoods for forecasting GDP inflation and GDP growth, however in the case of forecasting short-term interest rate, the DIHM-VAR(2) substantially outperforms the IHM-VAR(2), especially in the late 2000s.

It is important to highlight that a model including an infinite hidden Markov component does not guarantee a change in the model parameters. Instead, it only allows for potential changes to occur. This is very different from the TVP-SV models in which the parameters are always varying over time. One can think of the IHM models as an intermediate case of the constant parameters models and TVP-SV models. They ameliorate the potential over-parameterization concern of the TVP-SV models, while allowing for a more flexible framework than the constant parameters models in capturing parameters instabilities. Our forecasting results provide empirical support for this view. It is evident that the infinite hidden Markov component can act as a safeguard for structural breaks which is likely to improve rather than harm forecast accuracy. Moreover, models using a single infinite hidden Markov process to drive the changes of all parameters forecast poorer than the alternative versions in which two independent infinite hidden Markov processes separately govern the VAR coefficients and volatilities.

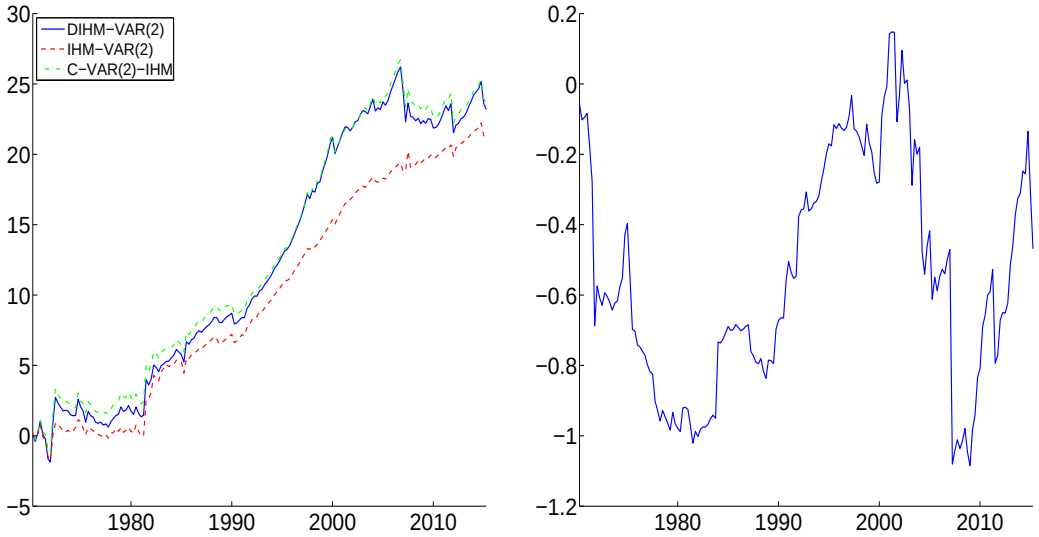

Figure 4.10: Cumulative log-predictive likelihoods for one-quarter-ahead forecasts of DIHM-VAR(2), IHM-VAR(2) and C-VAR(2)-IHM relative to VAR(1) (left panel), DIHM-VAR(2) relative to C-VAR(2)-IHM (right panel), for GDP inflation.

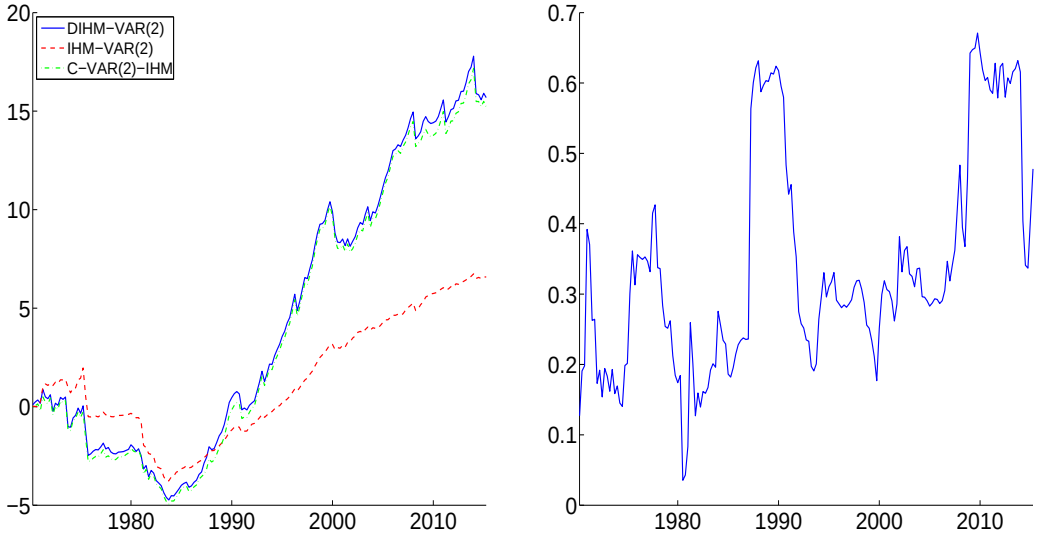

Figure 4.11: Cumulative log-predictive likelihoods for one-quarter-ahead forecasts of DIHM-VAR(2), IHM-VAR(2) and C-VAR(2)-IHM relative to VAR(1) (left panel), DIHM-VAR(2) relative to C-VAR(2)-IHM (right panel), for GDP growth.

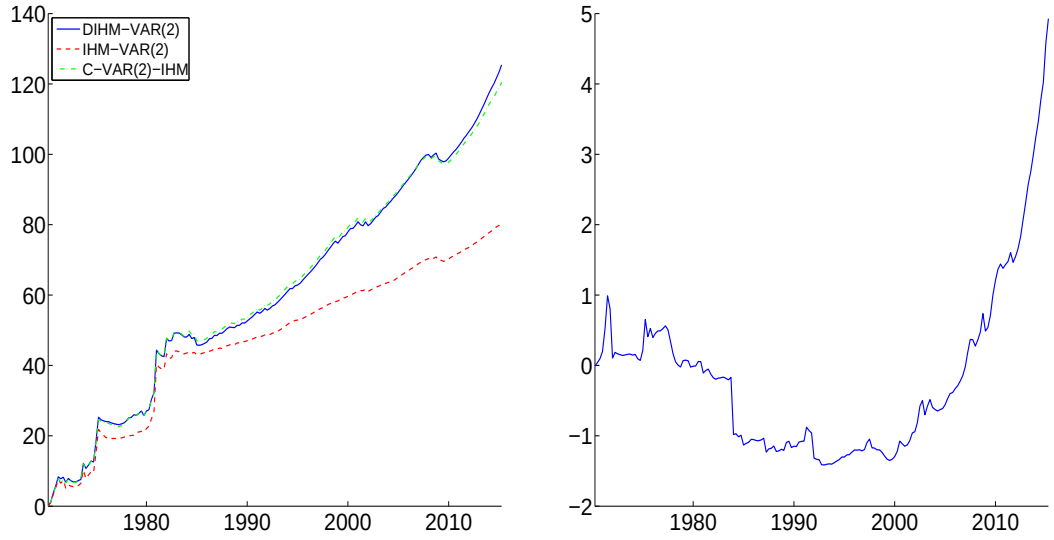

Figure 4.12: Cumulative log-predictive likelihoods for one-quarter-ahead forecasts of DIHM-VAR(2), IHM-VAR(2) and C-VAR(2)-IHM relative to VAR(1) (left panel), DIHM-VAR(2) relative to C-VAR(2)-IHM (right panel), for short-term interest rate.

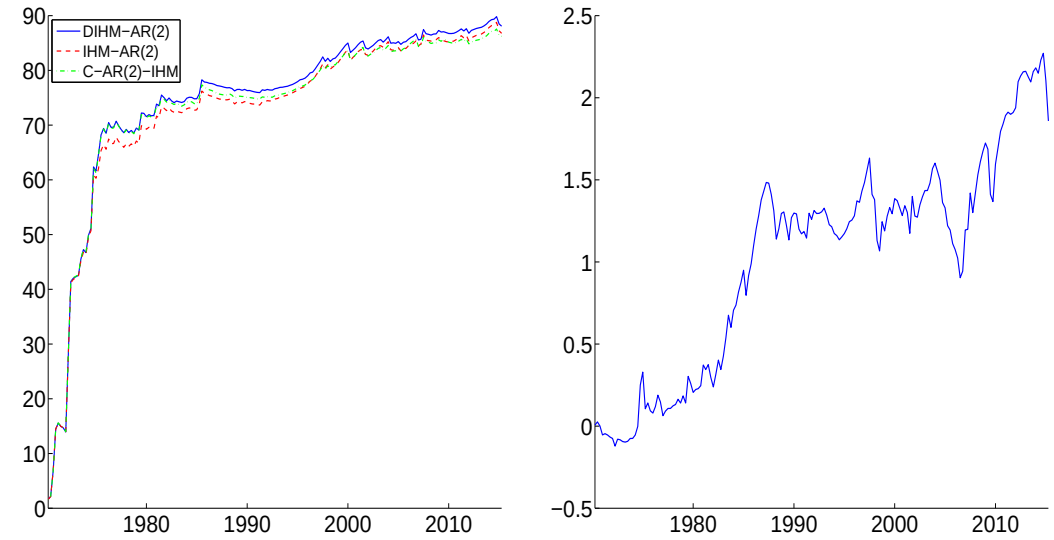

Figure 4.13: Cumulative log-predictive likelihoods for one-quarter-ahead forecasts of DIHM-AR(2), IHM-AR(2) and C-AR(2)-IHM relative to AR(1) (left panel), DIHM-AR(2) relative to C-AR(2)-IHM (right panel), for GDP inflation.

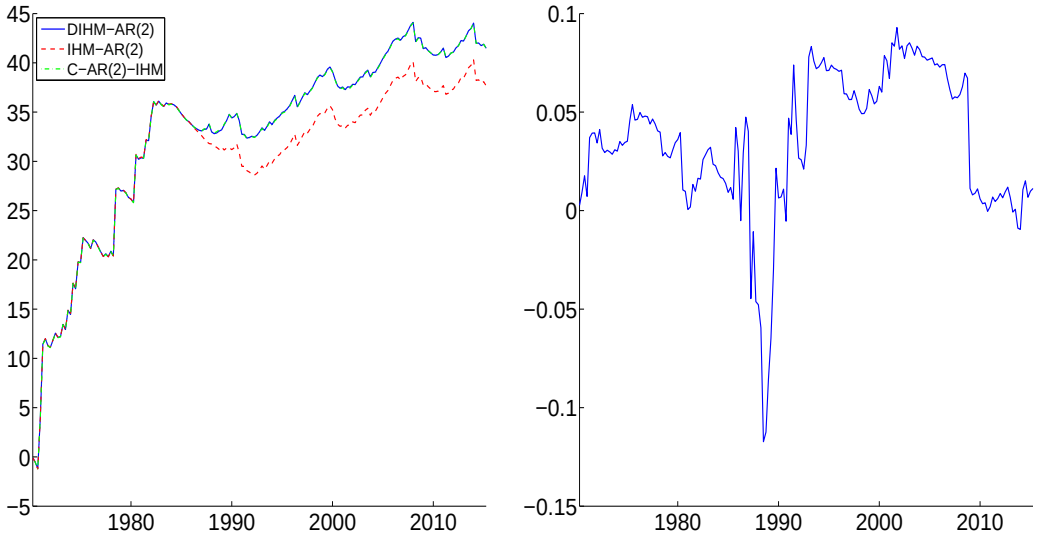

Figure 4.14: Cumulative log-predictive likelihoods for one-quarter-ahead forecasts of DIHM-AR(2), IHM-AR(2) and C-AR(2)-IHM relative to AR(1) (left panel), DIHM-AR(2) relative to C-AR(2)-IHM (right panel), for GDP growth.

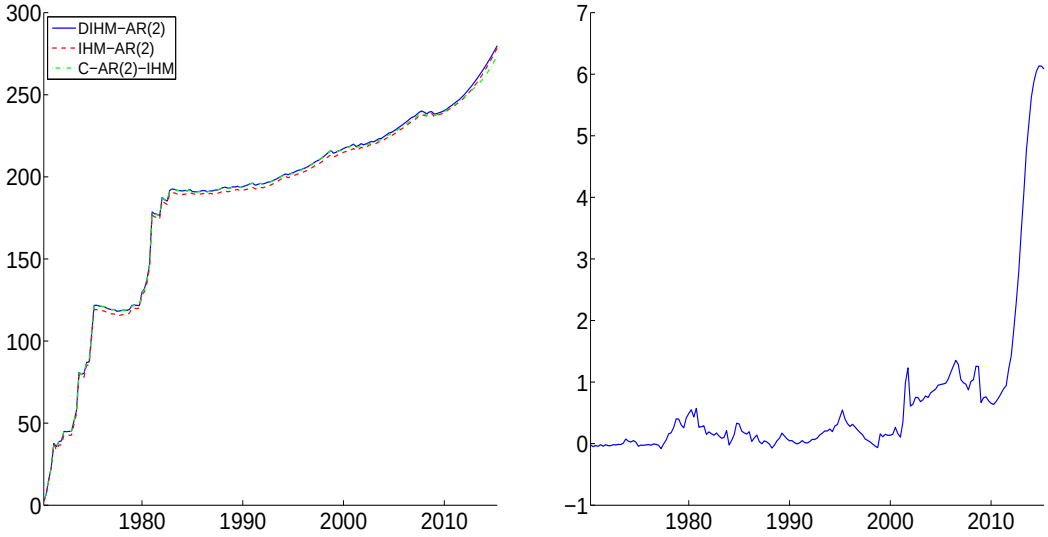

Figure 4.15: Cumulative log-predictive likelihoods for one-quarter-ahead forecasts of DIHM-AR(2), IHM-AR(2) and C-AR(2)-IHM relative to AR(1) (left panel), DIHM-AR(2) relative to C-AR(2)-IHM (right panel), for short-term interest rate.

4.6 Conclusion

This chapter compares the forecast performance of various specifications of infinite hidden Markov switching models. We first extend this model into the multivariate framework, and then propose an algorithm to improve the computational efficiency. The results of our recursive out-of-sample forecast exercise show that allowing for time-variation in model parameters is important in improving forecast accuracy, especially in density forecast. We find that none of the alternative specifications of infinite hidden Markov switching models consistently outperforms the model with two independent infinite hidden Markov process: one governs the changes of the conditional mean coefficients and one governs the changes of the volatilities. Models assuming a single infinite hidden Markov process which drives all model parameters forecasts poorer than the other specifications of infinite hidden Markov switching models. Furthermore, models incorporating an infinite hidden Markov component are tending to improve rather than harm the forecast accuracy.

Appendix

In this Appendix we provide the details of the posterior sampler stated in section 4.4 for model DIHM-VAR(q). Posterior draws can be obtained by sequentially sampling from Step 1. - Step 7. The sampling method is mainly based on Song [2014]. We first introduce some notation, let $\mathbf{p}_i^{s\prime} = (p_{i1}^s, \ldots, p_{iL}^s, p_{iL}^{sR})$ and $\boldsymbol{\pi}^s = (\pi_1^s, \ldots, \pi_L^s, \pi_L^{sR})$, where $p_{iL}^{sR} = \sum_{j=L+1}^{\infty} p_{ij}^s$ and $\pi_L^{sR} = \sum_{j=L+1}^{\infty} \pi_j^s$, with similar notation applying to $\mathbf{p}_i^{z\prime}$ and $\boldsymbol{\pi}^z$. An efficient way to obtain draws in Step 1. has been provided in the main content.

Step 2. Sample $\boldsymbol{\Phi}$

- $\boldsymbol{\beta}_0 \sim \mathcal{N}(\hat{\boldsymbol{\beta}}_0, \mathbf{V}_{\boldsymbol{\beta}_0})$, where

$$\mathbf{V}_{\boldsymbol{\beta}_0} = \left(L_s \mathbf{V}_0^{-1} + \mathbf{V}_{00}^{-1}\right)^{-1}, \quad \hat{\boldsymbol{\beta}}_0 = \mathbf{V}_{\boldsymbol{\beta}_0}(\mathbf{V}_0^{-1} \sum_{i=1}^{L_s} \boldsymbol{\beta}_i + \mathbf{V}_{00}^{-1} \boldsymbol{\beta}_{00})$$

- $\mathbf{V}_0^{-1} \sim \mathcal{W}(\widehat{\mathbf{A}}_{00}, \hat{a}_{00})$, where

$$\widehat{\mathbf{A}}_{00} = \left(\mathbf{A}_{00}^{-1} + \sum_{i=1}^{L_s} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_0)(\boldsymbol{\beta}_i - \boldsymbol{\beta}_0)'\right)^{-1}, \quad \hat{a}_{00} = a_{00} + L_s$$

- $\mathbf{\Sigma}_0 \sim \mathcal{W}(\widehat{\mathbf{Q}}_{00}, \hat{b}_{00})$, where

$$\widehat{\mathbf{Q}}_{00} = \left(\sum_{i=1}^{L_z} \mathbf{\Sigma}_i^{-1} + \mathbf{Q}_{00}^{-1} \right)^{-1}, \quad \hat{b}_{00} = b_{00} + L_z \nu_0$$

- the conditional distribution $p(\nu_0 | \mathbf{\Sigma}_0, \mathbf{\Sigma}_{1:L_z}, \lambda_{00})$ is non-standard, a Metropolis-Hastings step is implemented. Given the previous ν_0^o the current draw ν_0^c from the proposal distribution $q(\nu_0^c | \nu_0^o) \sim \mathcal{G}(\chi, \chi/\nu_0^o)$ is accepted with probability

$$\min \left\{ 1, \frac{p(\nu_0^c | \mathbf{\Sigma}_0, \mathbf{\Sigma}_{1:L_z}, \lambda_{00}) q(\nu_0^o | \nu_0^c)}{p(\nu_0^o | \mathbf{\Sigma}_0, \mathbf{\Sigma}_{1:L_z}, \lambda_{00}) q(\nu_0^c | \nu_0^o)} \mathbb{1}(\nu_0 > n) \right\},$$

  where the χ is a tuning parameter for adjusting the acceptant rate to be around 50%.

Step 3. Sample $\mathbf{u}^s, \mathbf{u}^z$

- $u_1^s \sim \mathbf{U}(0, \pi_{s_1}^s)$ and $u_t^s \sim \mathbf{U}(0, p_{s_{t-1}, s_t}^s)$ for $t = 2, \ldots, T$;
- $u_1^z \sim \mathbf{U}(0, \pi_{z_1}^z)$ and $u_t^z \sim \mathbf{U}(0, p_{z_{t-1}, z_t}^z)$ for $t = 2, \ldots, T$.

Check and expand $\boldsymbol{\pi}^s$ and $\mathbf{P}^s$ as follow: If

$$\max\{p_{1\bar{L}}^{sR}, \ldots, p_{\bar{L}\bar{L}}^{sR}\} > \min\{u_1^s, \ldots, u_T^s\} \tag{4.44}$$

then update $\bar{L}$ as follow

3.1 draw $\xi \sim \mathcal{B}(1, \gamma^s)$, and update

$$\boldsymbol{\pi}^s = (\pi_1^s, \ldots, \pi_{\bar{L}}^s, \xi \pi_{\bar{L}}^{sR}, (1-\xi)\pi_{\bar{L}}^{sR}) \equiv (\pi_1^s, \ldots, \pi_{\bar{L}}^s, \pi_{\bar{L}+1}^s, \pi_{\bar{L}+1}^{sR}),$$

3.2 For $i = 1, \ldots, \bar{L} - 1$, draw $\xi_i \sim \mathcal{B}(c^s(1-\rho^s)\pi_{\bar{L}+1}^s, c^s(1-\rho^s)\pi_{\bar{L}+1}^{sR})$ then update

$$\mathbf{p}_i^{s\prime} = (p_{i1}^s, \ldots, \xi_i p_{i\bar{L}}^s, (1-\xi_i) p_{i\bar{L}}^{sR}) \equiv \mathbf{p}_i^{s\prime} = (p_{i1}^s, \ldots, p_{i\bar{L}+1}^s, p_{i\bar{L}+1}^{sR}).$$

3.3 Draw

$$\mathbf{p}_{\bar{L}+1}^{s\prime} \equiv (p_{i1}^s, \ldots, p_{i\bar{L}+1}^s, p_{i\bar{L}+1}^{sR}) \sim \mathcal{D}(c^s(1-\rho^s)\pi_1^s, \ldots, c^s(1-\rho^s)\pi_{\bar{L}+1}^s + c^s\rho^s, c^s(1-\rho^s)\pi_{\bar{L}+1}^{sR}).$$

3.4 Set $\bar{L} = \bar{L} + 1$.

Repeat 3.1 - 3.4 until (4.44) is not hold. Updating of $\boldsymbol{\pi}^z$ and $\mathbf{P}^z$ are similar as those of $\boldsymbol{\pi}^s$ and $\mathbf{P}^s$.

For Step 4. - Step 7, we only provide sampling procedure for $\mathbf{s}, \boldsymbol{\pi}^s, \mathbf{P}^s, \mathbf{H}^s$ and only minor modification is required for sampling $\mathbf{z}, \boldsymbol{\pi}^z, \mathbf{P}^z, \mathbf{H}^z$.

Step 4. Sample $\mathbf{s}$

Given $\mathbf{u}^s$, there are only finitely many trajectories $\mathbf{s}$ with non-zeros probability, i.e. $s_t \in \{1, \dots, \bar{L}\}$. Hence Chib [1996] algorithm can be applied to obtain as follow

1) we first compute the conditional density of s_1 given u_1^s which is given by

$$p(s_1|u_1^s, \boldsymbol{\Theta}, \mathbf{P}^s) = p(s_1|u_1^s) \propto \mathbb{1}(u_1^s < \pi_{s_1}^s) \text{ for } s_t = 1, \dots, \bar{L};$$

2) given $p(s_{t-1}|\mathbf{y}_{1:t-1}, u_{1:t-1}^s, \boldsymbol{\Theta}, P^s)$, we compute $p(s_t|\mathbf{y}_{1:t}, u_{1:t}^s, \mathbf{P}^s, \boldsymbol{\Theta})$ by

$$\begin{aligned}
&p(s_t|\mathbf{y}_{1:t}, u_{1:t}^s, \mathbf{P}^s, \boldsymbol{\Theta}) \\
\propto &p(s_t, u_t^s, \mathbf{y}_{1:t}|\mathbf{y}_{1:t-1}, u_{1:t-1}^s, \boldsymbol{\Theta}, \mathbf{P}^s) \\
= &\sum_{s_{t-1}} p(\mathbf{y}_t, |s_t, \boldsymbol{\Theta}) p(u_t^s|s_t, s_{t-1}, \mathbf{P}^s) p(s_t|s_{t-1}, \mathbf{P}^s) p(s_{t-1}|\mathbf{y}_{1:t-1}, u_{1:t-1}^s, \boldsymbol{\Theta}, \mathbf{P}^s) \\
= &p(\mathbf{y}_t|s_t, \boldsymbol{\Theta}) \sum_{s_{t-1}} \mathbb{1}(u_t^s < p_{s_{t-1}s_t}^s) p(s_{t-1}|\mathbf{y}_{1:t-1}, u_{1:t-1}^s, \boldsymbol{\Theta}, \mathbf{P}^s)
\end{aligned}$$

until we get $p(s_T|\mathbf{y}_{1:T}, u_{1:T}^s, \boldsymbol{\Theta}, \mathbf{P}^s)$.

3) The backward sampling step can be implemented by first sample s_T from $p(s_T|\mathbf{y}_{1:T}, u_{1:T}^s, \boldsymbol{\Theta}, \boldsymbol{P}^s)$, then given s_{t+1}, sample s_t can be obtained from

$$\begin{aligned}
p(s_t|s_{t+1}, \mathbf{u}^s, \boldsymbol{\Theta}, \mathbf{P}^s, \mathbf{y}_{1:T}) &\propto p(s_t|\mathbf{y}_{1:t}, \mathbf{u}^s, \boldsymbol{\Theta}, \mathbf{P}^s) p(s_{t+1}|s_t, u_{t+1}) \\
&= p(s_t|\mathbf{y}_{1:t}, \mathbf{u}^s, \boldsymbol{\Theta}, \boldsymbol{P}_s) \mathbb{1}(u_{t+1}^s < p_{s_t s_{t+1}}^s).
\end{aligned}$$

After sampling $\mathbf{s}$ we find the number of active regime L_s and relabel $\mathbf{s}$, $\mathbf{P}^s$ and $\boldsymbol{\Theta}$ accordingly. Reconstruct $\mathbf{P}^s, \boldsymbol{\pi}^s$ by collapsing the non-active regimes and set $\bar{L} = L_s$.

Step 5. - Step 6. Sample $(\boldsymbol{\pi}^s, \mathbf{P}^s)$

To obtain sample $\boldsymbol{\pi}^s$ and $\mathbf{P}^s$ we use the algorithm proposed by Song [2014]. Three auxiliary variables (I_t, I_t', I_t'') are introduced to facilitate sampling. The output (I_t, I_t', I_t'') are also used to sample $\mathbf{H}^s$.

1. initialize $m = (m_1, \dots, m_{L_s})$ where $m_j = 1$ if $s_t = j$ and 0 otherwise. let n be $L_s \times L_s$ zeros matrix .

2. If $s_t \neq s_l$ for all $l = 1, \ldots, t-1$, set $(I_t, I_t', I_t'') = (1, 1, 1)$;
3. else
   - If $s_t = s_{t-1}$, sample $(I_t, I_t', I_t'') = \begin{cases} (0,0,0) \propto n_{s_{t-1}s_t}, \\ (1,0,0) \propto c^s \rho^s, \\ (1,1,0) \propto c^s(1-\rho^s)\frac{m_{s_t}}{\gamma^s + \sum_j m_j}, \end{cases}$
   - If $s_t \neq s_{t-1}$, sample $(I_t, I_t', I_t'') = \begin{cases} (0,0,0) \propto n_{s_{t-1}s_t}, \\ (1,0,0) \propto c^s(1-\rho^s)\frac{m_{s_t}}{\gamma^s + \sum_j m_j}, \end{cases}$
4. update $m_{s_t} = m_{s_t} + 1$ if $I_t' = 1$. Update $n_{s_{t-1}s_t} = n_{s_{t-1}s_{s_t}} + 1$.
5. repeat 2. – 3. for $t = 2, \ldots, T$.

Draw $\boldsymbol{\pi}^s \sim \mathcal{D}(m_1, \ldots, m_{L_s}, \gamma^s)$.

For $i = 1, \ldots, L_s$, draw

$$\mathbf{p}_i^{s\prime} \sim \mathcal{D}(c^s(1-\rho^s)\pi_1^s + n_{i1}, \ldots, c^s(1-\rho^s)\pi_i^s + c^s\rho^s + n_{ii}, \ldots, c^s(1-\rho^s)\pi_{L_s}^s + n_{iL_s}, c^s(1-\rho^s)\pi_{\bar{L}_s}^s).$$

Step 7. Sample $\mathbf{H}^s$

Given $(I_{1:T}, I_{1:T}', I_{1:T}'')$ obtained Step 4, we applied the method of Song [2014]. Samples from H_s can be obtained by

- Sample c^s
  1. we first compute $a_i = \sum_{t=1}^{T-1}(s_t = i)I_{t+1} + \mathbb{1}(s_1 = i)$ and $b_i = \sum_{t=1}^{T}\mathbb{1}(s_t = i)$ for $i = 1, \ldots, L_s$;
  2. sample $x_i \sim \mathcal{B}(c^s, b_i)$ for $i = 1, \ldots, L_s$;
  3. $c^s \sim \mathcal{G}(\omega_s + \sum_{i=1}^{L_s} a_i, \theta_s - \sum_{i=1}^{L_s}\log(x_i))$.
- Sample ρ^s
  1. compute $a' = \sum_{t=2}^{T} I_t(1 - I_t')$ and $b' = \sum_{t=2}^{T} I_t I_t'$;
  2. sample $\rho^s \sim \mathcal{B}(f_s + a', g_s + b')$.
- Sample γ^s
  1. compute $a'' = \sum_{t=1}^{T} I_t I_t'$;
  2. sample $x \sim \mathcal{G}(\gamma^s, a'')$;
  3. sample $\gamma^s \sim \mathcal{G}(h_s + L_s, \eta_s - \log(x))$.

Chapter 5

Conclusion

This thesis explores the performance of various types of Bayesian mixture models in macroeconomic analysis. Many studies have already highlighted the empirical importance of allowing for time-variation in model parameters. The main focus of this thesis is refining our understanding in modeling time-varying parameter processes to improve model performance in different empirical applications.

In Chapter 2, we investigate the G7 inflation forecast performance of autoregressive models with different error distributional assumptions. Our results show that both the heavy-tailed distributed errors and the stochastic volatility component are more likely to improve rather than harm forecast accuracy. These two components can help reduce the influence of outliers and capture the time varying volatility when these properties are in the data. Naturally, there is little improvement by including these two components on forecasting data that seemingly exhibit constant volatility and has few outliers.

Chapter 3 proposes a new mixture innovation model to re-examine the relationship between inflation and inflation uncertainty for the US, Germany, Canada and New Zealand. A simulation study shows that the proposed model produces more robust estimates compared with the conventional benchmark. Our empirical results show substantial time-variation in the relationship between inflation and inflation uncertainty over the last few decades for all countries. We also find strong evidence of the existence of abrupt changes in this relationship. From the results for the inflation targeting countries, there is evidence that the correlation between inflation and inflation uncertainty has been weakened since the adoption of inflation targeting policy.

Chapter 4 compares forecasting performance of univariate and multivariate infinite hidden Markov switching models. In addition, we develop a new MCMC method built upon the precision-based algorithm to improve computational efficiency. Our

results show that the conditional mean coefficients and volatilities are likely to be driven by independent infinite hidden Markov processes. Models assuming a single infinite hidden Markov process that drives all model parameters forecast poorer than more flexible specifications of allowing multiple hidden Markov states.

In sum, we conclude that although many recent macroeconomic studies have demonstrated the empirical success of standard time-varying parameter models, the task of carefully selecting a proper time-varying process remains an ongoing task in various applications. It requires researchers to grasp many recent emerging econometric techniques, and to have a deep understanding of the context of the empirical application.

Bibliography

Ball, L., 1992. Why does high inflation raise inflation uncertainty? *Journal of Monetary Economics*, 29, 3 (1992), 371–388. (cited on page 33)

Bauwens, L.; Carpantier, J.-F.; Dufays, A.; et al., 2015. Autoregressive moving average infinite hidden markov-switching models. *Journal of Business and Economic Statistics*, , forecoming (2015). (cited on page 60)

Bauwens, L.; Koop, G.; Korobilis, D.; and Rombouts, J. V., 2014. The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, (2014). (cited on page 60)

Berument, H.; Yalcin, Y.; and Yildirim, J., 2009. The effect of inflation uncertainty on inflation: Stochastic volatility in mean model within a dynamic framework. *Economic Modelling*, 26, 6 (2009), 1201–1207. (cited on pages 34 and 44)

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31, 3 (1986), 307–327. (cited on page 3)

Chan, J. C., 2013. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176, 2 (2013), 162–172. (cited on pages 3, 4, 10, and 61)

Chan, J. C., 2015a. Large Bayesian VARs: A flexible kronecker error covariance structure. *Available at SSRN 2688342*, (2015). (cited on pages 61 and 79)

Chan, J. C., 2015b. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business and Economic Statistics*, (2015). Forthcoming. (cited on pages 4, 22, 34, 35, 36, 37, 40, 42, 44, and 52)

Chan, J. C. and Eisenstat, E., 2015. Bayesian model comparison for time-varying parameter VARs with stochastic volatility. Technical report, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University. (cited on pages 4, 59, 65, and 72)

Chan, J. C. and Hsiao, C. Y., 2013. Estimation of stochastic volatility models with heavy tails and serial dependence. *Bayesian Inference in the Social Sciences. John Wiley and Sons, New york*, (2013). (cited on pages 4, 15, and 16)

Chan, J. C. and Jeliazkov, I., 2009. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 1-2 (2009), 101–120. (cited on pages 41, 54, 61, and 69)

Chan, J. C. and Strachan, R., 2014. The zero lower bound: Implications for modelling the interest rate. (2014). Working Paper. (cited on pages 40 and 52)

Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75, 1 (1996), 79–97. (cited on pages 67 and 87)

Chib, S., 1998. Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86, 2 (1998), 221–241. (cited on page 60)

Chib, S.; Nardari, F.; and Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108, 2 (2002), 281–316. (cited on page 4)

Clark, T. E., 2009. Is the Great Moderation over? An empirical analysis. *Economic Review*, 2009 (2009), Q4–5. (cited on page 3)

Clark, T. E., 2011. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29, 3 (2011). (cited on pages 4, 22, 59, and 61)

Clark, T. E. and Doh, T., 2011. A Bayesian evaluation of alternative models of trend inflation. *Federal Reserve Bank of Cleveland, working paper*, , 11-34 (2011). (cited on page 3)

Clark, T. E. and Ravazzolo, F., 2014. Macroeconomic forecasting performance under alternative specification of time-varying volatility. *Journal of Applied Econometrics*, (2014). (cited on pages 3, 4, 8, 10, 15, 16, 22, 23, 59, 61, and 79)

Cogley, T. and Sargent, T. J., 2002. Evolving post-World War II US inflation dynamics. In *NBER Macroeconomics Annual 2001, Volume 16*, 331–388. MIT Press. (cited on pages 34 and 59)

Cogley, T. and Sargent, T. J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic dynamics*, 8, 2 (2005), 262–302. (cited on pages 3, 10, 16, and 59)

Cukierman, A. and Meltzer, A. H., 1986. A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. *Econometrica: Journal of the Econometric Society*, (1986), 1099–1128. (cited on pages 33, 34, 46, and 47)

D'Agostino, A.; Gambetti, L.; and Giannone, D., 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28, 1 (2013), 82–101. (cited on pages 59 and 61)

Eisenstat, E. and Strachan, R. W., 2015. Modelling inflation volatility. *Journal of Applied Econometrics*, (2015). (cited on page 3)

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, (1982), 987–1007. (cited on page 3)

Ferguson, T. S., 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, (1973), 209–230. (cited on page 62)

Fonseca, T. C.; Ferreira, M. A.; and Migon, H. S., 2008. Objective bayesian analysis for the student-t regression model. *Biometrika*, 95, 2 (2008), 325–333. (cited on pages 4, 9, 15, and 27)

Fox, E. B.; Sudderth, E. B.; Jordan, M. I.; and Willsky, A. S., 2011. A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, (2011), 1020–1056. (cited on pages 61, 63, and 67)

Friedman, M., 1977. Nobel lecture: Inflation and unemployment. *The Journal of Political Economy*, (1977), 451–472. (cited on page 33)

Gerlach, R.; Carter, C.; and Kohn, R., 2000. Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, 95, 451 (2000), 819–828. (cited on pages 34 and 37)

Geweke, J., 1993. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, 8, S1 (1993), S19–S40. (cited on pages 4, 8, 15, and 16)

Geweke, J., 2007. Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics and Data Analysis*, 51, 7 (2007), 3529–3550. (cited on page 71)

Geweke, J. and Amisano, G., 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26, 2 (2010), 216–230. (cited on page 17)

Geweke, J. and Amisano, G., 2011. Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26, 1 (2011), 1–29. (cited on pages 50, 59, and 78)

Geweke, J. and Jiang, Y., 2011. Inference and prediction in a multiple-structural-break model. *Journal of Econometrics*, 163, 2 (2011), 172–185. (cited on page 75)

Giordani, P. and Kohn, R., 2008. Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business and Economic Statistics*, 26, 1 (2008). (cited on page 34)

Giordani, P. and Kohn, R., 2012. Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business and Economic Statistics*, (2012). (cited on page 60)

Gneiting, T. and Ranjan, R., 2012. Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business and Economic Statistics*, (2012). (cited on page 18)

Grier, K. B. and Perry, M. J., 1998. On inflation and inflation uncertainty in the G7 countries. *Journal of International Money and Finance*, 17, 4 (1998), 671–689. (cited on pages 33 and 34)

Grier, K. B. and Perry, M. J., 2000. The effects of real and nominal uncertainty on inflation and output growth: some GARCH-M evidence. *Journal of Applied Econometrics*, 15, 1 (2000), 45–58. (cited on pages 34 and 45)

Holland, A. S., 1995. Inflation and uncertainty: Tests for temporal ordering. *Journal of Money, Credit and Banking*, (1995), 827–837. (cited on pages 33, 34, and 44)

Hubrich, K. and Tetlow, R. J., 2015. Financial stress and economic dynamics: The transmission of crises. *Journal of Monetary Economics*, 70 (2015), 100–115. (cited on pages 59 and 60)

Jacquier, E.; Polson, N. G.; and Rossi, P. E., 2002. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 20, 1 (2002), 69–87. (cited on page 3)

Jacquier, E.; Polson, N. G.; and Rossi, P. E., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122, 1 (2004), 185–212. (cited on pages 4 and 15)

Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186, 1007 (1946), 453–461. (cited on page 15)

Jeliazkov, I., 2008. Specification and inference in nonparametric additive regression. (2008). (cited on pages 61 and 68)

Jensen, M. J. and Maheu, J. M., 2014. Estimating a semiparametric asymmetric stochastic volatility model with a dirichlet process mixture. *Journal of Econometrics*, 178 (2014), 523–538. (cited on page 60)

Jochmann, M., 2015. Modeling US inflation dynamics: A Bayesian nonparametric approach. *Econometric Reviews*, 34, 5 (2015), 537–558. (cited on pages 60, 64, and 67)

Jochmann, M. and Koop, G., 2015. Regime-switching cointegration. *Studies in Nonlinear Dynamics and Econometrics*, 19, 1 (2015), 35–48. (cited on page 60)

Kass, R. E. and Raftery, A. E., 1995. Bayes factors. *Journal of the American statistical association*, 90, 430 (1995), 773–795. (cited on page 50)

Kass, R. E. and Wasserman, L., 1996. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91, 435 (1996), 1343–1370. (cited on page 15)

Kim, C.-J.; Nelson, C. R.; and Piger, J., 2004. The less-volatile US economy: A Bayesian investigation of timing, breadth, and potential explanations. *Journal of Business and Economic Statistics*, 22, 1 (2004), 80–93. (cited on pages 34 and 59)

Kim, S.; Shephard, N.; and Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65, 3 (1998), 361–393. (cited on pages 3, 16, and 27)

Koop, G.; Leon-Gonzalez, R.; and Strachan, R. W., 2009. On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33, 4 (2009), 997–1017. (cited on pages 37, 59, and 65)

Koop, G.; Poirier, D. J.; and Tobias, J. L., 2007. *Bayesian econometric methods*, vol. 7. Cambridge University Press. (cited on page 16)

Koop, G. and Potter, S. M., 2007. Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies*, 74, 3 (2007), 763–789. (cited on pages 34, 59, and 60)

Koopman, S. J. and Hol Uspensky, E., 2002. The stochastic volatility in mean model: Empirical evidence from international stock markets. *Journal of Applied Econometrics*, 17, 6 (2002), 667–689. (cited on page 34)

Lange, K. L.; Little, R. J.; and Taylor, J. M., 1989. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84, 408 (1989), 881–896. (cited on page 4)

Liu, Y. and Morley, J., 2014. Structural evolution of the postwar US economy. *Journal of Economic Dynamics and Control*, 42 (2014), 50–68. (cited on pages 37 and 59)

Maheu, J. M. and Yang, Q., 2015. An infinite hidden Markov model for short-term Interest Rates. (2015). (cited on pages 60 and 75)

Maronna, R. A. and Yohai, V. J., 1998. Robust estimation of multivariate location and scatter. *Encyclopedia of Statistical Sciences*, (1998). (cited on page 4)

McCulloch, R. E. and Tsay, R. S., 1993. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, 88, 423 (1993), 968–978. (cited on pages 34 and 37)

Melino, A., 2012. Inflation targeting: A Canadian perspective. *International Journal of Central Banking*, 8, S1 (2012), 105–131. (cited on page 46)

Nakajima, J. and Omori, Y., 2012. Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew StudentâĂŹs t-distribution. *Computational Statistics and Data Analysis*, 56, 11 (2012), 3690–3704. (cited on page 4)

Primiceri, G. E., 2005. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72, 3 (2005), 821–852. (cited on pages 3, 4, 59, 65, and 72)

Romer, C. D. and Romer, D. H., 2002. A rehabilitation of monetary policy in the 1950's. *American Economic Review*, (2002), 121–127. (cited on page 44)

Sethuraman, J., 1994. A constructive definition of dirichlet priors. *Statistica Sinica*, 4 (1994), 639–650. (cited on page 62)

Sherwin, M., 1997. Inflation targeting: The New Zealand experience. In *Paper delivered to a Bank of Canada conference on Price Stability, Inflation Targets and Monetary Policy*. Citeseer. (cited on page 46)

Sims, C. A. and Zha, T., 2006. Were there regime switches in US monetary policy? *The American Economic Review*, (2006), 54–81. (cited on pages 59 and 72)

Song, Y., 2014. Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics*, 29, 5 (2014), 825–842. (cited on pages 60, 66, 70, 75, 85, 87, and 88)

Stock, J. H. and Watson, M. W., 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14, 1 (1996), 11–30. (cited on pages 34 and 59)

Stock, J. H. and Watson, M. W., 2003. Has the business cycle changed and why? In *NBER Macroeconomics Annual 2002, Volume 17*, 159–230. MIT press. (cited on pages 3 and 10)

Stock, J. H. and Watson, M. W., 2007. Why has US inflation become harder to forecast? *Journal of Money, Credit and banking*, 39, s1 (2007), 3–33. (cited on pages 3, 34, and 61)

Taylor, S. J., 1994. Modeling stochastic volatility: A review and comparative study. *Mathematical finance*, 4, 2 (1994), 183–204. (cited on page 3)

Teh, Y. W.; Jordan, M. I.; Beal, M. J.; and Blei, D. M., 2006. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101, 476 (2006). (cited on pages 60, 63, and 64)

Van Gael, J.; Saatci, Y.; Teh, Y. W.; and Ghahramani, Z., 2008. Beam sampling for the infinite hidden Markov model. In *Proceedings of the 25th International Conference on Machine Learning*, 1088–1095. ACM. (cited on page 67)

West, M., 1984. Outlier models and prior distributions in bayesian linear regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, (1984), 431–439. (cited on page 4)