Bayesian Applications in Financial Econometrics

by

Qiao Yang

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Economics
University of Toronto

© Copyright 2016 by Qiao Yang
Abstract

Bayesian Applications in Financial Econometrics

Qiao Yang
Doctor of Philosophy
Graduate Department of Economics
University of Toronto
2016

This thesis consists of three chapters in Bayesian financial econometrics. The three chapters apply both Bayesian nonparametric and parametric methods to financial market and macroeconomic time series.

Chapter 1 extends popular discrete time short-rate models to include Markov switching of infinite dimension. This is a Bayesian nonparametric model that allows for changes in the unknown conditional distribution over time. Applied to weekly U.S. data we find significant parameter change over time and strong evidence of non-Gaussian conditional distributions. Our new model with an hierarchical prior provides significant improvements in density forecasts as well as point forecasts. We find evidence of recurring regimes as well as structural breaks in the empirical application.

Chapter 2 studies the joint dynamic behaviour between stock market returns and real economic growth rates. Their relationship is an important empirical question in finance and macroeconomics. This chapter 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.

Chapter 3 studies the predictive power of oil price information for forecasting the U.S. industrial production. Oil price information is divided into two distinct categories: nominal oil price changes, and oil price shocks using four definitions proposed in the literature. Previous work has documented lack of predictive relationship of oil price changes but significant predictive power of oil price shocks for U.S. economic growth. However, existing studies focused only on predicting the mean of the economic growth using classical point forecast techniques. As a contribution to the existing literature, we propose a new forecasting model for the analyzed relationship where oil price shocks have the additional flexibility to influence both the mean and the variance of U.S. industrial production. Our forecast is well performed using the Bayesian predictive likelihood as opposed to classical point estimate. We show that the new model for oil price shocks outperforms existing models in terms of forecasting ability. We further confirm previous findings regarding the lack of predictive power of nominal oil price changes.
Acknowledgements

This thesis would not have been possible without the support of many people. I owe an enormous debt of gratitude to my co-supervisor, Professor John M. Maheu, whose guidance and encouragement helped me go through every step of my research. I am also thankful to another of my co-supervisor, Professor Martin Burda. I am very fortunate to have him as my co-advisor and grateful for his mentoring and inspiration. I am grateful to Professor Yuanyuan Wan, who are in my thesis committee, for their generous help and kind suggestions.

I would also like to thank Participants in University of Toronto seminar, SBIES, Rimini Bayesian Econometrics Seminar, CFE, Melbourne Bayesian Seminar.

The first chapter is based on joint work with John Maheu.
# Contents

1 An Infinite Hidden Markov Model for Short-term Interest Rates 1
   1.1 Introduction ................................................................. 2
   1.2 Benchmark Models ......................................................... 4
   1.3 Infinite Hidden Markov Model (IHMM) ................................. 5
      1.3.1 The Dirichlet Process Mixture Model .......................... 5
      1.3.2 Hierarchical Dirichlet process ................................. 6
      1.3.3 Polya Urn Process for HDP ........................................ 7
      1.3.4 Infinite Hidden Markov Model .................................... 9
      1.3.5 IHMM with Hierarchical Prior .................................... 10
   1.4 Posterior Sampling ........................................................ 11
      1.4.1 Predictive Density .................................................... 12
   1.5 Application to Short Term T-Bill Rate ............................... 13
      1.5.1 Data ................................................................. 13
      1.5.2 Model Priors .......................................................... 13
      1.5.3 Posterior Analysis .................................................. 13
      1.5.4 Out-of-Sample Analysis .......................................... 14
      1.5.5 Robustness ........................................................... 16
   1.6 Conclusion ................................................................. 16
   1.7 Appendix ................................................................. 17
      1.7.1 Sampler Steps .......................................................... 17

2 Stock Returns and Real Growth: A Bayesian Nonparametric Approach 31
   2.1 Introduction ................................................................. 32
   2.2 Benchmark Models ......................................................... 33
      2.2.1 Univariate Approach .............................................. 34
      2.2.2 Multivariate Approach ...................................... 36
   2.3 Infinite Hidden Markov Model (IHMM) ................................. 37
      2.3.1 Dirichlet Process (DP) .............................................. 37
2.3.2 Hierarchical Dirichlet Process (HDP) ........................................ 38
2.3.3 Infinite Hidden Markov Model (IHMM) ........................................ 39
2.4 Posterior Sampling of IHMM ............................................................. 41
2.5 Predictive Likelihood ................................................................. 42
  2.5.1 AR ................................................................. 43
  2.5.2 MS2-AR ............................................................. 44
  2.5.3 VAR ................................................................. 44
  2.5.4 MS2-VAR ............................................................. 45
  2.5.5 IHMM-AR ............................................................. 45
  2.5.6 IHMM-VAR ............................................................. 46
2.6 Empirical Results ................................................................. 47
  2.6.1 Data ................................................................. 47
  2.6.2 Model Priors ............................................................. 47
  2.6.3 Out-of-Sample Forecast ......................................................... 48
  2.6.4 The Out-of-Sample Forecast .................................................... 49
  2.6.5 Posterior Analysis and Prior Robustness ..................................... 49
  2.6.6 Structural Break at 1984 ......................................................... 50
  2.6.7 Empirical Results ............................................................. 51
2.7 Conclusion ................................................................. 52
2.8 Appendix ................................................................. 54
  2.8.1 Sampler Steps ............................................................. 54
3 A Bayesian Approach to Oil Price and Real Economy Relationship ........ 70
  3.1 Introduction ................................................................. 71
  3.2 Data ................................................................. 72
  3.3 Out-of-Sample Forecast ......................................................... 73
  3.4 Oil Price Changes ............................................................. 74
    3.4.1 Autoregression with Total Lags ............................................. 75
    3.4.2 Autoregression with Individual Lag ...................................... 75
    3.4.3 Autoregression with Average Lag ........................................ 76
  3.5 Oil Price Shock ............................................................. 76
    3.5.1 Oil Price Shock Measurements ........................................... 77
    3.5.2 Oil Price Shock and Real Growth ....................................... 78
  3.6 Priors and Posterior Sampling .................................................. 78
    3.6.1 Sampler ............................................................. 79
  3.7 Empirical Results ............................................................. 80
Chapter 1

An Infinite Hidden Markov Model for Short-term Interest Rates
1.1 Introduction

Models of the term structure of interest rates are important in finance. They are used to price contingent claims, manage financial risk and assess the cost of capital. In most models the short-rate plays a very important role (Lhabitant et al. 2001, Musiela & Rutkowski 2005, Canto 2008). The time-series dynamics of the short-rate are important and difficult to model over long periods due to changes in monetary regimes and economic shocks. In this paper we extend the popular short-rate models to include Markov switching of infinite dimension. This is a Bayesian nonparametric model that allows for changes in the unknown conditional distribution over time. Applied to weekly data we find significant parameter change over time and strong evidence of non-Gaussian conditional distributions. Our new model with an hierarchical prior provides significant improvements in density forecasts as well as point forecasts.

Markov switching model have been used extensively to model interest rates. Early applications include Hamilton (1988), Albert & Chib (1993a) and Garcia & Perron (1996). Markov switching and GARCH or stochastic volatility are combined by Cai (1994), Gray (1996) and Kalimipalli & Susmel (2004) to better capture volatility dynamics. However, Smith (2002) finds that stochastic volatility and Markov switching are substitutes with the latter being preferred. In related work Lanne & Saikkonen (2003) combine a mixture autoregressive process with time-varying transition probabilities and GARCH.

Ang & Bekaert (2002) show that a state dependent Markov switching model can capture the non-linearities in the drift and volatility function of the US short-rate. Evidence for nonlinear behaviour in the drift term is also found in Pesaran et al. (2006) using a model of structural change. In contrast, Durham (2003) finds no significant evidence of nonlinearity in the drift and concludes that volatility is the critical component. Guidolin & Timmermann (2009) use a four-state Markov switching model to capture the dynamics in US spot and forward rates. They improve point forecasts by combining forecasts of future spot rates with forecasts from time-series models or macroeconomic variables.

What is clear from this literature is that some form of regime switching is necessary to capture changes in the short-rate dynamics over time. Volatility clustering is important and simple two-state models are insufficient to deal with this. In addition, the papers that consider forecasting have focused on point forecasts and ignored density forecasts.

This paper contributes to this literature by designing an infinite hidden Markov model (IHMM) to capture the dynamics of U.S. short-term interest rate. IHMM can be thought of as a first-order Markov switching model with a countably infinite number of states. Given a finite dataset, the number of states is estimated along with all the other parame-
ters. This is essentially a nonparametric model and part of our focus is to flexibly model the conditional distributions of the short-rate and investigate density forecasts. The unbounded nature of the transition matrix allows for both recurring states from the past as well as new states to capture structure change. An advantage of this approach is that as new data arrives, if a new state is needed to capture new features of the conditional density, it is automatically introduced and incorporated into forecasts. These type of dynamic features cannot be captured by fixed state MS models.

The prior for the infinite transition matrix is a special case of the hierarchical Dirichlet process of Teh et al. (2006). Each row of the transition matrix is centered around a common draw from a top level Dirichlet process. This also aids in posterior simulation and centers the model around a standard Dirichlet process mixture model (Escobar & West 1995) which does not allow for time dependence. A sticky version of the infinite hidden Markov model which favours self transitions between states was introduced by Fox et al. (2011) and applied to inflation (Jochmann 2014, Song 2013a), ARMA models (Carpantier & Dufays 2014) and conditional correlations (Dufays 2012). The sticky version is less attractive for financial data in which rapid switching between states is necessary to capture the unknown distribution as well as changes in this distribution over time.

An IHMM extension is applied to the Vasicek (1977) (VSK) model and the Cox et al. (1985) (CIR) model. Applied to weekly data from 1954 to 2014, on average, the model uses about 8 states to capture the unknown conditional distribution. Overall, the CIR specification performs the best while the VSK version requires a few more states on average to fit the data. There is evidence of states reoccurring from the past as well as new unique states being introduced over time. This is especially true from 2008 on as this interest rate regime is historically unique and represents a structural break.

We find evidence of parameter change in both the conditional mean as well as the conditional variance with the latter showing the largest moves. Predictive density plots display significant asymmetry and fat tails that are frequently multimodal. For instance, in 2009 the predictive density has local modes in the right tail. These account for the small probability of returning to a higher interest rates regime.

Consistent with Song (2013a), adding a hierarchical prior for the data density parameters leads to gains in out-of-sample forecast accuracy. The model provides large gains in density forecasts compared to several finite state Markov switching models and a GARCH specification. All of the benchmark models we consider are strongly rejected by predictive Bayes factors in favour of the new nonparametric models. Point forecasts are competitive with existing models.
This paper is organized as follows. The next section discusses benchmark models used for model comparison and extension. Section 1.3 introduces the Dirichlet process, hierarchical Dirichlet process and the infinite hidden Markov model. Posterior sampling is discussed in Section 1.4 while empirical results are found in Section 1.5. The Appendix collects the details of posterior simulation.

1.2 Benchmark Models

Chan et al. (1992) show that the following specification for the short-term riskless rate \( y_t \) nests many popular models

\[
dy_t = (\lambda + \beta y_{t-1}) dt + \sigma y_{t-1}^x dW_t.
\]  

(1.1)

\( dW_t \) is a Brownian motion and both the drift term \( \lambda + \beta y_{t-1} \) and spot variance \( \sigma y_{t-1}^x \) can be a function of the level of \( y_{t-1} \). The discrete time version of the model, conditional on \( y_{1:t-1} = \{y_1, \ldots, y_{t-1}\} \), is

\[
\Delta y_t = \lambda + \beta y_{t-1} + \sigma y_{t-1}^x \epsilon_t, \quad \epsilon_t \sim N(0,1).
\]  

(1.2)

In the empirical application we will consider a rolling window version of this model and focus on \( x = 0 \) and \( 1/2 \) which correspond to the Vasicek (1977) (VSK) model and the Cox et al. (1985) (CIR) model, respectively.

The next specification is a finite state Markov switching model. Let \( s_t \in \{1, \ldots, K\} \) the model is

\[
\begin{align*}
\Delta y_t &= \lambda_{s_t} + \beta_{s_t} y_{t-1} + \sigma_{s_t} y_{t-1}^x \epsilon_t, \quad \epsilon_t \sim N(0,1), \\
s_t|s_{t-1} \sim P_{s_{t-1}}.
\end{align*}
\]  

(1.3) (1.4)

where \( P_{s_{t-1}} \) denotes row \( s_{t-1} \) of the \( K \times K \) transition matrix \( P \) and is the discrete distribution governing the move from state \( s_{t-1} \) to \( s_t \). This model can be estimated following Chib (1996). The VSK model \( (x = 0) \) is labelled as MS-K-VSK while the CIR version \( (x = 1/2) \) is MS-K-CIR. We also consider specifications with a hierarchical prior (MS-K-VSK-H,MS-K-CIR-H). This is discussed in more detail in the next section.
Chapter 1.

The final comparison model is a GARCH specification

\[ \Delta y_t = \lambda + \beta y_{t-1} + y_{t-1}^x \epsilon_t, \]  

\[ \epsilon_t = \sigma_t z_t \quad z_t \sim N(0, 1) \]  

\[ \sigma_t^2 = \omega_0 + \omega_1 \epsilon_{t-1}^2 + \omega_2 \sigma_{t-1}^2, \]

with \( \omega_0 > 0 \) and \( \omega_1 \geq 0, \omega_2 \geq 0 \). This gives the GARCH-VSK \((x = 0)\) and the GARCH-CIR \((x = 1/1)\) models.

1.3 Infinite Hidden Markov Model (IHMM)

An infinite Hidden Markov Model (IHMM) is a Bayesian nonparametric extension of the Markov-switching (MS) model. This builds on the Dirichlet process (DP), the hierarchical Dirichlet process (HDP) and their associated mixture models. We first discuss the Dirichlet process and the Dirichlet process mixture model before moving onto the HDP and the IHMM.

1.3.1 The Dirichlet Process Mixture Model

The Dirichlet process (DP) was introduced by Ferguson (1973) and is a distribution of probability measures over a measurable space \( \Theta \). \( G \sim DP(\alpha, H) \) denotes a distribution \( G \) drawn from the DP with base measure \( H \) and \( \alpha > 0 \) a concentration parameter. A key property of the DP is for any finite partition \( \{A_1, \ldots, A_K\} \) of \( \Theta \),

\[ G(A_1), \ldots, G(A_K) | \alpha, H \sim \text{Dir}\left(\alpha H(A_1), \ldots, \alpha H(A_K)\right), \]  

where \( \text{Dir}\left(\alpha H(A_1), \ldots, \alpha H(A_K)\right) \) denotes a Dirichlet distribution with parameter vector \( (\alpha H(A_1), \ldots, \alpha H(A_K)) \). Therefore, by the properties of the Dirichlet distribution \( E[G(A_i)] = H(A_i) \) and \( \text{Var}(G(A_i)) = G(A_i)(1 - G(A_i))/(1 + \alpha) \).

A constructive definition of the DP is due to Sethuraman (1994) who defined a stick-breaking representation. The stick-breaking representation considers a unit-length stick that has been divided into multiple sub-sticks, where each sub-stick \( (\pi_k) \) is a random proportion \( (v_k) \) of the remaining stick. Let \( \delta_{\theta_k} \) denote a probability measure concentrated at \( \theta_k \) then the stick-breaking construction of \( G \sim DP(\alpha, H) \) is,

\[ G = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i} \quad \text{where} \quad \theta_i \sim H \quad i = 1, 2, \ldots, \infty \]
\[ \pi_i = v_i \prod_{l=1}^{i-1} (1 - v_l), \quad v_i \overset{iid}{\sim} \text{Beta}(1, \alpha). \quad (1.10) \]

We denote the construction of the weights in (1.10) as \( \{\pi_i\}_{i=1}^{\infty} \sim \text{Stick}(\alpha) \) and they form a distribution over the natural numbers.

The parameter \( \alpha \) governs the distribution of the unit mass over the weights \( \pi_i \). Large values of \( \alpha \) spread the mass over many clusters \( \theta_i \) while small values concentrate most of the mass on a few clusters. The DP is a distribution over a discrete probability measure, which guarantees that the probability measure \( G \) from equation (2) is a subset of base distribution \( H \).

With this we can now define the Dirichlet process mixture (DPM) model for \( y_t, t = 1, 2, \ldots \), as

\[ s_t | \alpha \sim \text{Stick}(\alpha) \quad (1.11) \]
\[ y_t | s_t, \Theta \sim F(y_t | \theta_{s_t}) \quad (1.12) \]

where \( \Theta = \{\theta_i\}_{i=1}^{\infty} \), \( F(\cdot | \cdot) \) is the distribution of \( y_t \) given parameter \( \theta_{s_t} \) and \( \text{Stick}(\alpha) \) and \( \theta_{s_t} \) are defined above. This is the basic model for Bayesian density estimation (Escobar & West 1995) and is appropriate for modeling an unconditional distribution. However, if there is time variation in the distribution of \( y_t \) this model is not suitable. We will introduce time-varying weights in (1.9) that result in \( s_t \) following a first-order Markov chain.

### 1.3.2 Hierarchical Dirichlet process

The hierarchical Dirichlet process (HDP) prior is introduced by Teh et al. (2006) as an extension to the DP prior. The HDP is a family of Dirichlet processes that share a common base measure which is also distributed according to a DP prior. Thus, the HDP has a hierarchical structure which is constructed by two DPs,

\[ G_0 | \eta, H \sim DP(\eta, H) \quad (1.13a) \]
\[ G_j | \alpha, G_0 \overset{iid}{\sim} DP(\alpha, G_0), \quad j = 1, \ldots, \infty, \quad (1.13b) \]

where the process defines group-specific probability measures \( G_j \) conditional on a global probability measure \( G_0 \). \( \alpha \) and \( \eta \) are concentration parameters and \( H \) is the base measure.
Using the stick-breaking construction, we have the following representations of the HDP.

\[
G_0 = \sum_{i=1}^{\infty} \gamma_i \delta_{\theta_i}, \quad \Gamma = \{\gamma_i\}_{i=1}^{\infty} \sim \text{Stick}(\eta), \quad \theta_i \overset{iid}{\sim} H, \quad (1.14)
\]

\[
G_j = \sum_{i=1}^{\infty} \pi_{ji} \delta_{\theta_i}, \quad \Pi_j = \{\pi_{ji}\}_{i=1}^{\infty} \overset{iid}{\sim} \text{Stick2}(\alpha, \Gamma), \quad (1.15)
\]

where the weights of the latter measure \( G_j \), are denoted as \( \text{Stick2}(\alpha, \Gamma) \), and are constructed as

\[
\pi_{ji} = \hat{\pi}_{ji} \prod_{l=1}^{i-1} (1 - \hat{\pi}_{jl}), \quad \hat{\pi}_{ji} \overset{iid}{\sim} \text{Beta} \left( \alpha \gamma_i, \alpha \left( 1 - \sum_{l=1}^{i} \gamma_l \right) \right), \quad (1.16)
\]

\( i = 1, 2, \ldots \). Note that all \( G_j \) share the same atoms but have different weights, \( \pi_{ji} \). Each \( G_j \) can serve as a prior for row \( j \) of the transition matrix an an infinite Markov chain. That is, if \( s_t \in \{1, 2, \ldots \} \), then row \( j \) of the transition matrix is \( \Pi_j \) and directs the possible moves of \( s_t = j \) to \( s_{t+1} \). The model is completed with a conditional data distribution \( F(y_t|\theta_{s_t}) \).

### 1.3.3 Polya Urn Process for HDP

Before moving to the full specification of the model it is insightful to understand the conditional sampling distribution of the states in the IHMM induced by the HDP. First, we begin with the Polya urn process for the basic DP. In the DPM model of (1.11)-(1.12), the sequential sampling of the states \( s_t \) and their associated parameter \( \theta_{s_t} \) obeys a Polya urn sampling scheme (Blackwell & MacQueen 1973). The Polya urn sampling scheme is obtained by integrating out \( G \sim \text{DP}(\alpha, H) \). Let \( n_i \) be the current number of sampled states \( i \) (initially 0) and let the current number of different states be \( K \). If \( \delta_{i,j} \) denotes the Kronecker delta then draws from \( G \) are obtained as follows.

1. Set \( s_1 = 1, K = 1, n_1 = 1 \) and \( \theta_1 \sim H \).

2. Given \( s_{1:t-1} \) sample \( s_t \) as

\[
s_t|s_{1:t-1} \sim \sum_{i=1}^{K} \frac{n_i}{\alpha + \sum_j n_j} \delta_{i,s_t} + \frac{\alpha}{\alpha + \sum_j n_j} \delta_{K+1,s_t}. \quad (1.17)
\]

3. \( n_{s_t} \leftarrow n_{s_t} + 1 \) and if \( s_t = K + 1 \) draw \( \theta_{s_t} \sim H \) and set \( K \leftarrow K + 1 \). Increment \( t \) and go to 2.
This process is equivalent to starting with an empty urn and placing a ball of colour $\theta_1$ in and incrementing $n_1$. Thereafter, a ball of colour $\theta_i$, $i = 1, \ldots, K$ is randomly drawn with probability $n_i / (\alpha + \sum_j n_j)$ and otherwise a new ball is drawn from $\theta_i \sim H$ with probability $\alpha / (\alpha + \sum_j n_j)$. Then if an existing ball was selected it is replaced in the urn along with another copy, while if a new colour ball is selected it is put in the urn and its count set to 1 (that is, $n_i$ is incremented). This process is repeated. Note that states that are sampled frequently have a larger count $n_i$ and reinforce their likelihood of being sampled in the future.

The Polya urn process for the infinite Markov chain model in Teh et al. (2006) is closely related and results from integrating out $G_0$ and $G_j$ in (1.14) and (1.15). However, it involves a separate urn for each state (ball) sampled, as well as a top level oracle urn. Sampling states involves a two step process. First, we decide whether to sample from the existing urns (states) or to sample from the oracle urn. Sampling from an existing state is influenced by previous state counts and results in the recurrence of a past state. This is the next state in the Markov chain and we are done. On the other hand, if the oracle urn is selected it allows for previous states to be sampled, with probabilities related to previous counts from oracle draws, as well as new states. New states (balls) are only obtained from the oracle. Therefore, the next state in the Markov chain is either selected from the existing urns which contain pre-existing states or it is selected from the oracle urn in which case a previous state could be sampled or a new state sampled.

Let $c_i$ denote the counts of the sampled states (balls) $i = 1, 2, \ldots$, in the oracle urn and let $n_{ij}$ be the current number of times state $j$ was sampled conditional on being in urn $i$. Both $c_i$ and $n_{ij}$ are initialized to 0. New states $s_t$ and the associated parameters $\theta_{s_t}$ are generated as follows.

1. Set $s_1 = 1$, $K = 1$, $n_1 = 1$, $c_1 = 1$ and $\theta_1 \sim H$.

2. Sample $\bar{s}$ according to

$$
\bar{s} | s_{t-1} = i, s_{1:t-2} \sim \sum_{j=1}^{K} \frac{n_{ij}}{\alpha + \sum_i n_{ij}} \delta_{j, \bar{s}} + \frac{\alpha}{\alpha + \sum_i n_{ij}} \delta_{K+1, \bar{s}}.
$$

3. If $\bar{s} \leq K$ then $s_t = \bar{s}$, $n_{is_t} \leftarrow n_{is_t} + 1$ and go to 4. If $\bar{s} = K + 1$ then $s_t$ is obtained by sampling from the oracle.

$$
s_t | c_1, \ldots, c_K \sim \sum_{j=1}^{K} \frac{c_j}{\eta + \sum_l c_l} \delta_{j, s_t} + \frac{\eta}{\eta + \sum_l c_l} \delta_{K+1, s_t},
$$
\[ n_{is_t} \leftarrow n_{is_t} + 1 \text{ and } c_{st} \leftarrow c_{st} + 1. \text{ If } s_t = K + 1, \theta_{s_t} \sim H \text{ and } K \leftarrow K + 1. \]

4. Increment \( t \) and go to 2.

State transitions are largely governed by the past count of transitions \( n_{ij} \) and \( \alpha \), except for when a transition is queried from the oracle. In this case, the past counts of states sampled from the oracle affects the choices. As in the DP case, states that have a larger number of transitions between themselves will reinforce this in future moves. Clearly, the parameters \( \alpha \) and \( \eta \) play an important role in governing the likelihood of querying the oracle and the likelihood of new states being introduced over time. The Polya urn sampling scheme of the IHMM is important for setting priors on \( \alpha \) and \( \eta \) but it also plays a critical role in posterior sampling for these parameters and other components in the model.

Larger values of \( \alpha \) and \( \eta \) increase the likelihood of new states occurring. A small \( \alpha \) but large \( \eta \) will favor the infrequent introduction of new states over time. Small values of \( \alpha \) and \( \eta \) promote parsimony of states.

### 1.3.4 Infinite Hidden Markov Model

If \( s_t \in \{1, 2, 3, \ldots \} \) is the unobserved state variable and \( \Pi_j = (\pi_{j1}, \pi_{j2}, \ldots) \) is the \( jth \) row of the transition matrix, then \( \pi_{ji} \) becomes the prior probability of \( s_t \) moving from state \( j \) to \( i \) at time \( t + 1 \). The infinite hidden Markov (IHMM) model, a time dependent version of (1.11)-(1.12), can be written as

\[
\Gamma|\eta \sim \text{Stick}(\eta) \quad \theta_i \overset{iind}{\sim} H \quad i = 1, 2, \ldots, (1.20a)
\]

\[
\Pi_j|\alpha, \Gamma \overset{iind}{\sim} \text{Stick2}(\alpha, \Gamma), \quad j = 1, 2, \ldots, (1.20b)
\]

\[
s_t|s_{t-1}, \Pi_{st-1} \sim \Pi_{st-1}, \quad t = 1, \ldots, T (1.20c)
\]

\[
y_t|s_t, \Theta \sim F(y_t|\theta_{st}). (1.20d)
\]

Each row \( \Pi_j \) of the transition matrix \( \Pi \) is assumed to be drawn from a DP prior with a common base measure \( \Gamma = (\gamma_1, \gamma_2, \ldots) \) and precision \( \alpha \). As a result, it can be shown that \( E[\pi_{ji}] = E[\gamma_i] = \eta^{i-1}/(1 + \eta)^i \). In other words, the prior on the transition matrix is centered around equal values of row probabilities. This corresponds to the DPM model in (1.11)-(1.12) and is a natural starting point for inference.

The parameters \( \eta \) and \( \alpha \) play an important role in the distribution of the weights. Different combinations of \( \eta \) and \( \alpha \) can be used to enforce various prior beliefs about the Markov chain. As discussed in the last section, larger values of \( \eta \) favour more active states
while larger values of $\alpha$ allow for the consideration of new states more often. Rather that set these to specific values we place a prior on them and estimate them along with all other model parameters.

### 1.3.5 IHMM with Hierarchical Prior

When new states are introduced in the model there may be benefits to learning about the distributional features of the data density parameters, $\Theta = \{\theta_1, \theta_2, \ldots\}$ if they share a common distribution. To allow the prior to be centered and concentrated in empirically important regions of the parameter space we introduce a hierarchical prior for $H$.

In the empirical work we will focus our attention on the following extension of the CIR ($x = 1/2$) and VSK ($x = 0$) to the infinite hidden Markov model with hierarchical prior (IHMM-CIR-H and IHMM-VSK-H).

$$
\Gamma | \eta \sim \text{Stick}(\eta), \quad \theta_i \overset{iid}{\sim} H(\xi), \quad \xi \sim Q, \quad i = 1, 2, \ldots,
$$

(1.21a)

$$
\Pi_j | \alpha, \Gamma \overset{iid}{\sim} \text{Stick}2(\alpha, \Gamma), \quad j = 1, 2, \ldots,
$$

(1.21b)

$$
s_t | s_{t-1}, \Pi_{s_{t-1}} \sim \Pi_{s_{t-1}}, \quad t = 1, \ldots, T
$$

(1.21c)

$$
\Delta y_t | Y_{t-1}, s_t, \Theta \sim \mathcal{N}(\lambda_{s_t} + \beta_{s_t} y_{t-1}, \sigma_{s_t}^2 y_{t-1}^2),
$$

(1.21d)

where $\theta_i = (\lambda_i, \beta_i, \sigma_i)$. The hierarchical priors for $\eta$ and $\alpha$ are,

$$
\eta \sim \text{Gamma}(a_1, b_1) \quad \alpha \sim \text{Gamma}(a_2, b_2)
$$

(1.22)

where $\text{Gamma}(a, b)$ denotes a gamma distribution with mean $a/b$. Given $\xi$ the prior for $\theta_i$ is $H(\xi)$ and if $\vartheta_i = (\lambda_i, \beta_i)$ it follows

$$
\vartheta_i \sim N(\phi, B) \quad \sigma_i^{-2} \sim \text{Gamma}(\chi, \nu).
$$

(1.23)

Finally, the prior for $\xi = (\phi, B, \chi, \nu)$ follows

$$
\phi \sim \mathcal{N}(h_0, H_0) \quad B^{-1} \sim W(a_0, A_0) \quad \chi \sim \text{Exp}(\rho_0) \quad \nu \sim \text{Gamma}(c_0, d_0),
$$

(1.24)

where $W(a_0, A_0)$ is a Wishart distribution with scale matrix $A_0$, degree of freedom parameter $a_0$ and mean of $a_0 A_0$. $\text{Exp}(\rho_0)$ is the exponential distribution with parameter $\rho_0$. According to this parametrization $E(\sigma^{-2}) = \frac{\chi}{\nu}$ and $E(\chi) = \rho_0$. $\phi$ and $h_0$ are 2 × 1 vector, $H_0$ and $A_0$ are 2 × 2 matrices.
1.4 Posterior Sampling

If the IHMM model was replaced with a finite state hidden Markov model much of the posterior sampling complexity would be eliminated and existing methods of sampling such as Chib (1996) could be used. The beam sampler of Van Gael et al. (2008) is designed to exactly achieve this. It is a stochastic truncation that reduces the infinite state space to a finite one and allows for the forward-filter backward sampler (FFBS) of Chib (1996) to be applied.

The idea behind beam sampling is closely related to slice sampling for DPM models (Walker 2007) and involves the introduction of latent variables $u_t$, $t = 1, \ldots, T$ such that the conditional density of $u_t$ is

$$p(u_t|s_{t-1}, s_t, \Pi) = \frac{1}{\pi_{s_{t-1}, s_t}}(0 < u_t < \pi_{s_{t-1}, s_t}).$$

(1.25)

The $u_t$ are sampled along with the other parameters but the sampling of the states given $u_1:t$ in the filter step of the FFBS becomes,

$$p(s_t|y_1:t, u_1:t, \Pi) \propto p(y_t|y_{1:t-1}, s_t) \sum_{s_{t-1}=1}^{\infty} 1(u_t < \pi_{s_{t-1}, s_t})p(s_{t-1}|y_{1:t-1}, u_{1:t-1}, \Pi).$$

(1.26)

$$p(s_t|y_1:t, u_1:t, \Pi) \propto p(y_t|y_{1:t-1}, s_t) \sum_{s_{t-1}:u_t < \pi_{s_{t-1}, s_t}} p(s_{t-1}|y_{1:t-1}, u_{1:t-1}, \Pi).$$

(1.27)

The $u_t$ slices out states with small $\pi_{s_{t-1}, s_t}$ and results in a finite summation in (1.27) of dimension $K$, since the number of states $s_{t-1}$ that satisfy $u_t < \pi_{s_{t-1}, s_t}$ is finite. This turns the infinite summation into a finite one. Once the forward pass is computed for $t = 1, \ldots, T$ the backward pass follows from,

$$p(s_t|s_{t+1}, y_1:T, u_{1:T}) \propto p(s_t|y_{1:t}, u_{1:t})p(u_{t+1} < \pi_{s_t, s_{t+1}}), \quad t = T - 1, \ldots, 1.$$

(1.28)

This is initiated with a draw of $s_T$ from the last value of the filter $p(s_T|y_{1:T}, u_{1:T}, \Pi)$.

In each MCMC iteration the slice sampler effectively truncates the system to a dimension of size $K$. We order each of the states that receive a non-zero weight as the first $K$ states and keep track of the $K+1$ state as the residual probability. As such, posterior sampling is over the quantities $\Gamma = (\gamma_1, \ldots, \gamma_K, \gamma_{K+1})$ with $\gamma_{K+1} = \frac{1}{\sum_{l=K+1}^{\infty} \pi_l}$ and $\Pi_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jK+1})$, $j = 1, \ldots, K$, where $\pi_{jK+1} = \frac{1}{\sum_{l=K+1}^{\infty} \pi_{jl}}$.

After initializing the parameters the full MCMC routine involves the following steps:

1. sample $s_{1:T}|y_{1:T}, u_{1:T}, \Pi$
2. sample \( \Pi_j | s_{1:T}, \Gamma, j = 1, \ldots, K \).

3. sample \( u_{1:T} | s_{1:T}, \Pi \) and update \( K \).

4. sample \( \theta_j | s_{1:T}, y_{1:T}, \xi, j = 1, \ldots, K \).

5. sample \( \Gamma | s_{1:T}, \eta \).

6. sample \( \xi | \theta_1, \ldots, \theta_K, \eta | s_{1:T}, \Gamma \) and \( \alpha | s_{1:T}, \Gamma \).

Iterating over these sampling steps gives one draw from the posterior denoted as \( \Omega = \{ \Gamma, \Pi, K, \Theta, s_{1:T}, \xi, \eta, \alpha \} \). Full details of each of the posterior sampling steps can be found in the Appendix. After a suitable burn-in period a large number of draws are collected from which features of the posterior and predictive densities can be estimated. For example, given \( N \) posterior draws of \( \{ \theta^{(i)}_{s_t} \}_{i=1}^N \) then a simulation consistent estimate of \( E[\theta_{s_t} | y_{1:T}] \) is \( \frac{1}{N} \sum_{i=1}^N \theta^{(i)}_{s_t} \). For full sample estimates we drop the first 80,000 draws and collect the next 100,000 for posterior inference. For computing predictive likelihoods and predictive means sequentially for the out-of-sample period we use a burn-in of 5000 and use the next 20,000 for posterior/predictive inference.

### 1.4.1 Predictive Density

Given a sequence of posterior draws, \( \{ \Omega^{(i)} \}_{i=1}^N \) from the IHMM models using \( T \) observations, the predictive density can be computed at \( \Delta y_{T+1} \) (or \( y_{T+1} \)) as follows.

1. For each \( i \), randomly draw a state \( s_{T+1} \) according to the multinomial distribution \( \Pi_{s_{T+1}} \).

2. If \( s_{T+1} \leq K^{(i)} \) set \( (\lambda^{(i)}, \beta^{(i)}, \sigma^{2(i)}) \equiv \theta_{s_{T+1}} \) and otherwise set \( (\lambda^{(i)}, \beta^{(i)}, \sigma^{2(i)}) \equiv \theta \), where \( \theta \sim H(\xi^{(i)}) \).

The predictive density for \( y_{T+1} \) is estimated as

\[
p(y_{T+1} | y_{1:T}) \approx \frac{1}{N} \sum_{i=1}^{N} N(y_{T+1} | \lambda^{(i)} + (1 + \beta^{(i)})y_T, \sigma^{2(i)}y_T^2),
\]

where \( N(\cdot | \cdot, \cdot) \) is the normal probability density function. Similarly, the predictive mean of \( y_{T+1} \) can be estimated as

\[
E[y_{T+1} | y_{1:T}] \approx \frac{1}{N} \sum_{i=1}^{N} (\lambda^{(i)} + (1 + \beta^{(i)})y_T).
\]
The predictive density and prediction means can be used to compare and evaluate the performance of several models in the out-of-sample period.

1.5 Application to Short Term T-Bill Rate

1.5.1 Data

The data is the 3-month T-bill of secondary market obtained from the Board of Governors of the Federal Reserve System. The data is weekly, Friday to Friday from Jan-15-1954 to Mar-28-2014 (3142 observations). Figure 1.1 displays the data while Table 1.1 reports summary statistics.

1.5.2 Model Priors

For the IHMM-CIR-H and IHMM-VSK-H models $h_0$ is set to a vector of zeros, $H_0$ and $A_0$ to an identity matrix, $a_0 = 2$, $\rho_0 = 1$, $c_0 = 5$, and $d_0 = 1$. For the precision parameters $\eta$ and $\alpha$ the hyperparameters are $a_1 = a_2 = 5$ and $b_1 = b_2 = 1$. The first 3 columns of Table 1.2 summarize the prior. In some cases the 0.9 density intervals are obtained by simulation.

For the CIR, VSK, the MS-K-CIR, MS-K-VSK and GARCH-CIR and GARCH-VSK the priors are $\lambda \sim N(0, 25)$, $\beta \sim N(0, 1)$, $\sigma^2 \sim IG(2, 0.5)$. For the MS models, each row of the transition matrix has the prior $Dir(1, \ldots, 1)$ where $K$ is the fixed number of states. We also consider MS models with the identical hierarchical prior of the IHMM-CIR-H and IHMM-VSK-H specifications over $\theta$ and $\xi$. Finally, for the GARCH-CIR and GARCH-VSK models all priors are assumed to be independent $N(0, 1)$ with the restrictions $\omega_0 > 0, \omega_1 \geq 0$, and $\omega_2 \geq 0$ imposed.

1.5.3 Posterior Analysis

Table 1.2 reports posterior summary statistics for the IHMM-CIR-H and IHMM-VSK-H models. Compared to the prior (columns 2 and 3) there is significant learning in all the parameters as the density intervals change in location and their length decreases. On average, the two models are using 8 components in the mixture to capture the shape and changes in the distribution. Figure 1.2 is a histogram of the sampled active states $K$. Although the distribution is concentrated around 8 there is uncertainty in both models. Forecasts from this model automatically incorporate this regime uncertainty.
Figures 1.3 and 1.4 display the posterior mean of the model parameters subject to regime change ($E[\lambda_{st}|y_{1:T}]$, $E[\beta_{st}|y_{1:T}]$ and $E[\sigma_{st}|y_{1:T}]$). There is considerable time variation in all parameters. A fixed parameter model would have difficulty fitting such a long time period with these parameter shifts. Some of the state changes appear to return to previous values in the time history while some appear to be new and unique. This is most apparent for the VSK model in Figure 1.4 where somewhere in 2008 the model displays a combination of parameter estimates that are unique. This is like a structure break in that the model identifies dynamics for the short term interest rates that differ from all past states. The evidence for a structural change is much less for the CIR specification.

The second last panel in these figures displays $E[\beta_{st} > 0|y_{1:T}]$ over time. This corresponds to the probability of an explosive regime. This probability is non-negligible over the sample period for both models. In the case of the VSK model, Figure 1.4 shows that it is almost certainly explosive from 2008 and onward.

The final panel of the plot displays the posterior evidence for state changes at each point in the sample. There are a few periods where states persist but generally there is regular changes in states. This is expected if the Gaussian assumption for the data density does not fit the data as the model will resort to mixture of the Gaussian densities to approximate the unknown distribution.

Figures 1.5 and 1.6 display a heat map for the latent states $s_t$, $t = 1, \ldots, T$ for IHMM-CIR-H and IHMM-VSK-H. A heat map is an estimate of $P(s_i = s_j|y_{1:T})$ and reported in a table in which colour differences denote different probabilities over the range of $i = 1, \ldots, T$ and $j = 1, \ldots, T$. High probability values (light colour) on the main diagonal indicate new regimes that are unique to that time period. Many of these occur for both models and indicate a relatively high number of states. On the other hand, high probability values (light colour) off the main diagonal indicate regimes that reoccur historically. Both models display significant periods of unique regimes which is indicative of structure change, however, it is more pronounced in the VSK version. What these figures show is that there are new states being introduced into the model over time and there are frequent regime changes back to previous states. These type of dynamic features cannot be captured by fixed state MS models.

1.5.4 Out-of-Sample Analysis

The main out-of-sample results are based on the sample 1955-Dec-09 to 2014-Mar-28 (3043 observations). At each point $t$ in the sample, the models are estimated using data up to $t$ and model forecasts are computed for $y_{t+1}$. To compare models we focus on log-
predictive likelihood values and root-mean squared forecast errors based on predictive means.

The predictive likelihood of $y_{t+1}$ for the infinite hidden Markov models are estimated by plugging in the data $y_{t+1}$ into (2.33). Similarly, predictive likelihoods are estimated for other models. We report $LPL_M = \sum_{i=t_i}^{T} \log p(y_i | y_{1:i-1}, M)$ for model $M$. Several models can be compared by log-predictive Bayes factors. The log-predictive Bayes factors for model A against B is $LPL_A - LPL_B$ where positive values favour A and values in excess of 5 are considered strong support for A.

Forecast performance is reported in Table 1.3. Two sample periods are included, along with various benchmark comparison models. Log-predictive Bayes factors can be computed by subtracting any two log-predictive values in the table. The best model in terms of $LPL_M$, for the larger out-of-sample period is the IHMM-CIR-H. The log-predictive Bayes factor against the next best model (IHMM-VSK-H) is 44 representing a substantial improvement. The evidence for the IHMM-CIR-H against the best parametric model (MS-3-CIR-H ) is also strong with a log-Bayes factor of 52. The RMSE of the IHMM-CIR-H is also the best but the gains over the benchmark models are modest. It is 0.7% lower than the MS-3-CIR-H model.

The second portion of Table 1.3 record the forecast performance for a shorter more recent sample. Consistent with the previous results the quality of density forecasts are superior for the IHMM-CIR-H specification while the best point forecasts come from IHMM-VSK model, but again, gains are small in the RMSE.

Figures 1.7 and 1.8 display the cumulative log-predictive likelihoods for various models at each point in the out-of-sample period. All the specifications have difficulty with high volatility episodes. The IHMM hierarchical versions recover the fastest. The IHMM-CIR-H and IHMM-VSK-H models provide significant gains in density forecasts throughout the sample. The dominance of these models is not confined to any particular period or outliers but involves steady ongoing gains over time.

Closely related to this is the estimate of the number of effective states (posterior mean of $K$) at each point in the out-of-sample period. Figures 1.9 and 1.10 display these estimates. It shows a regular increase in the dimension of the model over time. However, there are periods in which the number of states level off or even drop (just before 97-July, Figure 1.9). The VSK version of the model requires more discrete jumps in the number of states to deal with heteroskedasticity. This can be seen after 72-Aug and after 2007 in Figure 1.10.

Finally, to see why the IHMM-CIR-H performs so well in density forecasts Figure 1.11 produces the predictive density for this model along with a rolling window version of the
basic CIR model. For each of the selected dates the density and log-density are displayed. It is clear that the IHMM-CIR-H is often very different than the parametric specification. The model displays fat tails and asymmetry. The log-density plots show the IHMM-CIR-H model to have several modes in the right tail. This is expected as the model allows for the possibility of moving from low interest rate regimes back to higher interest rate regimes.

1.5.5 Robustness

Table 1.4 reports the out-of-sample results for the shorter period for both IHMM models and various prior configurations. In general, the top panel presents results for priors that have an increased variance but centered in the same location. The second panel reports the impact from more concentrated priors on $\eta$ and $\alpha$. These favor fewer clusters in the model.

Different priors do lead to differences in the log-predictive likelihoods. However, the IHMM-CIR-H specification continues to be the best model for density forecasts and is always significantly better than the best performing benchmark model MS-3-CIR-H. For instance, the MS-3-CIR-H model has a log-predictive likelihood of 1306.158 while the least favorable prior gives the IHMM-CIR-H model a log-predictive likelihood value of 1324.724.

The density forecasts for the IHMM-VSK-H show a bit less variation. The RMSE for both models is very robust to changes in the prior. In summary, the new nonparametric models continue to dominate the benchmarks for different prior configurations.

1.6 Conclusion

This paper extends popular discrete time short interest rate models to include Markov switching of infinite dimension. This is a Bayesian nonparametric model that allows for changes in the unknown conditional distribution over time. Applied to weekly U.S. data we find significant parameter change over time and strong evidence of non-Gaussian conditional distributions. Our new model with an hierarchical prior provides significant improvements in density forecasts as well as point forecasts. The empirical study finds recurring regimes as well as unique regimes (structural breaks) in the US short-term rate.
1.7 Appendix

Several of the sampling steps are based on Teh et al. (2006), Van Gael et al. (2008) and Fox et al. (2011).

1.7.1 Sampler Steps

Recall the notation: full sample of $y_{1:T}$, state variables $s_{1:T}$, \Gamma = (\gamma_1, \ldots, \gamma_{K}, \gamma_{K+1})$ with $\gamma_{K+1} = \sum_{l=K+1}^{\infty} \gamma_l$, $\Pi_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jK+1})$, $j = 1, \ldots, K$, where $\pi_{jK+1} = \sum_{l=K+1}^{\infty} \pi_{jl}$ and $\theta = \{\lambda, \beta, \sigma\}$.

1. Sample $u_{1:T}$: $u_1 \sim U(0, \gamma_1)$ and $u_t \sim U(0, \pi_{s_{t-1}s_t})$, $t = 2, \ldots, T$

2. Expand $\Pi$ and $K$: If

$$\max\{\pi_{1K+1}, \ldots, \pi_{KK+1}\} > \min\{u_1, \ldots, u_T\},$$  \hspace{1cm} (1.31)

then repeat the following steps:

(a) Increment $K \leftarrow K + 1$ and $\theta_K \sim H(\xi)$,

(b) Update $\gamma$. Note that with the increment in $K$, $\gamma_K$ is the old residual probability which is broken as

$$\tau \sim \text{Beta}(1, \eta), \quad \gamma_{K+1}^r \leftarrow (1 - \tau) \gamma_K, \quad \gamma_K \leftarrow \tau \gamma_K.$$

(c) $\Pi_K = (\pi_{K1}, \ldots, \pi_{KK+1}) \sim \text{Dir}(\alpha \gamma_1, \ldots, \alpha \gamma_{K+1})$.

(d) Update $\Pi_j$, $j = 1, \ldots, K$ as

$$\tau_j \sim \text{Beta}\left(\alpha \gamma_K, \alpha \gamma_{K+1}\right), \quad \pi_{jK+1}^r \leftarrow (1 - \tau_j) \pi_{jK}, \quad \pi_{jK} \leftarrow \tau_j \pi_{jK}$$

Steps a) – d) are repeated until (1.31) does not hold. This process expands $\Pi$ until the possibility of new state is too small to consider in the FFBS steps next.

3. Forward filter of $s_{1:T}$. Let $f(\cdot|\cdot)$ be the density function of $y_t$.

(a) Initial forecast step for $s_1$:

$$p(s_1 = k|u_{1:T}, \Gamma, \Theta) \propto 1(u_1 < \gamma_k), \quad \text{for } k = 1, \ldots, K$$
(b) The updating step for $s_{1:T}$: and $k = 1, \ldots, K$.

$$p(s_t = k|y_{1:t}, u_{1:T}, \Pi, \Theta) \propto f(y_t|y_{1:t-1}, \theta_k)p(s_t = k|y_{1:t-1}, u_{1:T}, \Pi, \Theta)$$

(c) The forecasting step for $k = 1, \ldots, K$

$$p(s_{t+1} = k|y_{1:t}, u_{1:T}, \Pi, \Theta) \propto \sum_{l=1}^{K} 1(u_{t+1} < \pi_{lk})p(s_t = l|y_{1:t}, u_{1:T}, \Pi, \Theta)$$

(d) Iterate step b and c until $t = T$.

4. Backward Sampler for $s_{1:T}$:

(a) Sample the initial state of $s_T$ from $p(s_T|y_{1:T}, u_{1:T}, \Pi, \Theta)$.

(b) Sample $s_t$, $t = T - 1, T - 2, \ldots, 1$ recursively from

$$P(s_t = i|s_{t+1} = j, y_{1:t}, u_{1:T}, \Pi, \Theta) \propto 1(\pi_{ij} < u_{t+1})p(s_t = i|y_{1:t}, u_{1:T}, \Pi, \Theta).$$

After this step we perform a cleanup step to remove any empty states (states with no observations allocated to them). This results in a redefined $K$ and transition matrix $\Pi$, and other model parameters are adjusted accordingly so that states $1, \ldots, K$ are active states.

5. Sample $c_{1:K}$: Recall $c_i$ is number of counts of state $i$ sampled from the oracle. It is an auxiliary variable that facilitates several sampling steps below. Let $o_{ji}$ be the number of oracle draws out of the total transitions $n_{ji}$, of state $j$ to state $i$. Following Fox et al. (2011) we simulate the sequence of $o_{ji}$ to obtain a draw of $c_i$. For each $i = 1, \ldots, K$, $j = 1, \ldots, K$, do the following steps.

(a) set $o_{ji} = 0$

(b) draw $x_l \sim \text{Bernoulli}\left(\frac{\alpha_i}{\beta + c_i}\right)$, for $l = 1, \ldots, n_{ji}$: if $x_l = 1$ increment $o_{ji}$.

Then $c_i = \sum_{j=1}^{K} o_{ji}$.

6. Sample $\eta$: Two auxiliary variables $v$ and $\lambda$ are introduced to sample $\eta$. We apply a Gibbs sampler to sample $v$ and $\lambda$ conditional on $c = \sum_{i=1}^{K} c_i$ and previous $\eta$ first, then sample $\eta$ through a gamma distribution.

(a) $v \sim \text{Bernoulli}\left(\frac{c}{c+\eta}\right)$
(b) \( \Lambda \sim \text{Beta}(\eta + 1, c) \)

(c) \( \eta \sim \text{Gamma}(a_1 + K - \nu, b_1 - \log \Lambda) \)

7. Sample \( \alpha \): The derivations follows directly from Fox et al. (2011).

(a) \( \nu_j \sim \text{Bernoulli}(\frac{n_{j}^b}{n_{j}^{a} + \alpha}) \) for \( j = 1, \ldots, K \). \( n_j = \sum_{i=1}^{K} n_{ji} \)

(b) \( \lambda_j \sim \text{Beta}(\alpha + 1, n_j) \) for \( j = 1, \ldots, K \).

(c) \( \alpha \sim \text{Gamma}(a_2 + c - \sum_{j=1}^{K} \nu_j, b_2 - \sum_{j=1}^{K} \log(\lambda_j)) \)

8. Sample \( \Gamma \): Given the counts \( c_{1:K} \), which are from a sample from \( \Gamma \), conjugacy gives

the update for \( \Gamma \) as

\[ \Gamma|c_{1:K}, \eta \sim \text{Dir}(c_1, c_2, \ldots, c_K, \eta) \]

9. Sample \( \Pi_j \): for \( j = 1, \ldots, K \), given \( (n_{j1}, \ldots, n_{jK}) \) and the conjugate property of

the Dirichlet distribution,

\[ \Pi_j|\alpha, n_{j1:K} \sim \text{Dir}(\alpha \gamma_1 + n_{j1}, \ldots, \alpha \gamma_K + n_{jK}, \alpha \gamma_{K+1}) \]

10. Sample \( \theta_{1:K} \): Let \( \theta_k = (\vartheta_k, \sigma_k) \) and \( \vartheta_k = (\lambda_k, \beta_k)^T \), for \( k=1, \ldots, K \):

Let \( \hat{Y}_k \equiv \{ \Delta y_t | s_t = k \} \), \( \hat{X}_k \equiv \{(1, y_{t-1}) | s_t = k \} \) and \( T_k = \{ #t | s_t = k \} \). Therefore, \( \hat{Y}_k \) is a vector with dimension \( T_k \times 1 \) and \( \hat{X}_k \) is a matrix with dimension \( T_k \times 2 \). This gives the linear model

\[ \hat{Y}_k = \hat{X}_k \vartheta_k + u \quad u \sim N(0, \sigma_k^2 I) \]

with the following Gibbs sampling steps

\[ \vartheta_k \sim N \left( V \left( \frac{1}{\sigma^2} \hat{X}_k^T \hat{Y}_k + \phi B^{-1} \right), V \right) \]

\[ V = \left( \frac{1}{\sigma^2} \hat{X}_k^T \hat{X}_k + B^{-1} \right)^{-1} \]

\[ \sigma_k^{-2} \sim \text{Gamma} \left( \frac{T + 2\chi}{2}, \frac{(\hat{Y}_k - \hat{X}_k \vartheta_k)^T (\hat{Y}_k - \hat{X}_k \vartheta_k) + 2\nu}{2} \right) \]

11. Sample \( \xi \), hierarchical priors, which are \( \phi, B, \chi, \nu \). \( H(\xi) \) is defined as follows,

\[ \vartheta \sim N(\phi, B) \quad \sigma^{-2} \sim \text{Gamma}(\chi, \nu) \]

\( Q \) is defined as follows,

\[ \phi \sim N(h_0, H_0) \quad B^{-1} \sim W(a_0, A_0) \quad \chi \sim \text{Exp}(\rho_0) \quad \nu \sim \text{Gamma}(c_0, d_0) \]
(a) Sample $\phi \mid B, h_0, H_0, \vartheta_{1:K} \sim N(\mu_\phi, \Sigma_\phi)$ where

$$
\overline{\vartheta} = \frac{1}{K} \sum_{j=1}^{K} \vartheta_j, \quad \mu_\phi = \Sigma_\phi \left( H_0^{-1} h_0 + K B^{-1} \overline{\vartheta} \right), \quad \Sigma_\phi = \left( H_0^{-1} + K B^{-1} \right)^{-1}.
$$

(b) Sample $B^{-1} \mid \phi, a_0, A_0, \vartheta_{1:K} \sim \text{Wishart}(\omega_B, \Omega_B)$ where,

$$
\omega_B = K + \alpha_0 \quad \text{and} \quad \Omega_B = \left( A_0^{-1} + \sum_{j=1}^{K} (\vartheta_j - \phi)(\vartheta_j - \phi)^T \right)^{-1}
$$

(c) Sample $\nu \mid \chi, c_0, d_0, \sigma_{1:K}^{-2} \sim \text{Gamma} \left( c_0 + K \chi, d_0 + \sum_{j=1}^{K} \sigma_j^{-2} \right)$

(d) Sample $\chi \mid \nu, \rho_0, \sigma_{1:K}^{-2}$. There is no conjugate prior so we apply a Metropolis-Hastings step. The conditional posterior is

$$
\pi(\chi \mid \nu, \rho_0, \sigma_{1:K}^{-2}) = \text{Exp}(\chi \mid \rho_0) \prod_{j=1}^{K} G(\sigma_j^{-2} \mid \chi, \nu).
$$

The proposal is

$$
q(\chi^{new} \mid \chi^{old}) \sim \text{Gamma}(\zeta, \zeta / \chi^{old})
$$

and we choose $\zeta$ so the rejection rate is between 0.3 and 0.6. The proposal $\chi^{new}$, is accepted with probability

$$
\min \left[ \frac{\pi(\chi^{new} \mid \nu, \rho_0, \sigma_{1:K}^{-2}) / q(\chi^{new} \mid \chi^{old})}{\pi(\chi^{old} \mid \nu, \rho_0, \sigma_{1:K}^{-2}) / q(\chi^{old} \mid \chi^{new})}, 1 \right].
$$

12. Repeat 1-11.
Table 1.1: Statistical Summaries of 3-Month T-Bill

<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>25%Q</th>
<th>75%Q</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>4.658</td>
<td>3.030</td>
<td>4.615</td>
<td>2.720</td>
<td>6.110</td>
<td>0.840</td>
<td>1.260</td>
</tr>
<tr>
<td>Change</td>
<td>0.000</td>
<td>0.190</td>
<td>0.000</td>
<td>-0.050</td>
<td>0.050</td>
<td>-0.730</td>
<td>24.150</td>
</tr>
</tbody>
</table>

This table reports summary statistics for weekly 3 month T-bill rates from Jan-15-1954 to Mar-28-2014 (3142 observations).

Table 1.2: Posteriors and Prior Summary of Hierarchical-Priors

<table>
<thead>
<tr>
<th></th>
<th>Prior</th>
<th>Posterior of IHMM-CIR-H</th>
<th>Posterior of IHMM-VSK-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>η</td>
<td>5.000 (2.459, 8.110)</td>
<td>3.810 (2.258, 5.591)</td>
<td>4.104 (2.490, 5.944)</td>
</tr>
<tr>
<td>α</td>
<td>5.000 (2.541, 7.797)</td>
<td>0.768 (0.530, 1.023)</td>
<td>0.717 (0.503, 0.953)</td>
</tr>
<tr>
<td>χ</td>
<td>1.000 (0.996, 2.081)</td>
<td>0.678 (0.368, 1.046)</td>
<td>0.502 (0.283, 0.755)</td>
</tr>
<tr>
<td>ν</td>
<td>1.000 (0.093, 2.180)</td>
<td>8.07e-4 (3.11e-4, 1.41e-3)</td>
<td>7.02e-4 (2.6e-4, 1.23e-3)</td>
</tr>
<tr>
<td>φ₁</td>
<td>0.000 (-0.857, 0.892)</td>
<td>0.002 (-0.220, 0.229)</td>
<td>2.62e-4 (-0.196, 0.196)</td>
</tr>
<tr>
<td>φ₂</td>
<td>0.000 (-0.881, 0.980)</td>
<td>-0.003 (-0.211, 0.204)</td>
<td>6.23e-4 (-0.167, 0.170)</td>
</tr>
<tr>
<td>B₁₁</td>
<td>(0.227, 4.333)</td>
<td>0.396 (0.093, 0.770)</td>
<td>0.354 (0.079, 0.669)</td>
</tr>
<tr>
<td>B₂₂</td>
<td>(0.206, 4.572)</td>
<td>0.341 (0.081, 0.673)</td>
<td>0.203 (0.069, 0.332)</td>
</tr>
<tr>
<td>B₁₂</td>
<td>(-1.515, 1.603)</td>
<td>-0.157 (-0.428, 0.054)</td>
<td>-0.097 (-0.234, 0.048)</td>
</tr>
<tr>
<td>Regimes #</td>
<td>7.985 (7.000, 9.000)</td>
<td>8.771 (8.000, 10.000)</td>
<td></td>
</tr>
</tbody>
</table>

This table reports the mean and 0.90 density intervals (DI) from the benchmark prior and the posteriors of the IHMM-CIR-H and IHMM-VSK-H models.
Table 1.3: Forecast Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-Predictive Likelihood</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IHMM-CIR-H</td>
<td><strong>2860.079</strong></td>
<td><strong>0.1937</strong></td>
</tr>
<tr>
<td>IHMM-CIR</td>
<td>2432.647</td>
<td>0.1951</td>
</tr>
<tr>
<td>MS-3-CIR-H</td>
<td>2805.243</td>
<td>0.1947</td>
</tr>
<tr>
<td>MS-3-CIR</td>
<td>2435.441</td>
<td>0.1949</td>
</tr>
<tr>
<td>MS-2-CIR-H</td>
<td>2631.190</td>
<td>0.1949</td>
</tr>
<tr>
<td>MS-2-CIR</td>
<td>2422.794</td>
<td>0.1950</td>
</tr>
<tr>
<td>GARCH-CIR</td>
<td>2118.011</td>
<td>0.1946</td>
</tr>
<tr>
<td>CIR-Roll</td>
<td>1761.328</td>
<td>0.1954</td>
</tr>
<tr>
<td>iHMM-VSK-H</td>
<td>2815.764</td>
<td>0.1954</td>
</tr>
<tr>
<td>iHMM-VSK</td>
<td>2381.192</td>
<td>0.1956</td>
</tr>
<tr>
<td>MS-3-VSK-H</td>
<td>2647.157</td>
<td>0.1951</td>
</tr>
<tr>
<td>MS-3-VSK</td>
<td>2371.533</td>
<td>0.1956</td>
</tr>
<tr>
<td>MS-2-VSK-H</td>
<td>2177.383</td>
<td>0.1953</td>
</tr>
<tr>
<td>MS-2-VSK</td>
<td>2143.111</td>
<td>0.1952</td>
</tr>
<tr>
<td>GARCH-VSK</td>
<td>2055.751</td>
<td>0.1954</td>
</tr>
<tr>
<td>VSK-Roll</td>
<td>1332.411</td>
<td>0.1955</td>
</tr>
</tbody>
</table>

1955-Dec-09 to 2014-Mar-28 (3043 observations)

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-Predictive Likelihood</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IHMM-CIR-H</td>
<td><strong>1333.426</strong></td>
<td>0.09638</td>
</tr>
<tr>
<td>IHMM-CIR</td>
<td>1244.576</td>
<td>0.09717</td>
</tr>
<tr>
<td>MS-3-CIR-H</td>
<td>1306.158</td>
<td>0.09679</td>
</tr>
<tr>
<td>MS-3-CIR</td>
<td>1247.913</td>
<td>0.09744</td>
</tr>
<tr>
<td>MS-2-CIR-H</td>
<td>1244.722</td>
<td>0.09671</td>
</tr>
<tr>
<td>MS-2-CIR</td>
<td>1230.406</td>
<td>0.09681</td>
</tr>
<tr>
<td>GARCH-CIR</td>
<td>1054.613</td>
<td>0.09714</td>
</tr>
<tr>
<td>CIR-Roll</td>
<td>965.030</td>
<td>0.09677</td>
</tr>
<tr>
<td>IHMM-VSK-H</td>
<td>1309.384</td>
<td>0.09628</td>
</tr>
<tr>
<td>IHMM-VSK</td>
<td>1015.873</td>
<td><strong>0.09611</strong></td>
</tr>
<tr>
<td>MS-3-VSK-H</td>
<td>1205.062</td>
<td>0.09660</td>
</tr>
<tr>
<td>MS-3-VSK</td>
<td>1002.516</td>
<td>0.09650</td>
</tr>
<tr>
<td>MS-2-VSK-H</td>
<td>908.486</td>
<td>0.09694</td>
</tr>
<tr>
<td>MS-2-VSK</td>
<td>877.123</td>
<td>0.09686</td>
</tr>
<tr>
<td>GARCH-VSK</td>
<td>1112.603</td>
<td>0.09715</td>
</tr>
<tr>
<td>VSK-Roll</td>
<td>677.907</td>
<td>0.09671</td>
</tr>
</tbody>
</table>

2000-Jan-07 to 2014-Mar-28 (743 observations)

This table displays log-predictive likelihoods and root-mean squared forecast errors over two sample periods for various models. MS-K denotes a Markov switching model of dimension K, IHMM an infinite hidden Markov model, H a hierarchical prior, CIR-Roll and VSK-Roll refer models estimation with a rolling window of size 500. VSK models set $x = 0$ while CIR set $x = 0.5$. For additional details see the text.
Table 1.4: Sensitivity Analysis of 2000-Jan-07 to 2014-Mar-28 (743 observations)

<table>
<thead>
<tr>
<th>Changes to Prior</th>
<th>IHMM-CIR-H</th>
<th>IHMM-VSK-H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPL</td>
<td>RMSE</td>
</tr>
<tr>
<td>benchmark prior</td>
<td>1333.426</td>
<td>0.09638</td>
</tr>
<tr>
<td>(a_1 = 2.5, \ b_1 = 0.5)</td>
<td>1324.724</td>
<td>0.09723</td>
</tr>
<tr>
<td>(a_2 = 2.5, \ b_2 = 0.5)</td>
<td>1331.880</td>
<td>0.09630</td>
</tr>
<tr>
<td>(h_0 = (0,0)^T) \text{ Diag}(H_0) = 5</td>
<td>1327.240</td>
<td>0.09721</td>
</tr>
<tr>
<td>(a_0 = 2, \text{ Diag}(A_0) = 2)</td>
<td>1330.226</td>
<td>0.09626</td>
</tr>
<tr>
<td>(\rho_0 = 3, c_0 = 1, d_0 = 0.5)</td>
<td>1333.953</td>
<td>0.09534</td>
</tr>
<tr>
<td>All Above Combined</td>
<td>1324.741</td>
<td>0.09677</td>
</tr>
<tr>
<td>(a_1 = 2, \ b_1 = 8)</td>
<td>1326.314</td>
<td>0.09597</td>
</tr>
<tr>
<td>(a_2 = 2, \ b_2 = 8)</td>
<td>1333.778</td>
<td>0.09632</td>
</tr>
<tr>
<td>(a_1 = a_2 = 2, b_1 = b_2 = 8)</td>
<td>1328.790</td>
<td>0.09600</td>
</tr>
</tbody>
</table>

This table displays log-predictive likelihoods (LPL) and root-mean squared forecast errors (RMSE) for the two models, IHMM-CIR-H and IHMM-VSK-H, for various changes in the prior parameters from the assumed benchmark prior listed in Section 3.6.
Figure 1.1: Weekly 3-Month T-Bill Rate Level (Top) and Change (Bottom)

Figure 1.2: Histograms of States
Figure 1.3: Posterior of iHMM-CIR-H
Figure 1.4: Posterior of iHMM-VSK-H
Figure 1.5: Heat Map of states for IHMM-CIR-H

Figure 1.6: Heat Map of states for IHMM-VSK-H
Figure 1.7: Log-predictive Likelihood Comparison of CIR Based Models

Figure 1.8: Log-predictive Likelihood Comparison of VSK Based Models
Figure 1.11: Predictive, Log-Predictive Densities
Chapter 2

Stock Returns and Real Growth: A Bayesian Nonparametric Approach
2.1 Introduction

Studying the connection between stock market and real economic growth is an important empirical question\textsuperscript{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 illustrate 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-

\textsuperscript{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.
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 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 2.2 introduces benchmark models. Section 2.3 discusses infinite hidden Markov models and its links to Dirichlet process. Section 2.4 shows the posterior sampling steps. Section 2.5 describes how to compute out-of-sample density forecast accuracy from various models. Section 2.6 discusses empirical results, and Section 2.7 concludes the paper.

### 2.2 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

\(^2\)The lag-relation implies the lagged stock market returns should have predictive power for future real growth rates.
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.

### 2.2.1 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. Let \( r_t \) and \( g_t \) be stock market returns and real growth rates at time \( t \). We consider the following AR model augmented with lagged real growth rate:

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

(2.1)

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

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

(2.2)

Estimation of these models is carried out independently. Equation (2.1) and (2.2) are the unrestricted version. The priors for the above models are the following,

\[
    \vartheta \sim MN(a, A) \quad \frac{1}{\sigma^2} \sim Gamma(b_1, b_2),
\]

(2.3)

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 (2.1) and (2.2) 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. This approach has been suggested by Hamilton & Lin (1996) and Kanas & Ioannidis (2010). This paper estimates this model under the Bayesian approach. The following is the MS2-AR for stock market

\(^3\)A preliminary result shows the log-predictive Bayes factor based on two lags do not contribute the forecast performance under the IHMM-VAR.
returns:

\[ r_t = \mu_{st} + \beta_{1st}r_{t-1} + \beta_{2st}g_{t-1} + e_t \quad s_t \in \{1, 2\} \]  \hspace{1cm} (2.4a)
\[ e_t \overset{iid}{\sim} N(0, \sigma^2_{st}) \quad s_t | s_{t-1} \sim \Pi_{st-1} \]  \hspace{1cm} (2.4b)

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

\[ g_t = \mu_{st} + \beta_{1st}g_{t-1} + \beta_{2st}r_{t-1} + e_t \quad s_t \in \{1, 2\} \]  \hspace{1cm} (2.5a)
\[ e_t \overset{iid}{\sim} N(0, \sigma^2_{st}) \quad s_t | s_{t-1} \sim \Pi_{st-1} \]  \hspace{1cm} (2.5b)

The priors for the MS2-AR:

\[ \vartheta_{st} \sim MN(a, A) \quad \frac{1}{\sigma^2_{st}} \sim Gamma(b_1, b_2) \quad s_t \in \{1, 2\}, \]  \hspace{1cm} (2.6)

where \( \vartheta_{st} = (\mu_{st}, \beta_{1st}, \beta_{2st}) \). The restricted version of equations (2.4) and (2.5) are obtained with \( \beta_{2st} = 0 \). The prior for each row of transition matrix is the Dirichlet distribution. The posterior simulation of the MS2-AR is referred to Albert & Chib (1993b).

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 (2013b) and Maheu & Yang (2015a) 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 following are the hierarchical prior for the equation (2.6):

\[ a \sim MN(h_0, H_0) \quad A^{-1} \sim Wishart(d_0, D_0) \]  \hspace{1cm} (2.7a)
\[ b_1 \sim Gamma(\chi_0, \nu_0) \quad b_2 \sim Gamma(\chi_1, \nu_1), \]  \hspace{1cm} (2.7b)

where \( Wishart(d_0, D_0) \) denotes the Wishart distribution with \( d_0 \geq \dim(\vartheta_{st}) \) as degree of freedom and \( D_0 \) is the scale matrix with the same dimension of \( A^{-1} \).
2.2.2 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.

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

(2.8a)

\[
\Sigma = \begin{bmatrix}
    \sigma_r^2 & \rho \sigma_r \sigma_g \\
    \rho \sigma_r \sigma_g & \sigma_g^2
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
    e_{rt} \\
    e_{gt}
\end{bmatrix} \sim \text{iid } N(0, \Sigma)
\]

(2.8b)

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

\[\vartheta \sim MN(a, A), \quad \Sigma^{-1} \sim \text{Wishart}(b, B)\]

(2.9)

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 state to govern both time series.

\[
\begin{bmatrix}
    r_t \\
    g_t
\end{bmatrix} = \begin{bmatrix}
    \mu_{1s_t} \\
    \mu_{2s_t}
\end{bmatrix} + \begin{bmatrix}
    \beta_{1s_t} & \beta_{2s_t} \\
    \beta_{3s_t} & \beta_{4s_t}
\end{bmatrix} \begin{bmatrix}
    r_{t-1} \\
    g_{t-1}
\end{bmatrix} + \begin{bmatrix}
    e_{rt} \\
    e_{gt}
\end{bmatrix}, \quad \text{and } \begin{bmatrix}
    e_{rt} \\
    e_{gt}
\end{bmatrix} \sim \text{iid } N(0, \Sigma_{s_t})
\]

(2.10a)

\[
\Sigma = \begin{bmatrix}
    \sigma_{r_{s_t}}^2 & \rho \sigma_{r_{s_t}} \sigma_{g_{s_t}} \\
    \rho \sigma_{r_{s_t}} \sigma_{g_{s_t}} & \sigma_{g_{s_t}}^2
\end{bmatrix} \quad s_t | s_{t-1} \sim \Pi_{s_{t-1}} \quad s_t \in \{1, 2\}
\]

(2.10b)

The \(\Pi\) 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:

\[\vartheta_{s_t} \sim MN(a, A), \quad \Sigma_{s_t}^{-1} \sim \text{Wishart}(b, B) \quad s_t \in \{1, 2\}\]
The hierarchical priors for MS2-VAR are the following,

\[ a \sim MN(h_0, H_0) \quad A^{-1} \sim Wishart(d_0, D_0) \quad (2.11a) \]
\[ B \sim InvWhishart(e_0, E_0) \quad b \sim Gamma(\chi_0, \nu_0)I(b \geq 2) \quad (2.11b) \]

\[ \text{InvWhishart}(e_0, E_0) \] denotes an inverse Wishart distribution with the degree of freedom of \( e_0 \) and \( E_0 \) be scale square matrix with dimension of 2.

### 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, I will discuss the corresponding univariate and multivariate setting of the IHMM.

#### 2.3.1 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 \( \theta \) be the parameter set and \( \delta_{\theta_i} \) denote a probability mass at \( \theta_i \), in which case then the stick-breaking representation of \( G \sim DP(\eta, H) \) is,

\[ \pi_i = v_i \prod_{l=1}^{i-1} (1 - v_l), \quad v_i \sim Beta(1, \eta), \quad \theta_i \sim H \quad (2.12a) \]

\[ G = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i} \text{ for } i = 1, 2, \ldots, \infty \quad (2.12b) \]

The representation in equation (2.12a) is denoted as \( \{\pi_i\}_{i=1}^{\infty} \sim Stick(\eta) \). The parameter \( \eta \) governs the distribution of the weight \( (\pi_i) \) over a unit mass. A large value of \( \eta \) implies more numbers of different \( \theta_i \) but less weight is assigned to each \( \theta_i \). On the
other hand, a small value of $\eta$ indicates fewer numbers of $\theta_i$ but each $\theta_i$ 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.\footnote{Examples are Jensen & Maheu (2010), Jensen & Maheu (2013), Song (2013b) and Jensen & Maheu (2014).}

### 2.3.2 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 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,

\[ G_0|\eta, H \sim DP(\eta, H) \] (2.13a)
\[ G_j|\alpha, G_0 \sim DP(\alpha, G_0), \quad j = 1, \ldots, \infty, \] (2.13b)

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

\[ G_0 = \sum_{i=1}^{\infty} \gamma_i \delta_{\theta_i}, \quad \Gamma = \{\gamma_i\}_{i=1}^{\infty} \sim Stick(\eta), \quad \theta_i \overset{iid}{\sim} H, \] (2.14a)
\[ G_j = \sum_{i=1}^{\infty} \pi_{ji} \delta_{\theta_i}, \quad \{\pi_{ji}\}_{i=1}^{\infty} \sim Stick2(\alpha, \Gamma), \] (2.14b)

The $Stick2(\alpha, \Gamma)$\footnote{Using the notation of $Stick2(\alpha, \Gamma)$ to represent the second DP is firstly used in Maheu & Yang (2015a).} is constructed in the following way,

\[ \pi_{ji} = \hat{\pi}_{ji} \prod_{l=1}^{i-1} (1 - \hat{\pi}_{jl}), \quad \hat{\pi}_{ji} \overset{iid}{\sim} Beta \left( \alpha \gamma_i, \alpha \left( 1 - \sum_{l=1}^{i} \gamma_l \right) \right), \] (2.15)

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 $\Pi$ be the infinite Markov transition matrix. $G_j$ is served as the prior for $j$th row of $\Pi$. The $\{\pi_{ji}\}_{i=1}^\infty$ becomes the corresponding probability weights for $j$th row.

### 2.3.3 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 (2013b) 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 (2015a) 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).

#### IHMM Basics

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

\begin{align}
\{\gamma_i\}_{i=1}^\infty &= \Gamma | \eta \sim Stick(\eta) & \theta_i & \sim H & i = 1, 2, \ldots, \\
\Pi_j | \alpha, \Gamma & \sim Stick2(\alpha, \Gamma), & j = 1, 2, \ldots, \\
s_t | s_{t-1}, \Pi_{s_{t-1}} & \sim \Pi_{s_{t-1}}, & t = 1, \ldots, T \\
y_t | s_t, \Theta & \sim F(y_t | \theta_{s_t}).
\end{align}

In equations (2.16), two DPs construct the prior for Markov transition probability matrix with infinite dimension. Each row of $\Pi_j$ is drawn from a DP prior with a base measure. $H$ represents the priors over the parameter space $\Theta$. $F(\cdot)$ is the conditional density function for observation $y_t$. The $\eta$ and $\alpha$ govern the distribution of the weights. Various combinations of $\eta$ and $\alpha$ can enforce different prior beliefs on the dimension of Markov transition. For example, larger values of $\alpha$ and $\eta$ 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 $\alpha$ and $\eta$ is placed so we can infer them from the data rather than fixing them.

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

(2.17)

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 this 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 $y_{t+1}$,

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

(2.18)

**IHMM-VAR**

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

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

(2.19a)

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

(2.19b)

$$
\Sigma_{st} = \begin{bmatrix} \sigma_{rs}^2 & \rho \sigma_{rs} \sigma_{gs} \\ \rho \sigma_{rs} \sigma_{gs} & \sigma_{gs}^2 \end{bmatrix}, \quad s_t|s_{t-1}, \Pi_{s_{t-1}} \sim \Pi_{s_{t-1}}, \quad t = 1, \ldots, T
$$

(2.19c)

Let $s_t \in \{1, \ldots, \infty\}$ and $\vartheta_i = (\mu_{1i}, \mu_{2i}, \beta_{1i} \beta_{2i}, \beta_{3i}, \beta_{4i})$. The priors for $\vartheta_i$ and $\Sigma_i$ are the following.

$$
\vartheta_i \sim \text{MN}(a, A) \quad \Sigma^{-1}_i \sim \text{Wishart}(b, B) \quad i = 1, 2, \ldots
$$

(2.20)

The hierarchical prior for the IHMM is the same as equation (2.11) of MS2-VAR. The
restricted version is to set $\beta_{2s_t} = \beta_{3s_t} = 0$.

**IHMM-AR**

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

$$\Gamma|\eta \sim \text{Stick}(\eta), \quad \Pi_j|\alpha, \Gamma \sim \text{iid Stick}_2(\alpha, \Gamma), \quad j = 1, 2, \ldots,$$

$$g_t = \mu_{s_t} + \beta_{1s_t} g_{t-1} + \beta_{2s_t} r_{t-1} + e_t \quad e_t \sim N(0, \sigma_{s_t}^2)$$

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

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

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

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 (2.7) of MS2-AR.

For IHMM-AR on stock market returns, all of the parts are the same except equations (2.21b) is replaced by the following,

$$r_t = \mu_{s_t} + \beta_{1s_t} r_{t-1} + \beta_{2s_t} g_{t-1} + e_t \quad e_t \sim N(0, \sigma_{s_t}^2)$$

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

### 2.4 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 $\Pi$ and corresponding state variables $\{s_{1:T}\}$. The $u_t$ is
draw from the following density function,

\[ f(u_t|s_{t-1}, s_t, \Pi) = \frac{I(0 < u_t < \pi_{s_{t-1}, s_t})}{\pi_{s_{t-1}, s_t}}, \quad t = 1, \ldots, T \]  

(2.24)

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

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

(2.25)

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 \( t = 1, \ldots, T \), the backward step for sampling state variable \( s_t \) for \( t = T - 1, \ldots, 1 \), is the following,

\[ p(s_t|s_{t+1}, y_{1:T}, u_{1:T}) \propto p(s_t|y_{1:t}, u_{1:t})I(u_{t+1} < \pi_{s_t, s_{t+1}}) \]  

(2.26)

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 \( \Pi_j|s_{1:T}, \Gamma, j = 1, \ldots, K \)  
3. Sample \( u_{1:T}|s_{1:T}, \Pi \) and update \( K \)  
4. Sample \( \theta_j|s_{1:T}, y_{1:T}, \xi, j = 1, \ldots, K \)  
5. Sample \( \Gamma|s_{1:T}, \eta \)  
6. Sample \( \xi|\theta_1, \ldots, \theta_K, \eta|s_{1:T}, \Gamma \) 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 \( \Sigma_{s_t} \) which is \( E[\Sigma_{s_t}|y_{1:T}] \) at time \( t \) is \( \frac{1}{M} \sum_{l=1}^{M} \Sigma_{s_t}^{(l)} \).

### 2.5 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,

\[ \rho(y_{T+1}|y_{1:T}) = \int_{\Theta} f(y_{T+1}|\theta, y_{1:T}) \rho(\theta|y_{1:T}) d\theta, \quad \theta \in \Theta \tag{2.27} \]

where the marginalization is taken with respect to \( \rho(\theta|y_{1:T}) \) which is the predictive posterior distribution of \( \theta \). \( y_{1:T} \) are observations used for estimation and \( y_{T+1} \) is the observation to predict. The \( \Theta \) is the corresponding parameter set. Equation (3.1) 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 categories. 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 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.

### 2.5.1 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 rates \( g_{T+1} \) at time \( T + 1 \): At \( l \)th 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

\(^6\)The bivariate setting means the models of the VAR, MS2-VAR and IHMM-VAR
predictive likelihood for \( g_{T+1} \) is the following,

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

(2.28)

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

### 2.5.2 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 \( l \)th MCMC draw of \( \{\mu_{1,i}^{(l)}, \beta_{1,i}^{(l)}, \beta_{2,i}^{(l)}, \sigma_{i}^{2(l)}, s_{1:T,i}, \Pi^{(l)}\} \). Let \( \vartheta_i = \{\mu_{1,i}^{(l)}, \beta_{1,i}^{(l)}, \beta_{2,i}^{(l)}\} \) for \( i = 1, 2 \). We firstly draw the \( s_{T+1}^{(l)} \) through \( \Pi_{s|g}^{(l)} \) 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,

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

(2.29)

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:

\[
p(r_{T+1}, g_{T+1}|r_{1:T}, g_{1:T}) = p(r_{T+1}|r_{1:T}, g_{1:T})p(g_{T+1}|r_{1:T}, g_{1:T})
\]  

(2.30)

### 2.5.3 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 \( l \)th MCMC draw and a sequence of posterior draws is given by \( \{\mu_1^{(l)}, \mu_2^{(l)}, \beta_1^{(l)}, \beta_2^{(l)}, \sigma_g^{2(l)}, \sigma_r^{2(l)}, \beta_3^{(l)}, \beta_4^{(l)}, \Sigma^{(l)}\} \). 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,

\[ p(r_{T+1} | r_{1:T}, g_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} N(r_{T+1} | \varrho_r^{(l)}, \sigma_r^{2(l)}) \]  
\[ p(g_{T+1} | r_{1:T}, g_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} N(g_{T+1} | \varrho_g^{(l)}, \sigma_g^{2(l)}) \]  
\[ p(r_{T+1} | g_{T+1} | r_{1:T}, g_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} MN(r_{T+1}, g_{T+1} | \varrho_r^{(l)}, \varrho_g^{(l)}, \Sigma^{(l)}) \]  
\[ (2.31a) \]  
\[ (2.31b) \]  
\[ (2.31c) \]

### 2.5.4 MS2-VAR

Given \(l\)th 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 \(\varphi_{ri}^{(l)} = \{\mu_{1i}^{(l)}, \beta_{1i}^{(l)}, \beta_{2i}^{(l)}\}\), \(\varphi_{gi}^{(l)} = \{\mu_{2i}^{(l)}, \beta_{3i}^{(l)}, \beta_{4i}^{(l)}\}\) and \(\Sigma_i^{(l)} = \{\sigma_{ri}^{(l)}, \sigma_{gi}^{(l)}, \rho_i^{(l)}\}\).

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,

\[ p(r_{T+1} | r_{1:T}, g_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} N(r_{T+1} | \varrho_r^{(l)}, \sigma_r^{2(l)}) \]  
\[ (2.32a) \]  
\[ p(g_{T+1} | r_{1:T}, g_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} N(g_{T+1} | \varrho_g^{(l)}, \sigma_g^{2(l)}) \]  
\[ (2.32b) \]  
\[ p(r_{T+1} | g_{T+1} | r_{1:T}, g_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} MN(r_{T+1}, g_{T+1} | \varrho_r^{(l)}, \varrho_g^{(l)}, \Sigma_k^{(l)}) \]  
\[ (2.32c) \]

### 2.5.5 IHMM-AR

Let \(\{\varphi_i^{(l)}, \sigma_i^{(l)}, \Pi^{(l)}, s_{1:T}, K^{(l)}, \xi^{(l)}\}\), \(\varphi_i = \{\mu_{1i}, \beta_{1i}, \beta_{2i}\}\) and \(i \in \{1, \ldots, K^{(l)}\}\) be the \(l\)th posterior draws, where \(K^{(l)}\) is the total number of active states, which means at least one observation is assigned. \(s_{T+1}^{(l)} \in \{1, \ldots, K^{(l)}\}\). 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 \(s_{T+1}^{(l)} \leq K^{(l)}\), which suggests the \(g_{T+1}\) or its corresponding \(s_{T+1}\) belong to existing states, then set \((\varphi_{T+1}^{(l)}, \sigma_{T+1}^{2(l)}) \equiv (\varphi_k^{(l)}, \sigma_k^{2(l)})\), where \(k \in \{1, \ldots, K^{(l)}\}\). Otherwise, \((\varphi_k^{(l)}, \sigma_k^{2(l)}) \sim H(\xi(k))\) and thus, \((\varphi_{T+1}^{(l)}, \sigma_{T+1}^{2(l)}) \equiv (\varphi_k^{(l)}, \sigma_k^{2(l)})\). 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,

$$p(g_{T+1}|r_{1:T}, g_{1:T}) \approx \frac{1}{M} \sum_{l=1}^{M} N(g_{T+1}|\varphi_{k}^{(l)}, \sigma_{k}^{2(l)}),$$  \hspace{1cm} (2.33)

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.

### 2.5.6 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 \{\varphi_{ri}, \varphi_{gi}, \Sigma_{i}, \Pi, s_{1:T}, K, \xi^{(l)}\} be the $l$th posterior draw. Some extra notations: \varphi_{ri} = \{\mu_{1i}, \beta_{1i}, \beta_{2i}\} and \varphi_{gi} = \{\mu_{2i}, \beta_{3i}, \beta_{4i}\}, \Sigma_{i} = \{\sigma_{ri}^{2}, \sigma_{gi}^{2}, \rho_{i}\} and $i \in \{1, \ldots, K\}$, $s^{(l)}_t \in \{1, \ldots, K\}$, 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 $l$th 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 \{\varphi_{r_{sT+1}}^{(l)}, \varphi_{g_{sT+1}}^{(l)}, \Sigma_{sT+1}^{(l)}\} \equiv (\varphi_{rk}^{(l)}, \varphi_{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. \{\varphi_{rk}^{(l)}, \varphi_{gk}^{(l)}, \Sigma_{k}^{(l)}\} \sim H(\xi^{(l)}) and thus, \{\varphi_{r_{sT+1}}^{(l)}, \varphi_{g_{sT+1}}^{(l)}, \Sigma_{sT+1}^{(l)}\} \equiv (\varphi_{rk}^{(l)}, \varphi_{gk}^{(l)}, \Sigma_{k}^{(l)})

\begin{align*}
p(r_{T+1}|r_{1:T}, g_{1:T}) & \approx \frac{1}{M} \sum_{l=1}^{M} N(r_{T+1}|\varphi_{rk}^{(l)}, \sigma_{rk}^{2(l)}) \quad (2.34a) \\
p(g_{T+1}|r_{1:T}, g_{1:T}) & \approx \frac{1}{M} \sum_{l=1}^{M} N(g_{T+1}|\varphi_{gk}^{(l)}, \sigma_{gk}^{2(l)}) \quad (2.34b) \\
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}|\varphi_{rk}^{(l)}, \varphi_{gk}^{(l)}, \Sigma_{k}^{(l)}) \quad (2.34c)
\end{align*}
2.6 Empirical Results

2.6.1 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 3.1 shows the plot of corresponding stock market returns and real growth rates. Table 2.1 illustrates their statistical summaries.

2.6.2 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, \( \theta \sim MN(0, I_P) \) 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 \( \sigma^2_{a_i} \), 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 \( \eta \) and \( \alpha \), 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.

---

\(^7\)The hierarchical priors requires mixture models such as Markov switch with two states. The AR and VAR models implicitly assumes one state.
2.6.3 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 \( \tau_1 \) and \( \tau_2 \) 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 (\( LPL_{joint} \)) are the following,

\[
LPL_r = \log \prod_{t=\tau_1}^{\tau_2} p(r_{t+1}|r_{1:t}, g_{1:t}) = \sum_{t=\tau_1}^{\tau_2} \log p(r_{t+1}|r_{1:t}, g_{1:t}) \tag{2.35a}
\]

\[
LPL_g = \log \prod_{t=\tau_1}^{\tau_2} p(g_{t+1}|r_{1:t}, g_{1:t}) = \sum_{t=\tau_1}^{\tau_2} \log p(g_{t+1}|r_{1:t}, g_{1:t}) \tag{2.35b}
\]

\[
LPL_{joint} = \log \prod_{t=\tau_1}^{\tau_2} p(r_{t+1}, g_{t+1}|r_{1:t}, g_{1:t}) = \sum_{t=\tau_1}^{\tau_2} \log p(r_{t+1}, g_{t+1}|r_{1:t}, g_{1:t}) \tag{2.35c}
\]

The \( LPL_r \), \( LPL_g \) and \( LPL_{joint} \) for all the models are reported in Tables 2.2, 2.3 and 2.4. 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 2.2, 2.3, and 2.4, 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 (2.35a), 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 2.3 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.
2.6.4 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 2.2 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 2.3 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 2.4 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 2.2 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 2.2 show constant increasing trend, which suggests that the superior forecast accuracy of the IHMM-VAR in Table 2.2 is not due to any particular period or outliers, but steady ongoing gains. Similar outcomes are reflected in Figure 2.3 and Figure 2.4, which correspond to the cumulative log-predictive Bayes factor on stock market returns and real growth rates. In Figure 2.3, 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.

2.6.5 Posterior Analysis and Prior Robustness

Figure 2.5 and Figure 2.6 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 2.7 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 2.8 illustrates how frequently the IHMM-VAR introduces new states; the average posterior of states number in Figure 2.8 is computed under a recursive increasing sample from mid 1934 until the end of 2014.

Table 3.9 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 $\alpha$ and $\eta$, 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 3.9.

### 2.6.6 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 2.9 is a heat map and shows the probability of 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 2.5 and Figure 2.6 illustrate the posterior mean and interval of parameters in the IHMM-VAR, they reflect the impact of structural break on the parameters’ dynamics. Figure 2.5 indicates that the structural change at 1984 not only affects the posterior mean of $\mu_{1s_t}$, $\mu_{2s_t}$, $\beta_{1s_t}$ and $\beta_{2s_t}$, but also changes its posteriors’ interval after 1984. For example, the posterior of $\beta_{2s_t}$ 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 2.6, which are the posterior mean and intervals of $\sigma_{r_{st}}$, $\sigma_{g_{st}}$ and $\rho_{s_t}$. Both figure 2.5 and figure 2.6 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

---

$^8$Great Moderation.
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 2.10 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 favour 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 slope 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.

2.6.7 Empirical Results

The asset pricing literature suggests the stock market returns should have predictive power for future economic growth, while existing works suggest the lag stock market returns variables should have predictive power on real growth rates. 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 2.6 is the log-predictive Bayes factor between unrestricted and restricted versions of each model. Table 2.6 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, no evidence shows the $r_{t-1}$ has predictive power for $g_t$. Similar outcomes happen to MS2-VAR, and AR, which suggests inconsis-

---

9 According to dividend discount valuation and consumption capital asset pricing model.
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 2.2, Table 2.3 and Table 2.4.
tent results with respect to Fama (1990) and Choi et al. (1999). On the other side, Table 2.6 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 2.4, and it reveals that these models are poorly performed in term of out-of-sample forecast accuracy. Table 2.3 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 2.4, 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 2.4. 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.

2.7 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 breaks. 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.
2.8 Appendix

2.8.1 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 $\Gamma$. 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 $x_l \sim Bernoulli(\frac{\alpha \gamma_i}{1+\alpha \gamma_i})$, 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 $\alpha$ given $\chi_2$ and $\nu_2$.

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

   (a) $\bar{\nu}_j \sim Bernoulli(\frac{n_j}{n_j + \alpha})$ for $j = 1, \ldots, K$. $n_j = \sum_{i=1}^K n_{ji}$
   
   (b) $\bar{\lambda}_j \sim Beta(\alpha + 1, n_j)$ for $j = 1, \ldots, K$.
   
   (c) $\alpha \sim Gamma(\chi_2 + c. - \sum_{j=1}^K \bar{\nu}_j, \nu_2 - \sum_{j=1}^K log(\bar{\lambda}_j))$

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

3. Sample $\eta$ given $\chi_1$ and $\nu_1$.

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

4. Sample $\Gamma$:

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

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

5. Sample $\Pi_j$: for $j = 1, \ldots, K$

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

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

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

7. Adaptive Truncation of $\Pi$:

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

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

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

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

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

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

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 $\Pi$ using $u_{1:T}$ from infinite dimension into finite dimension denoted by $\tilde{\Pi}$ as well as the associated full parameter space $\Theta$ into a set of finite set is denoted by $\tilde{\Theta}$.

8. Sample $s_{1:T}$

(a) Initial step for $s_1$:

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

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

\[\text{footnote}{\text{A finite number of } \theta \text{ which will be considered in the FFBS, includes all alive states and finite number of unrepresented states.}}\]
(b) The forward-filtering part for $s_{2:T}$:

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

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

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

(c) Sample initial $s_T$:

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

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

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

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

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

$$y_t = Z_t \vartheta + \epsilon_t \quad \epsilon_t \sim MN(0, \Sigma_{st}),$$

where,

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

The posterior of $\vartheta_k$ is the following,

$$\vartheta_k|y_t, Z_t, \Sigma_k^{-1}, a, A \sim MN(a_1, A_1) \quad k = 1, \ldots, L$$

$$a_1 = A_1 \left( A^{-1} a + \sum_{s_t = k} Z_t^i \Sigma_k^{-1} y_t \right) \quad A_1 = \left( A^{-1} + \sum_{s_t = k} Z_t^i \Sigma_k^{-1} Z_t \right)^{-1}$$
The posterior of $\Sigma_k^{-1}$ is the following,

$$
\Sigma_k^{-1}|\vartheta_k, y_t, Z_t, b, B \sim \text{Wishart}(b_1, B_1)
$$

$$
b_1 = T_k + b_1 \quad B_1 = \left( B_1^{-1} + \sum_{s_t=k}(y_t - Z_t\vartheta_k)(y_t - Z_t\vartheta_k)' \right)^{-1}
$$

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

$$
\vartheta_k \sim \text{MN}(a, A) \quad \Sigma_k^{-1} \sim \text{Wishart}(b, B) \quad k = 1, \ldots, K
$$

The hierarchical priors the following:

$$
a \sim \text{MN}(h_0, H_0) \quad A^{-1} \sim \text{Wishart}(d_0, D_0)
$$

$$
B \sim \text{InvWhishart}(e_0, E_0) \quad b \sim \text{Gamma}(\chi_0, \nu_0) I(b \geq 2)
$$

(a) Sample $a$:

$$
a|h_0, H_0, A, \{\vartheta_k\}_{k=1}^K \sim \text{MN}(\mu_a, V_a)
$$

$$
\mu_a = V_a \left( H_0^{-1}h_0 + A^{-1}\sum_{k=1}^K \vartheta_k \right) \quad V_a = \left( H_0^{-1} + KA^{-1} \right)^{-1}
$$

(b) Sample $A$:

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

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

(c) Sample $B$:

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

(d) Sample $b$:

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

\[
\pi(b|\chi_0, \nu_0, B, \{\Sigma_k\}_{k=1}^K) = G(b|\chi_0, \nu_0) \prod_{1}^{K} \text{Wishart}(\Sigma_k^{-1}|b, B) \\
q(b^{new}|b^{old}) \sim \text{Gamma}(\zeta, \zeta/b^{old})I(b^{new} \geq M + 1)
\]

\[\text{AcceptProbability} = \min \left\{ \frac{\pi(b^{new}|\chi_0, \nu_0, \{\Sigma_k\}_{1}^{K})/q(b^{new}|b^{old})}{\pi(b^{old}|\chi_0, \nu_0, \{\Sigma_k\}_{1}^{K})/q(b^{old}|b^{new})}, 1 \right\} \]

11. Repeat 1-10.
Table 2.1: Statistical Summaries of stock market returns and Real Growth Rates

<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>Variance</th>
<th>Median</th>
<th>25%Q</th>
<th>75%Q</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>stock market returns</td>
<td>0.648</td>
<td>30.24</td>
<td>0.945</td>
<td>-2.043</td>
<td>3.595</td>
<td>0.409</td>
<td>9.603</td>
</tr>
<tr>
<td>Real Growth Rates</td>
<td>0.263</td>
<td>3.211</td>
<td>0.308</td>
<td>-0.304</td>
<td>0.846</td>
<td>0.354</td>
<td>14.87</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Unrestricted</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>-6591</td>
<td>-6589</td>
<td></td>
</tr>
<tr>
<td>MS2-AR</td>
<td>-4689</td>
<td>-4721</td>
<td></td>
</tr>
<tr>
<td>IHMM-AR</td>
<td>-4672</td>
<td>-4701</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Unrestricted</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR</td>
<td>-4659</td>
<td>-4672</td>
<td></td>
</tr>
<tr>
<td>MS2-VAR</td>
<td>-4230</td>
<td>-4165</td>
<td></td>
</tr>
<tr>
<td>IHMM-VAR</td>
<td>-4153</td>
<td>-4067</td>
<td></td>
</tr>
</tbody>
</table>

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 2.3: Log-Predictive Likelihood on Stock Market Returns (\(LPL_r\))

<table>
<thead>
<tr>
<th>Models</th>
<th>Unrestricted</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>-3986</td>
<td>-3986</td>
</tr>
<tr>
<td>MS2-AR</td>
<td>-2973</td>
<td>-2974</td>
</tr>
<tr>
<td>IHMM-AR</td>
<td>-2970</td>
<td>-2971</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>Unrestricted</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR</td>
<td>-2954</td>
<td>-2951</td>
</tr>
<tr>
<td>MS2-VAR</td>
<td>-2810</td>
<td>-2807</td>
</tr>
<tr>
<td>IHMM-VAR</td>
<td>-2796</td>
<td>-2794</td>
</tr>
</tbody>
</table>

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 2.4: Log-Predictive Likelihood on Real Growth Rates ($LPL_g$)

<table>
<thead>
<tr>
<th></th>
<th>Unrestricted</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>-2605</td>
<td>-2603</td>
</tr>
<tr>
<td>MS2-AR</td>
<td>-1716</td>
<td>-1747</td>
</tr>
<tr>
<td>IHMM-AR</td>
<td>-1702</td>
<td>-1730</td>
</tr>
<tr>
<td>VAR</td>
<td>-1711</td>
<td>-1724</td>
</tr>
<tr>
<td>MS2-VAR</td>
<td>-1416</td>
<td>-1360</td>
</tr>
<tr>
<td>IHMM-VAR</td>
<td>-1352</td>
<td>-1273</td>
</tr>
</tbody>
</table>

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 $LPL_g$ 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 2.5: Log-Predictive Likelihood on Joint Stock Returns and Growth Rates of IHMM-VAR (Restricted Version) with Different Hyper Priors

<table>
<thead>
<tr>
<th>Type</th>
<th>Hyper Priors</th>
<th>LPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Loose</td>
<td>$\chi_1 = 5, \nu_1 = 1$ and $\chi_1 = 5, \nu_1 = 1$</td>
<td>-4067</td>
</tr>
<tr>
<td>Less Loose</td>
<td>$\chi_1 = 5, \nu_1 = 1$ and $\chi_1 = 2.5, \nu_1 = 0.5$</td>
<td>-4066</td>
</tr>
<tr>
<td>Tight</td>
<td>$\chi_1 = 2, \nu_1 = 1$ and $\chi_1 = 5, \nu_1 = 1$</td>
<td>-4064</td>
</tr>
<tr>
<td>Less Tight</td>
<td>$\chi_1 = 5, \nu_1 = 1$ and $\chi_1 = 2, \nu_1 = 8$</td>
<td>-4068</td>
</tr>
<tr>
<td>Very Tight</td>
<td>$\chi_1 = 2, \nu_1 = 2$ and $\chi_1 = 2, \nu_1 = 8$</td>
<td>-4069</td>
</tr>
</tbody>
</table>

Table 2.6: Predictive Power of lagged stock market returns on Future Real Growth Rates

<table>
<thead>
<tr>
<th></th>
<th>Log-Predictive Bayes Factor</th>
<th>Do Lagged Stock Returns Contribute to Forecast Real Growths?</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>-2</td>
<td>No</td>
</tr>
<tr>
<td>MS2-AR</td>
<td>31</td>
<td>Yes</td>
</tr>
<tr>
<td>IHMM-AR</td>
<td>28</td>
<td>Yes</td>
</tr>
<tr>
<td>VAR</td>
<td>13</td>
<td>Yes</td>
</tr>
<tr>
<td>MS2-VAR</td>
<td>-6</td>
<td>No</td>
</tr>
<tr>
<td>IHMM-VAR</td>
<td>-18</td>
<td>No</td>
</tr>
</tbody>
</table>

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 2.1: 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 2.2: 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 2.3: 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 2.4: 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 2.5: 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 $\beta_{3s_t} = 0$. 
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 2.7

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

Figure 2.8: 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).
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 2.10: 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.
Chapter 3

A Bayesian Approach to Oil Price and Real Economy Relationship


3.1 Introduction

This paper studies the predictive power of oil price information for forecasting the U.S. industrial production. Oil price information is divided into two distinct categories: nominal oil price changes, and oil price shocks using four definitions proposed in the literature. Previous work has documented lack of predictive relationship of oil price changes but significant predictive power of oil price shocks\(^1\) for U.S. economic growth. However, existing studies focused only on predicting the mean of the economic growth using classical point forecast techniques (see Hamilton (1983), Hamilton (2011), Lutz & Vigfusson (2013), Hooker (1996), Ravazzolo & Rothman (2013), and Bachmeier et al. (2008)). As a contribution to the existing literature, we propose a new forecasting model for the analyzed relationship where oil price shocks have the additional flexibility to influence both the mean and the variance of U.S. industrial production. Our forecast is well performed using the Bayesian predictive likelihood as opposed to classical point estimate. We show that the new model for oil price shocks outperforms existing models in terms of forecasting ability. We further confirm previous findings regarding the lack of predictive power of nominal oil price changes.

There has been a great deal of interest in determining whether past oil price shocks are helpful in predicting U.S. real growth (see Hamilton (2003), Hamilton (2011), Lutz & Vigfusson (2011b), and Ravazzolo & Rothman (2013)). As a key part of the study of economic growth, we construct a model which accommodates the oil price shock to both expected mean and variance of economic growth. This parsimonious model allows the oil price shock to enter both the future expected mean and conditional variance of the economic growth. Our empirical results suggest that the new model significantly captures the dynamic relationship of oil price shock and economic growth better than benchmark models. Studying their empirical relationship through the first moment has been the only practice according to existing literature.

The existing works use point forecast to measure the out-of-sample forecast accuracy. For example, Bachmeier et al. (2008) calculate the mean-squared prediction error (MSPE) and apply to equal predictive power test\(^2\). Lutz & Vigfusson (2011b) and Lutz & Vigfusson (2013) use the MSPE ratio to compare the out-of-sample forecast performance. Ravazzolo & Rothman (2013) apply the CW test and HW test\(^3\) to evaluate the forecast gain and loss. This paper examines the forecast performance through the

---

\(^1\)The oil price shock is a special case of oil price change. Please refer to Section 3.5 for detailed definitions.

\(^2\)Chao et al. (2001) and Corradi & Swanson (2002)

density forecast. The advantage of density forecast is to allow us to evaluate the entire predictive distribution, where the point forecast only focuses the average of the predictive distribution. As a result, density forecast is able to do a better job than point forecast. The out-of-sample period is from October 1975 to September 2015, which is the largest out-of-sample period ever explored. Our empirical results indicate the oil price shock has a significant impact in forecasting real growth under the density forecast approach.

Early work suggests oil price changes are helpful in explaining future U.S. real growth (see Hamilton (1983)). Other works argue that oil price change is irrelevant to future economic growth (see Mork (1989) and Hooker (1996)). In this paper, we reinvestigate this empirical question by using density forecast. Our test results are consistent with Mork (1989) and Hooker (1996). Once the oil price changes are replaced by the oil price shock, our empirical results show that the past oil price shock is able to increase the forecast accuracy of economic growth; however, other works suggest a weak relationship only (see Bachmeier et al. (2008), Ravazzolo & Rothman (2013), and Lutz & Vigfusson (2011a)). One reason is due to the density forecast that we apply to this question instead of point forecast. Another reason is that our model allows the impact of oil price shock on expected change of real growth, but also on the variance of real growth.

The remainder of the article is organized as follows. Section 3.2 reviews the data. Section 3.3 explains the density forecast and its computation method. Section 3.4 analyzes the impact of oil price changes on future economic growth. Section 3.5 extends the analysis by proposing a new model to study the relationship of oil price shock and real growth; and it explains the definition of oil price shock and the proposed approach for modeling oil price shock. Section 3.6 contains the priors of our models and posterior sampling approach. In section 3.7, we analyze our empirical findings. The conclusion remarks are contained in Section 3.8.

### 3.2 Data

This paper chooses the industrial production index as an indicator of the real economy, where a lot of literature uses real GDP as a proxy for the real economy. We aim to shed light on the industrial production index which has a long history and is high frequency compared to GDP. For oil price, there is a fair amount of discussion regarding the "right" oil price data. We select the one of most representative oil price data, and RAC composite as our proxy to oil price. We apply the nominal oil price. It is without question

---

4Maheu & Yang (2015b)
an interesting extension to see how sensitive our empirical results respond to the different choice of the oil price. All data are monthly. The rate changes of oil price are log difference of monthly U.S. RAC (Refiners’ Acquisition Cost) composite, and real growth rates are the log difference of industrial production index of U.S. \( r_t \) is rate change of oil price and \( g_t \) is real growth rates at \( t \). There are 499 observations, and the out-of-sample size is 480 observations. The sample is dated from February 1974 to September 2015. Figure 3.1 plots two time series. The top one is the rate of changes of oil prices. The bottom plot is the real growth rates. The oil price shock is a special case of oil price change. Their distinguishable features are discussed later.

### 3.3 Out-of-Sample Forecast

Early work has documented the predictive power of oil price shock for future economic growth through the in-sample evidence. The out-of-sample forecast is the ultimate question of interest to the purpose of this paper. The mainstream approach for the out-of-sample forecast is the point forecast, which compares the total loss between the realization and expectation of mean\(^6\). In other words, the point forecast evaluates the forecast accuracy through a certain point of the predictive distribution. However, the density forecast evaluates the forecast accuracy through the entire predictive distribution rather than individual points.

The predictive likelihood measures the out-of-sample density forecast accuracy. In contrast to point forecast, such as root mean squared forecast errors (RMSFE), the predictive likelihood evaluate the predictive distribution as a whole, where the RMSFE only focuses on the central of the predictive distribution. Thus, it is not surprising that sometimes the two methods deliver contradictory outcomes. Let \( y_t \) be the observation at time \( t \). A representation of predictive likelihood of \( y_{t+1} \) is the following,

\[
p(y_{t+1}|y_{1:t}) = \int_{\Theta} f(y_{t+1} | \theta, y_{1:t}) p(\theta | y_{1:t}) d\theta, \quad \theta \in \Theta \tag{3.1}
\]

where the marginalization is taken with respect to \( \rho(\theta | y_{1:t}) \), which is the predictive posterior distribution of \( \theta \). \( y_{1:t} \) are observations used for estimation and \( y_{t+1} \) is the observation to predict, while the \( \Theta \) is the corresponding parameter set. Equation (3.1) can also be used for evaluating the model fitting since the parameter uncertainties are incorporated into the predictive likelihood computation because 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. After finishing computing $p(y_{t+1}|y_{1:t})$, we move to calculate $p(y_{t+2}|y_{1:t+1})$. We continue repeating the above process until the last observation in the out-of-sample period is predicted.

Next, we repeat the above process to evaluate the prediction performance at $y_{t+2}$ until the last observation in the out-of-sample period. Once the predictive likelihood of out-of-sample is computed, we evaluate the forecast performance through the aggregation:

$$\text{Log-predictive likelihood} = \log \prod_{t=\tau_1}^{\tau_2} p(y_{t+1}|y_{1:t}) = \sum_{t=\tau_1}^{\tau_2} \log p(y_{t+1}|y_{1:t})$$

The $\tau_1$ and $\tau_2$ are the beginning and the end of the out-of-sample period. This paper let $\tau_1 = 20$ and $\tau_2 = 499$. The out-of-sample size is 480 with a total sample size of 499. The following rule is used to make the decision of model computation:

$$\frac{\prod_{t=\tau_1}^{\tau_2-1} p(y_{t+1}|y_{1:t}, Model_A)}{\prod_{t=\tau_1}^{\tau_2-1} p(y_{t+1}|y_{1:t}, Model_B)} \geq 150$$

Similarly,

$$\sum_{t=\tau_1}^{\tau_2-1} \log p(y_{t+1}|y_{1:t}, Model_A) - \sum_{t=\tau_1}^{\tau_2-1} \log p(y_{t+1}|y_{1:t}, Model_B) \geq 5$$

Kass & Raftery (1995) suggest if the ratio is larger than 150 or 5 in terms of the logarithm, then $Model_A$ is strongly favoured. $Model_B$ is strongly favoured, when the above number is less than $\frac{1}{150}$ or less than $-5$.

### 3.4 Oil Price Changes

This section investigates the predictive power of lagged oil price change for future real growth rates. Some literature suggests the significant predictive power of lagged variables of oil price change for future real growth (see Hamilton (1983) and Hamilton (1996)). Also, some work suggest weak predictive power according to the point forecast (see Bachmeier et al. (2008), Ravazzolo & Rothman (2013), Lutz & Vigfusson (2013) ). This paper offers new insight into the study of their relationship through the density forecast. There are three categories of specifications to test. We compute the log-predictive likelihood for each specification and select the best performing model. Our empirical results suggest the oil price changes do not contribute to predicting future real growths, and the lagged real growth variables show significant predictive power for future real growth.
3.4.1 Autoregression with Total Lags

This approach includes the accumulated number of lags from most recent months to as far as 12 months. The \( q \) and \( p \) are the total number of lags for real growths and rate change of oil price. \( q \geq 1 \) and \( p \geq 1 \).

\[
g_t = \mu + \sum_{j=1}^{q} \alpha_j g_{t-j} + \sum_{j=1}^{p} \beta_j r_{t-j} + \sigma e_t \quad e_t \sim \text{N}(0, 1) \quad (3.2)
\]

The \( g_t \) and \( r_t \) are corresponded to real economic growths and rate change of oil price at time \( t \). At each MCMC draw, the predictive distribution of \( g_{t+1} \) is predicted, and it is a normal distribution. Then, the realization of \( g_{t+1} \) is plugged into the predictive distribution. Thus, the predictive likelihood of \( g_{t+1} \) is obtained. Given \( l \)th MCMC draw of \( \theta^{(l)} = \{\mu^{(l)}, \alpha^{(l)}_1, \alpha^{(l)}_2, \ldots, \alpha^{(l)}_q, \beta^{(l)}_1, \beta^{(l)}_2, \ldots, \beta^{(l)}_p, \sigma^{(l)}\} \) and corresponding lagged variables, the above steps are repeated over \( M \) iterations, and the predictive likelihood of \( g_{t+1} \) is achieved as the following,

\[
p(g_{t+1}|g_{1:t}, r_{1:t}) = \frac{1}{M} \sum_{l=1}^{M} N(g_{t+1}|g_{1:t}, r_{1:t}, \theta^{(l)})
\]

The Table 3.1 shows the log-predictive likelihood for a different combination of \( q \) and \( p \). The best-performed combination is at \( q = 12 \) and \( \beta_{1_p} = 0 \). The log-predictive likelihood is -500. The lagged real growth variables contribute to predicting future real growth while the lagged rate change of oil price hardly helps to explain the future real growth according to Table 3.1.

3.4.2 Autoregression with Individual Lag

We marginalize out parameters’ uncertainties when we compute the predictive likelihood. Because these explanatory variables are not contributing to explaining the left side of the equation, the density forecast accuracy will fall dramatically when these variables are included on the right side of the equation. As a result, this section only investigates the individual lags so that we can select the explanatory variables which make the most significant contribution. The \( q \) and \( p \) now correspond to \( q \)th lag and \( p \)the lag for real growths and rate change of oil price.

\[
g_t = \mu + \alpha g_{t-q} + \beta r_{t-p} + \sigma e_t \quad e_t \sim \text{N}(0, 1) \quad (3.3)
\]

Given each \( l \)th MCMC draw of \( \theta^{(l)} = \{\mu^{(l)}, \alpha^{(l)}, \beta^{(l)}, \sigma^{(l)}\} \) and corresponding lagged
variables, the predictive likelihood of $g_{t+1}$ is calculated as similar as the previous section except the mean of predictive distribution becomes $\mu^{(l)} + \alpha^{(l)}g_{t-q} + \beta^{(l)}r_{t-p}$ at each MCMC draw.

Table 3.2 shows the log-predictive likelihood for different combination of $q$ and $p$. It suggests $g_{t-3}$ and $r_{t-12}$ contribute the most to explain future real growth rates. The corresponding log-predictive likelihood is -495.

### 3.4.3 Autoregression with Average Lag

The use of average lag is a midway approach to test their relationship. This approach allows us to consider a large number of explanatory variables without letting each of them enter the right side of equation individually. The $q$ and $p$ now correspond to an average of past $q$ lags and $p$ lags. $q \geq 1$ and $p \geq 1$.

$$
g_t = \mu + \frac{1}{q} \sum_{j=1}^{q} g_{t-j} + \frac{1}{p} \sum_{j=1}^{p} r_{t-j} + \sigma e_t \quad e_t \sim iid \sim N(0,1) \quad (3.4)$$

Given each $l$th MCMC draw of $\theta^{(l)} = \{\mu^{(l)}, \alpha^{(l)}, \beta^{(l)}, \sigma^{(l)}\}$ and corresponding lagged variables, the predictive likelihood of $g_{t+1}$ is the average of all MCMC iterations.

Table 3.4 shows the log-predictive likelihood for a different combination of $q$ and $p$. The Table suggests $q = 3$ and $\beta = 0$ are a best-performed model. This implies lagged oil price changes variables are insignificant to be included. The log-predictive likelihood is -493.

According to Table 3.1, Table 3.2, and Table 3.4, we select autoregression with average lag at $q = 3$ and $\beta = 0$ as the best performed model to describe the relationship of real growth and oil price change. This version of the model is also selected as our benchmark model to compare to the same model with oil price shock included.

### 3.5 Oil Price Shock

A special case of oil price change is the oil price shock. Hamilton (2003) and Hamilton (2011) distinguish the oil price shock from the oil price change and claim past oil price shock is the key to predicting real growth. This section is organized into two parts. The first part is defined by various popular measurement of oil price shock. The second introduces our proposed model to study the relationship of oil price shock and real growth.
3.5.1 Oil Price Shock Measurements

This part defines several different approaches to measuring oil price shock, which is suggested by Lutz & Vigfusson (2013). There are four different measurements of oil price shocks. Let $d_{t-p}^n$ be the oil price shock at time $t$ and $OP_t$ be the oil price at time $t$. For all types of shocks, we let $n = 12$, and it suggests the shock is measured according to a recent 12-month.

Type-1

This is the most popular way of defining the oil price shock, and it is developed by Hamilton (1996). $d_{t-p}^n = \max \{0, \ln \frac{OP_{t-p}}{OP_{t-p}^{n+}}\}$. The $OP_{t-p}^{n+}$ is the highest oil price in the past $n$ months starting from $t - p$, such as $\{t - p - 1, \ldots, t - p - n\}$. This type of oil price shock implies the shock is identified when it exceeds the recent historical maximum price. Notably, the negative price shock does not matter to future economic growths. Moreover, the magnitudes of the shocks have impact according to type-1.

Type-2

This type of shock is the same as type-1 except the magnitude of the shock does not matter at all. For example: $d_{t-p}^n = I\{0, \ln \frac{OP_{t-p}}{OP_{t-p}^{n+}}\}$. Regardless of the size of the shock, the impact is always treated equally.

Type-3

This type of shocks studies the negative oil price shock instead of positive oil price shock such as type-1. Let $d_{t-p}^n = \min \{0, \ln \frac{OP_{t-p}}{OP_{t-p}^{n-}}\}$, where $OP_{t-p}^{n-}$ is the lowest oil price in the past $n$ months from $t - p - 1$ to $t - p - n$.

Type-4

This type of shock is determined by the standard deviation of historical rate change of oil price. For example, $d_{t-p}^n = I\{r_{t-p} > \frac{1}{2}std(r_{t-p-n:t-p})\}$. The $r_{t-p} = \ln \frac{OP_{t-p}}{OP_{t-p-1}}$. Type-4 suggests an alternative approach to measure the positive oil price shock; the rate of increase of oil price must be higher than the half of the standard deviation in past 12 months. Therefore, more shocks should be recognized under type-4 than type-1.
3.5.2 Oil Price Shock and Real Growth

As mentioned early, we model the transmission of oil price shock on real growth through two channels. One channel is the expected mean of real growth. Another channel is the variance of real growth. The existing literature has only focused on the connection of oil price shock and the expected mean of real growth. The shocks of oil price on the volatility of real growth have been largely ignored. This paper sheds light on another channel, which is the transmission of oil price shock through the volatility real growth.

Given the average lags of \( q = 3 \) and \( \beta = 0 \) to be the best performed version, we build on this version by including the oil price shock.

\[
g_t = \mu + \alpha \frac{1}{3} \sum_{j=1}^{3} g_{t-j} + \lambda d_{t-p} + e_t \quad e_t \sim iid \sim \mathcal{N}(0, \sigma^2 \exp(2\delta d_{t-p})) \tag{3.5}
\]

The \( \lambda \) represents the impact of oil price shock on the expected mean of \( g_t \), and \( \delta \) represents the impact of oil price shock on the volatilities of \( g_t \). The intuition of \( \delta \) is a multiplier, which could be either positive or negative. The \( \delta \) determines the magnitude of shrinkage or expansion on the \( \sigma^2 \) at a given oil price shock. The lag indicator \( p \) suggests the time it takes price shock to predicts the real growth. We will carry out the following out-of-sample density forecasts: within each type of oil price shock measurement, we investigate three versions of the model. They are \( \lambda \neq 0 \) and \( \delta \neq 0 \), \( \lambda = 0 \) and \( \delta \neq 0 \), and \( \lambda \neq 0 \) and \( \delta = 0 \). The purpose is to discern in which way the shock is more influential. Moreover, we explore the lagging effects, which are for testing the time it takes for the oil price shock to hit the future real growth.

3.6 Priors and Posterior Sampling

In order to produce robust results, we apply non-informative priors to all models. For equations (3.2), (3.3), and (3.4), the priors of intercepts and slope coefficients are multivariate normal with zero mean and identity matrix for variance covariance. For the prior of \( \sigma^2 \) is that \( \sigma^{-2} \sim \text{Gamma}(5, 1) \). The prior of equation (3.5) is \( \{\mu, \beta \lambda\} \sim \text{MN}(0, I) \), \( \text{MN} \) is multivariate normal distribution and \( I \) is identity matrix. \( \delta \sim \mathcal{N}(0, 1) \), and \( \sigma^{-2} \sim \text{Gamma}(3, 1) \).

In order to test the stability, we carry out the prior sensitivity test on equation (3.5). Table 3.9 represents log-predictive likelihood of equation (3.5) with \( \lambda = 0 \) at various priors. As is evident from Table 3.9, our model is very robust to different priors, which range from very loose prior to very tight prior.
3.6.1 Sampler

The sampling method is straightforward when we are facing equation (3.2), (3.3), and (3.4). They are regular autoregression with the perfect conjugate property. Once we introduce the $\delta$ to conditional variance, we lose partial conjugate property. As a result, we apply the Metropolis-Hasting (MH) algorithm to sample the non-conjugate parts.

$$g_t = \mu + \beta^T \sum_{j=1}^{3} g_{t-j} + \lambda d_l e_t \quad e_t \sim N(0, \sigma^2 \exp(2\delta d_t^a))$$

$$(\mu, \beta, \lambda) \sim MN(a, A), \quad \delta \sim N(b, B), \quad \sigma^{-2} \sim Gamma(\chi, \nu)$$

The $a$ is a vector with dimension of 3. $A$ is a square matrix with dimension of 3. The $b, B, \chi$ and $\nu$ are scalar. Let $x = \{a, A, b, B, \chi, \nu, g_{1:T}\}$. At each $i$th MCMC draw, we apply the Gibbs sampler to sample all the parameters as the following steps:

1. Draw $\sigma^{-2(i)} \sim p(\sigma^{-2}|\mu^{(i-1)}, \beta^{(i-1)}, \lambda^{(i-1)}, \delta^{(i-1)}, x)$
2. Draw $\mu^{(i)} \sim p(\mu|\beta^{(i-1)}, \lambda^{(i-1)}, \delta^{(i-1)}, \sigma^{-2(i)}, x)$
3. Draw $\beta^{(i)} \sim p(\beta|\mu^{(i)}, \lambda^{(i-1)}, \delta^{(i-1)}, \sigma^{-2(i)}, x)$
4. Draw $\lambda^{(i)} \sim p(\lambda|\mu^{(i)}, \beta^{(i)}, \delta^{(i-1)}, \sigma^{-2(i)}, x)$
5. Draw $\delta^{(i)} \sim p(\delta|\mu^{(i)}, \beta^{(i)}, \lambda^{(i)}, \sigma^{-2(i)}, x)$

To simplify the notations, we set $m(\mu, \beta, \delta) = \mu + \beta^T \sum_{j=1}^{3} g_{t-j} + \lambda d_l$. Due to the conjugate property, $\sigma^{-2}$ is sampled from the following,

$$\sigma^{-2}|\beta, \mu, \lambda, \delta, x \sim Gamma \left( \frac{\chi + \frac{T}{2}}{\nu + \frac{T}{2} \sum_{i=1}^{T} (g_t - m(\mu, \beta, \delta))^2} \right)$$

Step 2 to step 5 are sampled under the MH algorithm. The joint density function of $\mu, \beta, \lambda, \delta$ is the following,

$$p(\mu, \beta, \lambda, \delta|\sigma^2, x) \propto \exp \left( -\frac{(\delta - a)^2}{2B} \right) \exp \left( -\frac{\delta}{2} \sum_{i=1}^{T} d_i \right) \exp \left( \frac{1}{2\sigma^2} \sum_{i=1}^{T} \frac{(g_t - m(\mu, \beta, \delta))^2}{\exp(2\delta d_t)} \right)$$

The following example shows posterior sampling details of $\delta^{(i)}$: given $\mu^{(i)}, \beta^{(i)}, \lambda^{(i)}$ and $\sigma^{-2(i)}$, we first draw $\delta^{\text{new}} = \delta^{(i-1)} + N(0, s)$, where $s$ is the tuning parameter for adjusting
the acceptance probability, $\delta^{old} = \delta^{(i-1)}$. Then, we decide on accepting $\delta^{new}$ or keeping $\delta^{old}$ according to the following rule,

$$\theta = \min \left[ \frac{p(\delta^{new}|\mu^{(i)}, \beta^{(i)}, \lambda^{(i)}, \sigma^{-2(i)}, x)}{p(\delta^{old}|\mu^{(i)}, \beta^{(i)}, \lambda^{(i)}, \sigma^{-2(i)}, x)}, 1 \right]$$  \hspace{1cm} (3.7)

Next, we draw $u \sim Uniform(0, 1)$, if $u \leq \theta$, set $\delta^{(i)} = \delta^{new}$, otherwise set $\delta^{(i)} = \delta^{old}$. The rest of the parameters are sampled in the same manner.

### 3.7 Empirical Results

Table 3.5, 3.6, and 3.7 correspond to the log-predictive likelihood computed under various type of oil price shock. The first column is the lag indicators, which implies the time it takes for shock impact to appear. The top row indicates three different versions; the purpose of having different versions is to compare the impact of oil price shock on expected mean and conditional variance. Each log-predictive likelihood is calculated based on an out-of-sample size of 480.

Table 3.5 is the log-predictive likelihood of equation (3.5) for oil price shock is measured under Type-1. This measurement is the most popular approach that is developed by Hamilton (1996) and Hamilton (2003). Table 3.5 suggests two interesting outcomes: First, the oil price shock will affect the future real growth since the entries are significantly higher than the models without the oil price shock. However, it takes approximately 10-12 months for the changes to become feed through. This is why we observe that the log-predictive likelihood at $p = 10$ is significantly higher than $p = 1$ for all three versions. Second, the oil price shock has a stronger predictive power for future real growth in volatility than expected mean. Notably, such significant relationship is only distinguishable when $p \geq 6$.

Table 3.6 represents the log-predictive likelihood of equation (3.5) under type-2 oil price shock. Type-2 measures the oil price shock in the same way as type-1 except the magnitude of the shock is irrelevant. First, the oil price shocks of type-2 can significantly improve the forecast accuracy of real growth. The lagging period is approximately the same as type-1. Second, the shock can aid to predict real growth more accurately through volatility than expected mean. Third, there is no evidence to indicate that the magnitude of positive oil price shock can better forecast the real growth. In other words, the magnitude of positive oil price shock does not significantly contribute to predicting real growth. According to the results between Table 3.5 and Table 3.6, the type-2 forecast accuracy is improved more significantly than type-1 at $p = 2$, $p = 6$ and $p = 16$. Only
superior forecast of type-1 is found at $p = 10$.

Table 3.7 represents the log-predictive likelihood of equation (3.5) under type-3. Type-3 defines the oil price shock in the same way as type-1 except the type-3 studies the negative oil price shock. For example, type-3 implies a large fall of oil price (negative oil price shock), can affect the real economic growth. There are several interesting results in Table 3.7. First, by including the type-3 oil price shock oil on expected mean and volatility, the forecast accuracy is significantly improved, and the lag timing is similar as previous types. Second, the type-1 and type-2 oil price shock better predict the real economic growth than the type-3. In other words, positive oil price shock shows a much more significant impact on describing future real economic growth than negative oil price shock. Third, there is no drastic difference between imposing the type-3 oil price shock on either expected mean or volatility.

Table 3.8 represents the log-predictive likelihood of equation (3.5) under the type-4 shock. Again, we observe considerable improvement on density forecast accuracy by introducing type-4 oil price shock on expected mean and volatility of real growth; the lag effect is weak. Second, the type-4 oil price shock better captures future real growth through the volatility than the expected mean. Third, by comparing to type-1 and type-2 oil price shock, no evidence indicates type-4 shock can more accurately explain future real growth than other types.

The analysis above leads to two important findings. First, by using a Bayesian predictive density forecast instead of a frequentist point forecast to evaluate forecast performance, the out-of-sample density forecast accuracy is noticeably improved by including oil price shock variables. This empirical result is a new contribution to existing literature. A key reason is that existing works use the point forecast to test the model, whereas this paper evaluates the models by using the density forecast. Second, the oil price shock of type-2 best predicts future real growth than any other types. A key implications is that Hamilton (2003) is correct regarding type-1 oil price shock can affect real growth instead of oil price change, but the relationship between real growth and oil price shock relies on having the shock or not rather than the magnitude of the shock. As a result, we do not observe notable gain in the forecast when we compare Table 3.5 to 3.6.

3.8 Conclusion

We first examine the ability of existing models to use oil price changes to forecast U.S. real growth rates. We document that the rate change of oil prices does not have any predictive power on real growth. However, once we replace it with the oil price shock,
we document that oil price shock is able to help forecast real growth. We propose a new model to study the predictive relationship of oil price shock and economic growth. The new model allows the oil price shock to affect the expected mean of real growth, but also shapes its conditional variance. Our empirical results suggest the oil price shock can better predict real growth through conditional variance than expected mean. Our new model and density forecast shed light on a new direction of study on oil price shock and real economy relationship.
Chapter 3.

Figure 3.1: RAC Composite Monthly Rate of Change of Oil Price (Top)
U.S Industrial Production Monthly Growth Rates (Bottom)

Table 3.1: The Log-predictive Likelihood with Total Lags

<table>
<thead>
<tr>
<th></th>
<th>$\beta_{1:p} = 0$</th>
<th>$p = 1$</th>
<th>$p = 2$</th>
<th>$p = 3$</th>
<th>$p = 6$</th>
<th>$p = 12$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{1q} = 0$</td>
<td>-519</td>
<td>-519</td>
<td>-522</td>
<td>-527</td>
<td>-500</td>
<td>-546</td>
</tr>
<tr>
<td>$q = 1$</td>
<td>-511</td>
<td>-511</td>
<td>-515</td>
<td>-519</td>
<td>-533</td>
<td>-535</td>
</tr>
<tr>
<td>$q = 2$</td>
<td>-504</td>
<td>-505</td>
<td>-508</td>
<td>-512</td>
<td>-526</td>
<td>-529</td>
</tr>
<tr>
<td>$q = 3$</td>
<td>-501</td>
<td>-503</td>
<td>-506</td>
<td>-508</td>
<td>-503</td>
<td>-526</td>
</tr>
<tr>
<td>$q = 6$</td>
<td>-509</td>
<td>-511</td>
<td>-514</td>
<td>-517</td>
<td>-528</td>
<td>-533</td>
</tr>
<tr>
<td>$q = 12$</td>
<td><strong>-500</strong></td>
<td>-503</td>
<td>-507</td>
<td>-510</td>
<td>-523</td>
<td>-548</td>
</tr>
</tbody>
</table>

The out-of-sample size is 480, and the full sample size is 499.
The model: $g_t = \mu + \sum_{l=1}^{q} \alpha_l g_{t-l} + \sum_{l=1}^{p} \beta_l r_{t-l} + \sigma \epsilon_t$
Table 3.2: The Log-predictive Likelihood with Individual Lags

<table>
<thead>
<tr>
<th></th>
<th>$\beta = 0$</th>
<th>$p = 1$</th>
<th>$p = 2$</th>
<th>$p = 3$</th>
<th>$p = 6$</th>
<th>$p = 12$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0$</td>
<td>-519</td>
<td>-519</td>
<td>-522</td>
<td>-524</td>
<td>-525</td>
<td>-510</td>
</tr>
<tr>
<td>$q = 1$</td>
<td>-511</td>
<td>-511</td>
<td>-514</td>
<td>-515</td>
<td>-517</td>
<td>-498</td>
</tr>
<tr>
<td>$q = 2$</td>
<td>-507</td>
<td>-508</td>
<td>-510</td>
<td>-511</td>
<td>-512</td>
<td>-496</td>
</tr>
<tr>
<td>$q = 3$</td>
<td>-507</td>
<td>-507</td>
<td>-510</td>
<td>-510</td>
<td>-511</td>
<td><strong>-495</strong></td>
</tr>
<tr>
<td>$q = 6$</td>
<td>-523</td>
<td>-523</td>
<td>-526</td>
<td>-526</td>
<td>-525</td>
<td>-509</td>
</tr>
<tr>
<td>$q = 12$</td>
<td>-507</td>
<td>-507</td>
<td>-510</td>
<td>-510</td>
<td>-512</td>
<td>-513</td>
</tr>
</tbody>
</table>

The out-of-sample size is 480, and the full sample size is 499.

The model: $g_t = \mu + \alpha g_{t-q} + \beta r_{t-p} + \sigma e_t$

Table 3.3: The Log-predictive Likelihood with Average Lags

<table>
<thead>
<tr>
<th></th>
<th>$\beta = 0$</th>
<th>$p = 1$</th>
<th>$p = 2$</th>
<th>$p = 3$</th>
<th>$p = 6$</th>
<th>$p = 12$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0$</td>
<td>-519</td>
<td>-519</td>
<td>-520</td>
<td>-521</td>
<td>-522</td>
<td>-526</td>
</tr>
<tr>
<td>$q = 1$</td>
<td>-511</td>
<td>-511</td>
<td>-512</td>
<td>-513</td>
<td>-513</td>
<td>-514</td>
</tr>
<tr>
<td>$q = 2$</td>
<td>-499</td>
<td>-500</td>
<td>-502</td>
<td>-502</td>
<td>-502</td>
<td>-503</td>
</tr>
<tr>
<td>$q = 3$</td>
<td><strong>-493</strong></td>
<td>-494</td>
<td>-496</td>
<td>-497</td>
<td>-497</td>
<td>-498</td>
</tr>
<tr>
<td>$q = 6$</td>
<td>-496</td>
<td>-497</td>
<td>-498</td>
<td>-500</td>
<td>-501</td>
<td>-502</td>
</tr>
<tr>
<td>$q = 12$</td>
<td>-508</td>
<td>-508</td>
<td>-509</td>
<td>-511</td>
<td>-513</td>
<td>-514</td>
</tr>
</tbody>
</table>

The out-of-sample size is 480, and the full sample size is 499.

The model: $g_t = \mu + \alpha \frac{1}{q} \sum_{l=1}^{q} g_{t-l} + \beta \frac{1}{p} \sum_{l=1}^{p} r_{t-l} + \sigma e_t$

Table 3.4: The Log-predictive Likelihood with Average Lags (Updated)

<table>
<thead>
<tr>
<th></th>
<th>$\beta = 0$</th>
<th>$p = 1$</th>
<th>$p = 2$</th>
<th>$p = 3$</th>
<th>$p = 6$</th>
<th>$p = 12$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0$</td>
<td>-518</td>
<td>-518</td>
<td>-519</td>
<td>-520</td>
<td>-520</td>
<td>-520</td>
</tr>
<tr>
<td>$q = 1$</td>
<td>-492</td>
<td>-493</td>
<td>-494</td>
<td>-495</td>
<td>-494</td>
<td>-494</td>
</tr>
<tr>
<td>$q = 2$</td>
<td>-479</td>
<td>-480</td>
<td>-481</td>
<td>-482</td>
<td>-481</td>
<td>-481</td>
</tr>
<tr>
<td>$q = 3$</td>
<td><strong>-472</strong></td>
<td>-473</td>
<td>-474</td>
<td>-475</td>
<td>-474</td>
<td>-473</td>
</tr>
<tr>
<td>$q = 6$</td>
<td>-484</td>
<td>-485</td>
<td>-486</td>
<td>-487</td>
<td>-487</td>
<td>-485</td>
</tr>
<tr>
<td>$q = 12$</td>
<td>-503</td>
<td>-502</td>
<td>-504</td>
<td>-505</td>
<td>-506</td>
<td>-505</td>
</tr>
</tbody>
</table>

The out-of-sample size is 480, and the full sample size is 499.

The model: $g_t = \mu + \alpha \frac{1}{q} \sum_{l=1}^{q} g_{t-l} + \beta \frac{1}{p} \sum_{l=1}^{p} r_{t-l} + \sigma e_t$
Table 3.5: Type-1

<table>
<thead>
<tr>
<th></th>
<th>$\delta \neq 0$, $\lambda \neq 0$</th>
<th>$\delta \neq 0$, $\lambda = 0$</th>
<th>$\delta = 0$, $\lambda \neq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>-479</td>
<td>-475</td>
<td>-474</td>
</tr>
<tr>
<td>$p = 2$</td>
<td>-477</td>
<td>-474</td>
<td>-475</td>
</tr>
<tr>
<td>$p = 3$</td>
<td>-469</td>
<td>-468</td>
<td>-473</td>
</tr>
<tr>
<td>$p = 6$</td>
<td>-469</td>
<td>-467</td>
<td>-475</td>
</tr>
<tr>
<td>$p = 10$</td>
<td>-456</td>
<td>-457</td>
<td>-467</td>
</tr>
<tr>
<td>$p = 12$</td>
<td>-460</td>
<td>-457</td>
<td>-470</td>
</tr>
<tr>
<td>$p = 16$</td>
<td>-473</td>
<td>-475</td>
<td>-470</td>
</tr>
</tbody>
</table>

Shock Type: $d_{t-p}^n = \max \{0, \ln \frac{OP_{t-p}}{OP_{t-p}^+}\}$. The $OP_{t-p}^+$ is the highest oil price in the past $n$ months starting from $t-p$. The out-of-sample size is 480, and the full sample size is 499. $n = 12$. The Model:

$$g_t = \mu + \beta_1 \sum_{l=1}^3 g_{t-l} + \lambda d_{t-p}^n + e_t \quad e_t \overset{iid}{\sim} N(0, \sigma^2 \exp(2\delta d_{t-p}^n))$$

Table 3.6: Type-2

<table>
<thead>
<tr>
<th></th>
<th>$\delta \neq 0$, $\lambda \neq 0$</th>
<th>$\delta \neq 0$, $\lambda = 0$</th>
<th>$\delta = 0$, $\lambda \neq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>-478</td>
<td>-476</td>
<td>-474</td>
</tr>
<tr>
<td>$p = 2$</td>
<td>-469</td>
<td>-469</td>
<td>-475</td>
</tr>
<tr>
<td>$p = 3$</td>
<td>-467</td>
<td>-467</td>
<td>-475</td>
</tr>
<tr>
<td>$p = 6$</td>
<td>-464</td>
<td>-462</td>
<td>-477</td>
</tr>
<tr>
<td>$p = 10$</td>
<td>-461</td>
<td>-462</td>
<td>-466</td>
</tr>
<tr>
<td>$p = 12$</td>
<td>-456</td>
<td>-453</td>
<td>-469</td>
</tr>
<tr>
<td>$p = 16$</td>
<td>-464</td>
<td>-462</td>
<td>-470</td>
</tr>
</tbody>
</table>

The shock type is $d_{t-p}^n = I\{0, \ln \frac{OP_{t-p}}{OP_{t-p}^-}\}$. The out-of-sample size is 480, and the full sample size is 499. $n = 12$. The model:

$$g_t = \mu + \beta_1 \sum_{l=1}^3 g_{t-l} + \lambda d_{t-p}^n + e_t \quad e_t \overset{iid}{\sim} N(0, \sigma^2 \exp(2\delta d_{t-p}^n))$$

Table 3.7: Type-3

<table>
<thead>
<tr>
<th></th>
<th>$\delta \neq 0$, $\lambda \neq 0$</th>
<th>$\delta \neq 0$, $\lambda = 0$</th>
<th>$\delta = 0$, $\lambda \neq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>-477</td>
<td>-475</td>
<td>-472</td>
</tr>
<tr>
<td>$p = 2$</td>
<td>-477</td>
<td>-477</td>
<td>-477</td>
</tr>
<tr>
<td>$p = 3$</td>
<td>-475</td>
<td>-475</td>
<td>-478</td>
</tr>
<tr>
<td>$p = 6$</td>
<td>-482</td>
<td>-477</td>
<td>-479</td>
</tr>
<tr>
<td>$p = 10$</td>
<td>-477</td>
<td>-473</td>
<td>-472</td>
</tr>
<tr>
<td>$p = 12$</td>
<td>-474</td>
<td>-469</td>
<td>-471</td>
</tr>
<tr>
<td>$p = 16$</td>
<td>-474</td>
<td>-469</td>
<td>-472</td>
</tr>
</tbody>
</table>

The shock type is $d_{t-p}^n = \min \{0, \ln \frac{OP_{t-p}}{OP_{t-p}^-}\}$, where $OP_{t-p}^-$ is the lowest oil price in the past $n$ months from $t-p$. The out-of-sample size is 480, and the full sample size is 499. $n = 12$. The model:

$$g_t = \mu + \beta_1 \sum_{l=1}^3 g_{t-l} + \lambda d_{t-p}^n + e_t \quad e_t \overset{iid}{\sim} N(0, \sigma^2 \exp(2\delta d_{t-p}^n))$$
Chapter 3.

Table 3.8: Type-4

<table>
<thead>
<tr>
<th></th>
<th>$\delta \neq 0, \lambda \neq 0$</th>
<th>$\delta \neq 0, \lambda = 0$</th>
<th>$\delta = 0, \lambda \neq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>-474</td>
<td>-470</td>
<td>-477</td>
</tr>
<tr>
<td>$p = 2$</td>
<td>-472</td>
<td>-468</td>
<td>-475</td>
</tr>
<tr>
<td>$p = 3$</td>
<td>-478</td>
<td>-476</td>
<td>-475</td>
</tr>
<tr>
<td>$p = 6$</td>
<td>-479</td>
<td>-477</td>
<td>-478</td>
</tr>
<tr>
<td>$p = 10$</td>
<td>-471</td>
<td>-469</td>
<td>-470</td>
</tr>
<tr>
<td>$p = 12$</td>
<td>-468</td>
<td><strong>-466</strong></td>
<td>-469</td>
</tr>
<tr>
<td>$p = 16$</td>
<td>-473</td>
<td>-470</td>
<td>-472</td>
</tr>
</tbody>
</table>

The shock type is $d_{t-p} = I(r_{t-p} > \frac{1}{2}\text{std}(r_{t-p:n:t-p}))$. $n = 12$. The $r_{t-p} = \ln \frac{OP_t}{OP_{t-p}}$. The out-of-sample size is 480, and the full sample size is 499. The model:

$$g_t = \mu + \beta \sum_{l=1}^{3} g_{t-l} + \lambda d_{t-p} + \epsilon_t \sim N(0, \sigma^2 \exp(2\delta d_{t-p}))$$

Table 3.9: Priors Sensitivity Analysis

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$\chi$</th>
<th>$\nu$</th>
<th>$B$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Loose</td>
<td>$10 I$, $10$, $2$, $10$</td>
<td>-454</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less Loose</td>
<td>$5 I$, $6$, $1$, $5$</td>
<td>-454</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loose</td>
<td>$2 I$, $5$, $1$, $2$</td>
<td>-454</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less Tight</td>
<td>$I$, $5$, $1$, $1$</td>
<td>-453</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tight</td>
<td>$0.5 I$, $1$, $1$, $0.5$</td>
<td>-453</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a is a vector of zeros. $b$ is a scalar of zero.*


