Stock Returns and Real Growth: A Bayesian Nonparametric Approach

QIAO YANG

ShanghaiTech University

7 December 2016

Stock Returns and Real Growth: A Bayesian Nonparametric Approach

(Bill) Qiao Yang*

April 2016

Abstract

The study of the joint dynamic behaviour between stock market returns and real economic growth rates is an important empirical question in finance and macroeconomics. This paper investigates their linkage by proposing a vector autoregressive infinite hidden Markov model. Our model has two advantages over the existing approaches in the literatures. In contrast to Markov switching models with fixed states, our model will learn the number of states from the data rather than fixing it a priori. The vector autoregressive setting in our model allows the joint time series of stock market returns and real growth rates to share the same unobserved state variable. Compared to existing models, our model shows significant improvements in out-of-sample density forecast accuracy. This paper demonstrates the predictive power of stock market returns for future growth rates are better captured by the unobserved states variables, rather than the lagged stock market returns.

key words: hierarchical Dirichlet process prior, beam sampling, Markov switching, MCMC,

JEL: C58, C14, C22, C11

*ShanghaiTech University, yangqiao@shanghaitech.edu.cn. I would like to thank John Mahe, Martin Burda, Yuanyuan Wan for their valuable comments. I am grateful for comments from conference participants at midwest econometrics annual meeting, SBIES and seminar participants at University of Toronto.

1 Introduction

Studying the connection between stock market and real economic growth is an important empirical question[1]. In this paper, I study their relationship by using a flexible Bayesian nonparametric approach with several contributions. Firstly, I propose a vector autoregressive infinite hidden Markov model (IHMM-VAR). The new model allows us to investigate the nonlinear and contemporaneous relationship through a common unobserved state variable. Secondly, mixed evidence of predictive power of lagged stock market returns for future real growth rates is documented in contrast to the existing literature. Thirdly, I illustrates that the vector autoregressive dynamics coupled with Markov switching are essential for capturing the predictive power of stock market returns for real growth rates.

Fama (1990) and Schwert (1990) found lagged stock market returns have significant correlation with future real growth rates by using linear regression. Choi et al. (1999) build upon their work by using cointegration tests and error-correction models, in which only weak evidence for predictive power of lagged stock market returns on future growth rates is found. Lee (1992) and Hassapis & Kalyvitis (2002) find significant correlations between lagged stock market returns and future real growth rates by applying a vector autoregression (VAR). Kanas & Ioannidis (2010) continue their work by applying the nonlinear model. It suggests a model of Markov switching with two states and finds a nonlinear correlation between lagged stock market returns and future real growth rates. Kim & In (2003) find the predictive power of lagged stock market returns on future real growth rates is likely to be time-varying by using spectral and wavelet analysis. Hamilton & Lin (1996) suggest the hidden state variables are the main driving force for directing the dynamics of stock market returns and real growth rates.

What is clear from this literature is that the Markov switching and the vector autoregressive structure are necessary to capture changes in joint dynamics of stock market returns and real growth rates. The existing methods of dealing with this, such as simple Markov switching or vector autoregression, are insufficient. In addition, the existing papers examining model performance only focus on point forecasts and ignore density forecasts. This paper contributes to the literature by introducing a vector autoregressive infinite hidden Markov model (IHMM-VAR). In contrast to Markov switching models with fixed states, the IHMM-VAR allows the unbounded transition matrix to infer the number of states through the data rather than fixing as prior. Therefore, the IHMM-VAR can allow to introduce new states to capture any structural change and incorporate it into forecasts.

By applying our new model on joint time series of U.S monthly S&P 500 excess stock market returns and industrial production growth rates, a large gain in out-of-sample density forecast accuracy is delivered compared to benchmark models. An average of 20 states are used to model the joint time series in our model. Our model shows evidence

[1]Fama (1990), Schwert (1990), Choi et al. (1999) and Kanas & Ioannidis (2010) suggest real stock returns lead changes in real activity due to dividend discount valuation or consumption capital asset pricing model.

of capturing structural breaks as well as recurring states in their dynamic relationship. Our model shows evidence of the dynamic change in conditional mean and variance of stock market returns and real growth rates.

The prior for the transition matrix of the infinite hidden Markov model is constructed by two Dirichlet processes, which is a special case of hierarchical Dirichlet process of Teh et al. (2006). A common draw of the top Dirichlet process determines the prior for each row of the transition matrix. Then, each row of the transition matrix is a draw from the secondary Dirichlet process and it is centered around a common draw from the top level.

According to log-predictive Bayes factors, our model does not show any supporting evidence such that the lagged stock market returns should have predictive power for future real growth rates. Weak evidence of this lag-relation[2] is found in benchmark models. The empirical result in this paper suggests the unobserved Markov states shared by the joint time series are the main driving force to capture the predictive power of stock market returns for real growth rates, rather than the lag-relation.

This paper is organized as follows. Section 2 introduces benchmark models. Section 3 discusses infinite hidden Markov models and its links to Dirichlet process. Section 4 shows the posterior sampling steps. Section 5 describes how to compute out-of-sample density forecast accuracy from various models. Section 6 discusses empirical results, and Section 7 concludes the paper.

2 Benchmark Models

The benchmark models are the ones used in the exiting literatures for studying the joint behavior of stock market returns and real growth rates. One category is the univariate setting, which means to use autoregression (AR) to separately model stock market returns and real growth rates. Another category is the multivariate setting, which explicitly uses vector autoregression (VAR) to jointly model the stock market returns and real growth rates. In order to test the predictive power of lagged stock market returns on real growth rates and lag real growth rates on stock market returns, the unrestricted and restricted versions of each model are introduced. The difference in forecast performance between two versions reveals whether or not additional explanatory variables can contribute to the forecast accuracy. Hamilton & Lin (1996) suggests only one lag is necessary. This paper follows their approach[3]. Let r_t and g_t represents the corresponding stock market return and real growth rate at time t. The details of each benchmark model are outlined in the following sections.

[2]The lag-relation implies the lagged stock market returns should have predictive power for future real growth rates.

[3]A preliminary result shows the log-predictive Bayes factor based on two lags do not contribute the forecast performance under the IHMM-VAR.

2.1 Univariate Approach

The autoregression (AR) implicitly assumes there is only a single state to govern the whole time series. Fama (1990) and Schwert (1990) use a similar specification. This paper revisits this specification under the Bayesian approach. We consider the following AR model augmented with lagged real growth rate:

$$r_t = \mu + \beta_1 r_{t-1} + \beta_2 g_{t-1} + e_t \quad e_t \overset{iid}{\sim} N(0, \sigma^2), \tag{1}$$

For modeling the real growth rates, we use the following,

$$g_t = \mu + \beta_1 g_{t-1} + \beta_2 r_{t-1} + e_t \quad e_t \overset{iid}{\sim} N(0, \sigma^2), \tag{2}$$

Estimation of these models is carried out independently. Equation (1) and (2) are the unrestricted version. The priors for the above models are the following,

$$\vartheta \sim MN(a, A) \qquad \frac{1}{\sigma^2} \sim Gamma(b_1, b_2), \tag{3}$$

where $\vartheta = (\mu, \beta_1, \beta_2)$ and $Gamma(.)$ denotes gamma distribution with $E(\sigma^{-2}) = \frac{b_1}{b_2}$. $MN(a, A)$ is the multivariate normal distribution with the mean vector of a and variance covariance matrix A. The restricted version of equations (1) and (2) are obtained with $\beta_2 = 0$.

The autoregressive Markov switching model with 2 states (MS2-AR) allows the predictive power of explanatory variables to be regime-dependent. As suggested by Hamilton & Lin (1996) and Kanas & Ioannidis (2010). This paper estimated this model under the Bayesian approach. The following is the MS2-AR for stock market returns:

$$r_t = \mu_{s_t} + \beta_{1s_t} r_{t-1} + \beta_{2s_t} g_{t-1} + e_t \quad s_t \in \{1, 2\} \tag{4a}$$

$$e_t \overset{iid}{\sim} N(0, \sigma^2_{s_t}) \quad s_t | s_{t-1} \sim \Pi_{s_{t-1}} \tag{4b}$$

where Π is the 1st order Markov transition matrix with dimension of 2, and $\Pi_{s_{t-1}}$ represents the s_{t-1} row of transition matrix of Π. Similarly, the MS2-AR for real growth rates:

$$g_t = \mu_{s_t} + \beta_{1s_t} g_{t-1} + \beta_{2s_t} r_{t-1} + e_t \quad s_t \in \{1, 2\} \tag{5a}$$

$$e_t \overset{iid}{\sim} N(0, \sigma^2_{s_t}) \quad s_t | s_{t-1} \sim \Pi_{s_{t-1}} \tag{5b}$$

The priors for the MS2-AR:

$$\vartheta_{s_t} \sim MN(a, A) \qquad \frac{1}{\sigma^2_{s_t}} \sim Gamma(b_1, b_2) \quad s_t \in \{1, 2\}, \tag{6}$$

where $\vartheta_{s_t} = (\mu_{s_t}, \beta_{1s_t}, \beta_{2s_t})$. The restricted version of equations (4) and (5) are obtained with $\beta_{2s_t} = 0$. The prior for each row of transition matrix is the Dirichlet distribution. The posterior simulation of the MS2-AR is referred to Albert & Chib (1993).

Adding hierarchical priors to Markov switching models with finite or infinite states is a way of potentially improving the model performance. The hierarchical prior allows the priors to learn from each regime in a similar way as posterior parameters learn from the data rather than chosen by econometrician. Song (2013) and Maheu & Yang (2015) find significant gains in out-of-sample density forecasts accuracy by adding hierarchical priors. As well, they discover the model performance is more robust in the prior sensitivity test. The followings are the hierarchical prior for the equation (6):

$$a \sim MN(h_0, H_0) \qquad A^{-1} \sim Wishart(d_0, D_0) \tag{7a}$$

$$b_1 \sim Gamma(\chi_0, \nu_0) \qquad b_2 \sim Gamma(\chi_1, \nu_1), \tag{7b}$$

where $Wishart(d_0, D_0)$ denotes the Wishart distribution with $d_0 \geq dim(\vartheta_{s_t})$ as degree of freedom and D_0 is the scale matrix with the same dimension of A^{-1}.

2.2 Multivariate Approach

The vector autoregression (VAR) used by Lee (1992) and Hassapis & Kalyvitis (2002) is to model stock market returns and real growth rates jointly.

$$\begin{bmatrix} r_t \\ g_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 \\ \beta_3 & \beta_4 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} e_{rt} \\ e_{gt} \end{bmatrix} \tag{8a}$$

$$\Sigma = \begin{bmatrix} \sigma_r^2 & \rho\sigma_r\sigma_g \\ \rho\sigma_r\sigma_g & \sigma_g^2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} e_{rt} \\ e_{gt} \end{bmatrix} \overset{iid}{\sim} MN(0, \Sigma) \tag{8b}$$

This paper applies the independent Normal-Wishart prior, which is referred to Koop & Korobilis (2009).

$$\vartheta \sim MN(a, A), \quad \Sigma^{-1} \sim Wishart(b, B) \tag{9}$$

Let $\vartheta = (\mu_1, \mu_2, \beta_1, \beta_2, \beta_3, \beta_4)^T$ be the unrestricted version, where the restricted version is obtained by $\beta_2 = \beta_3 = 0$.

The vector autoregressive Markov switching with two states (MS2-VAR) is introduced by Hamilton & Lin (1996) and Kanas & Ioannidis (2010). They model the joint time series by using two regimes, which correspond to a high and low volatile periods. In contrast to the univariate Markov switch, an attractive feature of MS2-VAR is to allow a unique Markov states to govern both time series.

$$\begin{bmatrix} r_t \\ g_t \end{bmatrix} = \begin{bmatrix} \mu_{1s_t} \\ \mu_{2s_t} \end{bmatrix} + \begin{bmatrix} \beta_{1s_t} & \beta_{2s_t} \\ \beta_{3s_t} & \beta_{4s_t} \end{bmatrix} \begin{bmatrix} r_{t-1} \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} e_{rt} \\ e_{gt} \end{bmatrix}, \quad \text{and} \begin{bmatrix} e_{rt} \\ e_{gt} \end{bmatrix} \overset{iid}{\sim} MN(0, \Sigma_{s_t}) \tag{10a}$$

$$\Sigma = \begin{bmatrix} \sigma_{rs_t}^2 & \rho\sigma_{rs_t}\sigma_{gs_t} \\ \rho_{s_t}\sigma_{rs_t}\sigma_{gs_t} & \sigma_{gs_t}^2 \end{bmatrix} \qquad s_t|s_{t-1} \sim \Pi_{s_{t-1}} \qquad s_t \in \{1, 2\}, \tag{10b}$$

The Π and $\Pi_{s_{t-1}}$ are defined the same way as in MS2-AR. Let $\vartheta_{s_t} = (\mu_{1s_t}, \mu_{2s_t}, \beta_{1s_t}, \beta_{2s_t}, \beta_{3s_t}, \beta_{4s_t})^T$. The restricted version is obtained by $\beta_{2s_t} = \beta_{3s_t} = 0$. The priors for MS2-VAR is the following:

$$\vartheta_{s_t} \sim MN(a, A), \quad \Sigma_{s_t}^{-1} \sim Wishart(b, B) \quad s_t \in \{1, 2\}$$

The hierarchical priors for MS2-VAR are the following,

$$a \sim MN(h_0, H_0) \qquad A^{-1} \sim Wishart(d_0, D_0) \tag{11a}$$

$$B \sim InvWhishart(e_0, E_0) \qquad b \sim Gamma(\chi_0, \nu_0) I(b \geq 2) \tag{11b}$$

$InvWhishart(e_0, E_0)$ denotes an inverse Wishart distribution with the degree of freedom of e_0 and E_0 be scale square matrix with dimension of 2.

3 Infinite Hidden Markov Model (IHMM)

An infinite hidden Markov model (IHMM) is based on a Bayesian nonparametric prior introduced by Beal et al. (2002). Compared to the finite Markov switching model, the IHMM extends the transition probability matrix of Markov switching model from finite dimension to infinite dimension, which allows the model to learn the regime dynamics through the data rather than fixing as prior. The IHMM builds on the hierarchical Dirichlet process (HDP) priors, which is an extension of Dirichlet process (DP). I initially discuss the DP and HDP before moving to the IHMM. Afterward, we will discuss the corresponding univariate and multivariate setting of the IHMM.

3.1 Dirichlet Process (DP)

The Dirichlet process (DP) is introduced by Ferguson (1973). It is a distribution of probability measures over a base probability measure; a formal definition of DP is $G \sim DP(\eta, H)$. G is a random draw of a distribution based on H and $\eta > 0$ as a concentration parameter. A more helpful definition of the DP is introduced by Sethuraman (1994), who actually constructs the G, and it is called a stick-breaking representation.

Let θ be the parameter set and δ_{θ_i} denote a probability mass at θ_i, in which case then the stick-breaking representation of $G \sim DP(\eta, H)$ is,

$$\pi_i = v_i \prod_{l=1}^{i-1} (1 - v_l), \quad v_i \overset{iid}{\sim} Beta(1, \eta), \quad \theta_i \overset{iid}{\sim} H \tag{12a}$$

$$G = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i} \quad \text{for} \quad i = 1, 2, \ldots, \infty \tag{12b}$$

The representation in equation (12a) is denoted as $\{\pi_i\}_{i=1}^{\infty} \sim Stick(\eta)$. The parameter η governs the distribution of the weight (π_i) over a unit mass. A large value of η

implies more numbers of different θ_i but less weight is assigned to each θ_i. On the other hand, a small value of η indicates fewer numbers of θ_i but each θ_i is assigned with a larger weight. It does not matter if H is a continuous or discrete probability measure as G is always discrete. The DP is often applied as a Bayesian nonparametric prior in econometrics. The Dirichlet mixture models is the direct application of using DP priors. In recent years, a significant amount of empirical applications of DP have been employed in econometrics.[4]

3.2 Hierarchical Dirichlet Process (HDP)

The Hierarchical Dirichlet process (HDP) is introduced by Teh et al. (2006) as a combination of two Dirichlet processes (DP). The draw from the first DP is based on the initial the base measure of H. This draw is treated as a base measure for the second DP. Thus, both of the DP will share a common base measure. The HDP has a hierarchical structure and it is constructed as the following,

$$G_0|\eta, H \sim DP(\eta, H) \tag{13a}$$

$$G_j|\alpha, G_0 \sim DP(\alpha, G_0), \quad j = 1, \ldots, \infty, \tag{13b}$$

The G_j is conditional on a global probability measure G_0. The α and η are the corresponding concentration parameters. If we use the stick-breaking representation, we have the following,

$$G_0 = \sum_{i=1}^{\infty} \gamma_i \delta_{\theta_i}, \quad \Gamma = \{\gamma_i\}_{i=1}^{\infty} \sim Stick(\eta), \quad \theta_i \overset{iid}{\sim} H, \tag{14a}$$

$$G_j = \sum_{i=1}^{\infty} \pi_{ji} \delta_{\theta_i}, \quad \{\pi_{ji}\}_{i=1}^{\infty} \overset{iid}{\sim} Stick2(\alpha, \Gamma), \tag{14b}$$

The $Stick2(\alpha, \Gamma)$[5] is constructed in the following way,

$$\pi_{ji} = \hat{\pi}_{ji} \prod_{l=1}^{i-1} (1 - \hat{\pi}_{jl}), \qquad \hat{\pi}_{ji} \overset{iid}{\sim} Beta\left(\alpha\gamma_i, \alpha\left(1 - \sum_{l=1}^{i} \gamma_l\right)\right), \tag{15}$$

Let $j = 1, 2, \ldots.$ and $i = 1, 2, \ldots.$. Each G_j shares the same atoms with all other G_j as well as G_0, but all of the G_j and G_0 have different weights. Regardless, H is a continuous or discrete probability measure, where the G_j are always discrete, which becomes a suitable way for representing probability weights in each row of the transition matrix. Let Π be the infinite Markov transition matrix. G_j is served as the prior for jth row of Π. The $\{\pi_{ji}\}_{i=1}^{\infty}$ becomes the corresponding probability weights for jth row.

[4] examples are Jensen & Maheu (2010), Jensen & Maheu (2013), Song (2013) and Jensen & Maheu (2014)

[5] Using the notation of $Stick2(\alpha, \Gamma)$ to represent the second DP is firstly used in Maheu & Yang (2015).

3.3 Infinite Hidden Markov Model (IHMM)

An infinite hidden Markov model (IHMM) extends Markov switching model with finite dimension to infinite dimension. The key component of the IHMM is to construct the priors for the infinite transition matrix; it is a special case of the hierarchical Dirichlet process. This section discusses the general setting of the IHMM, the multivariate infinite hidden Markov model (IHMM-VAR), and the univariate infinite hidden Markov model (IHMM-AR). There are papers which apply the univariate IHMM in various empirical applications. For example, Song (2013) uses the IHMM to U.S. real interest rates. Similarly, Jochmann (2015) applies the IHMM to U.S inflation rates, and Dufays (2015) applies the IHMM to model stock volatilities. Maheu & Yang (2015) applies the IHMM to U.S. 3-month T-Bill rates. Shi & Song (2015) uses the IHMM to detect the speculative bubbles in NASDAQ stock market. Our model differs from their works by extending to the multivariate IHMM (IHMM-VAR).

3.3.1 IHMM Basics

The state variable s_t follows a 1st order Markov with an infinite transition matrix, such as $s_t \in \{1, 2, 3, \dots\}$ which is the state variable at time t. The Π_j is jth row of the transition matrix. An element in Π_j, such as π_{ji} is the probability of moving from state j to state i.

$$\{\gamma_i\}_{i=1}^{\infty} = \Gamma|\eta \sim Stick(\eta) \qquad \theta_i \overset{iid}{\sim} H \quad i = 1, 2, \dots, \tag{16a}$$

$$\Pi_j|\alpha, \Gamma \overset{iid}{\sim} Stick2(\alpha, \Gamma), \qquad j = 1, 2, \dots, \tag{16b}$$

$$s_t|s_{t-1}, \Pi_{s_{t-1}} \sim \Pi_{s_{t-1}}, \quad t = 1, \dots, T \tag{16c}$$

$$y_t|s_t, \Theta \sim F(y_t|\theta_{s_t}). \tag{16d}$$

In equations (16), two DPs construct the prior for Markov transition probability matrix with infinite dimension. Each row of Π_j is drawn from a DP prior with a base measure. H represents the priors over the parameter space Θ. $F(.)$ is the conditional density function for observation y_t. The η and α govern the distribution of the weights. Various combinations of η and α can enforce different prior beliefs on the dimension of Markov transition. For example, larger values of α and η allow for a higher possibility of considering a new state. In order to best capture the joint state dynamics of stock market returns and real growth rates, a hyper prior on α and η is placed so we can infer them from the data rather than fixing them.

$$\eta \sim Gamma(\chi_1, \nu_1) \quad \alpha \sim Gamma(\chi_2, \nu_2) \tag{17}$$

In contrast to finite Markov switching models, the unbounded transition matrix allows for recurring regimes from the past, as well as new states to capture structural changes in terms of estimating. In addition, the IHMM is able to introduce new states to capture any potential structural changes into forecasts. To put his another way, the flexible framework of IHMM allows the conditional distribution of y_t to be constructed by infinite number of Gaussian mixture component. Therefore, this feature significantly weakens the influence of distribution assumption on the innovation. For instance, with any given forms of the error term, the IHMM can always achieve it by adjusting number of mixture components as well as its corresponding parameters regardless either we use t-distribution or Gaussian define each mixture component. The following is the conditional distribution on y_{t+1},

$$f(y_{t+1}|\Theta, s_t) = \sum_{i=1}^{\infty} \pi_{s_t,i} p(y_{t+1}|\theta_i) \tag{18}$$

3.3.2 IHMM-VAR

The multivariate infinite hidden Markov model (IHMM-VAR) for jointly modeling stock market returns and real growth rates can be written as.

$$\Gamma|\eta \sim Stick(\eta), \quad \Pi_j|\alpha, \Gamma \overset{iid}{\sim} Stick2(\alpha, \Gamma), \quad j = 1, 2, \ldots, \tag{19a}$$

$$\begin{bmatrix} r_t \\ g_t \end{bmatrix} = \begin{bmatrix} \mu_{1s_t} \\ \mu_{2s_t} \end{bmatrix} + \begin{bmatrix} \beta_{1s_t} & \beta_{2s_t} \\ \beta_{3s_t} & \beta_{4s_t} \end{bmatrix} \begin{bmatrix} r_{t-1} \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} e_{rt} \\ e_{gt} \end{bmatrix}, \text{and} \begin{bmatrix} e_{rt} \\ e_{gt} \end{bmatrix} \overset{iid}{\sim} MN(0, \Sigma_{s_t}) \tag{19b}$$

$$\Sigma_{s_t} = \begin{bmatrix} \sigma^2_{rs_t} & \rho\sigma_{rs_t}\sigma_{gs_t} \\ \rho_{s_t}\sigma_{rs_t}\sigma_{gs_t} & \sigma^2_{gs_t} \end{bmatrix} \qquad s_t|s_{t-1}, \Pi_{s_{t-1}} \sim \Pi_{s_{t-1}}, \quad t = 1, \ldots, T \tag{19c}$$

Let $s_t \in \{1, \ldots, \infty\}$ and $\vartheta_i = (\mu_{1i}, \mu_{2i}, \beta_{1i}\beta_{2i}, \beta_{3i}, \beta_4)$. The priors for ϑ_i and Σ_i are the following.

$$\vartheta_i \sim MN(a, A) \qquad \Sigma_i^{-1} \sim Wishart(b, B) \quad i = 1, 2, \ldots, \tag{20}$$

The hierarchical prior for the IHMM is the same as equation (11) of MS2-VAR. The restricted version is to set $\beta_{2s_t} = \beta_{3s_t} = 0$.

3.3.3 IHMM-AR

The univariate IHMM is the infinite hidden Markov model (IHMM-AR) is the following example applies to the real growth rates,

$$\Gamma|\eta \sim Stick(\eta), \quad \Pi_j|\alpha, \Gamma \overset{iid}{\sim} Stick2(\alpha, \Gamma), \quad j = 1, 2, \ldots, \tag{21a}$$

$$g_t = \mu_{s_t} + \beta_{1s_t} g_{t-1} + \beta_{2s_t} r_{t-1} + e_t \qquad e_t \overset{iid}{\sim} N(0, \sigma^2_{s_t}) \tag{21b}$$

$$s_t|s_{t-1}, \Pi_{s_{t-1}} \sim \Pi_{s_{t-1}}, \quad t = 1, \ldots, T, \quad s_t \in \{1, \ldots, \infty\} \tag{21c}$$

Let $\vartheta_i = (\mu_i, \beta_{1i}, \beta_{2i})$ and the priors are:

$$\vartheta_i \sim MN(a, A) \qquad \frac{1}{\sigma_i^2} \sim Gamma(b_1, b_2) \quad i = 1, 2, \ldots, \tag{22}$$

The restricted version of IHMM-AR on real growth rates is to let the $\beta_{2s_t} = 0$. The hierarchical priors are the the same as equation (7) of MS2-AR.

For IHMM-AR on stock market returns, all of the parts are the same except equations (21b) is replaced by the following,

$$r_t = \mu_{s_t} + \beta_{1s_t} r_{t-1} + \beta_{2s_t} g_{t-1} + e_t \qquad e_t \overset{iid}{\sim} N(0, \sigma^2_{s_t}) \tag{23}$$

The restricted version is to set $\beta_{2s_t} = 0$.

4 Posterior Sampling of IHMM

Chib (1996) introduces the forward-filter backward sampler (FFBS) to sample state variables in the Markov switching models with fixed states. However, the FFBS is not feasible for sampling IHMM due to its infinite dimension. The beam sampler introduced by Van Gael et al. (2008) solves this issue. Basically, the beam sampler adaptively truncates the infinite transition matrix of IHMM to a finite one so that the FFBS can be applied.

The idea of the beam sampler is very close to the slice sampler by Walker (2007). The beam sampler involves the introduction of auxiliary variables $\{u_{1:T}\}$, which are stochastically generated based on Π and corresponding state variables $\{s_{t:T}\}$. The u_t is draw from the following density function,

$$f(u_t|s_{t-1}, s_t, \Pi) = \frac{\mathbf{I}(0 < u_t < \pi_{s_{t-1},s_t})}{\pi_{s_{t-1},s_t}}, \quad t = 1, \dots, T \tag{24}$$

Once the u_t is sampled, the forward step of the FFBS for s_t becomes the following,

$$p(s_t|y_{1:t}, u_{1:t}, \Pi) \propto p(y_t|y_{1:t-1}, s_t) \sum_{s_{t-1}=1}^{\infty} p(s_{t-1}|y_{1:t-1}, u_{1:t-1}, \Pi)\mathbf{I}(u_t < \pi_{s_{t-1},s_t}) \tag{25}$$

Let y_t be the observation at time t. The u_t turns the summation from infinite number of states into a finite one. Once the filter step is computed for $t = 1, \dots, T$, the backward step for sampling state variable s_t for $t = T-1, \dots, 1$, is the following,

$$p(s_t|s_{t+1}, y_{1:T}, u_{1:T}) \propto p(s_t|y_{1:t}, u_{1:t})\mathbf{I}(u_{t+1} < \pi_{s_t,s_{t+1}}) \tag{26}$$

In each Markov chain Monte Carlo (MCMC) iteration, the auxiliary variables, $u_{1:T}$, effectively reduce the transition probability matrix to a finite dimension. The full MCMC routine involves the following steps:

1. Sample $s_{1:T}|y_{1:T}, u_{1:T}, \Pi$
2. Sample $\Pi_j|s_{1:T}, \Gamma, j = 1, \dots, K$.
3. Sample $u_{1:T}|s_{1:T}, \Pi$ and update K
4. Sample $\theta_j|s_{1:T}, y_{1:T}, \xi, j = 1, \dots, K$
5. Sample $\Gamma|s_{1:T}, \eta$
6. Sample $\xi|\theta_1, \dots, \theta_K, \eta|s_{1:T}, \Gamma$ and $\alpha|s_{1:T}, \Gamma$.

The details of sampling steps are included in Appendix. For full sample estimates, the first 500,000 draws are the burn-in and the next 500,000 draws are for posterior inference. Let M be the number of posteriors draws after burn-in. Any features of posterior can be easily computed. For example, to compute posterior average of Σ_{s_t} which is $E[\Sigma_{s_t}|y_{1:T}]$ at time t is $\frac{1}{M}\sum_{l=1}^{M}\Sigma_{s_t}^{(l)}$.

5 Predictive Likelihood

The predictive likelihood measures the out-of-sample density forecast accuracy. In contrast to point forecast, such as root mean squared forecast errors, the predictive likelihood evaluate the predictive distribution as a whole, where point forecast only focus on the central of the predictive distribution. Thus it is not surprise that two methods deliver contradictory outcomes. A representation of predictive likelihood is the following,

$$\rho(y_{T+1}|y_{1:T}) = \int_{\Theta} f(y_{T+1}|\theta, y_{1:T})\rho(\theta|y_{1:T})d\theta, \qquad \theta \in \Theta \tag{27}$$

where the marginalization is taken with respect to $\rho(\theta|y_{1:T})$ which is the predictive posterior distribution of θ. $y_{1:T}$ are observations used for estimation and y_{T+1} is the observation to predict. The Θ is the corresponding parameter set. Equation (27) can also be used for evaluating the model fitting since the parameter uncertainties are incorporated into the predictive likelihood computation such as the marginalization is taken with respect $\rho(\theta|y_{1:T})$. For computing log-predictive likelihoods, the first 10,000 is the burn-in and the next 20,000 are for predictive inference. This paper uses a recursive method on predictive inference, that is the last draw for predicting $T+1$ will be used for the initial draw of predicting $T+2$.

Computing the predictive likelihood has two categorizes. One is under the univariate setting. Such as models of the AR, MS2-AR and IHMM-AR. Their predictive likelihoods on stock market returns and real growth rates are computed separately. Another category is the multivariate setting[6]. In contrast to the univariate setting, the joint predictive likelihood of stock market returns and real growth rates is feasible to compute, where it is an available in univariate setting. An alternative approach for computing the joint predictive likelihood under univariate setting is illustrated in next section. Moreover, the marginal predictive likelihood of stock market returns and real growth rates are feasible to compute under multivariate setting. All details are illustrated in next several subsections.

5.1 AR

Calculating the predictive likelihood of the AR model for stock market returns is the same as for real growth rates. For example, The predictive likelihood of real growth

[6]The bivariate setting means the models of the VAR, MS2-VAR and IHMM-VAR

rates g_{T+1} at time $T+1$: At lth MCMC draw that is estimated based on $\{g_{1:T}\}$, a sequence of posterior draws $\{\mu_1^{(l)}, \beta_1^{(l)}, \beta_2^{(l)}, \sigma^{2(l)}\}$ is sampled. Let $\vartheta^{(l)} = \{\mu_1^{(l)}, \beta_1^{(l)}, \beta_2^{(l)}\}$. M is the total number of MCMC draws that are used for forecast inference. At each MCMC iteration, the predictive distribution for g_{T+1} is actually a Gaussian distribution. In order to calculate the predictive likelihood at g_{T+1}, we plug-in the realization of g_{T+1}. Over all MCMC draws, we take the average. This routine applies to all models. The predictive likelihood for g_{T+1} is the following,

$$p(g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(g_{T+1}|\vartheta^{(l)}, \sigma^{2(l)}) \tag{28}$$

Similar steps are applied to calculating $p(r_{T+1}|r_{1:T}, g_{1:T})$.

5.2 MS2-AR

Computing the predictive likelihood of the MS2-AR models are the same as we do on AR models except we need one extra step to forecast the state variable. The predictive likelihood of g_{T+1} is calculated as the following steps. Give a sequence of lth MCMC draw of $\{\mu_{1i}^{(l)}, \beta_{1i}^{(l)}, \beta_{2i}^{(l)}, \sigma_i^{2(l)}, s_{1:T}^{(l)}, \Pi^{(l)}\}$. Let $\vartheta_i = \{\mu_{1i}^{(l)}, \beta_{1i}^{(l)}, \beta_{2i}^{(l)}\}$ for $i = 1, 2$. We firstly draw the $s_{T+1}^{(l)}$ through $\Pi_{s_T}^{(l)}$ since $s_T^{(l)}$ is given. Let the draw of $s_{T+1}^{(l)} = k$. The predictive likelihood of g_{T+1} is the following,

$$p(g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(g_{T+1}|\vartheta_k^{(l)}, \sigma_k^{2(l)}) \tag{29}$$

Similar steps are applied to calculating $p(r_{T+1}|r_{1:T}, g_{1:T})$.

For computing the joint predictive likelihood of $p(r_{T+1}, g_{T+1}|r_{1:T}, g_{1:T})$ under univariate models, we multiple the $p(r_{T+1}|r_{1:T}, g_{1:T})$ and $p(g_{T+1}|r_{1:T}, g_{1:T})$ due to their independence. For example:

$$p(r_{T+1}, g_{T+1}|r_{1:T}, g_{1:T}) = p(r_{T+1}|r_{1:T}, g_{1:T})p(g_{T+1}|r_{1:T}, g_{1:T}) \tag{30}$$

5.3 VAR

In contrast to the AR model, computing $p(r_{T+1}|r_{1:T}, g_{1:T})$, $p(g_{T+1}|r_{1:T}, g_{1:T})$ and $p(r_{T+1}, g_{T+1}|r_{1:T}, g_{1:T})$ can be done altogether in multivariate setting, such as VAR, MS2-VAR and IHMM-VAR. For computing the marginal predictive likelihood of stock market returns, real growth rates and joint predictive likelihood of them: we do the following steps. At lth MCMC draw and a sequence of posterior draws is given by $\{\mu_1^{(l)}, \mu_2^{(l)}, \beta_1^{(l)}, \beta_2^{(l)}, \beta_3^{(l)}, \beta_4^{(l)}, \Sigma^{(l)}\}$. By simplifying notations , we let $\vartheta_r^{(l)} = \{\mu_1^{(l)}, \beta_1^{(l)}, \beta_2^{(l)}\}$, $\vartheta_g^{(l)} = \{\mu_2^{(l)}, \beta_3^{(l)}, \beta_4^{(l)}\}$ and $\Sigma^{(l)} = \{\sigma_r^{2(l)}, \sigma_g^{2(l)}, \rho^{(l)}\}$. The marginal predictive likelihood for g_{T+1}, r_{T+1} and the joint of them are the following,

$$p(r_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(r_{T+1}|\vartheta_r^{(l)}, \sigma_r^{2(l)}) \tag{31a}$$

$$p(g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(g_{T+1}|\vartheta_g^{(l)}, \sigma_g^{2(l)}) \tag{31b}$$

$$p(r_{T+1}, g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} MN(r_{T+1}, g_{T+1}|\vartheta_r^{(l)}, \vartheta_g^{(l)}, \Sigma^{(l)}), \tag{31c}$$

5.4 MS2-VAR

Given lth MCMC posterior draw: $\{\mu_{1i}^{(l)}, \mu_{2i}^{(l)}, \beta_{1i}^{(l)}, \beta_{2i}^{(l)}, \beta_{3i}^{(l)}, \beta_{4i}^{(l)}, \Sigma_i^{(l)}, \Pi_i^{(l)}, s_{1:T}^{(l)}\}$ and $i \in \{1, 2\}$. Let $\vartheta_{ri}^{(l)} = \{\mu_{1i}^{(l)}, \beta_{1i}^{(l)}, \beta_{2i}^{(l)}\}$, $\vartheta_{gi}^{(l)} = \{\mu_{2i}^{(l)}, \beta_{3i}^{(l)}, \beta_{4i}^{(l)}\}$ and $\Sigma_i^{(l)} = \{\sigma_{ri}^{2(l)}, \sigma_{gi}^{2(l)}, \rho_i^{(l)}\}$. For computing the marginal predictive likelihood of r_{T+1}, g_{T+1}, and the joint of them; We initially draw the $s_{T+1}^{(l)}$ through $\Pi_{s_T}^{(l)}$ with given $s_T^{(l)}$. Let the draw of $s_{T+1}^{(l)} = k$ and $k \in \{1, 2\}$. We have the followings,

$$p(r_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(r_{T+1}|\vartheta_{rk}^{(l)}, \sigma_{rk}^{2(l)}) \tag{32a}$$

$$p(g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(g_{T+1}|\vartheta_{gk}^{(l)}, \sigma_{gk}^{2(l)}) \tag{32b}$$

$$p(r_{T+1}, g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{N} MN(r_{T+1}, g_{T+1}|\vartheta_{rk}^{(l)}, \vartheta_{gk}^{(l)}, \Sigma_k^{(l)}) \tag{32c}$$

5.5 IHMM-AR

Let $\{\vartheta_i^{(l)}, \sigma_i^{(l)}, \Pi^{(l)}, s_{1:T}^{(l)}, K^{(l)}, \xi^{(l)}\}$, $\vartheta_i = \{\mu_{1i}^{(l)}, \beta_{1i}^{(l)}, \beta_{2i}^{(l)}\}$ and $i \in \{1, \ldots, K^{(l)}\}$ be the lth posterior draws, where $K^{(l)}$ is the total number of active states, which means at least one observation is assigned. $s_t^{(l)} \in \{1, \ldots, K^{(l)}\}$. The steps are the following,

1. For each $l$the MCMC draw, simulate a state variable of $s_{T+1}^{(l)}$ given $s_T^{(l)}$ according to $\Pi_{s_T}^{(l)}$.
2. If $s_{T+1}^{(l)} \le K^{(l)}$, which suggests the g_{T+1} or its corresponding s_{T+1} belong to existing states, then set $(\vartheta_{s_{T+1}}^{(l)}, \sigma_{s_{T+1}}^{2(l)}) \equiv (\vartheta_k^{(l)}, \sigma_k^{2(l)})$, where $k \in (1, \ldots, K^{(l)})$. Otherwise, $(\vartheta_k^{(l)}, \sigma_k^{2(l)}) \sim H(\xi^{(l)})$ and thus, $(\vartheta_{s_{T+1}}^{(l)}, \sigma_{s_{T+1}}^{2(l)}) \equiv (\vartheta_k^{(l)}, \sigma_k^{2(l)})$. This implies the g_{T+1} belongs to a brand new state, which is a draw from the informative prior.

The predictive likelihood of g_{T+1} over all MCMC draws is the following,

$$p(g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(g_{T+1}|\vartheta_k^{(l)}, \sigma_k^{2(l)}), \tag{33}$$

The same rule is applied to computing predictive likelihood of r_{T+1} of IHMM-AR. One advantage of calculating the predictive likelihood under the IHMM is to allow the new states to be introduced for predicting any potential structural changes.

5.6 IHMM-VAR

The similar rule is applied in IHMM-VAR as we applied in IHMM-AR except all the parameters are jointly sampled in IHMM-VAR. Let $\{\vartheta_{ri}^{(l)}, \vartheta_{gi}^{(l)}, \Sigma_i^{(l)}, \Pi^{(l)}, s_{1:T}^{(l)}, K^{(l)}, \xi^{(l)}\}$ be the lth posterior draw. Some extra notations: $\vartheta_{ri}^{(l)} = \{\mu_{1i}^{(l)}, \beta_{1i}^{(l)}\beta_{2i}^{(l)}\}$ and $\vartheta_{gi}^{(l)} = \{\mu_{2i}^{(l)}, \beta_{3i}^{(l)}\beta_{4i}^{(l)}\}$, $\Sigma_i^{(l)} = \{\sigma_{ri}^{2(l)}, \sigma_{gi}^{2(l)}, \rho_i^{(l)}\}$ and $i \in \{1,\ldots,K^{(l)}\}$, $s_t^{(l)} \in \{1,\ldots,K^{(l)}\}$, and M is total number of MCMC draws that will be used for forecast inference. The predictive likelihood of r_{T+1}, g_{T+1} and the joint of them is the following,

1. For each $l$the MCMC draw, simulate a state variable for $s_{T+1}^{(l)}$ given $s_T^{(l)}$ according to $\Pi_{s_T}^{(l)}$.

2. If $s_{T+1}^{(l)} \leq K^{(l)}$, which suggests the g_{T+1} and r_{T+1} or their s_{T+1} belong to existing states, then set $(\vartheta_{rs_{T+1}}^{(l)}, \vartheta_{gs_{T+1}}^{(l)}, \Sigma_{s_{T+1}}^{(l)}) \equiv (\vartheta_{rk}^{(l)}, \vartheta_{gk}^{(l)}, \Sigma_k^{(l)})$, where $k \in \{1,\ldots,K^{(l)}\}$. Otherwise, it implies r_{T+1} and g_{T+1} to a brand new state. $(\vartheta_{rk}^{(l)}, \vartheta_{gk}^{(l)}, \Sigma_k^{(l)}) \sim H(\xi^{(l)})$ and thus, $(\vartheta_{rs_{T+1}}^{(l)}, \vartheta_{gs_{T+1}}^{(l)}, \Sigma_{s_{T+1}}^{(l)}) \equiv (\vartheta_{rk}^{(l)}, \vartheta_{gk}^{(l)}, \Sigma_k^{(l)})$.

$$p(r_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(r_{T+1}|\vartheta_{rk}^{(l)}, \sigma_{rk}^{2(l)}) \tag{34a}$$

$$p(g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} N(g_{T+1}|\vartheta_{gk}^{(l)}, \sigma_{gk}^{2(l)}) \tag{34b}$$

$$p(r_{T+1}, g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M}\sum_{l=1}^{M} MN(r_{T+1}, g_{T+1}|\vartheta_{rk}^{(l)}, \vartheta_{gk}^{(l)}, \Sigma_k^{(l)}) \tag{34c}$$

6 Empirical Results

6.1 Data

This paper reflects the data in a similar way to that of Hamilton & Lin (1996) and Fama (1990), but our sample period is largely extended. There is 1067 observations for each series, which are dated from February 1926 until December 2014. The monthly value-weighted return (includes dividend yield) of S&P500 minus the 3-month average of Fama risk-free rate (quoted at monthly rates) from CRSP are used to construct the monthly excess stock market returns (r_t), which is labelled as stock market returns in this paper. The industrial production (IP) index is from the Federal Reserve. The real growth rates (g_t) are the change of natural logarithm of IP index. The stock market returns and real growth rates are scaled by 100. Figure 1 shows the plot of corresponding stock market returns and real growth rates. Table 1 illustrates their statistical summaries.

6.2 Model Priors

Due to the fact that the hierarchical prior is not feasible on AR and VAR models[7], the priors of AR and VAR are chosen by preference. For AR model, a is a vector of zeros, and A is an identity matrix with a dimension of P. If the restricted version is applied, $P = 2$. Otherwise, $P = 3$. Let $b_1 = 5$ and $b_2 = 1$. For VAR, $\vartheta \sim MN(0, I_P)$ and $\Sigma \sim Wishart(3, I_2)$. I_P is an identity matrix with the size of P. $P = 4$ when the restricted version is applied, otherwise $P = 6$.

For MS2-AR, MS2-VAR, IHMM-AR, and IHMM-VAR models, it is feasible to implement hierarchical priors. For MS2-AR and IHMM-AR, $a \sim MN(0, I_P)$, $A^{-1} \sim Wishart(3, I_P)$, where $P = 4$ while restriction is imposed, otherwise $P = 6$. For the hyper-prior of $\sigma^2_{s_t}$, let $\chi_3 = \chi_4 = 5$ and $\nu_3 = \nu_4 = 1$.

For MS2-VAR and IHMM-VAR, h_0 is a vector of zeros, $H_0 = D_0 = I_P$, $d_0 = P + 1$, where $P = 4$ when the restricted version is applied; otherwise, $P = 6$. For other hyper priors, let $e_0 = 2$, $E_0 = I_2$, $\chi_3 = 5$ and $\nu_3 = 1$.

For hyper priors for η and α, let $\chi_1 = \chi_2 = 5$ and $\nu_1 = \nu_2 = 1$. The choices of priors and hierarchical priors are applied to both full sample estimation as well as out-of-sample forecast.

6.3 Out-of-Sample Forecast

The log-predictive likelihood is used as a measurement for model selection, and it evaluates the forecast accuracy based on a selected out-of-sample period. Let the τ_1 and τ_2 be the beginning and the end of the out-of-sample period. The log-predictive likelihood for real growth rates (LPL_g), stock market returns (LPL_r) and joint of them

[7]The hierarchical priors requires mixture models such as Markov switch with two states. The AR and VAR models implicitly assumes one state.

(LPL_{joint}) are the following,

$$LPL_r = \log \prod_{t=\tau_1}^{\tau_2} p(r_{t+1}|r_{1:t}, g_{1:t}) = \sum_{t=\tau_1}^{\tau_2} \log p(r_{t+1}|r_{1:t}, g_{1:t}) \tag{35a}$$

$$LPL_g = \log \prod_{t=\tau_1}^{\tau_2} p(g_{t+1}|r_{1:t}, g_{1:t}) = \sum_{t=\tau_1}^{\tau_2} \log p(g_{t+1}|r_{1:t}, g_{1:t}) \tag{35b}$$

$$LPL_{joint} = \log \prod_{t=\tau_1}^{\tau_2} p(r_{t+1}, g_{t+1}|r_{1:t}, g_{1:t}) = \sum_{t=\tau_1}^{\tau_2} \log p(r_{t+1}, g_{t+1}|r_{1:t}, g_{1:t}) \tag{35c}$$

The LPL_r, LPL_g and LPL_{joint} for all the models are reported in Tables 2, 3 and 4. The out-of-sample period is dated from $\tau_1 = 100$ (May, 1934) to $\tau_2 = 1067$ (December, 2014). The LPL_{joint} of the univariate setting is the summation of their LPL_r and LPL_g.

The log-predictive Bayes factor is formed by subtracting the log-predictive likelihood between any two models. For example, the log-predictive Bayes factor on stock market returns between IHMM-VAR and VAR is the LPL_r of IHMM-VAR subtracts the LPL_r of VAR, where values in excess of 5 are considered strongly in favour of the IHMM-VAR.

Instead of using a single entry to indicate the density forecast accuracy on the entire out-of-sample period, like the ones in Tables 2, 3, and 4, the cumulative log-predictive Bayes factor is a sequence of log-predictive Bayes factors and shows the predictive density accuracy on every single out-of-sample. For example, with respect equations (35a), we can compute LPL_r at $\tau_2 = 100, 101, 102 \ldots 1067$ while $\tau_1 = 100$ for any two models. It will be a a sequence of LPL_r with respect to an recursive increasing time period. Consequently, log-predictive Bayes factors with respect to $\tau_2 = 100, \ldots 1067$ are generated, and they represent the corresponding cumulative log-predictive Bayes factor for stock market returns. Figure 3 shows the cumulative log-predictive Bayes factor on stock market returns between IHMM-VAR and benchmark models. The cumulative log-predictive Bayes factor is able to tell whether or not the superior forecast performance between any two models is due to any certain period, to outliers or to steady ongoing gains.

6.4 The Out-of-Sample Forecast

The IHMM-VAR shows a significant gain in the log-predictive likelihood for real growth rates (LPL_g), stock market returns (LPL_r), and joint of them (LPL_{joint}) compared to benchmark models. The restricted version of the IHMM-VAR shows the most superior performance in terms of LPL_r, LPL_g and LPL_{joint} among all the models. These outcomes suggest the IHMM-VAR is the most preferable model among all benchmark models in terms of out-of-sample forecast accuracy.

Table 2 illustrates the LPL_{joint} of all models, where the IHMM-VAR outperform the second best model (MS2-VAR) by 77 and 98 units with respect to unrestricted and restricted versions. Similarly, table 3 shows that the IHMM-VAR is the best model in terms of LPL_r, a which shows substantial improvements with respect to second best model (MS2-VAR) by 14 and 13 with respect to unrestricted and restricted versions. Table 4 shows the LPL_g of each model, where the IHMM-VAR surpasses the second best models, the MS2-VAR, by 64 and 87 by corresponding to unrestricted and restricted versions.

The cumulative log-predictive Bayes factor are worth investigating since they can tell whether or not the superior forecast of the IHMM-VAR is due to a certain period, or to steady ongoing gains. Figure 2 illustrates the cumulative log-predictive Bayes factor on joint of stock market returns and real growth rates between the IHMM-VAR and all other benchmark models. All of the plots in Figure 2 show constant increasing trend, which suggests that the superior forecast accuracy of the IHMM-VAR in Table 2 is not due to any particular period or outliers, but steady ongoing gains. Similar outcomes are reflected in Figure 3 and Figure 4, which correspond to the cumulative log-predictive Bayes factor on stock market returns and real growth rates. In Figure 3, the subplot for the IHMM-VAR and the MS2-VAR indicates the IHMM-VAR perform better only in certain periods for predicting stock market returns.

6.5 Posterior Analysis and Prior Robustness

Figure 5 and Figure 6 shows the posterior average and interval for parameters in the IHMM-VAR of a restricted version. A lot of dynamic changes are captured in terms of posterior average and posterior interval. Notably, the Great Depression and Second World War are the two most volatile periods during the joint relationship of stock market returns and real growth rates.

The nonparametric nature of the IHMM-VAR allows the states space to be inferred through the data, so it can therefore always accommodate data with new features by introducing extra states. The histogram in Figure 7 represents the posterior distribution of states number. On average, the IHMM-VAR uses 20 states to characterize the data given the hyper-prior and hierarchical priors suggested in the previous section. Figure 8 illustrates how frequently the IHMM-VAR introduces new states; the average posterior of states number in Figure 8 is computed under a recursive increasing sample from mid 1934 until the end of 2014.

Table 5 investigates the priors sensitivity on the IHMM-VAR. The priors are inferred through the states variables under our hierarchical priors setting. We select various combination of hyper-priors on concentration parameters α and η, which governs the prior on the space of states number. The out-of-sample forecast performance of the IHMM-VAR is very robust to various choices hyper-priors in Table 5.

6.6 Structural Break at 1984

The IHMM-VAR not only detects a structural break at early 1984[8], but also shows its impact on out-of-sample forecast. Figure 9 is a heat map and shows the probability of

[8]Great Moderation.

sharing the same states between any two dates. The heat map allows us to recognize the structural break at exact dates. We observe that the months after 1984 have almost zero probability of sharing the same states with the months before 1984, which is indicative of a structural break at 1984. The issue is that the heap map relies on the state variables, which ignores actual parameters' changes underneath the state variable.

Figure 5 and Figure 6 illustrate the posterior mean and interval of parameters in the IHMM-VAR, they reflect the impact of structural break on the parameters' dynamics. Figure 5 indicates that the structural change at 1984 not only affects the posterior mean of μ_{1s_t}, μ_{2s_t}, β_{1s_t} and β_{2s_t}, but also changes its posteriors' interval after 1984. For example, the posterior of β_{2s_t} completely shifts downward with a more concentrated density after 1984. The shrinked posterior interval implies the lagged real growth rates have a more certain predictive power for future growth rates after 1984. The downward shift of the posterior mean suggests the lagged real growth rates changes its relation with future real growth rates from a positively correlated to a negatively correlated relation. As similar outcome is illustrated in Figure 6, which are the posterior mean and intervals of σ_{rs_t}, σ_{gs_t} and ρ_{s_t}. Both figure 5 and figure 6 suggest the significant impact of structural break on parameter changes; however, they do not imply that the structural break has a significant impact in an out-of-sample forecast. In other words, given that the 1984 structural break does exist, and that it is captured by the IHMM-VAR while other benchmark models which fail to do, the IHMM-VAR should indicate a significant improvement on the out-of-sample forecast accuracy with respect to approaches fail to model the structural break.

Figure 10 shows the cumulative log-predictive Bayes factor between the IHMM-VAR and the MS2-VAR on real growth rates. It is clear that the IHMM-VAR is constantly in favoured of the out-of-sample performance. Notably, the IHMM-VAR is in favour more, while the model predicts the period after 1984 structural change, and this is why the cumulative log-predictive Bayes factor shows a steeper slop after 1984. This implies that the IHMM-VAR not only captures the dynamic changes through the posterior of parameters and state variables, but also shows that what have been captured significantly contribute to out-of-sample density forecast accuracy. As mentioned early, the IHMM-VAR is able to automatically introduce new states to accommodate any structural changes, therefore compared to Markov switching with two-state model, the IHMM-VAR can deliver a much more superior out-of-sample forecast. In contrast to Kim & Nelson (1999), this paper documents the impact of the 1984 structural break in parameter changes as well as in out-of-sample forecast.

6.7 Empirical Results

The asset pricing literature suggests[9] the stock market returns should have predictive power for future economic growths, while existing works suggest the lag stock market

[9]According to dividend discount valuation and consumption capital asset pricing model.

returns variables should have predictive power on real growth rates[10]. This paper does not find any strong evidence to support this; however, we find other important empirical evidence to support that the stock market returns can help to predict future real growth rates. This paper documents that the common unobserved states variables actually capture the most predictive power for future growth rates, rather than the lagged stock market returns.

Table 6 is the log-predictive Bayes factor between unrestricted and restricted versions of each model. Table 6 shows whether or not adding lagged stock market returns variable can help to predict real growth rates. Mixed evidence of predictive power of lagged stock market returns for real growth rates are found in benchmark models. Given the most preferred model of IHMM-VAR[11], no evidence shows the r_{t-1} has predictive power for g_t. Similar outcomes happen to MS2-VAR, and AR, which suggests inconsistent results with respect to Fama (1990) and Choi et al. (1999). On the another side, Table 6 suggests that lagged stock market returns are significantly accounted for predicting future real growth rates in the MS2-AR, IHMM-AR and VAR models. They are consistent with Lee (1992), Hassapis & Kalyvitis (2002), and Kim & In (2003). However, these models are less reliable according to Table 4, and it reveals that these models are poorly performed in term of out-of-sample forecast accuracy. Table 3 suggests the lagged real growth rates (g_{t-1}) do not have any predictive power for stock market returns (r_t) for any models.

According to Table 4, the out-of-sample forecast on real growth rates is remarkably improved from the AR to the IHMM-VAR. The IHMM-AR shows a remarkable gain in LPL_g with respect to the AR model. The gains are 903 and 873 with respect to unrestricted and restricted versions. Under this univariate setting, the IHMM-AR fully explores the regimes dynamics without considering the contemporaneous relationship; the gain in density forecast accuracy implies the necessity of modeling the parameter regime-dependence. On the other hand, the VAR model shows a remarkable increase of 894 (unrestricted version) and 879 (restricted version) with respect to the AR model. This suggests modeling the contemporaneous relationship is a critical component for capturing the dynamic relationship of stock market returns and real growth rates. The IHMM-VAR delivers a significant gain in LPL_g with respect to the IHMM-AR and VAR in Table 4. Given that the restricted version of the IHMM-VAR is the best performed model, there is no evidence to support that lagged stock market returns should help to predict future real growth rates. But the IHMM-VAR illustrates that the predictive power of stock market returns for future real growth rates are captured by the the unobserved Markov states variables which are shared by two time series, rather than the lagged stock market returns variables.

[10] Fama (1990), Schwert (1990), Choi et al. (1999) and Kanas & Ioannidis (2010) suggest the lagged stock market returns should help to predict future real growth rate

[11] Based on the out-of-sample forecast performance in Table 2, Table 3 and Table 4.

7 Conclusion

This paper proposes a Bayesian nonparametric model that allows the conditional distribution of stock market returns and real growth rates to be an joint unknown distribution. Once having applied this new model to monthly U.S stock market excess returns and real growth rates, I discover significant parameter changes over time. The new model significantly improves the out-of-sample density forecast accuracy. The paper finds recurring regimes as well as structural break. This paper does not find robust evidence to support that the lagged stock market returns predict real growth rates. However, we find that the predictive power of stock market returns on real growth rates are captured by the unobserved Markov states variables, rather than the lagged stock market returns variables.

References

Albert, J. H. & Chib, S. (1993), 'Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts', *Journal of Business & Economic Statistics* **11**(1), 1–15.

Beal, M., Ghahramani, Z. & Rasmussen, C. (2002), 'The infinite hidden markov model', *In Advances in Neural Information Processing Systems* pp. 577–584.

Chib, S. (1996), 'Calculating posterior distributions and modal estimates in Markov mixture models', *Journal of Econometrics* **75**, 79–97.

Choi, J., Hauser, S. & Kopecky, K. (1999), 'Does the stock market predict real activity? Time series evidence from the G-7 countries', *Journal of banking and finance* **23**, 1771–1792.

Dufays, A. (2015), 'Infinite-state markov-switching for dynamic volatility', *Journal of Financial Econometrics* .

Fama, E. (1990), 'Stock returns, expected returns, and real activity', *Journal of Finance* **4**, 1089–1108.

Ferguson, T. S. (1973), 'A bayesian analysis of some nonparametric problems', *The Annals of Statistics* **1**, 209–230.

Hamilton, J. & Lin, G. (1996), 'Stock market volatility and the business cycle', *Journal of Applied Econometrics* **1**, 573–593.

Hassapis, C. & Kalyvitis, S. (2002), 'Investigating the links between growth and real stock price changes with empirical evidence form the G-7 economies', *The Quarterly Review of Economics and Finance* **42**, 543–575.

Jensen, M. & Maheu, M. (2010), 'Bayesian semiparametric stochastic volatility modeling', *Journal of Econometrics* **157**, 306–316.

Jensen, M. & Maheu, M. (2013), 'Bayesian semiparametric multivariate garch modeling', *Journal of Econometrics* **176**, 3–17.

Jensen, M. & Maheu, M. (2014), 'Estimating a semiparametric asymmetric stochastic volatility model with a dirichlet process mixture', *Journal of Econometrics* **178**, 523–538.

Jochmann, M. (2015), 'Modeling U.S. inflation dynamics: a bayesian nonparametric approach', *Econometric Reviews* **34**.

Kanas, A. & Ioannidis, C. (2010), 'Causality from real stock returns to real activity: evidence of regime-dependence', *International Journal of Finance and Economics* **15**, 180–197.

Kim, C. & Nelson, C. (1999), 'Has the US economy become more stable? a bayesian approach based on a markov-switching model of the business cycle', *The Review of Economics and Statistics* **81**, 608–616.

Kim, S. & In, F. (2003), 'The relationship between financial variables and real economic activity: evidence from spectral and wavelet analyses', *Studies in Nonlinear Dynamics and Econometrics* **7**.

Koop, G. & Korobilis, D. (2009), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics* **3**, 267–358.

Lee, B. (1992), 'Causal relations among stock returns, interest rates, real activity, and inflation', *The Journal of Finance* **48**, 1591–1604.

Maheu, M. & Yang, Q. (2015), An infinite hidden Markov model for short-term interest rates. working paper.

Schwert, G. (1990), 'Stock returns and real economic activity: A century of evidence', *Journal of Finance* **45**, 1237–1257.

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

Shi, S. & Song, Y. (2015), 'Identifying speculative bubbles using an infinite hidden markov model', *Journal of Financial Econometrics* .

Song, Y. (2013), 'Modelling regime switching and structural breaks with an infinite hidden Markov model', *Journal of Applied Econometrics* **29**.

Teh, Y., Jordan, M., Beal, M. & Blei, D. (2006), 'Hierarchical dirichlet processes', *Journal of the American Statistical Association* **101**, 1566–1581.

Van Gael, J., Saatci, Y., Teh, Y. & Ghahramani, Z. (2008), Beam sampling for the infinite hidden markov model, *in* 'Proceedings of the 25th International Conference on Machine Learning:', pp. 1088–1095.

Walker, S. (2007), 'Sampling the dirichlet mixture model with slices', *Communications in Statistics Simulation and Computation* **36**, 123–145.

8 Sampler Steps

The details of sampling steps are described in this section. It is for the vector autoregressive infinite hidden Markov model (IHMM-VAR). $y_{1:T}$ are observations and $y_t = (r_t, g_t)$. The 1st MCMC draw is randomly initialized state variables $s_{1:T}$ and Γ. Some notations: $\theta_i = (\mu_{1i}, \beta_{1i}, \beta_{2i}, \mu_{2i}, \beta_{3i}, \beta_{4i}, \Sigma_i)$. K is the total number of acting states.

1. Sample $c_{1:K}$,

   (a) Draw $x_l \sim Bernoulli(\frac{\alpha\gamma_i}{l-1+\alpha\gamma_i})$, for $l = 1, \ldots, n_{ji}$: if $x_l = 1$ increment o_{ji}, for $\{\{n_{ji}\}_{j=1}^{K}\}_{i=1}^{K}$.

   (b) Compute $c_j = \sum_{i=1}^{K} o_{ji}$

2. Sample α given χ_2, and ν_2:

   Two auxiliary variables, $\bar{\nu}$ and $\bar{\lambda}$.

   (a) $\bar{\nu}_j \sim Bernoulli(\frac{n_{j.}}{n_{j.}+\alpha})$ for $j = 1, \ldots, K$. $\quad n_{j.} = \sum_{i=1}^{K} n_{ji}$

   (b) $\bar{\lambda}_j \sim Beta(\alpha + 1, n_{j.})$ for $j = 1, \ldots, K$.

   (c) $\alpha \sim Gamma\big(\chi_2 + c_{..} - \sum_{j=1}^{K} \bar{\nu}_j, \nu_2 - \sum_{j=1}^{K} log(\bar{\lambda}_j)\big)$

   $c_{..} = \sum_{j=1}^{K} c_j$

3. Sample η given χ_1 and ν_1.

   Two auxiliary variables are used in here, , $\bar{\nu}$ and $\bar{\lambda}$.

   (a) $\bar{\nu} \sim Bernoulli(\frac{c_{..}}{c_{..}+\eta})$

   (b) $\bar{\lambda} \sim Beta(\eta + 1, c_{..})$

   (c) $\eta \sim Gamma(\chi_1 + \overline{K} - \bar{\nu}, \nu_1 - log\bar{\lambda})$,

4. Sample Γ:

$$\Gamma = (\gamma_1, \ldots, \gamma_K, \sum_{l=K+1}^{\infty} \gamma_l) | c_{1:K}, \eta \sim Dir(c_1, c_2, \ldots, c_K, \eta)$$

   $\gamma_{K+1} = \sum_{l=K+1}^{\infty} \gamma_l$

5. Sample Π_j: for $j = 1, \ldots, K$

$$\Pi_j = \pi_{j1}, \pi_{j2}, \ldots, \pi_{jK+1} | \alpha, n_{j,1:K} \sim Dir(\alpha\gamma_1 + n_{j1}, \ldots, \alpha\gamma_K + n_{jK}, \alpha\gamma_{K+1})$$

6. Sample $u_{1:T}$:

$$f(u_{t+1}|s_{t+1}, s_t) = \begin{cases} \frac{I(u_t < \pi_{s_t s_{t+1}})}{\pi_{s_t s_{t+1}}} & u_t \leq \pi_{s_t, s_{t+1}} \\ 0 & u_t > \pi_{s_t, s_{t+1}} \end{cases}$$

   Note, $u_1 \sim U[0, \gamma_{s_1}]$ and $u_t \sim Uniform[0, \pi_{s_{t-1}, s_t}]$

7. Adaptive Truncation of Π:
   If $\max(\{\pi_{j,K+1}\}_{j=1}^{K}) > \min(u_{1:T})$, we do the following steps:

   (a) $(\pi_{K+1,1}, \ldots, \pi_{K+1,K+1}) \sim Dir(\alpha\gamma_1, \ldots, \alpha\gamma_{K+1})$.

   (b) Increment the $\boldsymbol{\gamma}$ to size of K+2 with

   $$\tau \sim Beta(1, \eta) \quad and \quad \gamma_{K+2} = (1-\tau)\gamma_{K+1} \quad and \quad \gamma_{K+1} = \tau\gamma_{K+1}$$

   (c) Extend the $\left\{\{\pi_{ji}\}_{j=1}^{K}\right\}_{i=1}^{K+1}$ to $\left\{\{\pi_{ji}\}_{j=1}^{K+1}\right\}_{i=1}^{K+2}$ by

   $$\tau_j \sim Beta(\alpha\gamma_{K+1,}\alpha\gamma_{K+2}) \quad \pi_{j,K+1} = \tau_j\pi_{j,K+1} \quad \pi_{j,K+2} = (1-\tau_j)\pi_{j,K+1}$$

   Once the new state is generated, the corresponding new parameter set will be assigned, such as $\theta_{K+1} \sim H(\xi)$.

   $K \leftarrow K+1$ and keep repeating steps from a) to c) until the requirement is satisfied. Intuitively, we adaptively truncate Π using $u_{1:T}$ from infinite dimension into finite dimension denoted by $\bar{\boldsymbol{\Pi}}$ as well as the associated full parameter space Θ into a set of finite set is denoted by $\bar{\Theta}$[12].

8. Sample $s_{1:T}$

   (a) Initial step for s_1:

   $$p(s_1 = k|y_1, \theta_k) \propto f(y_1|s_1 = k, \theta_k) \sum_{i=1}^{K} I(\gamma_i > u_1) \quad \text{for } k = 1, \ldots, K$$

   $\pi_{s_{t-1}^i, s_t^k}$ stands for $\pi_{s_{t-1}=i, s_t=k}$

   (b) The forward-filtering part for $s_{2:T}$:

   $$p(s_t = k|y_t, \theta_k) \propto f(y_t|s_t = k, \theta_k) \sum_{i=1}^{K} I(\pi_{s_{t-1}^i, s_t^k} > u_t) p(s_{t-1} = i|y_{t-1}, \theta_i)$$

   do $k = 1, \ldots, K$ for each $t = 2, \ldots, T$.

   $$\mathbf{F} = \begin{pmatrix} p(s_1 = 1|y_1, \theta_1) & p(s_1 = 2|y_1, \theta_2) & \ldots & p(s_1 = K|y_1, \theta_K) \\ \vdots & \vdots & \ddots & \vdots \\ p(s_T = 1|y_T, \theta_1) & p(s_T = 2|y_T, \theta_2) & \ldots & p(s_T = K|y_T, \theta_K) \end{pmatrix}.$$

[12] A finite number of θ which will be considered in the FFBS, includes all alive states and finite number of unrepresented states.

(c) Sample initial s_T:

$$s_T \propto p(s_T = k|y_T, \theta_k) \qquad \text{for } k = 1, \dots, K$$

(d) Sample $s_{T-1:1}$ recursively. i indicates previous state:

$$P(s_t = k|s_{t+1} = i, y_t, \theta_k) \propto I(\pi_{s_t^k, s_{t+1}^i} > u_{t+1})F_{t,k}$$

for $t = T-1, \dots, 1$.

9. Sample $\theta_{1:K}$: Let $\vartheta_k = (\mu_{1k}, \beta_{1k}, \beta_{2k}, \mu_{2k}, \beta_{3k}, \beta_{4k})'$ for k=1..…,K. Let T_k be the total number observations are assigned to state k. $y_t = (r_t, g_t)'$.

$$y_t = Z_t \vartheta_{s_t} + \epsilon_t \qquad \epsilon_t \sim MN(0, \Sigma_{s_t}),$$

where,

$$Z_t = \begin{bmatrix} 1 & y_{t-1} & 0 & 0 \\ 0 & 0 & 1 & y_{t-1} \end{bmatrix}$$

The posterior of ϑ_k is the following,

$$\vartheta_k|y_t, Z_t, \Sigma_k^{-1}, a, A \sim MN(a_1, A_1) \quad k = 1, \dots, L$$
$$a_1 = A_1\Big(A^{-1}a + \sum_{s_t=k} Z_t'\Sigma_k^{-1}y_t\Big) \quad A_1 = \Big(A^{-1} + \sum_{s_t=k} Z_t'\Sigma_k^{-1}Z_t\Big)^{-1}$$

The posterior of Σ_k^{-1} is the following,

$$\Sigma_k^{-1}|\vartheta_k, y_t, Z_t, b, B \sim Wishart(b_1, B_1)$$
$$b_1 = T_k + b \quad B_1 = \Big(B^{-1} + \sum_{s_t=k}(y_t - Z_t\vartheta_k)(y_t - Z_t\vartheta_k)'\Big)^{-1}$$

10. Sample the priors through hierarchical priors. The priors are following:

$$\vartheta_k \sim MN(a, A) \qquad \Sigma_i^{-1} \sim Wishart(b, B) \quad k = 1, , \dots, K$$

The hierarchical priors the following:

$$a \sim MN(h_0, H_0) \qquad A^{-1} \sim Wishart(d_0, D_0)$$
$$B \sim InvWhishart(e_0, E_0) \qquad b \sim Gamma(\chi_0, \nu_0)I(b \geq 2)$$

(a) Sample a:

$$a|h_0, H_0, A, \{\vartheta_k\}_{k=1}^K \sim MN\big(\mu_a, V_a\big)$$
$$\mu_a = V_a\big(H_0^{-1}h_0 + A^{-1}\sum_1^K \vartheta_k\big) \qquad V_a = \big(H_0^{-1} + KA^{-1}\big)^{-1}$$

(b) Sample A:

$$A^{-1}|d_0, D_0, a, \{\vartheta_k\}_{k=1}^{K} \sim Wishart(\omega_A, \Omega_A)$$

$$\omega_A = K + d_0 \qquad \Omega_A = \Big(D_0^{-1} + \sum_{k=1}^{K}(\vartheta_k - a)(\vartheta_k - a)'\Big)^{-1}$$

(c) Sample B:

$$B|b, e_0, E_0, \{\Sigma_i\}_1^K \sim InvWishart(bK + e_0, E_0 + \sum_{k=1}^{K}\Sigma_k^{-1})$$

(d) Sample b:

Due to non-conjugate property, the Metropolis-Hastings is applied with proposal a density. We choose ζ to arrives at reasonable accept frequencies. M is the dimension of the data.

$$\pi(b|\chi_0, \nu_0, B, \{\Sigma_k\}_{k=1}^{L}) = G(b|\chi_0, \nu_0)\prod_{1}^{K} Wishart(\Sigma_k^{-1}|b, B)$$

$$q(b^{new}|b^{old}) \sim Gamma(\zeta, \zeta/b^{old})I(b^{new} \geq M+1)$$

$$AcceptProbability = \min\Big[\frac{\pi(b^{new}|\chi_0, \nu_0, \{\Sigma_k\}_1^K)/q(b^{new}|b^{old})}{\pi(b^{old}|\chi_0, \nu_0, \{\Sigma_k\}_1^K)/q(b^{old}|b^{new})}, 1\Big]$$

11. Repeat 1-10.

Table 1: Statistical Summaries of stock market returns and Real Growth Rates

Name	Mean	Variance	Median	25%Q	75%Q	Skewness	Kurtosis
stock market returns	0.648	30.24	0.945	-2.043	3.595	0.409	9.603
Real Growth Rates	0.263	3.211	0.308	-0.304	0.846	0.354	14.87

This table reports summary statistics for monthly S&P 500 stock excess returns and U.S. industrial production growth rates from February 1926 to December 2014 (1067 observations).

Table 2: Log-Predictive Likelihood on Joint of Stock Market Returns and Real Growth (LPL_{joint})

Univariate	Unrestricted	Restricted
AR	-6591	-6589
MS2-AR	-4689	-4721
IHMM-AR	-4672	-4701
Bivariate	**Unrestricted**	**Restricted**
VAR	-4659	-4672
MS2-VAR	-4230	-4165
IHMM-VAR	-4153	-4067

The out-of-sample period is from May 1934 to December 2014 (with a size of 967). The full sample size is 1067. The IHMM and MS2 are corresponding to the infinite Hidden Markov models and two-state Markov switching model. The AR and VAR imply autoregression and vector autoregression.

Table 3: Log-Predictive Likelihood on Stock Market Returns (LPL_r)

Models	Unrestricted	Restricted
AR	-3986	-3986
MS2-AR	-2973	-2974
IHMM-AR	-2970	-2971
VAR	-2954	-2951
MS2-VAR	-2810	-2807
IHMM-VAR	-2796	-2794

The out-of-sample period is from May 1934 to December 2014 (with a size of 967). The full sample size is 1067. The difference in LPL_r between the unrestricted and restricted versions are to distinguish if the past real growth rates have predictive power on future stock market returns. The AR, MS2-AR and IHMM-AR are under univariate setting. The VAR, MS2-VAR, and IHMM-VAR belong to a bivariate setting.

Table 4: Log-Predictive Likelihood on Real Growth Rates (LPL_g)

	Unrestricted	Restricted
AR	-2605	-2603
MS2-AR	-1716	-1747
IHMM-AR	-1702	-1730
VAR	-1711	-1724
MS2-VAR	-1416	-1360
IHMM-VAR	-1352	-1273

The out-of-sample period is from May 1934 to December 2014 (with a size of 967). The full sample size is 1067. The difference in LPL_r between the unrestricted and restricted versions is to distinguish if the past real growth rates have predictive power on the future stock market returns. The AR, MS2-AR and IHMM-AR are under a univariate setting. The VAR, MS2-VAR and IHMM-VAR are belong to a bivariate setting.

Table 5: Log-Predictive Likelihood on Joint Stock Returns and Growth Rates of IHMM-VAR (Restricted Version) with Different Hyper Priors

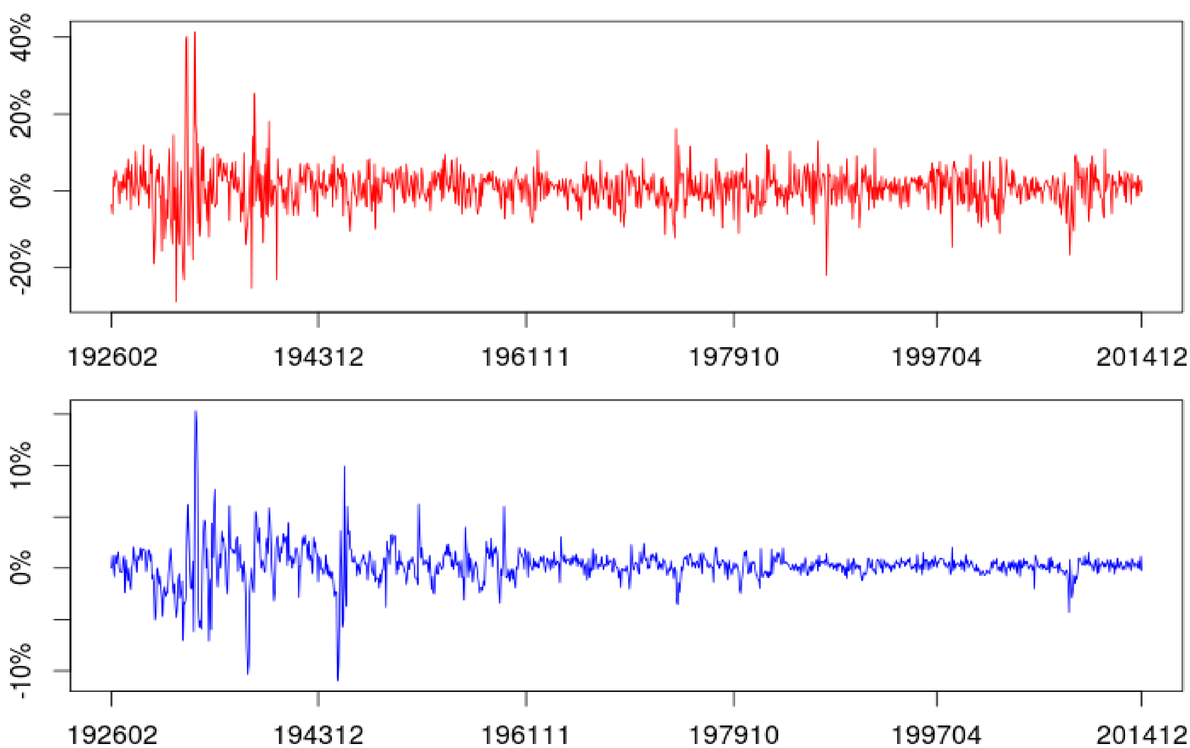
Very Loose	$\chi_1 = 5, \nu_1 = 1$ and $\chi_1 = 5, \nu_1 = 1$	-4067
Less Loose	$\chi_1 = 5, \nu_1 = 1$ and $\chi_1 = 2.5, \nu_1 = 0.5$	-4066
Tight	$\chi_1 = 2, \nu_1 = 1$ and $\chi_1 = 5, \nu_1 = 1$	-4064
Less Tight	$\chi_1 = 5, \nu_1 = 1$ and $\chi_1 = 2, \nu_1 = 8$	-4068
Very Tight	$\chi_1 = 2, \nu_1 = 2$ and $\chi_1 = 2, \nu_1 = 8$	-4069

Table 6: Predictive Power of lagged stock market returns on Future Real Growth Rates

	Log-Predictive Bayes Factor	Do Lagged Stock Returns Contribute to Forecast Real Growths?
AR	-2	No
MS2-AR	31	Yes
IHMM-AR	28	Yes
VAR	13	Yes
MS2-VAR	-6	No
IHMM-VAR	-18	No

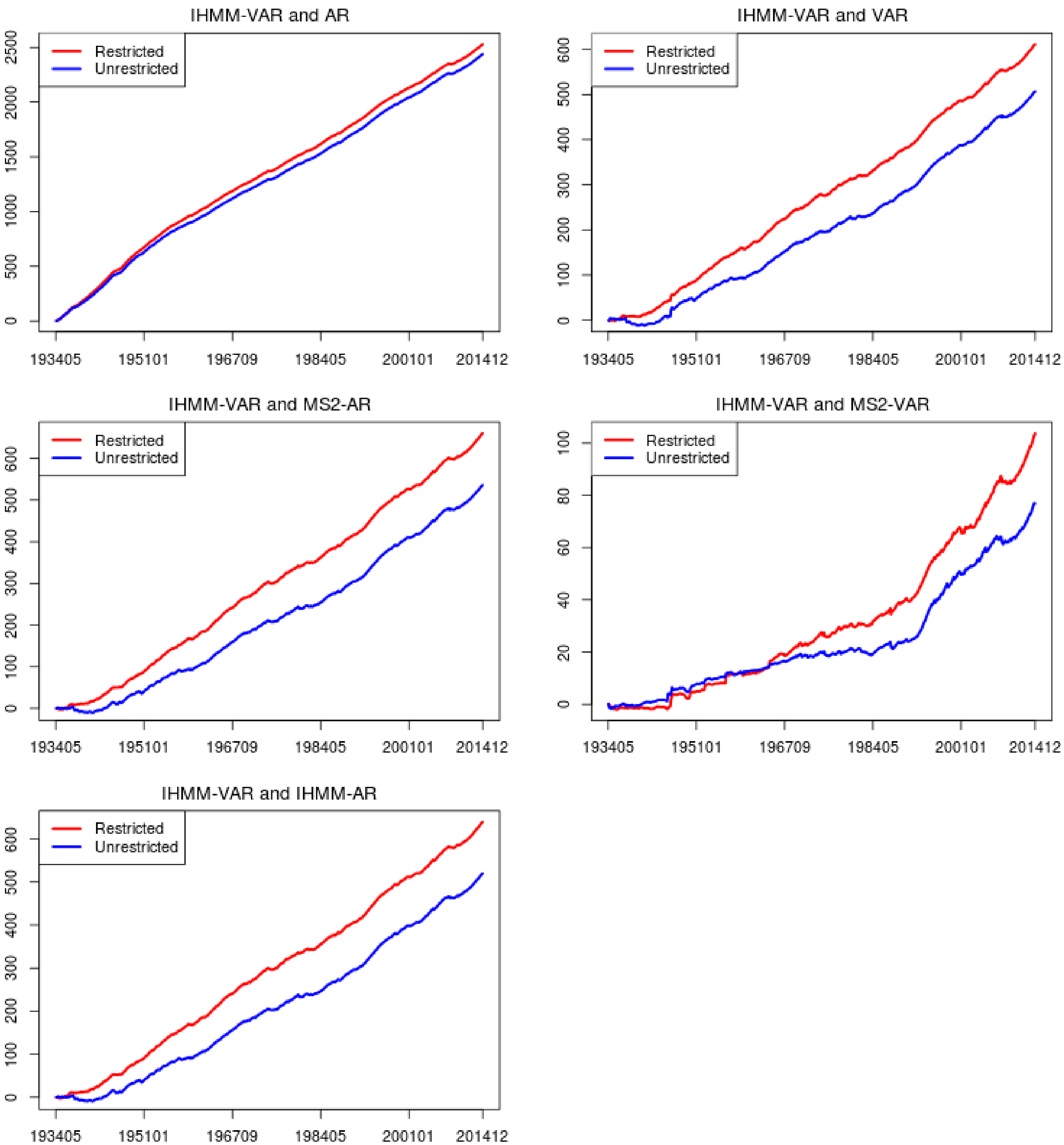
The log-predictive Bayes factor is the log-predictive likelihood of the unrestricted version subtracted by the restricted version. If the log-predictive Bayes factor is less than 5, this suggests the lagged stock market returns DOES NOT have predictive power on the future real growth rates. The out-of-sample period is from May 1934 until December 2014 (with a size of 967). The full sample size is 1067.

Figure 1: S&P 500 Monthly Excess stock market returns (Top)
U.S Industrial Production Growth Rates (Bottom)



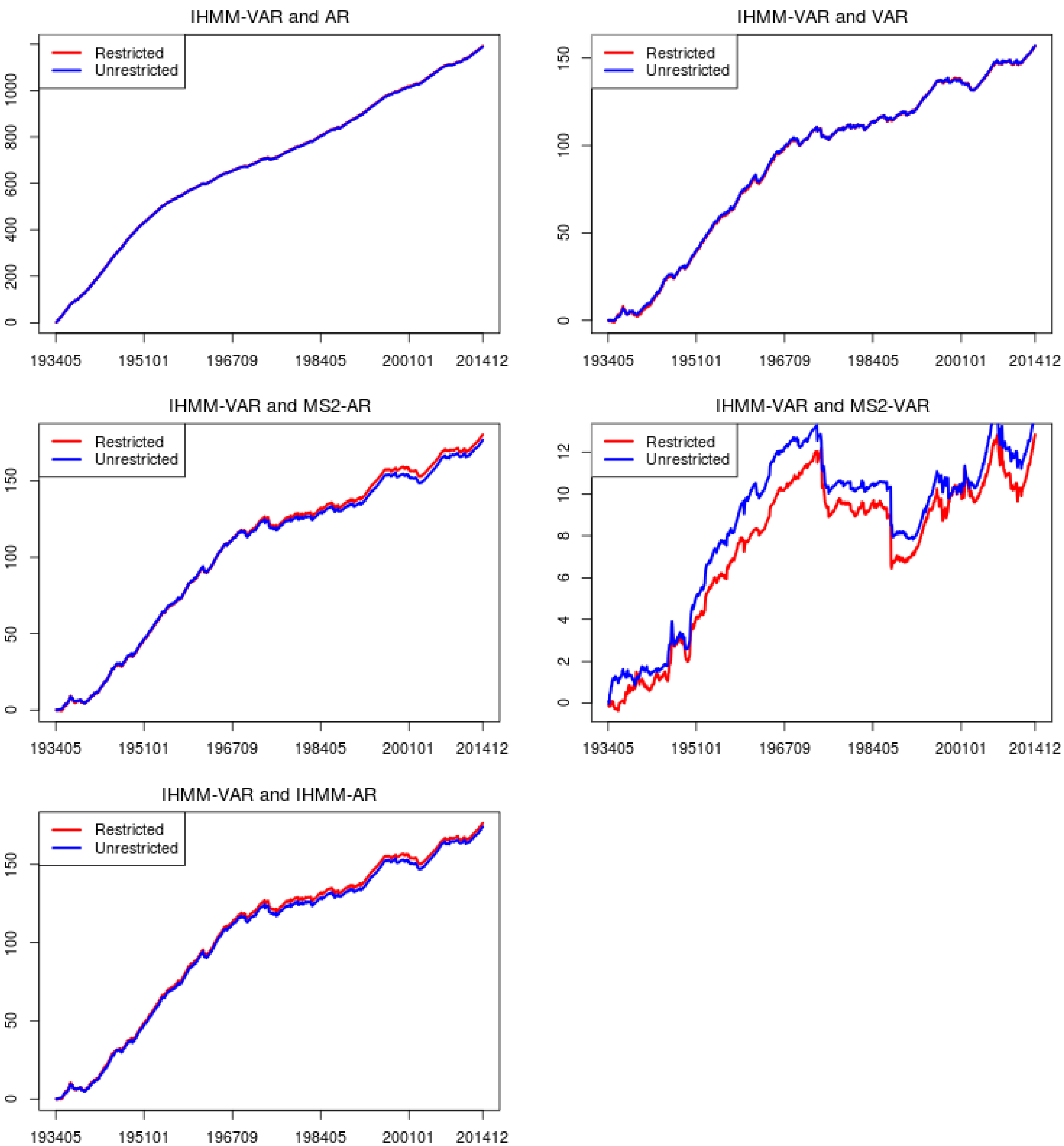
Note: The horizontal axis represents the years and months.

Figure 2: Cumulative Log-Predictive Bayes Factor on Joint Stock Market Returns and Real Growth Rates



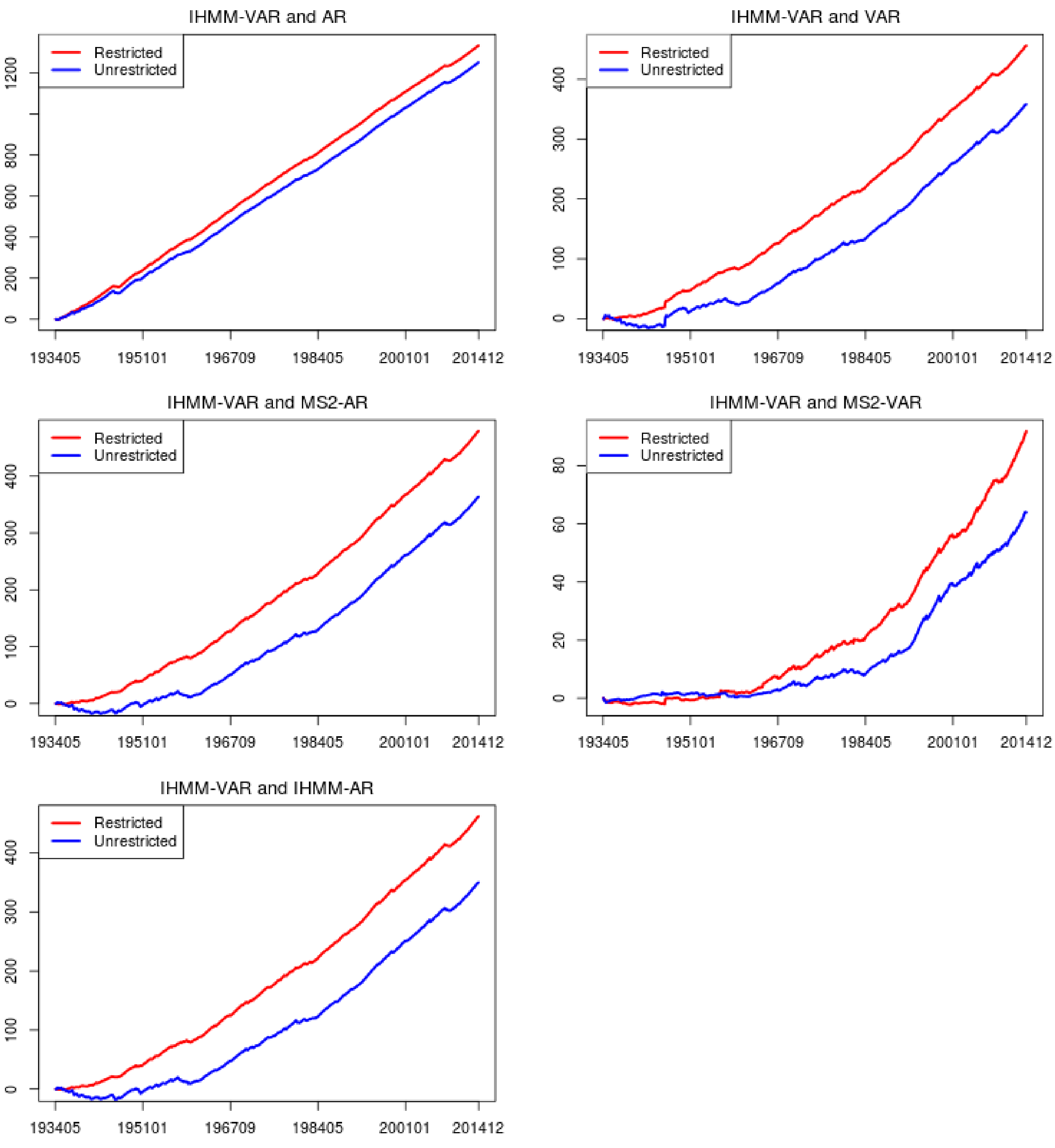
The horizontal and vertical axis represent corresponding dates and log-likelihood. Each plot shows the cumulative log-predictive Bayes factor between IHMM-VAR and the corresponding benchmark model. The out-of-sample period is from May 1934 to December 2014 (with a size of 967). The full sample size is 1067.

Figure 3: Cumulative Log-Predictive Bayes Factor on Stock Market Returns



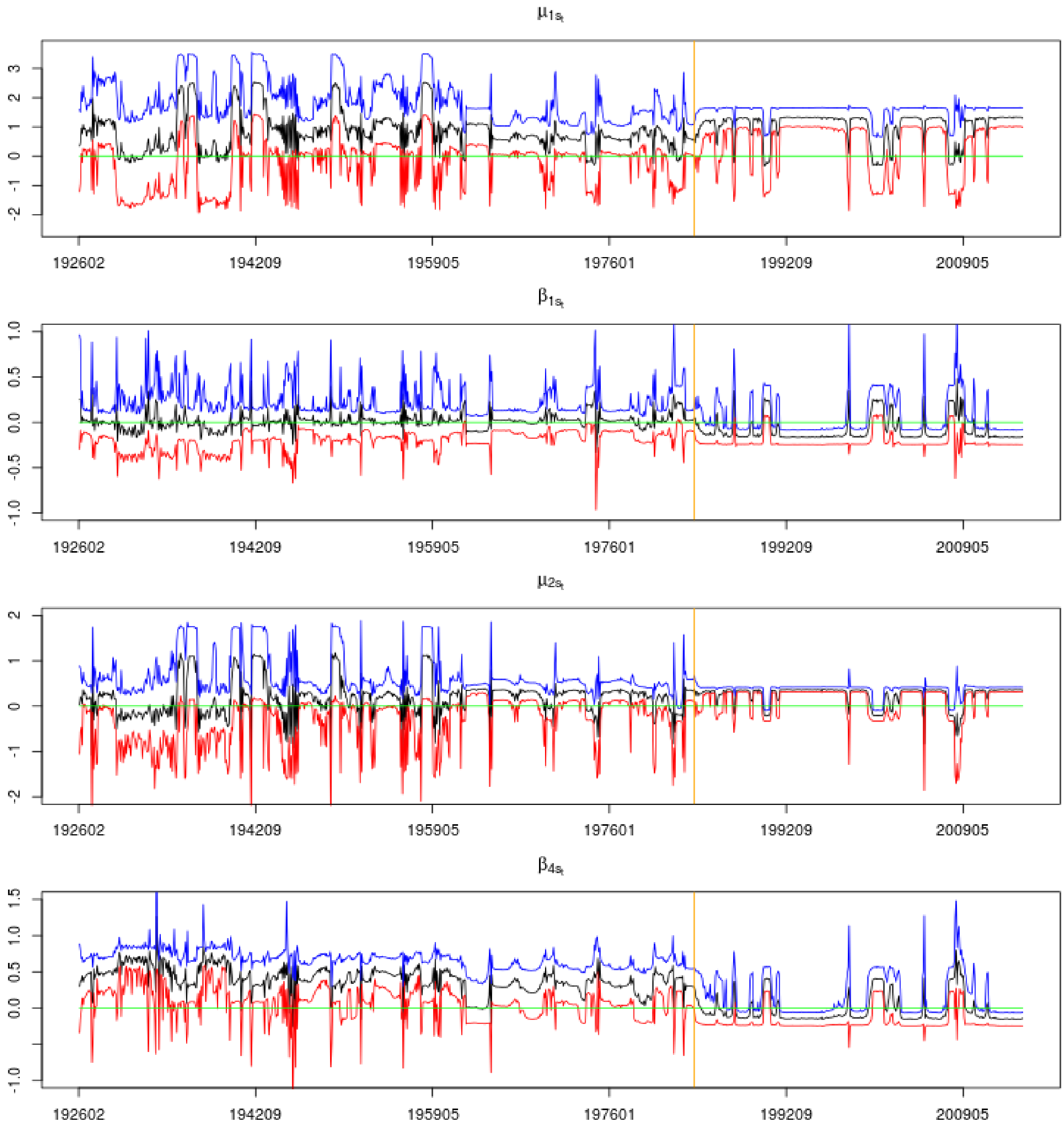
The horizontal and vertical axis represent corresponding dates and log-likelihood. Each plot shows the cumulative log-predictive Bayes factor between IHMM-VAR and the corresponding benchmark model. The out-of-sample period is from May 1934 to December 2014 (with a size of 967). The full sample size is 1067.

Figure 4: Cumulative Log-Predictive Bayes Factor on Real Growth Rates



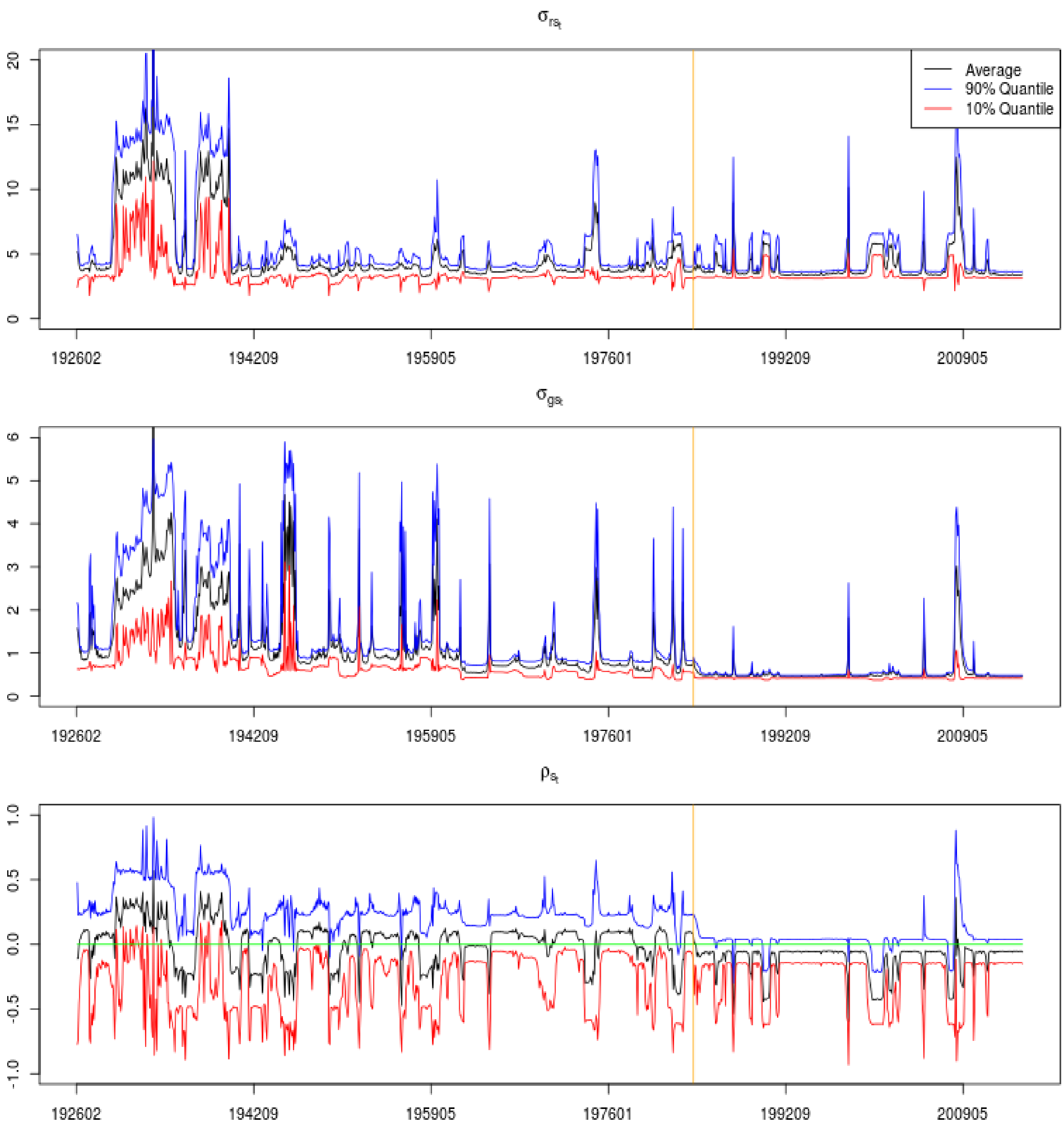
The horizontal and vertical axis represent corresponding dates and log-likelihood. Each plot shows the cumulative log-predictive Bayes factor between IHMM-VAR and the corresponding benchmark model. The out-of-sample period is from May 1934 to December 2014 (with a size of 967). The full sample size is 1067.

Figure 5: Posterior of Parameter for the IHMM-VAR (Restricted Version)



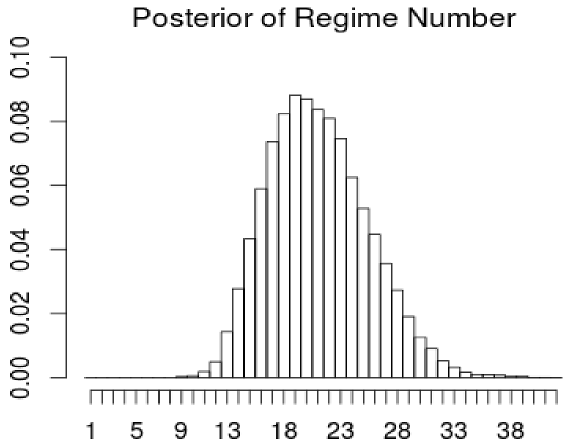
The red and blue lines are corresponding 10% and 90% posterior interval. The black line is the posterior average. The vertical orange line indicates the structural break date. The horizontal green line indicates zero. The estimations are under the IHMM-VAR (Restricted Version) and based on the full sample. Comparing with the unrestricted version of IHMM-VAR, the restricted version $\beta_{2s_t} = 0$ and $\beta_{3s_t} = 0$.

Figure 6: Posterior of Parameter for the IHMM-VAR (Restricted Version)



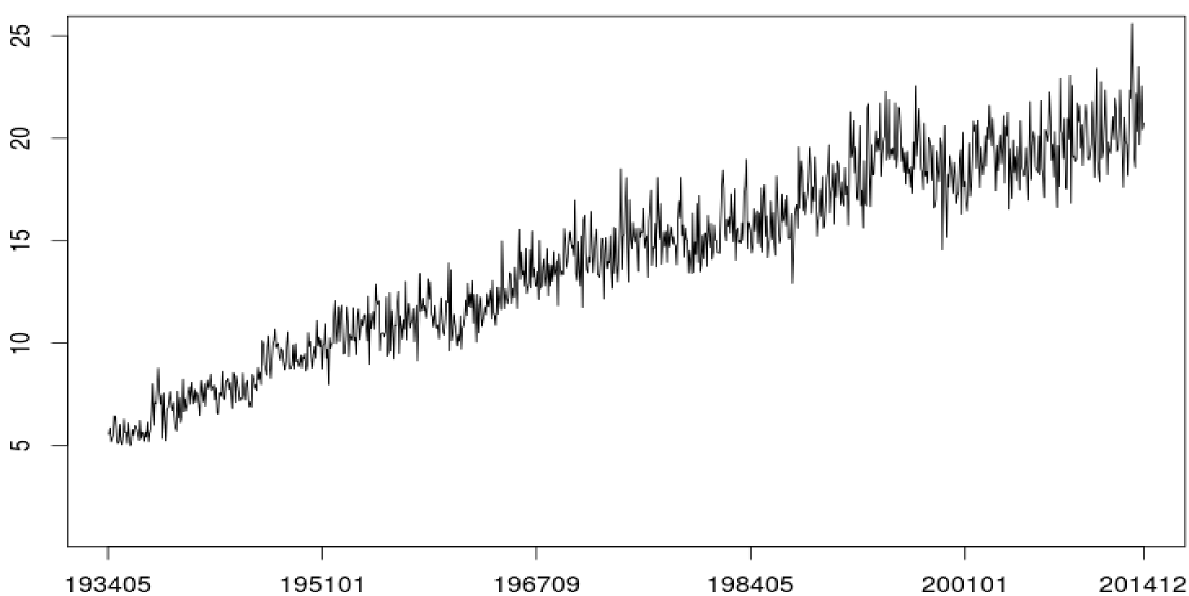
The red and blue lines are corresponding 10% and 90% posterior interval. The black line is the posterior average. The vertical orange line indicates the structural break date. The horizontal green line indicates zero. The estimations are under the IHMM-VAR (Restricted Version) and based on the full sample.

Figure 7



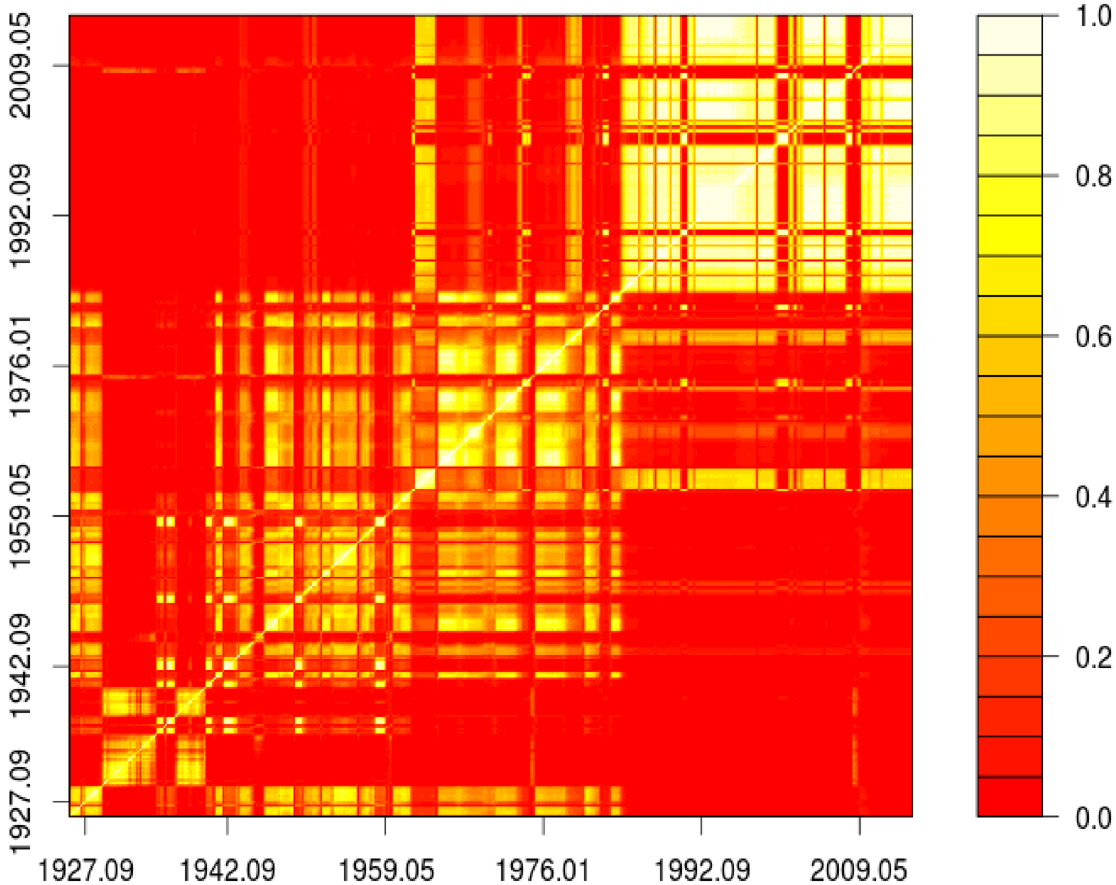
The histogram shows the posterior of regime number. The estimations are based on IHMM-VAR(Restricted Version) and full samples.

Figure 8: Regime Dynamics under Recursive Increasing Sample



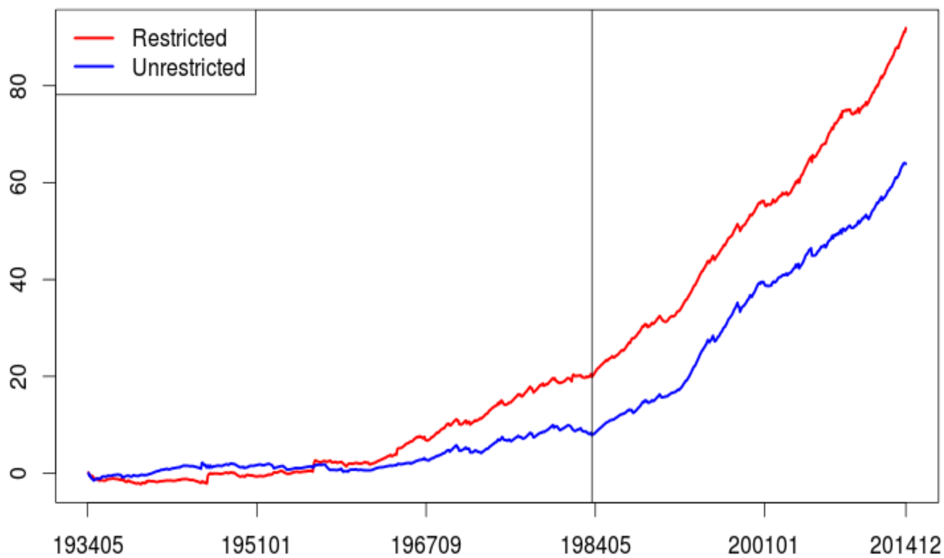
This figure shows the posterior average of the regime number under a recursive increasing sample as indicated on the x-axis. The estimations are based on IHMM-VAR (Restricted Version).

Figure 9: Heat Map



The color represents the probability level. The map is symmetric through the 45 degree diagonal line originated from the bottom left. The label switching does not matter here since we only care if the two dates share the same state at each MCMC. A dummy variable is used to indicate if they share the same regime. Then, the summation of the dummies for those two particular dates is divided by the total number of MCMC, which represents the probability of sharing the same state of those two dates. This heat map is estimated based on IHMM-VAR (Restricted Version).

Figure 10: Cumulative Log-Predictive Bayes Factor Plot Between IHMM-VAR and MS2-VAR



Each plot has a vertical black line, which indicates the month of structural change. Restricted implies the cumulative log-predictive Bayes factor of restricted version between the IHMM-VAR and MS2-VAR. The same rule applies to the unrestricted version.