Journal of Empirical Finance 38 (2016) 202–220

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

An infinite hidden Markov model for short-term interest rates☆

John M. Maheu*, [a], [b], Qiao Yang [c]

[a] DeGroote School of Business, McMaster University,1280 Main Street W., Hamilton, ON L8S4M4, Canada
[b] RCEA, Italy
[c] School of Entrepreneurship and Management, ShanghaiTech University, Shanghai, China

ARTICLE INFO

Article history:
Received 11 March 2015
Received in revised form 9 June 2016
Accepted 13 June 2016
Available online 20 June 2016

JEL classification:
C58
C14
C22
C11

Keywords:
Hierarchical Dirichlet process prior
Beam sampling
Markov switching
MCMC

ABSTRACT

The time-series dynamics of short-term interest rates are important as they are a key input into pricing models of the term structure of interest rates. In this paper we extend 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 a 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.

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 and 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 a hierarchical prior provides significant improvements in density forecasts as well as point forecasts.

☆ We are grateful for helpful comments from an anonymous referee, Yong Song and seminar participants at the University of Toronto, the 2015 RCEA Bayesian Econometric workshop and the 2015 Melbourne Bayesian Econometrics workshop. Maheu thanks the SSHRC for financial support.

* Corresponding author.
E-mail addresses: maheujm@mcmaster.ca (J. Maheu), gy631223@gmail.com (Q. Yang).

http://dx.doi.org/10.1016/j.jempfin.2016.06.006
0927-5398/

Markov switching model has been used extensively to model interest rates. Early applications include Hamilton (1988), Albert and Chib (1993) and Garcia and Perron (1996). Markov switching and GARCH or stochastic volatility are combined by Cai (1994), Gray (1996) and Kalimipalli and 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 and Saikkonen (2003) combine a mixture autoregressive process with time-varying transition probabilities and GARCH.

Ang and 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 behavior 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 and 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 parameters. 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 types 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 and West, 1995) which does not allow for time dependence. A *sticky* version of the infinite hidden Markov model which favors self transitions between states was introduced by Fox et al. (2011) and applied to inflation (Jochmann, 2015, Song, 2014), ARMA models (Carpantier and 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.[1]

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 (2014), 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 favor 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 3 introduces the Dirichlet process, hierarchical Dirichlet process and the infinite hidden Markov model. Posterior sampling is discussed in Section 4 while empirical results are found in Section 5. The Appendix collects the details of posterior simulation.

2. Benchmark models

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

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

dW_t is a Brownian motion and both the drift term $\lambda + \beta y_{t-1}$ and spot variance σy_{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

$$\Delta y_t = \lambda + \beta y_{t-1} + \sigma y_{t-1}^{x} \epsilon_t, \quad \epsilon_t \sim N(0, 1). \tag{2}$$

[1] In the empirical applications we did not find any significant improvements from using the sticky version of the model.