

A Unified HJM Approach to Non-Markov Gaussian Dynamic Term Structure Models: International Evidence ^{*}

HAITAO LI[†]

Cheung Kong Graduate School of Business

XIAOXIA YE[‡]

Stockholm Business School

FAN YU[§]

Claremont McKenna College

Abstract

We show that international government bond yields exhibit a strong non-Markov property, in the sense that moving averages of long-lagged yields significantly improve the predictability of excess bond returns. We then develop a systematic approach of constructing non-Markov Gaussian dynamic term structure models (GDTSMs) under the Heath-Jarrow-Morton (HJM) framework. Compared to the current literature, our approach is more flexible and parsimonious, enabling us to estimate an economically significant non-Markov effect that helps predict excess bond returns both in-sample and out-of-sample.

Keywords: Non-Markov; Gaussian Dynamic Term Structure Models; Excess Returns; International Bond Markets; Moving Averages; Forecasting

JEL classification: C61; E43; E44; G12

^{*}We are grateful to Areski Cousin, Jin-Chuan Duan, Koji Inui, Masaaki Kijima, Shaoyu Li, Linlin Niu, Haoxi Yang, and participants at the Young Researchers Workshop in Finance at Tokyo Metropolitan University and the Workshop in Fixed Income and Bond Markets at Xiamen University for their helpful comments. The computation in this paper was partially performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). All remaining errors are our own.

[†]htli@ckgsb.edu.cn

[‡]xiaoxia.ye@sbs.su.se

[§]fyu@cmc.edu

1 Introduction

The term structure of interest rates is one of the most widely studied topic in economics and finance. Built on the pioneering works of Vasicek (1977) and Cox et al. (1985), a large number of dynamic term structure models (DTSMs) have been developed in the finance literature in the past two decades. According to Dai and Singleton (2003), DTSMs assume that the evolution of the spot rate and the yield curve depends on a finite number of state variables. By judiciously choosing the dynamics of the state variables and their relation with the spot rate, the DTSMs that have been developed to date are empirically flexible and analytically tractable. Among the most prominent classes of DTSMs are the affine term structure models (ATSMs) of Duffie and Kan (1996) and Dai and Singleton (2000), in which the spot rate is a linear function of the state variables that follow affine diffusions. These models have been extensively studied in the literature to address a wide range of term structure related issues.

One of the fundamental assumptions common to all DTSMs is that bond yields follow Markov processes. That is, changes in bond yields and excess bond returns depend on current but not lagged yields. In most DTSMs, the state variables are assumed to follow either continuous-time Markov processes or discrete-time AR(1) processes. Contrary to this important assumption, however, there is increasing evidence that bond yields do not follow Markov processes. For example, based on nonparametric methods, Chen and Hong (2011) show that the seven-day Eurodollar rates strongly violate the Markov assumption. Cochrane and Piazzesi (2005) and Feunou and Fontaine (2014) also find that, in addition to current yields (forward rates), lagged yields explain a significant portion of the variation of bond excess returns, a clear violation of the Markov assumption.

In light of the above evidence, existing DTSMs in the literature have been extended to capture non-Markov bond yields. For example, Ang and Piazzesi (2003) and Jardet et al. (2013) consider macro term structure models, in which the state variables follow a discrete-time Gaussian VAR(p) process under both the physical (\mathbb{P}) and risk-neutral (\mathbb{Q}) measures to incorporate lagged information into bond pricing. Joslin et al. (2013) and Feunou and Fontaine (2014) also consider non-Markov Gaussian DTSMs, in which the state variables follow a VAR(p) or VARMA(1,1) process under the \mathbb{P} measure. However, both models require the state variables to be Markov under the \mathbb{Q} measure. Under this setup, lagged information only

affects the expected change of yields, but has no direct impact on current yields. As a result, lagged information is left unspanned by current yields.

While these non-Markov models provide interesting new insights on term structure dynamics, it is fair to say that non-Markov term structure models are still severely understudied compared to the large number of Markov DTSMs in the literature. In this paper, we develop a systematic approach of constructing continuous-time non-Markov Gaussian DTSMs under the Heath, Jarrow, and Morton (1992, HJM) framework. Compared to extant DTSMs, HJM models are particularly convenient for modelling the non-Markov features in the data for the following reasons.

First, interest rates generally follow infinite dimensional non-Markov processes under the HJM framework, and are Markov only under very restrictive assumptions. This feature offers great flexibility for modelling bond yields because we can start with a large set of non-Markov models, from which we choose a subset of models that fit the data well. In contrast, DTSMs are Markov by construction and can be extended in only limited ways, such as adding lags to the state variables to obtain non-Markov models. If the current yields depend on yields in the distant past, for example, such non-Markov models would need many lags and can become overly complicated to parameterize.¹

Second, under the HJM framework, the volatility function of the forward rates completely determines the dynamics of the yield curve. For certain volatility functions, the term structure has a finite-dimensional representation in the sense that forward rates are linear functions of a finite number of state variables, which follow Gaussian processes. We obtain Finite-Dimensional Representations (FDRs) of Gaussian HJM models based on the linear system approach of Björk and Gombani (1999). One important advantage of this approach is that it leads to infinitely many equivalent FDRs for the same HJM model, similar to the “invariant transforms” of Dai and Singleton (2000).²

Third, the HJM approach offers potentially richer non-Markov dynamics for the term structure than the current non-Markov DTSMs. Based on appropriate specifications of the volatility function, we can obtain non-Markov Gaussian DTSMs by allowing the number of state vari-

¹ Neither $\text{VAR}(p)$ nor $\text{VARMA}(1,1)$ is flexible enough to be consistent with the empirical data. For example, for three-factor models, $\text{VAR}(p)$ becomes intractable when $p > 12$. For monthly data, however, $p = 12$ only accounts for lags up to one year, which does not go far enough to capture the strong non-Markov property exhibited by the data.

² Dai and Singleton (2000) use invariant transforms to classify all affine term structure models into a few subclasses of canonical representations.

ables to be larger than the number of factors, i.e., the dimension of the Wiener process. By imposing additional constraints, we show that our specification reduces to VARMA state variables in discrete time. Moreover, by specifying an appropriate pricing kernel or the market prices of risk as in Joslin et al. (2014), we are able to reproduce the unspanning property of current yields.

We conduct a comprehensive analysis of non-Markov Gaussian term structure models based on government bond yields from nine countries, which we divide into three groups based on geographical locations or the similarity of economic systems: Asia Pacific (Australia, Japan, and New Zealand), Continental Europe (Germany, Sweden, and Switzerland), and North America and UK (Canada, the United Kingdom, and the United States). Our empirical analysis progresses in three steps.

First, we provide compelling evidence that government bond yields from the nine countries exhibit a strong non-Markov property. Similar to Cochrane and Piazzesi (2005), we find that forward rates with up to two-month lags have strong predictive power of excess bond returns. We also find that moving averages of bond yields with up to 60-month lags explain a large portion of the variation of excess bond returns. To examine the economic significance of the out-of-sample forecasting power of the Cochrane and Piazzesi (2005) specification and our specification with long-lagged moving averages, we compute their trading profits according to the risk-adjusted average return of a trading rule that exploits the predictability of bond excess returns. We find that our specification with long-lagged moving averages results in much higher trading profits than the specification of Cochrane and Piazzesi (2005). These results demonstrate that bond yields strongly violate the Markov assumption and extremely long lags are needed to capture the dynamics of bond yields.

Second, we study whether the Markov property is satisfied under the \mathbb{Q} measure as assumed in Joslin et al. (2013) and Feunou and Fontaine (2014). Since we only observe bond yields under the \mathbb{P} measure, we need specific assumptions on the market prices of risk to estimate the term structure dynamics under the \mathbb{Q} measure. To avoid such assumptions, we focus on the term structure of U.S. repo rates, for which Longstaff (2000) and Della Corte et al. (2008) have shown that the expectations hypothesis cannot be rejected, and hence the dynamics of the repo rates under the \mathbb{P} and \mathbb{Q} measures would coincide. We provide strong evidence that the Markov assumption is violated by the repo rates under the \mathbb{Q} measure.

Third, using the bond yields of the nine countries, we compare the in-sample and out-of-sample performance of the non-Markov models constructed using our approach with that of the non-Markov models in the existing literature, specifically Joslin et al. (2013)'s VAR(4) model and Feunou and Fontaine (2014)'s VARMA(1,1) model. Upon inspecting the in-sample fit to the data using various statistical criteria, such as AIC, BIC, and HIC, we find that our non-Markov models can fit the data in-sample better than both the three-factor Markov model and the non-Markov models of VAR(4) and VARMA (1,1). We also examine the out-of-sample performance of all models in predicting the excess bond returns. The results show that in seven of the nine bond markets, the Markov model fails to capture the excess bond returns out-of-sample in an economically significant way.³ In contrast, in all countries except for one, there are at least three non-Markov models producing economically significant trading profits. We also find that even though VAR(4) and VARMA(1,1) have similar in-sample trading profits as the continuous-time Markov and non-Markov models, they fail to match the out-of-sample performance of our non-Markov models. Looking at the results across the nine countries, different markets call for different non-Markov specifications. These observations suggest that the non-Markov property is a staple feature of the bond market. We contribute to the literature by offering a flexible and parsimonious modelling framework that can accommodate this feature.

The structure of the paper is as follows. In Section 2, we show that moving averages of bond yields (over extended windows) explain a large portion of the variation of excess bond returns in all nine bond markets considered. In Section 3, we develop a systematic approach to building non-Markov GDTSMs. Section 4 provides empirical evidence on the non-Markov property under the Q measure. Section 5 carries out an out-of-sample exercise that identifies the non-Markov model that delivers the best performance in capturing the excess bond returns. Section 6 concludes the paper. Appendix A to Appendix D contain proofs and technical details. In Appendix E, we discuss the specification of unspanned risks within our framework.

2 Non-Markov property of international bond yields

In this section, we provide extensive evidence that government bond yields in different countries strongly violate the Markov assumption. Specifically, we show that bond yields with

³ Since we do not consider transaction costs, we require that the associated trading rule to yield a risk-adjusted average return of 30 percent or more to be qualified as being economically significant.

extensive lags can significantly improve the predictability of excess bond returns. This evidence provides strong motivation for developing non-Markov term structure models beyond those in the existing literature.

2.1 Data

The data used in this paper are bond yields from nine industrialized countries with different monetary policies constructed by Wright (2011), and made available through an online appendix.⁴ We categorize the nine countries into three groups based on their geographical locations and economic similarities: Asia Pacific (Australia, Japan, and New Zealand), Continental Europe (Switzerland, Sweden, and Germany), and North America and UK (Canada, the United Kingdom, and the United States).

The data contain zero yields with maturities ranging from one year to ten years with increments of one year, and observed at the monthly frequency. The sample periods start at different dates for different countries and all end in April 2009. In our analysis, we divide the data into two subsamples. For countries with sample periods starting earlier than January 1980 (DE, GB, and US), the in-sample period ends in April 1994. For the rest of the countries (AU, JP, NZ, CH, SE, and CA), the in-sample period ends in April 2000 or April 2004. The details about the sample periods of different countries are summarized in Table 1.

[Insert Table 1 about here]

2.2 Moving averages explain excess returns

In this section, we examine whether bond yields in the nine countries violate the Markov assumption, in the sense that excess bond returns depend not only on current yields but also on historical yields.

Following Cochrane and Piazzesi (2005), we focus on the average log excess holding period

⁴ In Wright (2011)'s dataset, zero-coupon yield curves of Norway are also available. However, due to their relatively short history, we exclude Norway's data from our analysis.

return across maturities between two to five years:

$$\begin{aligned}\bar{r}x_{t+1} &\equiv \frac{1}{4} \sum_{n=2}^5 rx_{t+1}^{(n)}, \\ rx_{t+1}^{(n)} &\equiv r_{t+1}^{(n)} - y_t^{(1)}, \\ r_{t+1}^{(n)} &\equiv ny_t^{(n)} - (n-1)y_{t+1}^{(n-1)},\end{aligned}$$

where $y_t^{(n)}$ is the n -year zero yield at time t .

We consider two forecasting models for bond excess returns that include lagged yield information. The first one is the regression model of Cochrane and Piazzesi (2005), which we refer to as the CP approach:

$$\bar{r}x_{t+1} = \alpha_0 + \alpha_1 f_t + \alpha_2 f_{t-\frac{1}{12}} + \alpha_3 f_{t-\frac{2}{12}} + \varepsilon_{t+1},$$

where $f_t = \begin{bmatrix} y_t^{(1)} & f_t^{(2)} & \dots & f_t^{(10)} \end{bmatrix}^\top$ and $f_t^{(n)} \equiv ny_t^{(n)} - (n-1)y_t^{(n-1)}$, that is, f_t represents one-year forward rates up to ten years in maturity.⁵ The CP approach incorporates both current forward rates and forward rates with one- and two-month lags as predictors.

The second model we consider incorporates both current zero yields and moving averages of lagged zero yields as predictors, which we refer to as the moving average (MA) approach:

$$\bar{r}x_{t+1} = \alpha_0 + \alpha_1 y_t + \alpha_2 \bar{y}_{-t}^l + \varepsilon_{t+1}, \quad (1)$$

where $y_t = \begin{bmatrix} y_t^{(1)} & y_t^{(2)} & \dots & y_t^{(10)} \end{bmatrix}^\top$, $\bar{y}_{-t}^l = \begin{bmatrix} \bar{y}_{-t}^{(1),l} & \bar{y}_{-t}^{(2),l} & \dots & \bar{y}_{-t}^{(10),l} \end{bmatrix}^\top$, $\bar{y}_{-t}^{(n),l} = \frac{1}{l} \sum_{j=1}^l y_{t-\frac{j}{12}}^{(n)}$, and l is the number of lags in months. Therefore, while y_t represents the current yields, \bar{y}_{-t}^l represents moving averages of historical yields up to l months in the past.

Besides the statistical fit of the above forecasting models, we also use the profits generated by a trading strategy based on these models to gauge the economic significance of the lagged predictors. Specifically, we report the “adjusted return” and “cumulative return” of the trading strategy considered in Cochrane and Piazzesi (2006).

For simplicity, we assume that we are able to trade a portfolio that gives a log annual return of $\bar{r}x_t$. We invest in $\mathbb{E}_t(\bar{r}x_{t+1})$ units of this portfolio every month and close the position after

⁵ The original CP paper uses forward rates up to five years in maturity. We extend the maturities to ten years because our dataset includes zero yields up to ten years in maturity.

holding it for 12 months.⁶ The return of this “trading rule” is the product of the excess return and its expected value one year prior:

$$rn_t \equiv \bar{r}x_t \times \mathbb{E}_{t-1}(\bar{r}x_t).$$

In this section, we use $\mathbb{E}_t(\bar{r}x_{t+1}) = \hat{\alpha}_0 + \hat{\alpha}_1 f_t + \hat{\alpha}_2 f_{t-\frac{1}{12}} + \hat{\alpha}_3 f_{t-\frac{2}{12}}$ for the CP approach, and $\mathbb{E}_t(\bar{r}x_{t+1}) = \hat{\alpha}_0 + \hat{\alpha}_1 y_t + \hat{\alpha}_2 y_{-t}^l$ for the MA approach. The “adjusted return” is the average of rn_t divided by its standard deviation (it can be regarded as the Sharpe ratio of the trading strategy):

$$\text{AdjRn} = \frac{\langle rn_t \rangle}{\text{std}(rn_t)}.$$

The “cumulative return” is simply the sum of rn_t over monthly intervals (with $\Delta t = 1/12$):

$$\text{CumRn}_t = \sum_{i=\Delta t}^{t/\Delta t} rn_i.$$

Table 2 reports the in-sample R^2 's and the in- and out-of-sample adjusted returns of the CP and MA models.⁷ The model parameters in all out-of-sample tests are fixed at their in-sample estimates, i.e., as time progresses, we do not update the model parameters in real time.⁸ For the MA approach, we have the number of lags as an extra free parameter. We use lags of 12 to 60 months to run the test, and report results based on the number of lags that gives the best out-of-sample AdjRn among all lags considered.

From Table 2, we see that adding moving averages of historical yields in the regression improves the in-sample R^2 's significantly relative to the CP approach. Moreover, while the two approaches have comparable in-sample adjusted returns, the MA approach seems to have better out-of-sample adjusted returns. The same table also shows that the performance of the CP approach deteriorates sharply when moving from in-sample to out-of-sample, suggesting that the CP approach fails to capture the non-Markov feature inherent in bond yields. In contrast,

⁶ We long the portfolio if $\mathbb{E}_t(\bar{r}x_{t+1}) > 0$ and short the portfolio if $\mathbb{E}_t(\bar{r}x_{t+1}) < 0$.

⁷ Out-of-sample R^2 's are not reported because they are not as informative as the adjusted returns in measuring the out-of-sample performance of a trading rule.

⁸ While this might be inconsistent with industry practice, our goal is not to develop real-time trading strategies. By deliberately keeping the model parameters fixed, we are more interested in features of the data that different model specifications are unable to reproduce in an out-of-sample setting.

the performance of the MA approach does not deteriorate as much as that of the CP approach. Graphically, Figure 1 shows that the MA approach delivers higher in-sample cumulative returns than the CP approach, and Figure 3 shows that the MA approach largely delivers better out-of-sample cumulative returns than the CP approach as well. Similar results were also identified by Feunou and Fontaine (2014).

[Insert Table 2 about here]

[Insert Figure 1 about here]

While the above results for the MA approach are based on the number of lags that gives the best out-of-sample performance, Figure 2 reports the in- and out-of-sample performances of the MA approach for different lags. For many countries, the MA approach performs similarly across different lags. For others, variations of performance across different lags are bigger. The last row in Table 2 reports the number of lags for different countries that gives the MA approach its best out-of-sample performance (as measured by the out-of-sample AdjRn).

[Insert Figure 2 about here]

Although the MA approach does outperform the CP approach both in-sample and out-of-sample for some countries, we notice that there are still others, such as CH, DE, JP, and US, where both approaches fail out-of-sample. The significant improvement of performance brought by the MA approach suggests that the non-Markov property does exist in the bond market. The long lags for the best MA model also suggest that the non-Markov feature is complex and cannot be easily captured by adding a small number of lags. However, both of these specifications likely suffer from over-fitting, as evidenced by the dramatic drop in performance from in-sample to out-of-sample for CH, DE, JP, and US.⁹ Therefore, these regression-based approaches are unable to capture the non-Markov property consistently in all cases. In the next section, we introduce a systematic approach of constructing non-Markov continuous-time GDTSMs, which are able to capture the non-Markov property in all nine bond markets in an out-of-sample context.

[Insert Figure 3 about here]

⁹ This could be due to the large number of explanatory variables in both the MA and the CP regressions.

3 A systematic approach to non-Markov Gaussian DTSMs

The evidence presented in the previous section shows that the Markov assumption is overwhelmingly violated by international bond yields, and that regression-based models have difficulty capturing the non-Markov feature of the data in a consistent manner. In this section, we develop a systematic approach of constructing non-Markov Gaussian DTSMs under the Heath, Jarrow, and Morton (1992, HJM) framework. We first discuss the advantages of the HJM approach to non-Markov term structure modeling. Then, we describe the linear systems approach of Björk and Gombani (1999), which allows us to construct infinitely many Markov and non-Markov Gaussian term structure models from the same HJM model. Specializing to a deterministic specification of the forward rate volatility function, we present an algorithm for constructing both Markov and non-Markov term structure models. Finally, we show that our non-Markov models are more flexible than those in the current literature. These new models will be implemented in the subsequent empirical analysis.

3.1 The HJM framework

We begin by briefly introducing the HJM framework. Let $f(t, T)$ represent the instantaneous forward rate at time t for a future date $T > t$, which represents the rate that can be contracted at time t for instantaneous risk-free borrowing or lending at time T . Given $f(t, T)$ for all maturities between t and T , the price at time t of a zero-coupon bond with maturity T can be obtained as:

$$P(t, T) = \exp \left\{ - \int_t^T f(t, s) ds \right\}.$$

The spot interest rate at time t is simply $r_t = f(t, t)$.

HJM model term structure dynamics through the stochastic evolution of the forward rates:

$$df(t, T) = \mu(t, T) dt + \sigma^f(t, T) dW_t,$$

where $W_{m \times 1}$ is an m -dimensional Wiener process under the Q measure, and $\mu(t, T)$ and $\sigma^f(t, T)_{1 \times m}$ are the drift and volatility of the forward rate, respectively. HJM also establish the following

no-arbitrage restriction on the drift of the forward rate process:

$$\mu(t, T) = \sigma^f(t, T) \left(\int_t^T \sigma^f(t, s) ds \right)^\top.$$

Therefore, the volatility function $\sigma^f(t, T)$ completely determines the drift of the forward rate under the Q measure.

For convenience, we consider the Musiela parameterization, which uses the time to maturity (denoted by x), rather than the time of maturity, to parameterize bonds and forward rates:

Definition 1 For all $x \geq 0$, the forward rate in the Musiela parameterization, $r(t, x)$, is defined as:

$$r(t, x) = f(t, t+x) \text{ and } P(t, T) = \exp \left\{ - \int_0^{T-t} r(t, s) ds \right\}.$$

Following Brace and Musiela (1994), the standard HJM drift condition can be rewritten as:

$$\begin{aligned} dr(t, x) &= \mu_r(t, x) dt + \sigma(t, x) dW_t, \\ \mu_r(t, x) &= \frac{\partial}{\partial x} r(t, x) + \sigma(t, x) \int_0^x \sigma(t, s)^\top ds, \end{aligned}$$

where $W_t = [[w_i]_{i=1}^m]^\top$, $\sigma(t, x) = \sigma^f(t, t+x) = [\sigma_i(t, x)]_{i=1}^m$, and $[\bullet_i]_{i=1}^m$ is a compact notation for a row vector $[\bullet_1, \bullet_2, \dots, \bullet_m]$. We then have:

$$r(t, x) = r(0, t+x) + \Theta(t, x) + r_0(t, x), \quad (2)$$

$$\Theta(t, x) = \int_0^t \sigma(s, x+t-s) \int_0^{x+t-s} \sigma(s, \tau)^\top d\tau ds, \quad (3)$$

$$r_0(t, x) = \int_0^t \sigma(s, x+t-s) dW_s, \quad (4)$$

$$dr_0(t, x) = \frac{\partial r_0(t, x)}{\partial x} dt + \sigma(t, x) dW_t, \quad r_0(0, x) = 0. \quad (5)$$

From (2) to (5), we can see that $\Theta(t, x) + r_0(t, x)$ is the time-varying stochastic component of the forward rate $r(t, x)$. When $\sigma(t, x)$ is time-invariant, i.e., $\sigma(t, x) = \sigma(x)$, which is the main focus of this paper, the non-Markov property is reflected in the drift term of $dr_0(t, x)$, $\frac{\partial r_0(t, x)}{\partial x} = \int_0^t \frac{\partial \sigma(x+t-s)}{\partial x} dW_s$, which requires integrating over the entire history of the underlying Wiener process.

3.2 Finite-Dimensional Representation of HJM models

Interest rates under HJM models are generally non-Markov and infinite dimensional (see Björk and Gombani, 1999), which makes their empirical implementation difficult. Consequently, a large literature has been developed to identify conditions that allow Finite-Dimensional Representations (FDRs) of HJM models. One widely recognized sufficient condition for HJM models to exhibit FDRs is a time-invariant volatility that is a deterministic function of time to maturity, which satisfies a multi-dimensional linear ODE with constant coefficients (e.g., Björk and Svensson, 2001, Corollary 5.1 or Chiarella and Kwon, 2003, Assumption 1).¹⁰ Since the time and maturity dependent components of the volatility function are separable, we can obtain Markov state variables by integrating over the historical Brownian shocks.

Following the approach of obtaining FDRs of HJM models developed by Björk and Gombani (1999) based on linear systems theory, we start from the following definition:

Definition 2 A triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$, where \mathbf{A} is an $n \times n$ matrix, \mathbf{B} is an $n \times m$ full rank matrix, and $\mathbf{C}(x)$ is an n -dimensional row-vector function, is called an n -dimensional realization of the system $r_0(t, x)$ if $r_0(t, x)$ has the following representation:

$$r_0(t, x) = \mathbf{C}(x) Z_t, \quad (6)$$

$$dZ_t = \mathbf{A}Z_t dt + \mathbf{B}dW_t, \quad Z_0 = \mathbf{0}, \quad (7)$$

where Z_t is an n -dimensional column vector of state variables.

Björk and Gombani (1999) show that for an HJM model to have an FDR as in (6)-(7), the volatility function must be written as:

$$\sigma(x) = \mathbf{C}(x)\mathbf{B} = \mathbf{C}_0 \exp(\mathbf{A}x)\mathbf{B},$$

where \mathbf{A} , \mathbf{B} , and $\mathbf{C}(x)$ are given in Definition 2, and $\mathbf{C}_0 = \mathbf{C}(0)$. Also, one can apply invariant transforms to the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ to construct a new realization $\{M\mathbf{A}M^{-1}, M\mathbf{B}, \mathbf{C}(x)M^{-1}\}$

¹⁰ Specifically, the i th component of σ , $\sigma_i(x)$ (which is n times differentiable with respect to x), satisfies an n th ODE of the form

$$\frac{\partial^n}{\partial x^n} \sigma_i(x) - \sum_{j=0}^{n-1} \kappa_{ij}(x) \frac{\partial^j}{\partial x^j} \sigma_i(x) = 0$$

where $\kappa_{ij}(x)$'s are continuous deterministic functions.

given a nonsingular $n \times n$ matrix M . Then:

$$r_0(t, x) = \mathbf{C}(x) M^{-1} (MZ_t),$$

$$d(MZ_t) = M \mathbf{A} M^{-1} (MZ_t) dt + M \mathbf{B} dW_t,$$

is another FDR for the same HJM model with a new state vector MZ .

Traditional GDTSMs are time-homogeneous. Although the FDR we obtain for $r_0(t, x)$ is time-homogeneous, the first two components of the forward curve in (2), $r(0, t+x)$ and $\Theta(t, x)$, are not. To preserve the time-homogeneity feature, we construct GDTSMs from HJM models by replacing

$$r(0, t+x) + \Theta(t, x)$$

with

$$\lim_{t \rightarrow \infty} r(0, t+x) + \Theta(t, x),$$

essentially assuming that the model has evolved from the distant past. De Jong and Santa-Clara (1999) and Trolle and Schwartz (2009) adopt a similar treatment. Details of this derivation are presented in Appendix A and Appendix B.

3.3 Non-Markov and Markov states

An important feature of this approach is that the number of states is allowed to be larger than the number of factors, i.e., the dimension of the Wiener process. In this subsection, using an invariant transform, we show that any model having more states than factors exhibits a *non-Markov property*. We emphasize that by this we mean that some of the states are non-Markov on their own, i.e., their conditional forecast depends on their current as well as lagged values. However, when we examine all states as a whole, they form a Markov system because the requisite lagged information is included as additional state variables.

Let us consider a model with \mathbf{B} being $n \times m$ and lower trapezoidal, where $n > m$.¹¹ Parti-

¹¹ A full rank \mathbf{B} with $n > m$ can always be transformed into a lower trapezoidal matrix via an invariant transformation.

tion \mathbf{B} as follows:

$$\mathbf{B} = \begin{bmatrix} \underbrace{B_1}_{m \times m} \\ \underbrace{B_2}_{(n-m) \times m} \end{bmatrix}_{n \times m}.$$

Since \mathbf{B} is a full rank and lower trapezoidal matrix, B_1 is a nonsingular and lower triangular matrix by construction. Define a transform matrix M as:

$$M = \begin{bmatrix} \underbrace{\alpha_1 \mathbb{1}}_{m \times m} & \underbrace{\alpha_2 \mathbb{1}}_{m \times (n-m)} \\ \underbrace{-B_2 B_1^{-1}}_{(n-m) \times m} & \underbrace{\mathbb{1}}_{(n-m) \times (n-m)} \end{bmatrix}_{n \times n},$$

where α_1 and α_2 are free scalar parameters, and $\mathbb{1}$ is an identity matrix if it is square, and otherwise an identity matrix with a proper number of appended zero rows at the bottom or columns on the right. Then, the last $n - m$ rows of

$$\mathbf{B}_{\text{new}} \equiv M\mathbf{B} = \begin{bmatrix} \underbrace{\alpha_1 B_1 + \alpha_2 B_2 (1:m, 1:m)}_{m \times m} \\ \underbrace{\mathbf{0}}_{(n-m) \times m} \end{bmatrix}$$

are zeros rows, where $B_2 (1:m, 1:m)$ is the first m rows of B_2 when $n - m \geq m$ or B_2 appended below with zero rows when $n - m < m$.

Therefore, the last $n - m$ state variables in $Z_{\text{new},t} \equiv MZ_t$ are not directly subject to contemporaneous Gaussian shocks, and are instead exponentially-weighted averages of the first m state variables. In other words, these $n - m$ state variables by design summarize the history of the first m states. Therefore, the first m states are non-Markov by themselves, in the sense that their drift can depend on both current and lagged values of the first m states. We formalize this observation using the following proposition, a proof of which is provided in [Appendix C](#):

Proposition 1 *For any FDR with $n > m$, there exists at least one m -dimensional subsystem of the FDR that is non-Markov.*

It is worth noting that when $n = m$, all transformations of the m -dimensional FDR are

Markov. Therefore, Proposition 1 clarifies the difference between an FDR with $n = m$ and one with $n > m$. In what follows, we will refer to FDRs with $n = m$ as Markov models, and those with $n > m$ as non-Markov models.

3.4 Volatility specification and base realization

One important advantage of using the HJM framework is that the volatility function completely determines model specification under the \mathbb{Q} measure. In this subsection, we introduce a volatility function that guarantees FDRs of HJM models. The volatility function we consider is a special case of the most general deterministic function that allows FDRs according to Björk and Svensson (2001) and consists mainly of polynomials and exponentials.¹²

Specifically, for an m -factor Gaussian HJM model, the volatility function is

$$\sigma(x) \equiv \underbrace{\left[\begin{bmatrix} 1 & x & \dots & x^{n_i-1} \end{bmatrix} e^{-k_i x} \right]_{i=1}^I}_{1 \times \sum_{i=1}^I n_i} \times \mathbf{\Omega}_{\left(\sum_{i=1}^I n_i\right) \times m'} \quad (8)$$

where n_i is a natural number, $\begin{bmatrix} 1 & x & \dots & x^{n_i-1} \end{bmatrix}$ is a $1 \times n_i$ row vector, k_i is a positive real number with $k_i < k_j$ for $i < j$, and $\left[\begin{bmatrix} 1 & x & \dots & x^{n_i-1} \end{bmatrix} e^{-k_i x} \right]_{i=1}^I$ is a $1 \times \sum_{i=1}^I n_i$ row vector.¹³ The matrix $\mathbf{\Omega}$ satisfies the following restrictions:

1. $\mathbf{\Omega}$ is an $\sum_{i=1}^I n_i \times m$ full rank matrix with $\sum_{i=1}^I n_i \geq m$, i.e., $\text{Rank}(\mathbf{\Omega}) = m$.¹⁴
2. In $\mathbf{\Omega}$, any $\sum_{i=1}^j n_i$ th row is not a zero row, for $j = 1, 2, \dots, I$.¹⁵
3. $\mathbf{\Omega}(1, j) + \sum_{k=1}^{I-1} \mathbf{\Omega}\left(1 + \sum_{i=1}^k n_i, j\right) \geq 0$, for $j = 1, 2, \dots, m$.¹⁶

¹² Björk and Svensson (2001) show that the most general deterministic volatility function that allows FDRs of HJM models is the so-called “quasi-exponential” (or QE) function that has the following general form:

$$\sigma_{\text{QE}}(x) = \sum_i e^{\lambda_i x} + \sum_j e^{\alpha_j x} [p_j(x) \cos(\omega_j x) + q_j(x) \sin(\omega_j x)],$$

where λ_i , α_j , and ω_j are real numbers, and p_j and q_j are polynomials. Moreover, $\sigma_{\text{QE}}(x)$ can be written as $\mathbf{C}_0 \exp(\mathbf{A}x) \mathbf{B}$.

¹³ Though the k_i 's could take complex values, Joslin et al. (2011) show that a model with two complex conjugate eigenvalues is empirically equivalent to a model with two real eigenvalues equal to the real and imaginary part of the complex eigenvalues. Therefore, without loss of generality, we restrict the k_i 's to be real in our specification.

¹⁴ If $m > \sum_{i=1}^I n_i$, some of the parameters in $\mathbf{\Omega}$ are unidentifiable, as the m -factor model degenerates to a $\sum_{i=1}^I n_i$ -factor model.

¹⁵ If it were, then n_i needs to be reduced until this is no longer the case.

¹⁶ This restriction ensures that the volatility of the derived short rate is non-negative.

4. Ω is set to a lower trapezoidal matrix (a generalization of the lower triangular form for a non-square matrix) for the purpose of identification.¹⁷

Given the volatility function (8), a base realization is presented in the following theorem, a proof of which is provided in Appendix C.

Theorem 1 For the HJM volatility function defined in (8), one realization triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is:

$$\mathbf{C}(x) = \left[\left[\begin{array}{cccc} 1 & x & \dots & x^{n_i-1} \end{array} \right] e^{-k_i x} \right]_{i=1}^I, \mathbf{B} = \Omega,$$

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{A}_1 & & & \\ & \mathbf{A}_2 & & \\ & & \dots & \\ & & & \mathbf{A}_I \end{array} \right]_{n \times n}, \mathbf{A}_i = \left[\begin{array}{cccc} -k_i & 1 & & \\ & -k_i & 2 & \\ & & -k_i & \dots \\ & & & \dots & n_i - 1 \\ & & & & -k_i \end{array} \right]_{n_i \times n_i},$$

where $n = \sum_{i=1}^I n_i$, and \mathbf{A} is in block diagonal form with each block given by \mathbf{A}_i , whose non-zero elements are indicated above.

Using this base realization, we present a concrete example of a non-Markov model using an invariant transform. Specifically, we consider a one-factor HJM model with its volatility function given by:

$$\sigma(x) = [\Omega_1 + \Omega_2 x] e^{-kx}. \quad (9)$$

The base realization of this model is:

$$dZ_t = \begin{bmatrix} -k & 1 \\ 0 & -k \end{bmatrix} Z_t dt + \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix} dW_t.$$

Suppose that $0 < k < 1/4$, and we set:

$$M = \begin{bmatrix} \frac{(\sqrt{1-4k}-1)\Omega_2}{2\Omega_1} & \frac{-(\sqrt{1-4k}-1)\Omega_1-2\Omega_2}{2\Omega_1} \\ -\frac{\Omega_2}{\Omega_1} & 1 \end{bmatrix},$$

¹⁷ In our empirical estimation, Ω appears only through $\Omega\Omega^\top$. Therefore, only the lower part of Ω is identifiable.

then the new state variables $Z_{\text{new},t} = MZ_t$ have the following dynamics:

$$dZ_{\text{new},t} = \begin{bmatrix} \nu & -\nu \\ 1 & -\vartheta \end{bmatrix} Z_{\text{new},t} dt + \begin{bmatrix} -\frac{\Omega_2^2}{\Omega_1} \\ 0 \end{bmatrix} dW_t.$$

where $\nu = \frac{1}{2} - k - \frac{1}{2}\sqrt{1-4k}$ and $\vartheta = k + \frac{1}{2} - \frac{1}{2}\sqrt{1-4k}$. It can be easily shown that the second state variable in $Z_{\text{new},t}$ is an exponentially-weighted average of the history of the first state variable, i.e.,

$$Z_{\text{new},t}(2) = \int_0^t e^{-\vartheta(t-s)} Z_{\text{new},s}(1) ds,$$

and the first state variable “extrapolates” from the second:¹⁸

$$dZ_{\text{new},t}(1) = \nu [Z_{\text{new},t}(1) - Z_{\text{new},t}(2)] dt - \frac{\Omega_2^2}{\Omega_1} dW_t.$$

Apparently, $Z_{\text{new},t}(1)$ by itself is non-Markov since its drift contains its own history. However, $Z_{\text{new},t}$ as a whole is a Markov system.

3.5 Concise model notations

For convenience, we establish some concise notations to distinguish the different models within our framework. The previous subsection shows that the specification of the model is completely determined by two entities: the vector $[[1, x, \dots, x^{n_i-1}] e^{-k_i x}]_{i=1}^I$ and the matrix $\mathbf{\Omega}$, where the former determines the number of state variables and the latter the number of factors. We use $\mathbf{N} = [n_1, n_2, \dots, n_I]$, a $1 \times I$ row vector with $n_i - 1, i = 1, 2, \dots, I$, being the highest order in the polynomial associated with k_i , to designate the number of state variables ($\sum_{i=1}^I n_i$) and the base transfer matrix \mathbf{A} . We use m to designate the number of factors. For example, a model with $\mathbf{N} = [2, 2, 1]$ and $m = 3$ has 5 state variables and 3 factors, and its base

¹⁸ As documented in Barberis et al. (2015), investors often use extrapolation to form beliefs about future state variables. Though this simple example features extrapolation, it is just as easy to construct an alternative specification in which the first state variable mean-reverts to the second.

realization is:

$$\mathbf{C}(x) = \begin{bmatrix} e^{-k_1 x} & x e^{-k_1 x} & e^{-k_2 x} & x e^{-k_2 x} & e^{-k_3 x} \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} -k_1 & 1 & 0 & 0 & 0 \\ 0 & -k_1 & 0 & 0 & 0 \\ 0 & 0 & -k_2 & 1 & 0 \\ 0 & 0 & 0 & -k_2 & 0 \\ 0 & 0 & 0 & 0 & -k_3 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 & \Omega_6 & 0 \\ \Omega_3 & \Omega_7 & \Omega_{10} \\ \Omega_4 & \Omega_8 & \Omega_{11} \\ \Omega_5 & \Omega_9 & \Omega_{12} \end{bmatrix}.$$

3.6 Comparing with existing non-Markov models

The recent term structure literature has shown an increasing interest in non-Markov models. Two studies in this aspect are Joslin et al. (2013) and Feunou and Fontaine (2014). Following Joslin et al. (2013)'s notations, these two models can be summarized as follows:

$$r_t = \rho_0 + \rho_1 Z_t,$$

$$Z_t = K_{0Z}^Q + K_{1Z}^Q Z_{t-\Delta t} + \sqrt{\Sigma_Z} \epsilon_t^Q, \epsilon_t^Q \sim N(0, \mathbf{1}),$$

where r_t is the short-term interest rate, Δt denotes the time interval, and Z_t is a vector of observable states. For example, Z_t can consist of the principal components of bond yields and macro variables.¹⁹

The difference between these models lies in their specification of the dynamics of Z_t under the \mathbb{P} measure. In Joslin et al. (2013), Z_t is a VAR(p) process under the \mathbb{P} measure:

$$Z_t = K_{0Z}^{\mathbb{P}} + K_{1Z}^{\mathbb{P}} \mathcal{Z}_{t-\Delta t}^p + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{P}}, \epsilon_t^{\mathbb{P}} \sim N(0, \mathbf{1}),$$

where $\mathcal{Z}_{t-\Delta t}^p \equiv (Z_{t-\Delta t}, \dots, Z_{t-p\Delta t})$. In Feunou and Fontaine (2014), Z_t is a VARMA(1,1) process under the \mathbb{P} measure:

$$(Z_t - v) = K_{1Z}^{\mathbb{P}} (Z_{t-\Delta t} - v) + \sqrt{\Sigma_Z} \epsilon_t^{\mathbb{P}} + M_1 \sqrt{\Sigma_Z} \epsilon_{t-\Delta t}^{\mathbb{P}}, \epsilon_t^{\mathbb{P}} \sim N(0, \mathbf{1}),$$

¹⁹ Joslin et al. (2013) show that this model is observationally equivalent to a canonical representation of affine models, in which r_t is a linear function of some latent state variables.

where v is the unconditional \mathbb{P} -mean of Z_t .

Apparently, under the \mathbb{Q} measure, both Joslin et al. (2013) and Feunou and Fontaine (2014) restrict the state variables to be Markov processes, i.e., VAR(1). Our framework, however, does not impose this restriction at all. In fact, our approach is more general than the existing non-Markov approach in two aspects: a) we allow the dynamics of the state variables under both \mathbb{Q} and \mathbb{P} to be non-Markov; b) we can incorporate unspanned risks if we restrict the pricing state variables (bond market specific states) to be a proper subset of the entire system of state variables. This unifies the unspanned risk specifications of Joslin et al. (2014), Feunou and Fontaine (2014), and Joslin et al. (2013), and is illustrated in Appendix E.

As shown in Section 2.2 and Feunou and Fontaine (2014), the MA component (over long lags) is more crucial than the AR component (over short lags) in explaining the risk premia. Therefore, the VARMA specification seems to be better than the VAR specification under the discrete-time setting. In the rest of this subsection, we show that our volatility specification can generate VARMA representations.

It is shown in Bergstrom (1983) that the discrete-time representation of a continuous-time vector autoregressive process of order p (CVAR(p)) is precisely a VARMA($p, p - 1$) process. Under our framework, we can easily specify the volatility function to generate a CVAR realization. This is illustrated in the following example, in which we consider a two-dimensional system with $p = 2$.

Following earlier notations, we specify a model with $\mathbf{N} = [1, 1, 1, 1]$, $m = 2$, and the base triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ given by:

$$\mathbf{C}(x) = \begin{bmatrix} e^{-k_1 x} & e^{-k_2 x} & e^{-k_3 x} & e^{-k_4 x} \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} -k_1 & 0 & 0 & 0 \\ 0 & -k_2 & 0 & 0 \\ 0 & 0 & -k_3 & 0 \\ 0 & 0 & 0 & -k_4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \Omega_1 & 0 \\ -\frac{\Omega_1 k_2}{k_1} & 0 \\ \Omega_2 & \Omega_3 \\ -\frac{\Omega_2 k_4}{k_3} & -\frac{\Omega_3 k_4}{k_3} \end{bmatrix}.$$

Considering an invariant transform with:

$$M = \begin{bmatrix} -\frac{1}{k_1} & -\frac{1}{k_2} & 0 & 0 \\ 0 & 0 & -\frac{1}{k_3} & -\frac{1}{k_4} \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

we have a CVAR(2) realization as:

$$\mathbf{C}_{\text{CVAR}}(x) = \begin{bmatrix} \frac{k_1 k_2 e^{-xk_1} - k_1 k_2 e^{-xk_2}}{k_1 - k_2} & \frac{k_3 k_4 e^{-xk_3} - k_3 k_4 e^{-xk_4}}{k_3 - k_4} & \frac{k_1 e^{-xk_1} - k_2 e^{-xk_2}}{k_1 - k_2} & \frac{k_3 e^{-xk_3} - k_4 e^{-xk_4}}{k_3 - k_4} \end{bmatrix},$$

$$\mathbf{A}_{\text{CVAR}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -k_1 k_2 & 0 & -k_1 - k_2 & 0 \\ 0 & -k_3 k_4 & 0 & -k_3 - k_4 \end{bmatrix}, \mathbf{B}_{\text{CVAR}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{\Omega_1(k_1 - k_2)}{k_1} & 0 \\ \frac{\Omega_2(k_3 - k_4)}{k_3} & \frac{\Omega_3(k_3 - k_4)}{k_3} \end{bmatrix},$$

and denoting the resulting state variables by Z_{CVAR} .

If we normalize the time interval Δt to 1, by Bergstrom (1983, Theorem 2), the first two states in Z_{CVAR} has a VARMA(2,1) representation:

$$Z_{\text{CVAR},t}^{1:2} = F_1 Z_{\text{CVAR},t-1}^{1:2} + F_2 Z_{\text{CVAR},t-2}^{1:2} + \epsilon_t + G\epsilon_{t-1},$$

$$\mathbb{E}(\epsilon_t) = 0, \mathbb{E}(\epsilon_t \epsilon_t^\top) = K, \mathbb{E}(\epsilon_s \epsilon_t^\top) = 0 \quad (s \neq t),$$

where $Z_{\text{CVAR},t}^{1:2}$ denotes the first two states in Z_{CVAR} ,

$$F_1 = \begin{bmatrix} e^{-k_1} + e^{-k_2} & 0 \\ 0 & e^{-k_3} + e^{-k_4} \end{bmatrix}, F_2 = \begin{bmatrix} -e^{-(k_1+k_2)} & 0 \\ 0 & -e^{-(k_3+k_4)} \end{bmatrix},$$

and K and G satisfy the equations:

$$K + GKG^\top = \Gamma_0, \quad GK = \Gamma_1,$$

where

$$\begin{aligned}\Gamma_0 &= \int_0^1 P(s) \mathbf{B}_{\text{CVAR}} \mathbf{B}_{\text{CVAR}}^\top P^\top(s) + Q(s) \mathbf{B}_{\text{CVAR}} \mathbf{B}_{\text{CVAR}}^\top Q^\top(s) ds, \\ \Gamma_1 &= \int_0^1 P(s) \mathbf{B}_{\text{CVAR}} \mathbf{B}_{\text{CVAR}}^\top Q^\top(s) ds, \\ P(h) &= \begin{bmatrix} \frac{e^{-(k_1+k_2)(h+1)}(e^{k_1+hk_2}-e^{k_2+hk_1})}{k_1-k_2} & 0 \\ 0 & \frac{e^{-(k_3+k_4)(h+1)}(e^{k_3+hk_4}-e^{k_4+hk_3})}{k_3-k_4} \end{bmatrix}, \\ Q(h) &= \begin{bmatrix} \frac{e^{-hk_2}-e^{-hk_1}}{k_1-k_2} & 0 \\ 0 & \frac{e^{-hk_4}-e^{-hk_3}}{k_3-k_4} \end{bmatrix}.\end{aligned}$$

Under this specification, K and G cannot be solved in closed-form. However, this should not be a concern, since the purpose of this illustration is merely to show that our general framework is capable of generating VARMA representations. Once we have estimated the structural parameters of the triplet, we can solve K and G numerically.

It is worth noting that the volatility function for this VARMA(2,1) representation is constrained. While the lower trapezoidal matrix \mathbf{B} has seven free parameters, we constrain four of these parameters to be functions of other parameters or zero so as to achieve the VARMA representation. If we do not impose these constraints, we will have more flexibility in capturing the non-Markov property under the \mathbb{Q} measure, although the discrete-time state variables will no longer follow a VARMA process. Moreover, the VARMA model presented here should not be directly compared with that of Feunou and Fontaine (2014), as their model is unrestricted under the \mathbb{P} measure and of order (1,1), while the one presented here is restricted under the \mathbb{Q} measure and can be readily extended to higher orders.²⁰ However, as mentioned in Wymer (1993), a large efficiency gain can be expected of the estimates of the restricted VARMA representation of a continuous-time system relative to the unrestricted VARMA estimates when it comes to prediction.

²⁰ Under the \mathbb{P} measure, our model will have more degrees of freedom thanks to the market price of risk parameters, although the VARMA representation is usually lost.

4 Empirical analysis I: Non-Markov property under the Q measure

In this section, we formally compare the Markov and non-Markov models. Following Joslin et al. (2014), we consider the Akaike (1973) (AIC), Hannan and Quinn (1979) (HQIC), and Schwarz et al. (1978) Bayesian information criteria (SBIC).

In order to evaluate these information criteria, we must compute the likelihood for each model under the Q measure given the full-information ML estimates of its parameters. The fact that we only observe the yields under the P measure, however, imposes a great challenge for this exercise, as the (inferred) state variables simply do not evolve according to their Q-dynamics unless their P- and Q-dynamics coincide. Fortunately, at the very short end of the yield curve, the Q-dynamics of the short rate becomes exceptionally close to its counterpart under the P measure. To support this claim, we refer to the empirical evidence in Longstaff (2000) and Della Corte et al. (2008), showing that the expectations hypothesis (EH) cannot be rejected using US repo rates. If the pure EH holds, as Longstaff (2000) finds to be the case, our task will be made much simpler.²¹

Therefore, we use US repo rates to estimate both Markov and non-Markov models, assuming that the state variables have the same dynamics under both P and Q. Our data are obtained from Bloomberg, similar to Longstaff (2000) and Della Corte et al. (2008), and consist of daily observations of the closing overnight, 1-week, 2-week, 3-week, 1-month, 2-month, and 3-month general collateral government repo rates from May 21, 1991 to January 17, 2014. The total number of observations is 39,935.

Since we focus only on the very short end of the yield curve, we consider only one-factor models with $m = 1$. The Markov model is $\mathbf{N} = [1]$, and the non-Markov models are $\mathbf{N} = [1, 1]$, $[2]$, $[1, 1, 1]$, $[1, 2]$, $[2, 1]$, and $[3]$. The models are estimated using the Kalman filter in conjunction with ML. Following standard practice (Duffee, 1999), we assume that the repo rates are observed with noise, and that these measurement errors are independent over time and normally distributed with zero means and a diagonal covariance matrix with distinct diagonal elements. The resulting maximized values of the log likelihood function are used to evaluate

²¹ Under the continuous-time framework, this is equivalent to assuming that the Local EH (Cox et al., 1981) holds. Although Cox et al. (1981) show that many traditional forms of the EH are incompatible with each other in a theoretical sense, Campbell (1986) demonstrates that the differences among them are typically of little empirical significance. This has been confirmed for US repo rates by Longstaff (2000).

the three different information criteria: AIC, HQIC, and SBIC.

We report the estimates of the model parameters (excluding the covariance matrix of the pricing errors) in Table 3 and our assessment of the models in Table 4. The latter shows that all three criteria unanimously prefer the non-Markov models to the Markov model. Among the non-Markov models, those with three state variables generally perform better than those with only two. While this may be attributed to a greater number of parameters that come with an additional state variable, it is not the sole determinant of model performance because the functional form of the volatility $\sigma(x)$ clearly makes a difference as well. For example, a variety of different k_i 's are preferred to a single value ($[1, 1, 1]$ vs. $[3]$), and having the term xe^{-kx} in the volatility function associated with the larger k_2 is preferred to one with the smaller k_1 ($[1, 2]$ vs. $[2, 1]$). Interestingly, having an extra k_i at the cost of giving up the xe^{-kx} term in the case of $\mathbf{N} = [1, 1, 1]$ does not improve upon the performance of $\mathbf{N} = [1, 2]$. In fact, Table 3 shows that all parameters of $[1, 2]$ are significantly different from zero, while some of the parameters of $[1, 1, 1]$ have large standard errors. In any case, these results suggest that adding just a few more deterministically varying states to the Markov model, with a carefully chosen volatility specification, can increase the Q likelihood without overfitting the data, confirming that non-Markov state variables are required even under the Q measure.

[Insert Tables 3 and 4 about here]

5 Empirical analysis II: An out-of-sample exercise

5.1 Non-Markov models with a Markov origin

We have shown that any model with \mathbf{B} being $n \times m$, where $n > m$ (regardless of how \mathbf{N} is specified), exhibits a non-Markov property. However, in order to take advantage of Joslin et al. (2011) (JSZ)'s estimation method for Markov GDTSMs, in the empirical analysis we focus on models with not only \mathbf{B} being $n \times m$, but also \mathbf{N} being $1 \times m$. Non-Markov models with this type of specifications have a Markov model as their origin. In other words, they all reduce to a Markov model when certain parameters are set to zero.

To see this point more clearly, let us consider again the model with $\mathbf{N} = [2, 2, 1]$ and $m = 3$.

If we set $\Omega_2, \Omega_4, \Omega_6, \Omega_8, \Omega_{10}$, and Ω_{11} in \mathbf{B} to zero, i.e.,

$$\mathbf{B} = \begin{bmatrix} \Omega_1 & 0 & 0 \\ 0 & 0 & 0 \\ \Omega_3 & \Omega_7 & 0 \\ 0 & 0 & 0 \\ \Omega_5 & \Omega_9 & \Omega_{12} \end{bmatrix},$$

the model reduces to $\mathbf{N} = [1, 1, 1], m = 3$, because the volatility functions of $\mathbf{N} = [2, 2, 1], m = 3$ and $\mathbf{N} = [1, 1, 1], m = 3$ are exactly the same:

$$\begin{aligned} & \mathbf{C}_0 \exp(\mathbf{A}x) \mathbf{B} \\ &= \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \end{bmatrix} \exp \left(\begin{bmatrix} -k_1 & 1 & 0 & 0 & 0 \\ 0 & -k_1 & 0 & 0 & 0 \\ 0 & 0 & -k_2 & 1 & 0 \\ 0 & 0 & 0 & -k_2 & 0 \\ 0 & 0 & 0 & 0 & -k_3 \end{bmatrix} x \right) \begin{bmatrix} \Omega_1 & 0 & 0 \\ 0 & 0 & 0 \\ \Omega_3 & \Omega_7 & 0 \\ 0 & 0 & 0 \\ \Omega_5 & \Omega_9 & \Omega_{12} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \exp \left(\begin{bmatrix} -k_1 & 0 & 0 \\ 0 & -k_2 & 0 \\ 0 & 0 & -k_3 \end{bmatrix} x \right) \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_3 & \Omega_7 & 0 \\ \Omega_5 & \Omega_9 & \Omega_{12} \end{bmatrix}. \end{aligned}$$

The reason why the model $\mathbf{N} = [2, 2, 1], m = 3$ has a Markov origin is that its matrix \mathbf{A} is 5×5 but has only three distinct eigenvalues. Thus, by setting some of the Ω_i 's to zero, we manage to degenerate \mathbf{A} to 3×3 , which will be for a Markov model with the three original eigenvalues. In contrast, the model $\mathbf{N} = [1, 1, 1, 1], m = 3$ does not have a Markov model as its origin because its matrix \mathbf{A} has four distinct eigenvalues and therefore cannot degenerate to a 3×3 matrix.

5.2 Estimation

In this empirical analysis, we consider non-Markov models with four to six state variables, which all have the same three-factor Markov model as their origin. Specifically, using the notations introduced in Section 3.5, $m = 3$ for all these non-Markov models, and their \mathbf{N}

vectors are laid out in the third column of Table 5. In total, we will be testing 19 different non-Markov models for each country.

To estimate the model parameters, we need to specify the dynamics of the state variables under the \mathbb{P} measure. Adopting an essentially affine specification of the market price of risk (Duffee, 2002):

$$dW_t^{\mathbb{P}} = (\lambda_1 + \lambda_2 Z_t) dt + dW_t,$$

the \mathbb{P} -dynamics of Z_t is:

$$dZ_t = \mathbf{A}^{\mathbb{P}} (-\mu + Z_t) dt + \mathbf{B} dW_t^{\mathbb{P}},$$

where

$$\mathbf{A}^{\mathbb{P}} = \mathbf{A} - \mathbf{B}\lambda_2, \mu = \left(\mathbf{A}^{\mathbb{P}}\right)^{-1} \mathbf{B}\lambda_1.$$

In order to reduce the chances of being stuck in some local optima far away from the global optimum, the parameters of these models are estimated in a two-step procedure. First, the majority of the parameters from the underlying three-factor Markov model are estimated using JSZ's method. Second, using these estimates and zeros for the rest of the parameters as initial values, all parameters are estimated using the Kalman filter in conjunction with ML. These two steps are elaborated in the following two subsections. The ten maturities of the term structure (one to ten years) from the in-sample periods (see Table 1) are used to estimate the parameters.

5.2.1 Applying JSZ's method to continuous-time Markov GDTSMs

The purpose of applying JSZ's method is not to estimate the \mathbb{P} and \mathbb{Q} parameters separately, but to use the results as initial values to speed up the convergence of the full estimation.

Following JSZ (Case P in their paper), we assume that the first three principal components (PCs) of the yields are observed perfectly, while the yields themselves are observed with error. These measurement errors are assumed to be normally distributed with zero mean and

variance σ^2 :

$$\underbrace{\mathbf{e}_t}_{10 \times 1} \sim N \left(\underbrace{\mathbf{0}}_{10 \times 1}, \underbrace{\sigma^2 \mathbf{1}}_{10 \times 10} \right),$$

and independent across time. Then, the conditional log likelihood function (under \mathbb{P}) of the observed yields has two parts:

$$\log ll_{t|t-\Delta t} = \log ll_t^{\text{errors}} + \log ll_{t|t-\Delta t}^{\text{states}}.$$

The first part of the likelihood function represents the log likelihood of measurement errors,

$$\log ll_t^{\text{errors}} \propto -\frac{1}{2} \log \left(\det \left(\sigma^2 \mathbf{1} \right) \right) - \frac{\text{error}_t^\top \text{error}_t}{2\sigma^2},$$

where

$$\begin{aligned} \text{error}_t &= y_t - \{C_m + B_m Z_t\}, \\ Z_t &= (wm \cdot B_m)^{-1} (\text{PC}_t - wm \cdot C_m), \\ C_m &= \underbrace{\left[\varphi + \frac{\int_0^1 \Theta^*(s) ds}{1}, \dots, \varphi + \frac{\int_0^{10} \Theta^*(s) ds}{10} \right]^\top}_{10 \times 1}, \\ B_m &= \underbrace{\left[\left(\frac{\int_0^1 \mathbf{C}(s) ds}{1} \right)^\top, \dots, \left(\frac{\int_0^{10} \mathbf{C}(s) ds}{10} \right)^\top \right]^\top}_{10 \times 3}. \end{aligned}$$

Here, $\underbrace{y_t}_{10 \times 1}$ is a vector of zero yields of ten different maturities, and $\underbrace{wm}_{3 \times 10}$ is the weighting matrix for the principal components, i.e., $\underbrace{\text{PC}_t}_{3 \times 1} = wm \cdot y_t$. The above expression makes use of the relation between the model-implied zero yield and the state variables backed out from the principal components when forward rates are assumed to be time-homogeneous.²²

The second part of the likelihood function represents the log transition density of the state variables,

$$\log ll_{t|t-\Delta t}^{\text{states}} \propto -\frac{1}{2} \log (\det (cv)) - \frac{(\alpha + \beta Z_{t-\Delta t} - Z_t)^\top cv^{-1} (\alpha + \beta Z_{t-\Delta t} - Z_t)}{2},$$

²² See (A5) in Appendix A. The parameter φ and function $\Theta^*(\cdot)$ can also be found therein.

where $cv = \int_0^{\Delta t} \exp(\mathbf{A}^{\mathbb{P}}(\Delta t - s)) \mathbf{B}\mathbf{B}^{\top} \exp(\mathbf{A}^{\mathbb{P}}(\Delta t - s))^{\top} ds$ is the conditional variance of Z_t given $Z_{t-\Delta t}$, and $\mathbf{A}^{\mathbb{P}}$, α , and β are explicit functions of the data and the Q parameters, which are determined as follows: First, given the Q parameters, the state variables are backed out from the principal components. Then, α and β are obtained as the OLS estimates of the following regression:

$$Z_t = \alpha + \beta Z_{t-\Delta t} + \epsilon_t.$$

Finally, we calculate $\mathbf{A}^{\mathbb{P}}$ as $\frac{\log(\beta)}{\Delta t}$, where \log denotes the matrix logarithm operator.²³

Therefore, $\sum_{t=2\Delta t}^{\mathbf{T}\Delta t} \log ll_{t|t-\Delta t}$ is a function of the data and the Q parameters only. This simplification reduces the number of parameters in the optimization and speeds up the convergence of the estimation. Given the ML estimates of the Q parameters, $\mathbf{A}^{\mathbb{P}}$ and μ (hence λ_2 and λ_1) are explicitly determined from the values of α and β . By the results of JSZ, such estimates are also ML estimates.

5.2.2 Estimating the continuous-time non-Markov GDTSMs

It is possible to use the Kalman filter in conjunction with ML to estimate all parameters without any priors. However, recent literature (e.g., Bauer et al., 2012) has found that the log likelihood function of GDTSMs can be badly behaved (very flat in the parameter space), exhibiting many local maxima. Therefore, it is time-consuming to search for the global optimum from uninformative initial values when the number of parameters is large.

To alleviate the difficulty in estimating the non-Markov models, we initialize the parameters of a non-Markov model that have counterparts in its Markov origin (we call them *Markov parameters*) to their estimates using the JSZ method, and the other parameters (we call them *non-Markov parameters*) to zero. Then, all parameters are estimated using the Kalman filter in conjunction with ML. This separation between the Markov and non-Markov parameters is feasible because a) all non-Markov models we consider here have a Markov origin, and b) we use the essentially affine market price of risk specification.

We use the model $\mathbf{N} = [1, 1, 2]$, $m = 3$ as an example to illustrate this separation. To

²³ In all of our estimations, the ML parameter estimates result in β having a unique $\log(\beta)$, i.e., β is nonsingular with no negative eigenvalues, and every eigenvalue of $\log(\beta)$ has an imaginary part lying strictly between $-\pi$ and π . See, e.g., Higham (2008, Theorem 1.31).

conserve space, we denote the models $\mathbf{N} = [1, 1, 2], m = 3$ and $\mathbf{N} = [1, 1, 1], m = 3$ by “nMrkv” and “Mrkv”, respectively. From Mrkv to nMrkv, there is no change in the k_i 's, although $\mathbf{A}^{\text{nMrkv}}$ is now 4×4 :

$$\mathbf{A}^{\text{nMrkv}} = \begin{bmatrix} -k_1 & 0 & 0 & 0 \\ 0 & -k_2 & 0 & 0 \\ 0 & 0 & -k_3 & 1 \\ 0 & 0 & 0 & -k_3 \end{bmatrix}.$$

In addition, there are three more non-Markov parameters in $\mathbf{B}^{\text{nMrkv}}$:

$$\underbrace{\mathbf{B}^{\text{nMrkv}}}_{4 \times 3} = \begin{bmatrix} \underbrace{\mathbf{B}^{\text{Mrkv}}}_{3 \times 3} \\ \underbrace{\mathbf{B}_4^{\text{nMrkv}}}_{1 \times 3} \end{bmatrix}.$$

Finally, there are three more parameters in λ_2^{nMrkv} , but there is no change in λ_1^{nMrkv} :

$$\underbrace{\lambda_2^{\text{nMrkv}}}_{3 \times 4} = \begin{bmatrix} \underbrace{\lambda_2^{\text{Mrkv}}}_{3 \times 3} & \underbrace{\lambda_{2,4}^{\text{nMrkv}}}_{3 \times 1} \end{bmatrix},$$

$$\lambda_1^{\text{nMrkv}} = \lambda_1^{\text{Mrkv}}.$$

Therefore,

$$\mathbf{A}^{\text{P,nMrkv}} = \mathbf{A}^{\text{nMrkv}} - \mathbf{B}^{\text{nMrkv}} \lambda_2^{\text{nMrkv}},$$

$$\mu^{\text{nMrkv}} = \left(\mathbf{A}^{\text{P,nMrkv}} \right)^{-1} \mathbf{B}^{\text{nMrkv}} \lambda_1^{\text{Mrkv}}.$$

5.3 Finding the best non-Markov model

Following the procedures outlined in the previous subsections, we estimate a three-factor Markov model along with the nineteen non-Markov models. We also estimate Joslin et al. (2013)'s VAR(4) model and Feunou and Fontaine (2014)'s VARMA(1,1) model, the details of which are presented in Appendix D. All models are estimated using only data from the in-sample periods. The AIC values of the different models are reported in Table 5. Among the three benchmark models (the three-factor Markov model, the VAR(4), and the VARMA(1,1)),

the VAR(4) consistently has the lowest AIC values. However, we find that the AIC values of most of the non-Markov models to be even lower, especially those with five or six state variables. Generally speaking, non-Markov models with a larger number of states have lower AIC values. This indicates that the data exhibit a strong non-Markov property.²⁴

[Insert Table 5 about here]

Given the parameter estimates from the in-sample periods, for the 20 continuous-time models (the Markov and non-Markov models specified under our framework), we filter the state variables from the observed zero yields in the full sample. For the two discrete-time models (VAR(4) and VARMA(1,1)), the state variables are the three principal components constructed using the weighting matrix estimated from the in-sample periods. Therefore, the first subsample of the state variables are in-sample values, while the second subsample are out-of-sample ones. Given the parameters and the state variables, the expected excess return $\mathbb{E}_t(\bar{r}\bar{x}_{t+1})$ is computed following its definition in Section 2.2:

$$\begin{aligned}\mathbb{E}_t(\bar{r}\bar{x}_{t+1}) &= \left(\frac{1}{4} \sum_{j=2}^5 j y_t^{(j)} - y_t^{(1)} \right) \\ &\quad - \frac{1}{4} \sum_{j=1}^4 \left(j \varphi + \int_0^j \Theta^*(s) ds + \int_0^j \mathbf{C}(s) ds \left(\mathbb{1} - \exp(\mathbf{A}^{\mathbb{P}}) \right) \mu \right) \\ &\quad - \frac{1}{4} \sum_{j=1}^4 \int_0^j \mathbf{C}(s) ds \exp(\mathbf{A}^{\mathbb{P}}) Z_t,\end{aligned}$$

where we have used the relation between zero yields and state variables in (A5) and the conditional mean of Z_{t+1} given Z_t . Using the model-implied expected excess return $\mathbb{E}_t(\bar{r}\bar{x}_{t+1})$, we can calculate trading returns of the form $rn_{t+1} \equiv \bar{r}\bar{x}_{t+1} \times \mathbb{E}_t(\bar{r}\bar{x}_{t+1})$, just as we did in Section 2.2 for the CP and MA approaches. We can then evaluate the 22 models considered here based on their in- and out-of-sample adjusted returns ($\text{AdjRn} = \langle rn_t \rangle / \text{std}(rn_t)$) and cumulative returns ($\text{CumRn}_t = \sum_{i=\Delta t}^{t/\Delta t} rn_i$), as well as their in-sample R^2 's from projecting the realized excess return onto its model-implied expectation.

We present the in-sample R^2 's and the adjusted returns in Tables 6 and 7, respectively. Intuitively, a high R^2 means that the model can explain the excess bond returns well, and this should be a necessary condition for obtaining a mostly positive trading return rn_t and a high

²⁴ Results using other information criteria, such as the SBIC and HQIC, are similar. These additional results are available upon request.

value for its risk-adjusted mean, which we denote as AdjRn. Indeed, this is confirmed upon a careful inspection of both tables. For example, Table 6 shows that for Germany (DE), the model [1,3,2] has the highest in-sample R^2 among all models considered ($R^2 = 0.410$). In Table 7, [1,3,2] also has the highest adjusted returns (AdjRn= 0.622) among all models for DE. The same inspection also reveals that there is no single model that outperforms for all countries. For example, in Table 6, VAR(4) has the highest in-sample R^2 for three countries (NZ, CH, and SE), while non-Markov models do so for the other six; in Table 7, the three-factor Markov model has the highest in-sample AdjRn for three countries (NZ, CH, and US), while VARMA(1,1) does so for one (AU) and non-Markov models round out the other five.

[Insert Tables 6 and 7 about here]

Figure 4 plots the cumulative returns (CumRns) for the in-sample periods. We include the CP and MA models studied earlier, as well as the three benchmark models (three-factor Markov, VAR(4), and VARMA(1,1)), and the best non-Markov specification according to its out-of-sample AdjRn (see Table 8). The graphs show that the simple regression-based CP and MA approaches actually perform better in-sample than the GDTSMs, in the sense that they produce larger cumulative trading returns over time. Putting aside these two regression-based models, we see no single GDTSM specification that consistently outperforms among the 22 continuous-time and discrete-time GDTSMs. The best performing model seems to be spread evenly across the countries among VAR(4), VARMA(1,1), the three-factor Markov model, and the best three-factor non-Markov specification. Therefore, while the information criteria clearly show an advantage in using non-Markov models to describe bond yields, this advantage is less obvious based on the trading profits from a strategy that exploits the predictability of bond excess returns.

[Insert Figure 4 about here]

In model selection, however, one must be cautious about relying too much on in-sample fitting. Ultimately, whether a model successfully captures the true data generating process can only be tested with out-of-sample model predictions. It is in this context that problems with overfitting are easily revealed. Therefore, we also examine the adjusted returns and cumulative returns of the trading strategy in an out-of-sample context, keeping model parameters fixed

at their in-sample estimates while continuing to filter the state variables and predict the bond excess returns. Since our out-of-sample periods are five, nine, or 15 years long, depending on the country, we subject the models to a rather severe form of out-of-sample testing. This is intentional, however, since our objective is to see which model is capable of capturing the true term structure dynamics in the long-run.²⁵

We present the out-of-sample AdjRns in Table 8, in which we highlight, for each country, one of the 22 GDTSMs that yields the highest AdjRn. In two of the countries this is the three-factor Markov specification, while one of our non-Markov specifications is the best for the other seven countries. Comparing the in-sample AdjRns from Table 7 with the out-of-sample AdjRns from Table 8, we find that the performance of the three benchmark models declines significantly from in-sample to out-of-sample. Since we do not consider transaction costs when computing the trading returns, we impose a somewhat arbitrary risk-adjusted return of 30 percent or more for it to be qualified as economically significant (other cutoffs yield similar insights). Consequently, while VAR(4), VARMA(1,1), and the three-factor Markov model all yield economically significant in-sample AdjRns for all nine countries, VAR(4) no longer does so for any country out-of-sample, VARMA(1,1) does so for only one (US), and the three-factor Markov model, two (NZ and SE).

[Insert Table 8 about here]

In contrast, the situation with non-Markov models is more encouraging. Take the model $[4, 1, 1]$ as an example. It generates economically significant in-sample AdjRns for six countries (AU, JP, DE, SE, CA, and US). Out of these six cases, five (AU, DE, SE, CA, and US) continue to have economically significant out-of-sample AdjRns. While the performance of the other non-Markov specifications is slightly weaker, even a randomly chosen non-Markov specification among the 19 seems to perform better out-of-sample than the three benchmark models. Another way to see the superior out-of-sample performance of the non-Markov models is to count the number of cases in which they generate economically significant CumRns. This is equal to 3 (AU), 7 (JP), 0 (NZ), 14 (CH), 14 (DE), 5 (SE), 9 (CA), 8 (GB), and 13 (US). Hence, with

²⁵ The common practice for out-of-sample testing is to do re-calibration using a rolling window of historical observations. However, the re-calibration of model parameters tends to mask the differences among the models. Our specifications are also more time-consuming to estimate compared to, say, the Joslin et al. (2013) model, and it would be infeasible for us to conduct rolling window estimation across the 19 non-Markov models and nine countries.

the exception of NZ, we have a large number of non-Markov specifications to choose from that can potentially capture the underlying term structure dynamics.

The advantage of our non-Markov setup is confirmed by the out-of-sample CumRn plots in Figure 5, which look distinctly different from their in-sample counterparts in Figure 4. For the majority of the countries, the non-Markov specification selected to yield the best out-of-sample AdjRn also yields the highest CumRn. We mention two additional observations based on this figure. First, when the non-Markov model is not the best in terms of its CumRn (AU, NZ, and CA), the best model turns out to be the MA model, which shows that the non-Markov feature still plays an important role. Second, when the MA model fails to perform well out-of-sample due to potential overfitting (Table 2 shows that this occurs for JP, CH, DE, and US), a non-Markov model is the best-performing one. In summary, although our non-Markov models perform similarly in-sample as the three benchmark models, their superior out-of-sample performance indicates that the non-Markov feature captures a crucial aspect of term structure dynamics.

[Insert Figure 5 about here]

5.4 More states than factors: A potential feature embedded in fundamentals

As Table 8 shows, a three-factor non-Markov model with six states delivers the best out-of-sample performance in four countries (CH, DE, CA, and US), a model with five states does so in one country (GB), and a model with four states does so in two countries (AU and JP). This suggests that our flexible and parsimonious framework for specifying GDTSMs is indeed helpful for capturing the non-Markov property across many different bond markets.

Under a general equilibrium setting, the dynamics of interest rates is explicitly determined by the dynamics of the fundamental economic variables, such as investment and production (e.g., Cox et al., 1985; Longstaff and Schwartz, 1992). Indeed, Jin and Glasserman (2001) show that every HJM model can arise as the equilibrium term structure in a Cox-Ingersoll-Ross production economy. Therefore, the empirical fact that deterministically varying states in the dynamics of interest rates can significantly forecast excess bond returns indicates that a similar non-Markov structure might also be present in the economic fundamentals. One way in which

this can occur has been suggested by Cox et al. (1981) (footnote 34): “One possible justification (for why past interest rates are plausible state variables in a rational expectations equilibrium) arises when investment is not readily reversible so that past interest rates are still reflected in the current production function. Changes in the interest rate will then be affected by past rates as these investments disappear or are abandoned.” Merton (1973) also mentions that the Markov property of “the stochastic processes describing the opportunity set and its changes” is rather general in the sense that the stochastic processes describing returns can be non-Markov, but by including supplementary variables, the entire (expanded) set is once again Markov. The extra states in our models echo the very idea of “supplementary variables.” However, to the best of our knowledge, there are only a few studies, mostly theoretical, in this direction.²⁶

6 Conclusion

In this paper, we confirm the presence of a non-Markov property among international government bond yields, i.e., the moving averages of long-lagged yields significantly improve the forecasts of one-year excess bond returns. Motivated by this evidence, we develop a systematic approach to building non-Markov GDTSMs. This approach not only inherits the canonical features emphasized in the recent literature, e.g., Joslin et al. (2011), but also offers great flexibility in specifying non-Markov dynamics for the state variables under both \mathbb{Q} and \mathbb{P} . A non-Markov specification for the \mathbb{Q} -dynamics is called for by the data, but is not typically implemented by the modeling approaches we currently have. Exploiting the flexibility of our approach, we conduct a specification analysis to examine the ability of our non-Markov GDTSMs to forecast bond excess returns out-of-sample. We find that in a majority of the bond markets (seven out of nine, including the U.S.), the traditional three-factor Markov model cannot produce economically significant trading profits. In contrast, in most of the markets (eight out of nine), there are at least three non-Markov specifications producing economically significant trading profits, with different specifications producing the best results in different bond markets. In five of the nine markets, at least two or three deterministically varying state variables are needed to capture the non-Markov property in a model with three independent random sources (factors). This suggests that the non-Markov property can be strong and non-trivial to model.

²⁶ One such example is Dumas et al. (2009). In their model, four state variables are driven by two independent random sources.

Collectively, the empirical evidence presented in our paper suggests that the exploration of non-Markov properties within a general equilibrium framework can be a fruitful avenue for future research.

Table 1: Data summary

This table summarizes the start and end dates of the data. The dates are of the “yyyymm” format. The lengths of the data in terms of the number of months are also reported for the in-sample and the full sample periods.

	Asia Pacific			Continental Europe			NA and UK Group		
	AU	JP	NZ	CH	DE	SE	CA	GB	US
Start	198702	198501	199001	198801	197301	199212	198601	197901	197111
In-sample End	200004	200004	200404	200004	199404	200404	200004	199404	199404
Full-sample End	200904	200904	200904	200904	200904	200904	200904	200904	200904
In-sample mths	159	184	172	148	256	137	172	184	270
Full-sample mths	268	293	233	257	437	198	281	365	451

Table 2: CP vs MA (best out-of-sample performance)

This table compares the in-sample and out-of-sample performance of the CP and MA approaches. The MA results are based on the number of lags that gives rise to the best out-of-sample *AdjRns* between 12 to 60 months. This number is reported on the last row. “In-sample” and “out-of-sample” are indicated by “(in)” and “(out)”, respectively.

		Asia Pacific			Continental Europe			NA and UK Group		
		AU	JP	NZ	CH	DE	SE	CA	GB	US
CP	R^2 (in)	0.58	0.57	0.53	0.87	0.45	0.75	0.75	0.75	0.56
	AdjRn (in)	0.81	0.65	0.50	0.85	0.71	1.10	0.94	0.47	0.61
	AdjRn (out)	-0.02	-0.07	0.43	-0.39	-0.31	0.30	-0.47	-0.22	-0.20
MA	R^2 (in)	0.86	0.91	0.66	0.96	0.66	0.90	0.87	0.86	0.71
	AdjRn (in)	0.78	0.73	0.58	0.62	0.84	0.76	0.74	0.52	0.60
	AdjRn (out)	0.57	-0.17	1.05	0.08	-0.38	0.46	0.93	0.58	-0.19
	lag (mths)	29	25	12	60	12	49	51	35	12

Table 3: Estimation results of one-factor Markov and non-Markov Models

This table reports the parameter estimates (excluding the seven diagonal elements in the covariance matrix of the pricing errors) of the one-factor Markov and non-Markov models. These models are estimated using Kalman filter in conjunction with MLE. The standard errors are in parentheses.

	states	N	k_1	k_2	k_3	Ω_1	Ω_2	Ω_3	ϕ	Z_t dynamic
Markov	1	[1]	0.013 (0.010)	-	-	0.009 (0.001)	-	-	0.000 (0.230)	$dZ_t = -k_1 Z_t dt + \Omega_1 dW_t$
	2	[1 1]	0.061 (0.010)	2.114 (0.091)	-	0.017 (0.002)	-0.010 (0.001)	-	0.112 (0.013)	$dZ_t = \begin{bmatrix} -k_1 & 0 \\ 0 & -k_2 \end{bmatrix} Z_t dt + \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix} dW_t$
	2	[2]	0.580 (0.018)	-	-	0.008 (0.001)	0.020 (0.002)	-	0.062 (0.002)	$dZ_t = \begin{bmatrix} -k_1 & 1 \\ 0 & -k_1 \end{bmatrix} Z_t dt + \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix} dW_t$
non-Markov	3	[1 1 1]	0.038 (0.014)	6.572 (1.084)	6.658 (1.434)	0.015 (0.002)	-0.828 (6.383)	0.823 (6.382)	0.133 (0.018)	$dZ_t = \begin{bmatrix} -k_1 & 0 & 0 \\ 0 & -k_2 & 0 \\ 0 & 0 & -k_3 \end{bmatrix} Z_t dt + \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{bmatrix} dW_t$
	3	[1 2]	0.038 (0.013)	6.614 (1.094)	-	0.015 (0.002)	-0.006 (0.001)	-0.071 (0.019)	0.133 (0.018)	$dZ_t = \begin{bmatrix} -k_1 & 0 & 0 \\ 0 & -k_2 & 1 \\ 0 & 0 & -k_2 \end{bmatrix} Z_t dt + \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{bmatrix} dW_t$
	3	[2 1]	0.061 (0.010)	2.114 (0.091)	-	0.017 (0.002)	0.000 (0.000)	-0.010 (0.001)	0.112 (0.013)	$dZ_t = \begin{bmatrix} -k_1 & 1 & 0 \\ 0 & -k_1 & 0 \\ 0 & 0 & -k_2 \end{bmatrix} Z_t dt + \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{bmatrix} dW_t$
	3	[3]	0.352 (0.144)	-	-	0.008 (0.001)	0.016 (0.006)	-0.004 (0.005)	0.000 (0.242)	$dZ_t = \begin{bmatrix} -k_1 & 1 & 0 \\ 0 & -k_1 & 2 \\ 0 & 0 & -k_1 \end{bmatrix} Z_t dt + \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{bmatrix} dW_t$

Table 4: Information criteria for one-factor Markov and non-Markov Models

This table compares the one-factor Markov and non-Markov models in terms of their AIC, HQIC, and SBIC. The table also reports the number of states and parameters, and the log likelihood values for different models. The number of parameters includes the seven diagonal elements in the covariance matrix of the pricing errors. The model specification is indicated by **N** in the third column.

	states	N	# of parameters	log likelihood	AIC	HQIC	SBIC
Markov	1	[1]	10	229206.7	-45.839	-45.837	-45.831
	2	[1 1]	12	232028.4	-46.403	-46.400	-46.393
	2	[2]	11	231592.3	-46.316	-46.313	-46.307
non-Markov	3	[1 1 1]	14	232668.3	-46.531	-46.527	-46.519
	3	[1 2]	13	232668.3	-46.531	-46.528	-46.520
	3	[2 1]	13	232028.4	-46.403	-46.400	-46.392
	3	[3]	12	231709.5	-46.340	-46.336	-46.329

Table 5: In-sample AIC

This table reports the in-sample AIC of the Markov and non-Markov models. The number of states and factors is indicated in the first two columns. The model specification is indicated in the third column.

			Asia Pacific			Continental Europe			NA and UK Group		
			AU	JP	NZ	CH	DE	SE	CA	GB	US
Markov	3-factor	[1 1 1]	-2.636	-3.101	-2.918	-2.481	-3.914	-2.303	-2.642	-2.729	-4.380
		[1 1 2]	-2.639	-3.021	-2.969	-2.524	-4.176	-2.314	-2.459	-2.582	-4.585
	4-state	[1 2 1]	-2.777	-3.100	-2.988	-2.565	-4.055	-2.243	-2.652	-2.621	-4.665
[2 1 1]		-2.843	-3.191	-2.838	-2.526	-4.118	-2.263	-2.588	-2.724	-4.698	
non-Markov (3-factor)	5-state	[1 1 3]	-2.835	-3.399	-3.098	-2.492	-3.929	-2.254	-2.575	-2.743	-4.736
		[1 2 2]	-2.831	-3.496	-3.070	-2.593	-4.313	-2.301	-3.012	-2.891	-5.093
		[1 3 1]	-3.010	-3.430	-3.009	-2.827	-4.466	-2.145	-3.124	-3.011	-4.880
		[2 1 2]	-3.016	-3.296	-3.021	-2.593	-4.613	-2.400	-2.795	-2.963	-5.137
		[2 2 1]	-2.997	-3.064	-3.089	-2.834	-4.724	-2.342	-2.857	-2.882	-5.176
	6-state	[3 1 1]	-2.979	-3.454	-3.130	-2.678	-4.522	-2.521	-2.887	-2.950	-5.029
		[1 1 4]	-3.290	-3.675	-3.152	-2.821	-4.215	-2.537	-2.741	-3.267	-4.864
		[1 2 3]	-3.050	-3.735	-2.856	-2.603	-4.435	-2.131	-3.149	-3.031	-5.338
		[1 3 2]	-3.137	-3.743	-3.030	-2.911	-4.787	-2.060	-3.355	-3.193	-5.345
		[1 4 1]	-3.380	-3.765	-3.183	-3.123	-4.529	-2.176	-3.268	-3.391	-5.605
Discrete-time non-Markov	3-factor	[2 1 3]	-3.060	-3.477	-3.174	-2.808	-4.364	-2.327	-3.299	-2.983	-5.166
		[2 2 2]	-3.162	-3.598	-3.152	-3.138	-4.770	-2.598	-3.150	-3.175	-5.439
	[2 3 1]	-3.185	-3.578	-3.341	-2.750	-4.900	-2.509	-3.165	-3.418	-5.686	
	[3 1 2]	-3.171	-3.712	-3.407	-3.016	-4.606	-2.527	-2.964	-3.141	-5.358	
	[3 2 1]	-3.140	-3.450	-3.311	-3.054	-5.071	-2.568	-3.113	-3.252	-5.396	
		[4 1 1]	-3.396	-3.758	-4.106	-3.295	-4.545	-2.357	-3.208	-3.322	-5.739
Discrete-time non-Markov	3-factor	VAR(4)	-2.750	-3.228	-3.087	-2.579	-4.005	-2.409	-2.693	-2.826	-4.615
		VARMA(1,1)	-2.259	-2.663	-2.453	-2.154	-3.419	-1.971	-2.385	-2.473	-3.822

Table 6: In-sample R-squared's

This table reports the in-sample R^2 's of the Markov and non-Markov models for the nine bond markets. The number of states and factors is indicated in the first two columns. The model specification is indicated in the third column. The numbers reported represent the R^2 's of projecting the realized excess return onto the model-implied expected excess return.

			Asia Pacific			Continental Europe			NA and UK Group		
			AU	JP	NZ	CH	DE	SE	CA	GB	US
Markov	3-factor	[1 1 1]	0.087	0.188	0.297	0.594	0.175	0.116	0.203	0.179	0.304
	4-state	[1 1 2]	0.011	0.032	0.090	0.383	0.099	0.066	0.151	0.108	0.151
		[1 2 1]	0.041	0.147	0.002	0.193	0.095	0.097	0.106	0.062	0.119
		[2 1 1]	0.007	0.225	0.223	0.095	0.239	0.274	0.133	0.116	0.208
	5-state	[1 1 3]	0.009	0.111	0.058	0.389	0.106	0.206	0.127	0.417	0.197
		[1 2 2]	0.057	0.253	0.001	0.171	0.097	0.035	0.130	0.005	0.226
		[1 3 1]	0.039	0.089	0.177	0.107	0.009	0.127	0.137	0.121	0.246
		[2 1 2]	0.026	0.069	0.211	0.019	0.095	0.028	0.191	0.240	0.073
		[2 2 1]	0.038	0.282	0.024	0.017	0.046	0.050	0.066	0.031	0.216
		[3 1 1]	0.009	0.062	0.276	0.040	0.008	0.000	0.092	0.044	0.092
non-Markov (3-factor)	6-state	[1 1 4]	0.092	0.012	0.408	0.287	0.143	0.012	0.127	0.004	0.227
		[1 2 3]	0.028	0.162	0.155	0.175	0.159	0.051	0.114	0.040	0.121
		[1 3 2]	0.029	0.120	0.239	0.144	0.410	0.121	0.103	0.061	0.091
		[1 4 1]	0.004	0.084	0.176	0.225	0.000	0.021	0.032	0.001	0.241
		[2 1 3]	0.010	0.147	0.341	0.086	0.167	0.001	0.035	0.430	0.092
		[2 2 2]	0.108	0.167	0.009	0.001	0.068	0.003	0.060	0.027	0.204
		[2 3 1]	0.011	0.142	0.013	0.013	0.026	0.005	0.079	0.065	0.325
		[3 1 2]	0.000	0.046	0.043	0.299	0.325	0.194	0.222	0.067	0.232
		[3 2 1]	0.178	0.021	0.352	0.023	0.003	0.043	0.050	0.003	0.224
		[4 1 1]	0.021	0.251	0.056	0.424	0.072	0.323	0.150	0.025	0.181
Discrete time non-Markov	3-factor	VAR(4)	0.064	0.207	0.505	0.649	0.242	0.359	0.155	0.207	0.287
		VARMA(1,1)	0.098	0.058	0.453	0.347	0.226	0.024	0.078	0.178	0.281

Table 7: In-sample AdjRns

This table reports the in-sample *AdjRns* of the Markov and non-Markov models for the nine bond markets. The number of states and factors is indicated in the first two columns. The model specification is indicated in the third column. The results presented are based on the model-implied expected excess return.

			Asia Pacific			Continental Europe			NA and UK Group			
			AU	JP	NZ	CH	DE	SE	CA	GB	US	
40	Markov	3-factor	[1 1 1]	0.476	0.670	0.464	0.815	0.438	0.423	0.516	0.303	0.433
		4-state	[1 1 2]	0.134	0.534	0.350	0.733	0.372	0.313	0.473	0.258	0.370
			[1 2 1]	0.257	0.050	0.010	0.491	0.400	0.213	0.412	0.027	0.322
			[2 1 1]	0.287	0.681	0.138	-0.247	0.476	0.469	0.459	-0.311	0.383
		5-state	[1 1 3]	0.379	0.095	0.285	0.659	0.379	0.153	0.420	0.404	0.408
			[1 2 2]	0.415	0.436	-0.112	0.461	0.370	-0.283	0.436	-0.069	0.278
			[1 3 1]	0.355	-0.290	0.393	0.352	0.270	0.214	0.451	0.227	0.285
			[2 1 2]	0.102	0.648	0.191	-0.104	-0.378	-0.291	0.452	0.333	0.283
			[2 2 1]	0.358	-0.349	-0.368	0.141	0.354	0.227	0.337	-0.224	0.313
			[3 1 1]	0.369	-0.540	-0.226	0.217	0.338	0.613	0.377	0.240	-0.251
	non-Markov (3-factor)	6-state	[1 1 4]	0.366	0.695	0.272	0.537	0.399	0.290	0.399	-0.043	0.359
			[1 2 3]	0.278	0.267	0.386	0.470	0.451	0.163	0.429	-0.242	0.256
	[1 3 2]		0.391	0.150	0.372	0.375	0.622	0.265	0.390	0.175	0.251	
	[1 4 1]		0.255	-0.043	0.049	0.527	0.216	-0.311	0.357	-0.129	0.280	
	[2 1 3]		0.288	0.788	0.444	-0.324	-0.301	0.396	0.300	0.468	0.274	
	[2 2 2]		0.050	0.500	-0.176	-0.006	0.384	0.278	0.121	0.202	0.324	
	[2 3 1]		-0.050	0.280	-0.276	-0.071	0.351	0.379	0.143	-0.110	0.388	
	[3 1 2]		0.213	0.552	-0.252	-0.177	0.593	0.647	0.544	0.213	0.309	
	[3 2 1]		-0.179	-0.600	-0.032	-0.165	0.315	0.176	0.385	0.135	0.329	
	[4 1 1]		0.391	0.337	-0.151	-0.665	0.399	0.629	0.460	0.124	0.314	
Discrete time non-Markov	3-factor	VAR(4)	0.385	0.488	0.386	0.747	0.444	0.595	0.385	0.331	0.413	
		VARMA(1,1)	0.481	0.477	0.458	0.377	0.459	0.381	0.402	0.309	0.379	

Table 8: Out-of-sample AdjRns

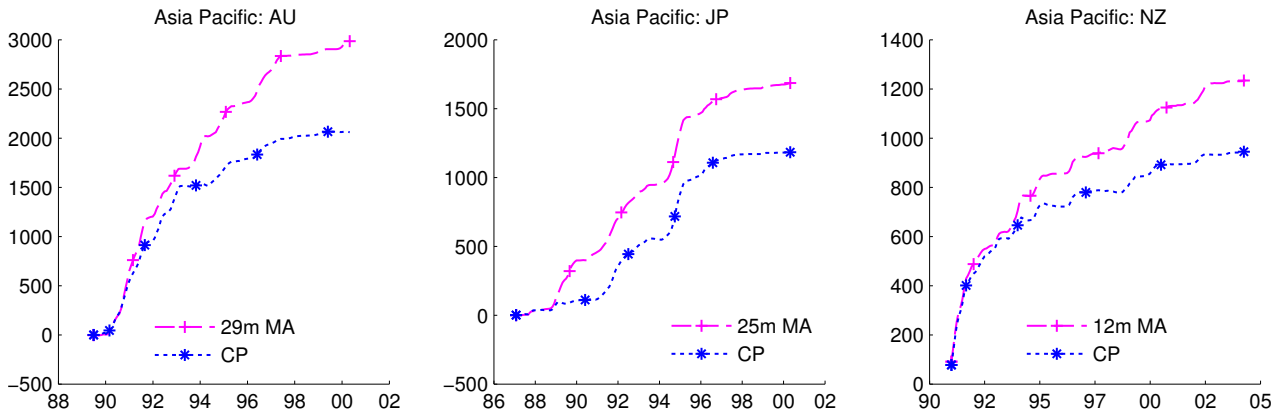
This table reports the out-of-sample *AdjRns* of the Markov and non-Markov models for the nine bond markets. The number of states and factors is indicated in the first two columns. The model specification is indicated in the third column. The results presented are based on the model-implied expected excess return. The best result in each column is highlighted in gray.

			Asia Pacific			Continental Europe			NA and UK Group		
			AU	JP	NZ	CH	DE	SE	CA	GB	US
Markov	3-factor	[1 1 1]	-0.004	-0.438	0.633	-0.260	0.196	0.469	-0.338	-0.120	-0.330
		[1 1 2]	0.274	0.802	0.016	0.133	0.386	-0.256	0.299	-0.247	-0.137
	4-state	[1 2 1]	-0.101	-0.781	0.076	0.591	0.540	-0.325	0.462	-0.363	0.364
		[2 1 1]	0.377	0.877	0.293	0.429	0.440	-0.170	0.447	0.394	0.470
		[1 1 3]	0.036	-0.776	0.107	0.471	0.634	-0.346	0.159	-0.252	-0.222
	5-state	[1 2 2]	0.328	-0.666	0.040	0.436	0.146	-0.336	0.301	0.443	0.565
		[1 3 1]	-0.052	-0.797	0.096	0.555	0.657	-0.274	0.622	0.135	0.303
		[2 1 2]	0.238	0.839	0.037	-0.523	-0.125	-0.365	-0.045	-0.314	-0.135
		[2 2 1]	0.096	0.721	0.018	0.567	0.635	0.436	-0.463	0.428	-0.348
		[3 1 1]	0.265	-0.812	0.029	0.464	0.664	0.406	0.366	0.524	0.464
non-Markov (3-factor)	6-state	[1 1 4]	-0.183	0.816	0.065	0.622	0.546	-0.376	0.074	0.201	0.144
		[1 2 3]	-0.141	-0.755	0.069	0.405	0.462	0.081	0.202	-0.448	0.499
		[1 3 2]	0.270	-0.783	0.076	0.572	-0.118	-0.448	0.188	0.068	0.529
		[1 4 1]	-0.228	-0.793	0.008	0.634	0.625	-0.351	0.994	0.425	0.329
	[2 1 3]	0.215	0.555	0.073	0.459	0.298	0.430	0.904	-0.181	-0.374	
	[2 2 2]	-0.167	-0.678	-0.006	-0.558	0.528	-0.100	-0.398	-0.458	0.547	
	[2 3 1]	-0.215	-0.762	0.011	-0.537	0.582	0.410	-0.766	0.394	0.578	
	[3 1 2]	-0.179	0.763	0.070	-0.700	-0.449	0.209	0.003	-0.176	0.534	
	[3 2 1]	-0.234	-0.788	0.033	0.525	0.737	0.253	0.950	0.479	0.596	
	[4 1 1]	0.349	-0.189	0.033	0.373	0.610	0.452	0.709	0.519	0.501	
Discrete time non-Markov	3-factor	VAR(4)	-0.124	-0.538	0.296	-0.514	-0.272	-0.028	-0.506	-0.069	-0.378
		VARMA(1,1)	0.248	-0.418	0.126	0.284	0.140	0.228	-0.493	0.206	0.424

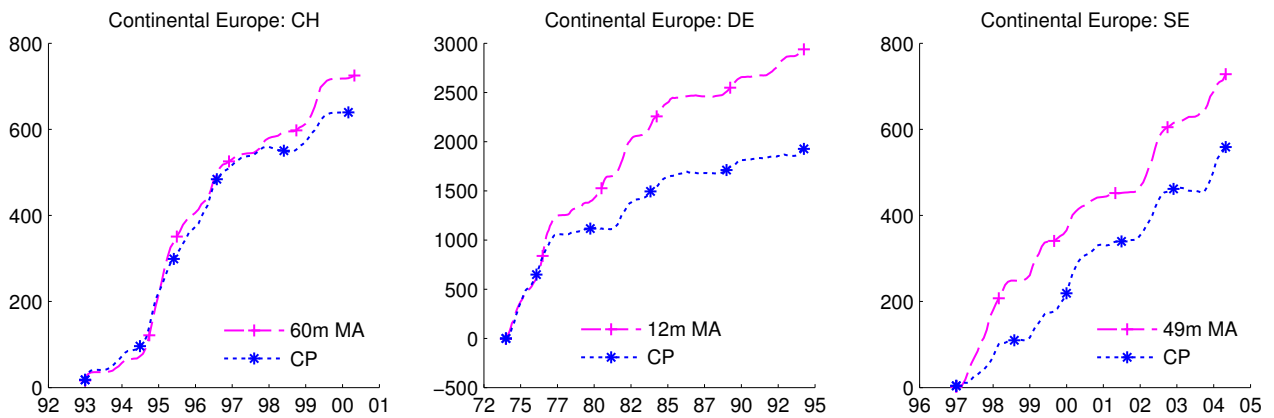
Figure 1: CP vs. MA: In-sample CumRns

The plots in this figure present the in-sample *CumRns* over time of the CP and MA approaches for the nine bond markets. Panels (a), (b), and (c) present results of the Asia Pacific, Continental Europe, and NA and UK Group, respectively. The MA results are based on the number of lags that gives rise to the best out-of-sample *AdjRns* between 12 to 60 months. That number is shown in the legends. The *CumRns* are in basis points. All the x-axes are year in the format of “yy”.

(a) Asia Pacific



(b) Continental Europe



(c) NA and UK Group

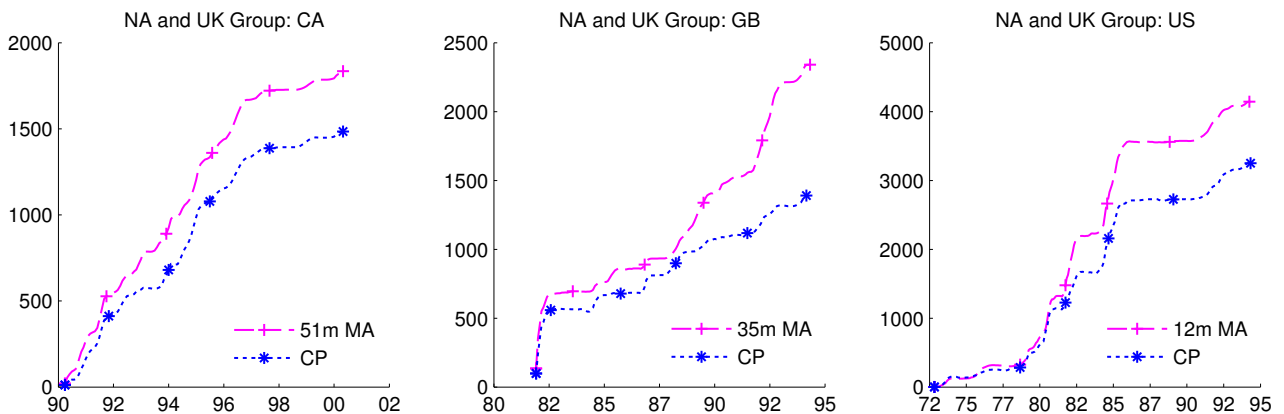
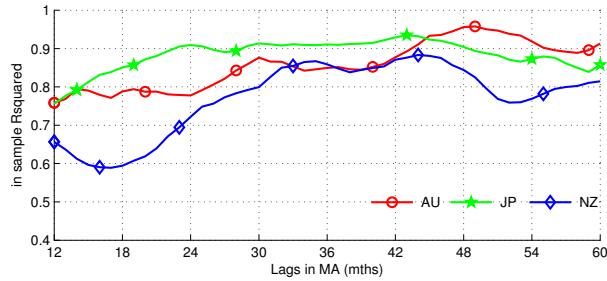


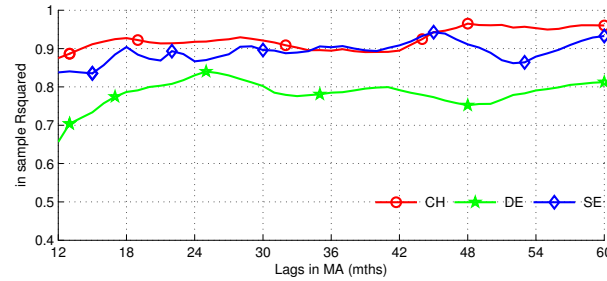
Figure 2: MA approach with varying lags

The panels in the upper row present the in-sample R^2 's; the panels in the middle row present the in-sample $AdjRns$; the panels in the bottom row present the out-of-sample $AdjRns$. The left, middle, and right columns correspond to the results of Asia Pacific, Continental Europe, and NA and UK groups, respectively.

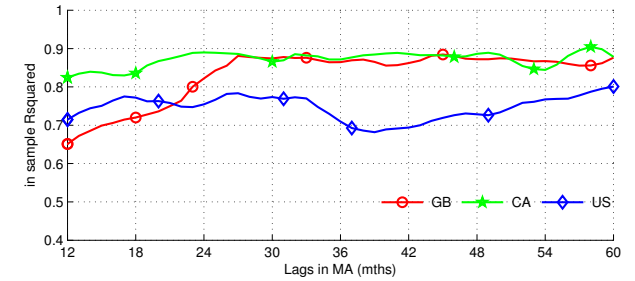
(a) In-sample R^2 s: AP Group



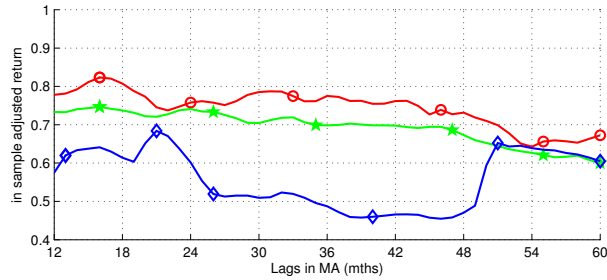
(b) In-sample R^2 s: CE Group



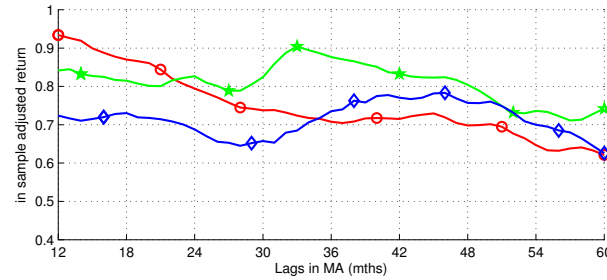
(c) In-sample R^2 s: NA and UK Group



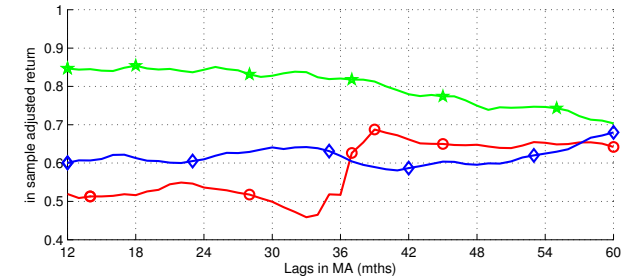
(d) In-sample $AdjRns$: AP Group



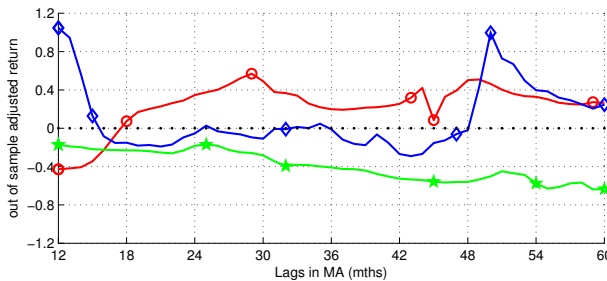
(e) In-sample $AdjRns$: CE Group



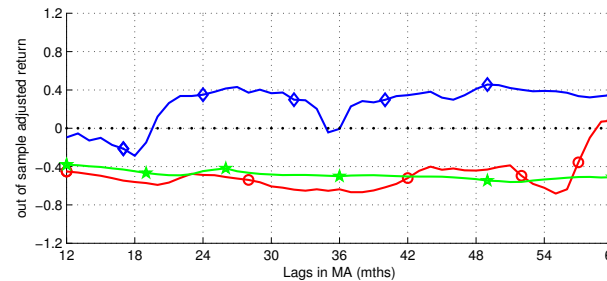
(f) In-sample $AdjRns$: NA and UK Group



(g) Out-of-sample $AdjRns$: AP Group



(h) Out-of-sample $AdjRns$: CE Group



(i) Out-of-sample $AdjRns$: NA and UK Group

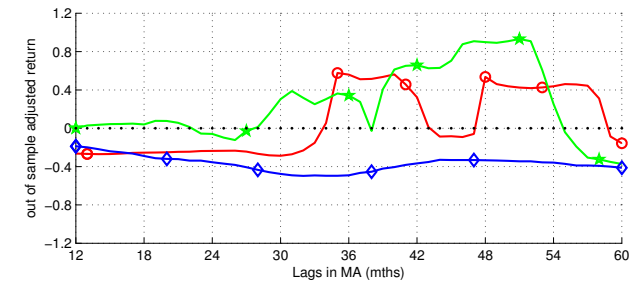
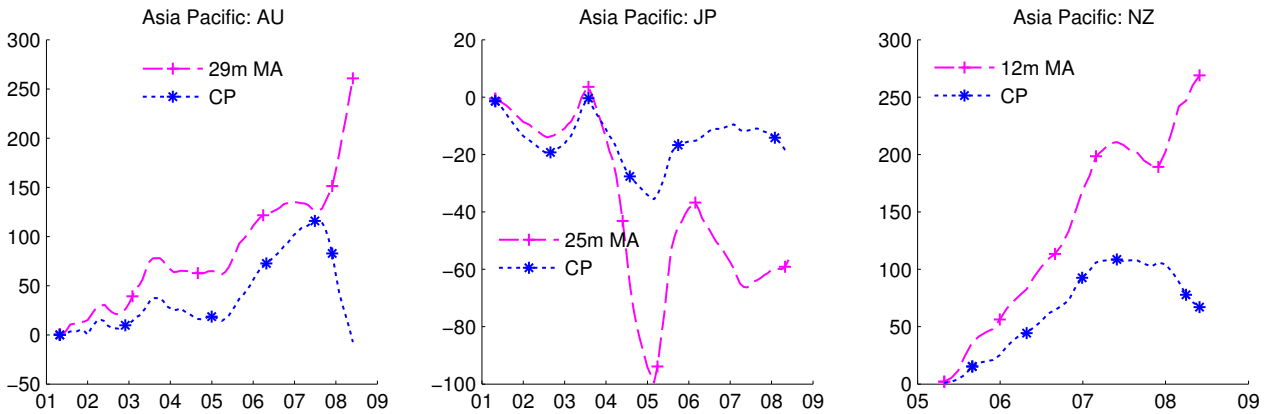


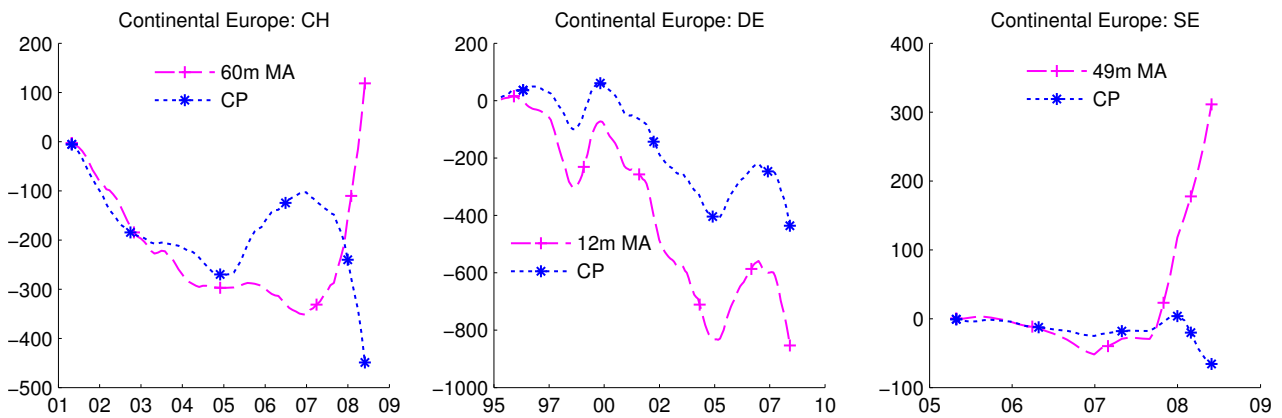
Figure 3: CP vs. MA: Out-of-sample CumRns

The plots in this figure present the out-of-sample *CumRns* over time of the CP and MA approaches for the nine bond markets. Panels (a), (b), and (c) present results of the Asia Pacific, Continental Europe, and NA and UK Group, respectively. The MA results are based on the number of lags that gives rise to the best out-of-sample *AdjRns* between 12 to 60 months. That number is shown in the legends. The *CumRns* are in basis points. All the x-axes are year in the format of “yy”.

(a) Asia Pacific



(b) Continental Europe



(c) NA and UK Group

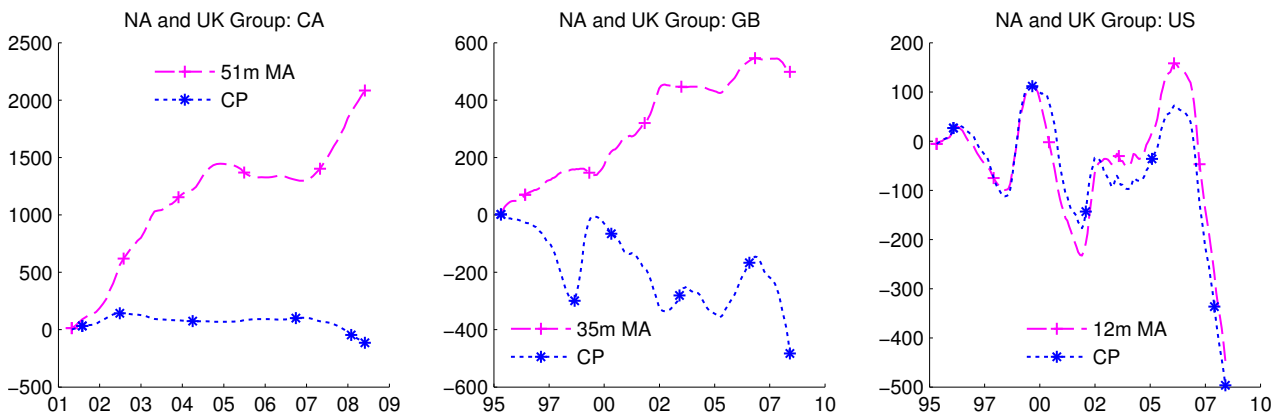
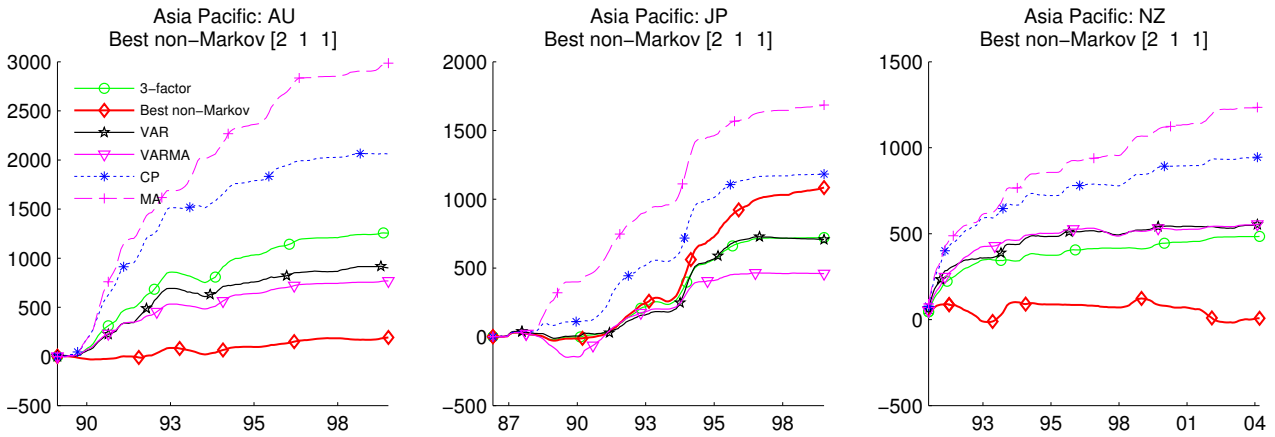


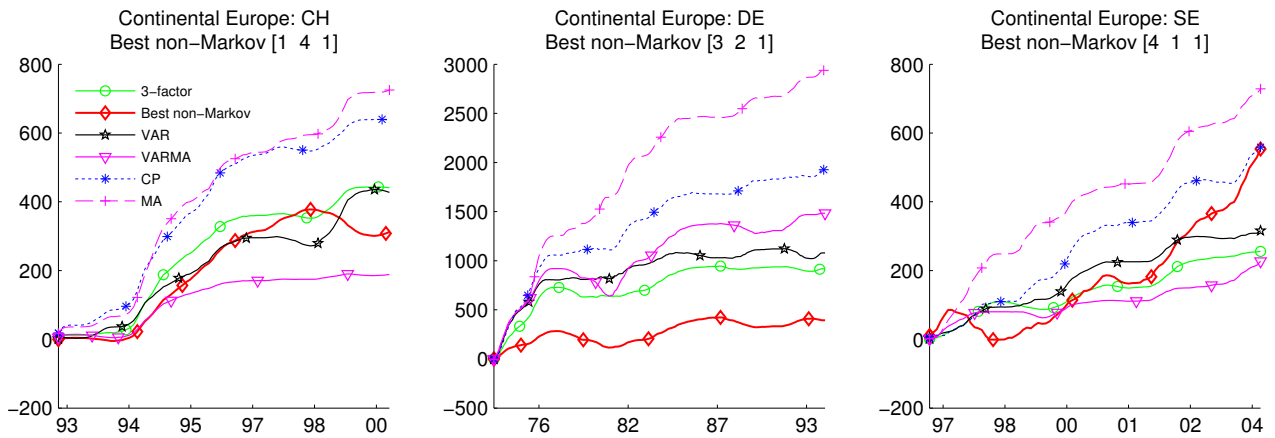
Figure 4: In-sample CumRns: Best non-Markov model vs. benchmarks vs. CP and MA

The plots in this figure present the in-sample *CumRns* over time for the best non-Markov model (among the 19 specifications), the benchmark models (three-factor Markov model, VAR(4), and VARMA(1,1)), and the CP and MA approaches for the nine bond markets. Panels (a), (b), and (c) present results of the Asia Pacific, Continental Europe, and NA and UK Group, respectively. The *CumRns* are in basis points. All the x-axes are year in the format of “yy”.

(a) Asia Pacific



(b) Continental Europe



(c) NA and UK Group

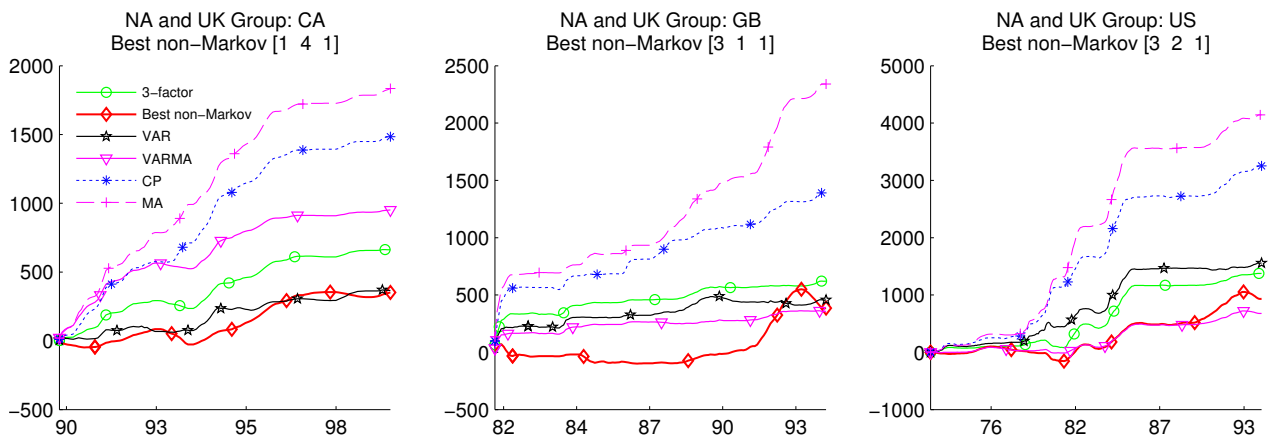
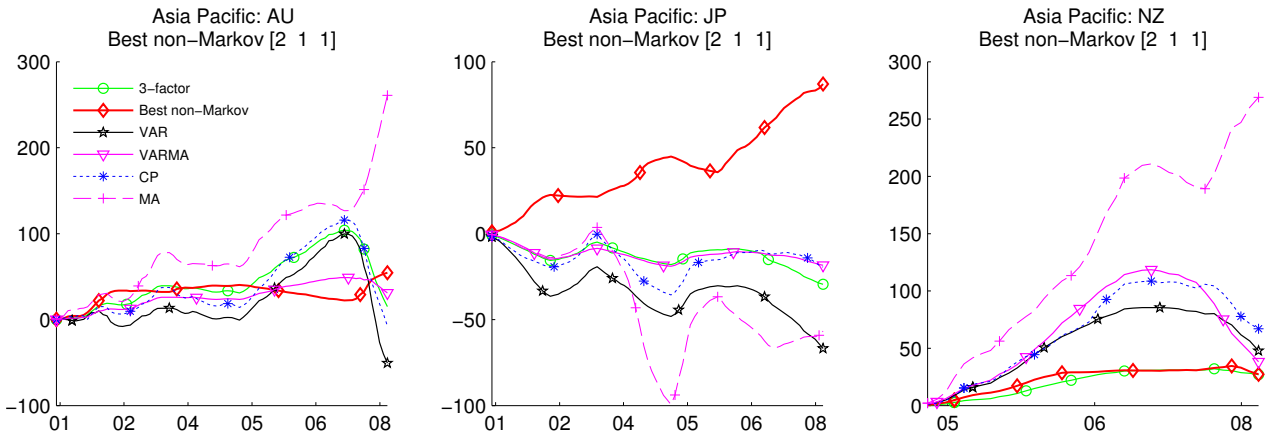


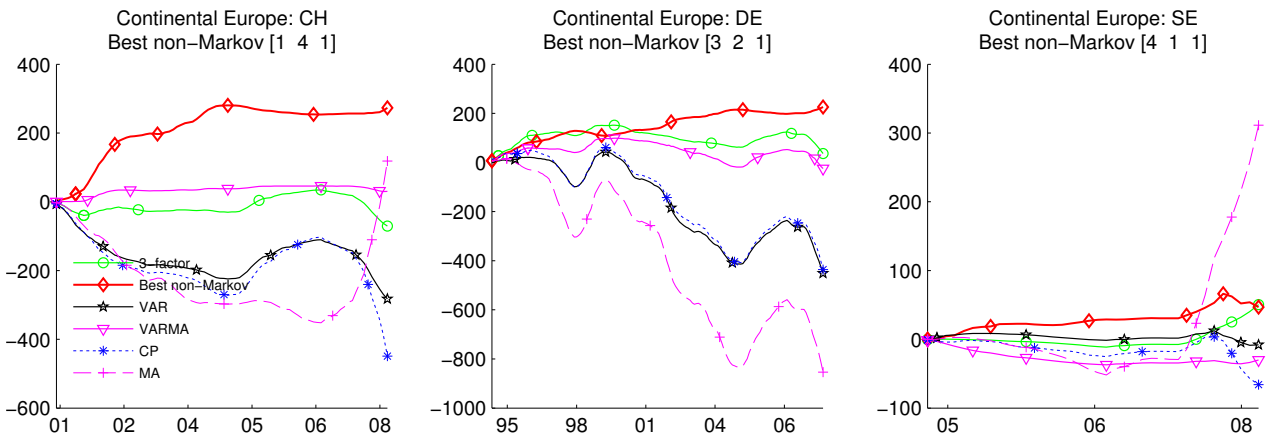
Figure 5: Out-of-sample CumRns: Best non-Markov model vs. benchmarks vs. CP and MA

The plots in this figure present the out-of-sample *CumRns* over time for the best non-Markov model (among the 19 specifications), the benchmark models (three-factor Markov models, VAR(4), and VARMA(1,1)), and the CP and MA approaches for the nine bond markets. Panels (a), (b), and (c) present results of the Asia Pacific, Continental Europe, and NA and UK Group, respectively. The *CumRns* are in basis points. All the x-axes are year in the format of “yy”.

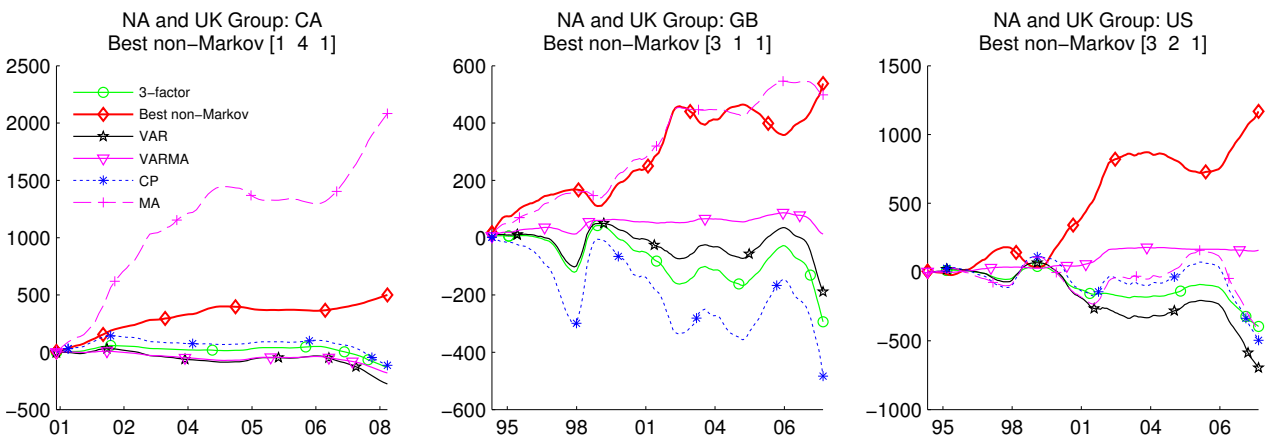
(a) Asia Pacific



(b) Continental Europe



(c) NA and UK Group



Appendices

A Time-homogeneous forward curves

By taking the limit as $t \rightarrow \infty$ in (2), the forward rate can be rewritten as $r(t, x) = \varphi + \Theta^*(x) + \mathbf{C}_0 \exp(\mathbf{A}x) Z_t$, where

$$\Theta^*(x) \equiv \lim_{t \rightarrow \infty} \Theta(t, x) \quad (\text{A1})$$

$$= \mathbf{C}(x) \left(\mathbf{A}^{-1} \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^{-1} \right) \mathbf{C}_0^\top - \frac{1}{2} \mathbf{C}(x) \left(\mathbf{A}^{-1} \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^{-1} \right) \mathbf{C}(x)^\top, \quad (\text{A2})$$

$$\varphi \equiv \lim_{t \rightarrow \infty} r(0, t+x). \quad (\text{A3})$$

In Appendix B, we show that the second equality holds for any invertible \mathbf{A} .²⁷ The zero-coupon bond price is given by:

$$P(t, t+x) = \exp\left(-\int_0^x r(t, s) ds\right) = \exp(\mathbf{H}(x) - \mathbf{F}(x)^\top Z_t), \quad (\text{A4})$$

where

$$\mathbf{H}(x) = -\varphi x - \int_0^x \Theta^*(s) ds,$$

$$\mathbf{F}(x)^\top = \mathbf{C}_0 \int_0^x \exp(\mathbf{A}s) ds = (\mathbf{C}(x) - \mathbf{C}_0) \mathbf{A}^{-1}.$$

These results can also be derived using the traditional GDTSM approach by starting from the short rate specification. That is, the short rate is:

$$r(t, 0) = \varphi + \Theta^*(0) + \mathbf{C}_0 Z_t, \text{ where } dZ_t = \mathbf{A}Z_t dt + \mathbf{B}dW_t, \quad Z_0 = \mathbf{0}.$$

Then, $\mathbf{H}(x)$ and $\mathbf{F}(x)$ can be solved from the following ordinary differential equations:

$$\frac{d\mathbf{H}(x)}{dx} = \frac{1}{2} \mathbf{F}(x)^\top \mathbf{B} \mathbf{B}^\top \mathbf{F}(x) - (\varphi + \Theta^*(0)), \quad \frac{d\mathbf{F}(x)}{dx} = \mathbf{A}^\top \mathbf{F}(x) + \mathbf{C}_0^\top,$$

with the boundary conditions $\mathbf{H}(0) = 0$, and $\mathbf{F}(0) = \mathbf{0}_{m \times 1}$.

²⁷ This requirement is automatically satisfied when the k_i 's are restricted to be positive.

Therefore, following (A4), the model-implied n -year zero yield at time t can be expressed as:

$$y_t^{(n)} = \varphi + \frac{\int_0^n \Theta^*(s) ds}{n} + \frac{\int_0^n \mathbf{C}(s) ds}{n} Z_t. \quad (\text{A5})$$

B Derivation of $\Theta^*(x)$

By $\Theta(t, x) = \int_0^t \sigma(x+t-s) \int_0^{x+t-s} \sigma(\tau)^\top d\tau ds$ and $\sigma(x) = \mathbf{C}(x) \mathbf{B}$, we have:

$$\begin{aligned} \Theta(t, x) &= \mathbf{C}(x) \left[\int_0^t \exp(\mathbf{A}(t-s)) \mathbf{B} \mathbf{B}^\top \int_0^{x+t-s} \exp(\mathbf{A}^\top \tau) d\tau ds \right] \mathbf{C}_0^\top \\ &= \mathbf{C}(x) \int_0^t \exp(\mathbf{A}(t-s)) \mathbf{B} \mathbf{B}^\top \left[\begin{array}{c} \int_0^x \exp(\mathbf{A}^\top \tau) d\tau + \\ \int_0^{t-s} \exp(\mathbf{A}^\top \tau) d\tau \exp(\mathbf{A}^\top x) \end{array} \right] ds \mathbf{C}_0^\top \\ &= \mathbf{C}(x) \int_0^t \exp(\mathbf{A}(t-s)) ds \mathbf{B} \mathbf{B}^\top \int_0^x \exp(\mathbf{A}^\top \tau) d\tau \mathbf{C}_0^\top \\ &\quad + \mathbf{C}(x) \int_0^t \exp(\mathbf{A}(t-s)) \mathbf{B} \mathbf{B}^\top \int_0^{t-s} \exp(\mathbf{A}^\top \tau) d\tau ds \mathbf{C}(x)^\top \\ &= \mathbf{C}(x) \int_0^t \exp(\mathbf{A}(t-s)) \mathbf{B} \mathbf{B}^\top \exp(\mathbf{A}^\top(t-s)) ds (\mathbf{A}^\top)^{-1} \mathbf{C}(x)^\top \\ &\quad - \mathbf{C}(x) \int_0^t \exp(\mathbf{A}(t-s)) \mathbf{B} \mathbf{B}^\top ds (\mathbf{A}^\top)^{-1} \mathbf{C}_0^\top. \end{aligned}$$

Then,

$$\begin{aligned} \Theta^*(x) &= \mathbf{C}(x) \left[\lim_{t \rightarrow +\infty} \int_0^t \exp(\mathbf{A}(t-s)) \mathbf{B} \mathbf{B}^\top \exp(\mathbf{A}^\top(t-s)) ds \right] (\mathbf{A}^\top)^{-1} \mathbf{C}(x)^\top \\ &\quad - \mathbf{C}(x) \left[\lim_{t \rightarrow +\infty} \int_0^t \exp(\mathbf{A}(t-s)) ds \right] \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^{-1} \mathbf{C}_0^\top \\ &= \mathbf{C}(x) \left(\mathbf{A}^{-1} \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^{-1} \right) \mathbf{C}_0^\top - \frac{1}{2} \mathbf{C}(x) \left(\mathbf{A}^{-1} \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^{-1} \right) \mathbf{C}(x)^\top. \end{aligned}$$

The second equation above is by the fact that $Y \equiv \int_0^\infty \exp(\mathbf{A}(t-s)) \mathbf{B} \mathbf{B}^\top \exp(\mathbf{A}^\top(t-s)) ds$ satisfies the Lyapunov equation:

$$\mathbf{A}Y + Y\mathbf{A}^\top = -\mathbf{B}\mathbf{B}^\top.$$

Therefore,

$$\mathbf{C}(x) \mathbf{Y} (\mathbf{A}^\top)^{-1} \mathbf{C}(x)^\top + \mathbf{C}(x) \mathbf{A}^{-1} \mathbf{Y} \mathbf{C}(x)^\top = -\mathbf{C}(x) \left(\mathbf{A}^{-1} \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^{-1} \right) \mathbf{C}(x)^\top,$$

which means that

$$\mathbf{C}(x) \mathbf{Y} (\mathbf{A}^\top)^{-1} \mathbf{C}(x)^\top = -\frac{1}{2} \mathbf{C}(x) \left(\mathbf{A}^{-1} \mathbf{B} \mathbf{B}^\top (\mathbf{A}^\top)^{-1} \right) \mathbf{C}(x)^\top,$$

since $\mathbf{C}(x) \mathbf{Y} (\mathbf{A}^\top)^{-1} \mathbf{C}(x)^\top$ and $\mathbf{C}(x) \mathbf{A}^{-1} \mathbf{Y} \mathbf{C}(x)^\top$ are both scalars, and are transposes of each other.

C Proofs

C.1 Proof of Proposition 1

It is trivial that any FDR with $n > m$ can be transformed into one with \mathbf{B} being

$$\begin{bmatrix} \underbrace{B_1}_{m \times m} \\ \underbrace{\mathbf{0}}_{(n-m) \times m} \end{bmatrix}.$$

So, without loss of generality, we consider the following Z_t as representative of any FDR with $n > m$:

$$dZ_t = d \begin{bmatrix} \underbrace{Z_{1,t}}_{m \times 1} \\ \underbrace{Z_{2,t}}_{(n-m) \times 1} \end{bmatrix} = \begin{bmatrix} \underbrace{A_{11}}_{m \times m} & \underbrace{A_{12}}_{m \times (n-m)} \\ \underbrace{A_{21}}_{(n-m) \times m} & \underbrace{A_{22}}_{(n-m) \times (n-m)} \end{bmatrix} \begin{bmatrix} \underbrace{Z_{1,t}}_{m \times 1} \\ \underbrace{Z_{2,t}}_{(n-m) \times 1} \end{bmatrix} dt + \begin{bmatrix} \underbrace{B_1}_{m \times m} \\ \underbrace{\mathbf{0}}_{(n-m) \times m} \end{bmatrix} \underbrace{dW_t}_{m \times 1}.$$

Therefore, the dynamics of $Z_{2,t}$ is given by an ODE:

$$dZ_{2,t} = (A_{21}Z_{1,t} + A_{22}Z_{2,t}) dt, \quad Z_{2,0} = 0. \tag{A6}$$

That is, $Z_{2,t} = \int_0^t A_{21} Z_{1,s} ds + \int_0^t A_{22} Z_{2,s} ds$. Given this, the dynamic of $Z_{1,t}$ is:

$$dZ_{1,t} = \left(A_{11} Z_{1,t} + A_{12} A_{21} \int_0^t Z_{1,s} ds + A_{12} A_{22} \int_0^t Z_{2,s} ds \right) dt + B_1 dW_t.$$

Apparently, the drift term of $Z_{1,t}$ contains integrals of its own and $Z_{2,t}$'s historical values.

We now show that $Z_{2,t}$ is an exponentially-weighted average of $Z_{1,t}$. We guess, and later verify, that $Z_{2,t}$ takes the form

$$Z_{2,t} = e^{A_{22}t} \phi(t).$$

Therefore, we have

$$dZ_{2,t} = A_{22} e^{A_{22}t} \phi(t) dt + e^{A_{22}t} d\phi(t). \quad (\text{A7})$$

Comparing (A7) with (A6), it is clear that $\phi(t)$ solves the following ODE:

$$d\phi(t) = e^{-A_{22}t} A_{21} Z_{1,t} dt.$$

Thus $\phi(t) = \int_0^t e^{-A_{22}s} A_{21} Z_{1,s} ds$ and $Z_{2,t} = \int_0^t e^{A_{22}(t-s)} A_{21} Z_{1,s} ds$. Therefore, the drift of $Z_{1,t}$ generally depends on its own history, showing that the first m states are non-Markov on their own.

C.2 Proof of Theorem 1

To show that the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is a realization, according to Björk and Gombani, 1999, Proposition 3.1, we only need to demonstrate that:

$$\mathbf{C}(0) \exp(\mathbf{A}x) = \mathbf{C}(x) = \left[\left[1, x, \dots, x^{n_i-1} \right] e^{-k_i x} \right]_{i=1}^I.$$

First, we rewrite $\left[\left[1, x, \dots, x^{n_i-1} \right] e^{-k_i x} \right]_{i=1}^I$ in matrix form:

$$\left[\left[1, x, \dots, x^{n_i-1} \right] e^{-k_i x} \right]_{i=1}^I = \text{Poly}(\mathbf{N}, x) \text{ExpM}(\mathbf{K}, \mathbf{N}, x),$$

where $\mathbf{N} = [n_1, n_2, \dots, n_I]$ is a $1 \times I$ row vector with elements of natural numbers, $\mathbf{K} =$

$[k_1, k_2, \dots, k_I]$ is a $1 \times I$ row vector with elements of distinct positive real numbers, $\text{ExpM}(\mathbf{K}, \mathbf{N}, x) : (\mathbf{R}^{1 \times I}, \mathbf{N}^{1 \times I}, \mathbf{R}_+) \rightarrow \mathbf{R}_+^{n \times n}$, $n = \sum_{i=1}^I n_i$, is a matrix function defined as:

$$\text{ExpM}(\mathbf{K}, \mathbf{N}, x) \equiv \begin{bmatrix} \text{ExpM}_1 & \cdots & 0 \\ & \text{ExpM}_2 & \vdots \\ \vdots & & \ddots \\ 0 & \cdots & \text{ExpM}_I \end{bmatrix},$$

$$\text{ExpM}_i \equiv \exp \left\{ - \underbrace{\text{diag}[k_i, \dots, k_i]}_{n_i \times n_i} x \right\},$$

where $\text{diag}[\dots]$ is a compact notation for a diagonal matrix, and $\text{Poly}(\mathbf{N}, x) : (\mathbf{N}^{1 \times I}, \mathbf{R}_+) \rightarrow \mathbf{R}_+^{1 \times n}$ is a vector function defined as:

$$\text{Poly}(\mathbf{N}, x) \equiv [1, x, \dots, x^{n_i-1}]_{i=1}^I.$$

It is then enough to show that for any $i = 1, 2, \dots, I$,

$$\underbrace{[1, 0, \dots, 0]}_{1 \times n_i} \exp(\mathbf{A}_i x) = [1, x, \dots, x^{n_i-1}] \text{ExpM}_i$$

$$\iff [1, 0, \dots, 0] \exp(\mathbf{A}_i x) (\text{ExpM}_i)^{-1} = [1, x, \dots, x^{n_i-1}]. \quad (\text{A8})$$

Since ExpM_i is a diagonal matrix with identical elements on the diagonal, $\exp(\mathbf{A}_i x) (\text{ExpM}_i)^{-1}$

becomes:

$$\exp \left(\begin{bmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & 0 & \ddots & \\ & & & \ddots & n_i - 1 \\ & & & & 0 \end{bmatrix} x \right) = UPas^x = UPas \circ UToe(x),$$

where $UPas$ is an n_i -dimensional upper-triangular Pascal matrix (which has a first row of ones), $UToe(x)$ is an n_i -dimensional upper-triangular Toeplitz matrix of the power series of x :

$$UToe(x) = \begin{bmatrix} 1 & x & x^2 & \dots & x^{n_i-1} \\ & 1 & x & \dots & x^{n_i-2} \\ & & \ddots & & \vdots \\ & & & 1 & x \\ & & & & 1 \end{bmatrix},$$

and \circ denotes the Hadamard product. In fact, $UPas^x$ is also referred to as the transpose of the generalized Pascal matrix of x . For example, see Yang and Micek (2007) and Stefan (2011).

Therefore, the left hand side of (A8) is the first row of the Hadamard product of $UPas$ and $UToe(x)$, which equals the right hand side of (A8).

D Estimation and forecast of VAR(4) and VARMA(1,1) models

We estimate three-factor VAR and VARMA models, where the three factors are the first three principal components of the zero yields. The estimation for Joslin et al. (2013)'s VAR(4) model is exactly the same as that in JSZ with the number of lags being four instead of one. The estimation for Feunou and Fontaine (2014)'s VARMA(1,1) model is again very similar to that in JSZ, except that the VARMA parameters directly enter the likelihood function and are part of the ML estimation, because they cannot be estimated using OLS regressions. The likelihood function for an unrestricted VARMA(1,1) model can be found in Lütkepohl (2007, p. 464).

Suppose that the VAR(4) model is described by:

$$PC_t = v + A_1 PC_{t-\frac{1}{12}} + \dots + A_4 PC_{t-\frac{4}{12}} + \sqrt{\Sigma} \epsilon_t,$$

and the VARMA(1,1) model by:

$$(PC_t - v) = A_1 \left(PC_{t-\frac{1}{12}} - v \right) + \sqrt{\Sigma} \epsilon_t + M_1 \sqrt{\Sigma} \epsilon_{t-\frac{1}{12}}.$$

Given the parameters estimates, for VAR(4), the h -month ahead forecast is given by:

$$\widehat{PC}_t \left(\frac{h}{12} \right) = \hat{v} + \hat{A}_1 \widehat{PC}_t \left(\frac{h-1}{12} \right) + \dots + \hat{A}_4 \widehat{PC}_t \left(\frac{h-4}{12} \right).$$

For VARMA(1,1), the h -month ahead forecast is given by:

$$\widehat{PC}_t \left(\frac{h}{12} \right) = \hat{v} + \sum_{i=1}^{12(t-1)+h} \hat{\Pi}_i \left[\widehat{PC}_t \left(\frac{h-i}{12} \right) - \hat{v} \right],$$

where $\widehat{PC}_t(j) = PC_{t+j}$ for $j < 0$, and $\hat{\Pi}_i = (-1)^{i-1} \left(\hat{M}_1^i + \hat{M}_1^{i-1} \hat{A}_1 \right)$.

Denote the model-implied j -year zero yield by:

$$y_t^{(j)} = C_m^{\text{PC}}(j) + B_m^{\text{PC}}(j) PC_t.$$

Then, the model-implied expected excess return is given by:

$$\mathbb{E}_t(\bar{r}_{t+1}) = \left(\frac{1}{4} \sum_{j=2}^5 j y_t^{(j)} - y_t^{(1)} \right) - \frac{1}{4} \sum_{j=1}^4 \left(j C_m^{\text{PC}}(j) + j B_m^{\text{PC}}(j) \widehat{PC}_t(1) \right).$$

E Unspanned risk specifications

In this appendix, we show that our framework can be easily extended to cover unspanned risk specifications in the sense of Joslin et al. (2014).

The key to having the unspanned risk property is to allow the stochastic discount factor, hence the market price of risk, to depend on a broader set of state variables, some of which are unspanned by the bond market specific states.²⁸ Denote these broader state variables by $Z_t^E \in \mathbf{R}^N$, which are driven by an M -dimensional Wiener process W_t^E under the \mathbb{P} measure. Recall that the bond market specific states $Z_t \subseteq Z_t^E$ are driven by $W_t^{\mathbb{P}} \subseteq W_t^E$, each with dimensionality n and m , respectively, where $n \geq m$, $N \geq n$, $M \geq m$, and $N \geq M$. Since the market price of risk is an affine function of Z_t^E , $W_t^{\mathbb{P}}$ is linked to W_t via the following relation:

$$dW_t^{\mathbb{P}} = \left(\underbrace{\lambda_1}_{m \times 1} + \underbrace{\lambda_2}_{m \times N} Z_t^E \right) dt + dW_t. \quad (\text{A9})$$

In light of the recent literature, non-Markov states (Joslin et al., 2013; Feunou and Fontaine, 2014) and additional risk factors, such as macro variables (Joslin et al., 2014), might be unspanned by the bond market specific states. We present two simple examples below to demonstrate how our framework can accommodate these unspanning features.

E.1 Unspanned non-Markov states

For simplicity, we set $M = m = 1$, $N = 2$, and $n = 1$. Therefore, Z_t under the \mathbb{Q} measure follows:

$$dZ_t = -k_1 Z_t dt + \Omega_1 dW_t. \quad (\text{A10})$$

Since there is no unspanned risk factor in this case, i.e., $W_t = W_t^E$, we specify the dynamics

²⁸ Joslin et al. (2014) call these broader state variables the “states of the economy.”

of Z_t^E under \mathbb{Q} as:

$$dZ_t^E = \begin{bmatrix} -k_1 & 1 \\ 0 & -k_1 \end{bmatrix} Z_t^E dt + \begin{bmatrix} \Omega_2 \\ \Omega_1 \end{bmatrix} dW_t.$$

Thus the second element in Z_t^E is Z_t , i.e., $Z_t^E(2) = Z_t$. As for $Z_t^E(1)$, which captures the lagged information about Z_t , it is unspanned by the current yields because it is excluded from the pricing state variable Z_t .

Given (A9), the \mathbb{P} -dynamics of Z_t^E is given by:

$$dZ_t^E = - \begin{bmatrix} \Omega_2 \\ \Omega_1 \end{bmatrix} \lambda_1 dt + \left(\begin{bmatrix} -k_1 & 1 \\ 0 & -k_1 \end{bmatrix} - \begin{bmatrix} \Omega_2 \\ \Omega_1 \end{bmatrix} \lambda_2 \right) Z_t^E dt + \begin{bmatrix} \Omega_2 \\ \Omega_1 \end{bmatrix} dW_t^{\mathbb{P}}.$$

E.2 Unspanned additional risk factors

In this example, we assume $M = 2$, $m = 1$, $N = 2$, and $n = 1$. Although Z_t still follows (A10), W_t^E includes an additional random source other than $W_t^{\mathbb{P}}$:

$$W_t^E = \begin{bmatrix} W_t^{\mathbb{P}} \\ W_t^U \end{bmatrix}.$$

The \mathbb{P} -dynamics of Z_t^E is specified as:

$$dZ_t^E = \mathbf{A}^{E,\mathbb{P}} \left(Z_t^E - \mu^E \right) dt + \begin{bmatrix} \Omega_1 & 0 \\ \Omega_2 & \Omega_3 \end{bmatrix} dW_t^E.$$

Given the assumption in Joslin et al. (2011) and others about the perfect observability of portfolios of yields and macro variables, the parameters $\mathbf{A}^{\text{E,P}}$, μ^{E} , Ω_1 , Ω_2 , and Ω_3 can be estimated using standard maximum likelihood. Once they are estimated, the market prices of risk λ_1 and λ_2 are given by the following equations:

$$\lambda_1 = \frac{(\mathbf{A}^{\text{E,P}} \mu^{\text{E}}) (1)}{\Omega_1},$$

$$\lambda_2 = \frac{\begin{bmatrix} -k_1 & 0 \end{bmatrix} - \mathbf{A}^{\text{E,P}} (1, :)}{\Omega_1},$$

where $(\mathbf{A}^{\text{E,P}} \mu^{\text{E}}) (1)$ and $\mathbf{A}^{\text{E,P}} (1, :)$ denote the first rows of $\mathbf{A}^{\text{E,P}} \mu^{\text{E}}$ and $\mathbf{A}^{\text{E,P}}$, respectively.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, pp. 267–281. Akademinai Kiado.
- Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics* 50(4), 745–787.
- Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-capm: An extrapolative capital asset pricing model. *Journal of Financial Economics* 115(1), 1 – 24.
- Bauer, M. D., G. D. Rudebusch, and J. C. Wu (2012). Correcting estimation bias in dynamic term structure models. *Journal of Business & Economic Statistics* 30(3), 454–467.
- Bergstrom, A. R. (1983). Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica: Journal of the Econometric Society*, 117–152.
- Björk, T. and A. Gombani (1999). Minimal realizations of interest rate models. *Finance and Stochastics* 3(4), 413–432.
- Björk, T. and L. Svensson (2001). On the Existence of Finite-Dimensional Realizations for Non-linear Forward Rate Models. *Mathematical Finance* 11(2), 205–243.
- Brace, A. and M. Musiela (1994). A MULTIFACTOR GAUSS MARKOV IMPLEMENTATION OF HEATH, JARROW, AND MORTON. *Mathematical Finance* 4(3), 259–283.
- Campbell, J. Y. (1986). A defense of traditional hypotheses about the term structure of interest rates. *The Journal of Finance* 41(1), 183–193.
- Chen, B. and Y. Hong (2011). Testing for the markov property in time series. *Econometric Theory* 28(1), 130.
- Chiarella, C. and O. Kwon (2003). Finite Dimensional Affine Realisations of HJM Models in Terms of Forward Rates and Yields. *Review of Derivatives Research* 6(2), 129–155.
- Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *The American Economic Review* 95(1), 138–160.

- Cochrane, J. H. and M. Piazzesi (2006). Appendix to “bond risk premia”. *Chicago Booth and Stanford GSB*.
- Cox, J., J. Ingersoll Jr, and S. Ross (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society* 53(2), 385–407.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross (1981). A re-examination of traditional hypotheses about the term structure of interest rates. *The Journal of Finance* 36(4), 769–799.
- Dai, Q. and K. Singleton (2000). Specification analysis of affine term structure models. *Journal of Finance* 55(5), 1943–1978.
- Dai, Q. and K. Singleton (2003). Term structure dynamics in theory and reality. *Review of Financial Studies* 16(3), 631–678.
- De Jong, F. and P. Santa-Clara (1999). The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables. *Journal of Financial and Quantitative Analysis* 34(1), 131–157.
- Della Corte, P., L. Sarno, and D. L. Thornton (2008). The expectation hypothesis of the term structure of very short-term rates: statistical tests and economic value. *Journal of Financial Economics* 89(1), 158–174.
- Duffee, G. (1999). Estimating the price of default risk. *Review of Financial Studies* 12(1), 197–226.
- Duffee, G. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405–443.
- Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance* 6(4), 379–406.
- Dumas, B., A. Kurshev, and R. Uppal (2009). Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility. *The Journal of Finance* 64(2), 579–629.
- Feunou, B. and J.-S. Fontaine (2014). Gaussian term structure models and bond risk premia. *Available at SSRN*.
- Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 190–195.

- Heath, D., R. Jarrow, and A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* 60(1), 77–105.
- Higham, N. J. (2008). *Functions of matrices: theory and computation*. Siam.
- Jardet, C., A. Monfort, and F. Pegoraro (2013). No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth. *Journal of Banking & Finance* 37(2), 389 – 402.
- Jin, Y. and P. Glasserman (2001). Equilibrium positive interest rates: a unified view. *Review of Financial Studies* 14(1), 187–214.
- Joslin, S., A. Le, and K. J. Singleton (2013). Gaussian macro-finance term structure models with lags. *Journal of Financial Econometrics* 11(4), 581–609.
- Joslin, S., M. Priebsch, and K. J. Singleton (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*, Forthcoming.
- Joslin, S., K. Singleton, and H. Zhu (2011). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies*.
- Longstaff, F. A. (2000). The term structure of very short-term rates: New evidence for the expectations hypothesis. *Journal of Financial Economics* 58(3), 397–415.
- Longstaff, F. A. and E. S. Schwartz (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *The Journal of Finance* 47(4), 1259–1282.
- Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Springer.
- Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867–887.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics* 6(2), 461–464.
- Stefan, S. (2011). A generalization of the pascal matrix and its properties. *Facta universitatis-series: Mathematics and Informatics* (26), 17–27.

- Trolle, A. and E. Schwartz (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *Review of Financial Studies* 22(5), 2007.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5(2), 177–188.
- Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *The American Economic Review* 101(4), 1514–1534.
- Wymer, C. R. (1993). Continuous-time models in macroeconomics: specification and estimation. In *Continuous-Time Econometrics*, pp. 35–79. Springer.
- Yang, Y. and C. Micek (2007). Generalized pascal functional matrix and its applications. *Linear algebra and its applications* 423(2-3), 230–245.