AFFINE AND QUADRATIC MODELS WITH MANY FACTORS AND FEW PARAMETERS

Dr Marco Realdon

Swansea School of Management

Swansea University, Bay Campus, Fabian Way, Swansea, UK

email: marco.realdon@gmail.com; tel.: 0044/07951893423.

17/9/2018 (this version 15/10/2019)

Abstract

"Classic" affine and quadratic term structure models in the literature usually have three or four factors and tens of parameters. However affine and quadratic term structure models with many factors and few parameters (MFFP), i.e. with up to twenty factors and **with six to seven** parameters, fit and predict US **and** Euro sovereign yields better than "classic" affine and quadratic models. **MFFP** models also fit **the volatility of and the** correlations between changes in yields of different maturities better than "classic" models. MFFP models outperform because fewer parameters **reduce in sample over-fitting** and because more factors give models more flexibility to match yields of different matu-

rities. **Among MFFP models, a type of affine model with stochastic volatility is usually preferable to the homoschedastic affine model, but for US yields the quadratic model seems preferable among five factor MFFP models**.

Key words: affine term structure models, quadratic term structure models, **discrete time**, **squared** Gaussian shocks, **Giacomini-White tests**.

JEL classification: G12, G13.

1 Introduction

The vast majority of the literature on affine and quadratic term structure models uses three or four stochastic factors, which are either latent factors or linear combinations of observed Government bond yields. These models usually have many parameters, typically between twenty and sixty. Canonical affine and quadratic models have also been proposed, which have the maximum number of econometrically identifiable parameters for a given number of latent factors. Thus the literature has effectively focused on affine and quadratic term structure models with few factors and many parameters. However affine and quadratic models remain tractable even with ten, twenty or more stochastic factors, while tens of parameters are often burdensome to estimate and pose the risk of over-fitting the model to sample data. Few parameters reduce the risk of over-fitting, of statistically insignificant parameters and of unstable parameters as samples are updated with new data. Fewer parameters also alleviate the considerable

likelihood optimisation burden in model estimation. Therefore focusing on just three or four factors does not fully exploit the tractability of affine and quadratic models, and tens of parameters may be an unnecessary complication. For these reasons affine and quadratic models with few parameters and many factors, i.e. with **six to seven** parameters and with **five to twenty** factors, seem promising, but have been little explored.

A recent exception is Calvet and others (2018), who showed the considerable merits of affine Gaussian models with up to ten factors and as few as six parameters. The factors enter their model through a **cascade structure** whereby one factor reverts to a mean that is itself another factor, which in turn reverts to a mean that is itself yet another factor, and so on. **Some** factors in this chain revert to their respective means more quickly, **due** to greater mean reversion speed **parameters**. **Thus** their model can mimic the high persistence of long term interest rates. While Calvet and others (2018) concentrated on affine Gaussian models, this paper shows that the merits of the many factors-few parameters (MFFP) approach extend beyond affine Gaussian models to affine models with stochastic volatility, **to affine models that rule out negative yields and to quadratic Gaussian models**. Using US Treasury bond yields and Euro AAA rated sovereign bond yields, this paper finds that affine **and quadratic** models with **five to twenty** factors and six to **seven** parameters respectively outperform "classic" affine **and quadratic models** with three factors and up to about forty parameters. With MFFP and a mean reversion chain, quadratic models **and affine models** with stochastic volatility fit and **predict** US and

Euro Government bond yields very accurately, as yields observation errors are often around one or two basis points. **Affine models with MFFP perform so well even without the cascade parameter structure of Calvet and others (2018)**. **The** very good fit to observed yields is welcome since Adrian and others (2013), Golinski and Spencer (2017) and others noted that yield observation errors imply negatively serially correlated bond returns, which are not observed in the data. While Golinski and Spencer (2017) address this problem by modelling excess returns rather than yields, this paper uses models with MFFP to drastically reduce yield observation errors. Small observation errors do not eliminate, but greatly reduce the problem that yield observation errors imply negatively serially correlated bond returns. Small yield observation errors also make MFFP models suitable to price interest rate derivatives. Moreover MFFP models fit **the volatilities of yields changes and the** correlations between **yields changes** of different maturities better than "classic" models. MFFP models outperform "classic" models because fewer parameters **reduce in sample over-fitting** and because more factors give models more flexibility to match yields of different maturities. "Mean reversion chains" preserve the tractability of affine **and quadratic** term structure models, which is why this paper focuses on these models and not on **others**.

In sample and out of sample tests show that, for US yields, quadratic models seem preferable among MFFP models with five factors, while affine models with stochastic volatility of the type $\mathbb{A}_1(n)$ are usually preferable to homoschedastic affine models. **For Euro yields, among**

MFFP models with five or more factors affine models of the type $\mathbb{A}_1(n)$ **are preferable**.

Next the term structure models are presented. Then the empirical evidence shows **the merits of MFFP models in fitting and predicting US and** Euro sovereign yields.

2 Review of the models

This section presents novel extensions of discrete time affine and quadratic models from past literature. These extensions consist in increasing the number of factors up to twenty and in reducing the number of parameters to six or seven, so as to obtain new specifications with many factors and few parameters (MFFP). The many factors enter the models through long "mean reversion chains", whereby one factor reverts to a long term mean that is itself another factor, which in turn reverts to a long term mean that is itself another factor and so on. After the models are estimated through Kalman Filters, empirical tests compare various MFFP models and "classic" models. The "classic" models feature three factors and tens of parameters. The empirical tests, which are the econometric contribution of this paper, are in sample Vuong tests and out of sample Giacomini-White (2006) tests. The Giacomini-White tests compare out of sample conditional densities of US and Euro sovereign yields. The conditional densities

are those forecast by the different affine and quadratic models and are twenty-variate since they refer to yields of all yearly maturities from one year to twenty years.

2.1 Discrete time affine term structure models (DTASTM) with squared Gaussian shocks (SGS)

The discrete time affine models tested in this paper are special cases of DTASTM-SGS, either in their "classic" versions or in their versions with many factors and few parameters (MFFP). DTASTM-SGS are alternative to the DTASTM based on auto-regressive Gamma processes proposed by Le, Singleton and Dai (2010) and others, and are just as tractable. DTATSM-SGS were recently proposed in Realdon (2018) and are chosen **for two reasons:** they need no Feller conditions restricting market prices of risk **and need fewer parameters for** $\mathbb{A}_1(n)$ **models**.

We divide time into weekly steps each of length $\Delta = 1/52$, since time is measured in years. Let t and m be **integer** numbers. $P_{m,t}$ is the value at time $t \cdot \Delta$ of a default-free discount bond with unit face value and with maturity at time $(t+m)\Delta$. $r_t = -\frac{\ln(P_{1,t})}{\Delta}$ is the one week default-free interest rate during $[t\Delta, (t+1)\Delta]$. Similarly the m-week discount bond yield is $-\frac{\ln(P_{m,t})}{m\Delta}$.

DTASTM-SGS assume that

$$r_t = \boldsymbol{\rho}' \mathbf{z}_t$$

$$\mathbf{z}_{t+1} = \mathbf{z}_t + (\boldsymbol{\mu} - \boldsymbol{\kappa}\mathbf{z}_t)\,\Delta + \mathbf{S} \cdot diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right) \boldsymbol{\xi}_{t+1}^{\mathbb{Q}} \sqrt{\Delta} + diag\left(\psi_i \left(\mathbf{S}^{(i)} \boldsymbol{\xi}_{t+1}^{\mathbb{Q}}\right)^2\right) \boldsymbol{\iota}_n \Delta \tag{1}$$

$$\mathbf{z}_{t+1} = \mathbf{z}_t + (\boldsymbol{\mu}^* - \boldsymbol{\kappa}^*\mathbf{z}_t)\,\Delta + \mathbf{S} \cdot diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right) \boldsymbol{\xi}_{t+1}^{\mathbb{P}} \sqrt{\Delta} + diag\left(\psi_i \left(\mathbf{S}^{(i)} \boldsymbol{\xi}_{t+1}^{\mathbb{P}}\right)^2\right) \boldsymbol{\iota}_n \Delta \tag{2}$$

$$\boldsymbol{\xi}_{t+1}^{\mathbb{Q}} \backsim \mathbb{N}\left(\mathbf{0}_{n\times 1}, \mathbf{I}_n\right), \quad \boldsymbol{\xi}_{t+1}^{\mathbb{P}} \backsim \mathbb{N}\left(\mathbf{0}_{n\times 1}, \mathbf{I}_n\right)$$

$$\boldsymbol{\xi}_{t+1}^{\mathbb{Q}} = \left(\xi_{1,t+1}^{\mathbb{Q}}, .., \xi_{n,t+1}^{\mathbb{Q}}\right)', \quad \boldsymbol{\xi}_{t+1}^{\mathbb{P}} = \left(\xi_{1,t+1}^{\mathbb{P}}, .., \xi_{n,t+1}^{\mathbb{P}}\right)'.$$

The vector $\mathbf{z}_t$ denotes the value of n latent stochastic factors at time t. $\boldsymbol{\rho}$ is an $n \times 1$ vector of parameters. The $\mathbf{z}$ process is specified under both the risk-neutral measure $\mathbb{Q}$ and the physical measure $\mathbb{P}$. $\boldsymbol{\kappa}$, $\boldsymbol{\kappa}^*$ and $\mathbf{S}$ are $n \times n$ matrixes of parameters. $\boldsymbol{\mu}$, $\boldsymbol{\mu}^*$ are $n \times 1$ vectors of parameters. $\boldsymbol{\kappa}$, $\boldsymbol{\mu}$ denote parameters under $\mathbb{Q}$. $\boldsymbol{\kappa}^*$, $\boldsymbol{\mu}^*$ denote parameters under $\mathbb{P}$. $diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right)$ is an $n \times n$ diagonal matrix with i-th diagonal entry equal to $\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}$. $\mathbf{h}_i$ is an $n \times 1$ vector of constants and k_i are scalar parameters for $i = 1, .., n$. $\boldsymbol{\xi}_{t+1}^{\mathbb{Q}}$ is an $n \times 1$ Gaussian vector under $\mathbb{Q}$ at time $t+1$ with mean $\mathbf{0}_{n\times 1}$ and covariance $\mathbf{I}_n$. $\mathbf{0}_{n\times 1}$ is an $n \times 1$ vector of zeros and $\mathbf{I}_n$ is the $n \times n$ identity matrix. $\xi_{i,t+1}^{\mathbb{Q}}$ for $i = 1, .., n$ are time $t+1$ values of scalar Gaussian shocks under the $\mathbb{Q}$. $\boldsymbol{\xi}_{t+1}^{\mathbb{P}}$ has similar meaning under the $\mathbb{P}$ measure. $\mathbf{S}$ is lower triangular in some models and diagonal in others. $diag\left(\psi_i \left(\mathbf{S}^{(i)} \boldsymbol{\xi}_{t+1}^{\mathbb{Q}}\right)^2\right)$ is an $n \times n$ diagonal matrix whose i-th diagonal element is $\psi_i \left(\mathbf{S}^{(i)} \boldsymbol{\xi}_{t+1}^{\mathbb{Q}}\right)^2$. ψ_i is a scalar parameter and $\mathbf{S}^{(i)}$ is the i-th row vector of $\mathbf{S}$. $\boldsymbol{\iota}_n$ is an $n \times 1$ vector whose elements are all equal to 1.

Appendix A.4 explains the market price of risk that links process 1 under $\mathbb{Q}$ and process 2 under $\mathbb{P}$. It can be shown that according to this DTATSM-SGS $P_{m,t} = \exp\left(A_m + \mathbf{B}'_m \mathbf{z}_t\right)$ where A_m is a scalar function of m and $\mathbf{B}_m$ is an $n \times 1$ vector of functions of m. An Appendix provides the Riccati difference equations for $A_m, \mathbf{B}_m$.

2.2 $\mathbb{A}_1(n)$ family of models

Some of the empirical tests below concern DTATSM-SGS whose continuous time limits are the $\mathbb{A}_1(n)$ models of Dai and Singleton (2000, 2002). We refer to these DTATSM-SGS simply as $\mathbb{A}_1(n)$ models and, when they have MFFP, they are such that: $\boldsymbol{\rho} = \mathbf{e}_n$; $\mathbf{e}_i$ is the i-th column of $\mathbf{I}_n$, therefore $r_t = z_{n,t}$; $\boldsymbol{\mu} = \mathbf{e}_2 \cdot \mu_2$ and $\boldsymbol{\mu}^* = \mathbf{e}_2 \cdot \mu_2^*$ where $\mu_2, \mu_2^* \geq 0$ are scalar **parameters**; $\mathbf{h}_i = \mathbf{e}_1$ for all i so that $z_{1,t}$ is the only factor that drives the volatility of all factors; we impose the restrictions $k_1 = \frac{\mu_1 \Delta}{1 - \kappa_{1,1}\Delta}$ and $\psi_1 = \frac{1}{4(1-\kappa_{1,1}\Delta)}$ to ensure that $z_{1,t} \geq 0$ for all t; for $i > 1$, $k_i = k_1$ and $\psi_i = 0$; **$\boldsymbol{\kappa}$ is an $n \times n$ matrix whose elements are all zero except for the elements in the main diagonal and in the main sub-diagonal;** $\kappa_{i,j}$ is the element in the i-th row and j-th column of $\boldsymbol{\kappa}$ and **therefore**

$$\boldsymbol{\kappa} = \begin{pmatrix} \kappa_{1,1} & 0 & 0 & .. & 0 \\ \kappa_{2,1} & \kappa_{2,2} & 0 & .. & 0 \\ 0 & \kappa_{3,2} & \kappa_{3,3} & .. & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & .. & \kappa_{n,n-1} & \kappa_{n,n} \end{pmatrix}; \tag{3}$$

in $\mathbb{A}_1(n)$ **models** $\kappa_{i,i} = \kappa_{2,2}$ **for** $i = 2, .., n$ and $\kappa_{2,1} = 0$, $\kappa_{i,i-1} = -\kappa_{i,i}$ for $i = 3, .., n$; $\mathbf{S} = \begin{pmatrix} s_1 & \mathbf{0}_{1\times(n-1)} \\ \mathbf{0}_{(n-1)\times 1} & s_2 \cdot \mathbf{I}_{n-1} \end{pmatrix}$ where s_2 **is a scalar and the normalisation** $s_1 = 1$ **is needed to identify the latent factor** $z_{1,t}$ **in estimation.** $\boldsymbol{\kappa}$ **features a "mean reversion chain", since factor** $z_{n,t}$ **reverts to its mean** $z_{n-1,t}$**, which in turn reverts to its mean** $z_{n-2,t}$ **and so on until** $z_{2,t}$ **reverts to its mean** μ_2**.** $z_{1,t}$ **reverts to its mean** $\frac{\psi_1}{\kappa_{1,1}}$**.** $\kappa^*_{i,j}$ **is the element in the** i**-th row and** j**-th column of** $\boldsymbol{\kappa}^*$ **and** $\kappa^*_{1,1} = \kappa_{1,1}, \kappa^*_{i,j} = \kappa_{i,j} \cdot \varphi$ **for** $i \geq 2$ **where** φ **is a scalar parameter**. This $\mathbb{A}_1(n)$ model specification proved preferable to other specifications within the same family. In the empirical tests DTASTM-SGS with MFFP are compared with "classic" three factor DTASTM-SGS. One such "classic" three factor DTATSM-SGS is $\mathbb{A}_1(3)$, which has more parameters than its counterpart models $\mathbb{A}_1(n)$ with MFFP, partly because of more general risk premia. The specification of $\mathbb{A}_1(3)$ is detailed in the Appendix.

2.3 $\mathbb{A}_n(n)$ family of models

$\mathbb{A}_1(n)$ models do not rule out negative bond yields, which may be a concern especially as we model the entire term structure of yields with maturities up to twenty years. Therefore we also test DTATSM-SGS whose factors and bond yields are always non-negative and whose continuous time limits are the $\mathbb{A}_n(n)$ models of Dai and Singleton (2000, 2002). We refer to these DTATSM-SGS as $\mathbb{A}_n(n)$ models and, in their MFFP version, they assume: $\boldsymbol{\rho} = \mathbf{e}_n$; $\boldsymbol{\mu} = \mathbf{e}_1 \cdot \mu_1$; $\boldsymbol{\mu}^* = \mathbf{e}_1 \cdot \mu_1^*$; $\mathbf{h}_i = \mathbf{e}_i$ for all i; $\mathbf{S} = s_1 \cdot \mathbf{I}_n$; moreover $\boldsymbol{\kappa}$ is given by 3 with $\kappa_{i,i} = \kappa_{1,1}$ and $\kappa_{i,i-1} = -\kappa_{i,i}$ for $i = 2, .., n$; again feedback matrix $\boldsymbol{\kappa}$ features a "**mean reversion chain**"; $\kappa^*_{i,j} = \kappa_{i,j} \cdot \varphi$ **for all** i, j; to ensure that $\mathbf{z}_t \geq \mathbf{0}_{n\times 1}$ at all times, we impose that $\mu_i, \mu_i^* \geq 0$, $\psi_i = \frac{1}{4(1-\kappa_{i,i}\Delta)}$ and $k_i = \frac{\mu_i \Delta}{1-\kappa_{i,i}\Delta}$ for all i.

In the empirical tests models $\mathbb{A}_n(n)$ with MFFP are compared with their "classic" three factor model counterpart $\mathbb{A}_3(3)$, which has more parameters and only three factors. The specification of $\mathbb{A}_3(3)$is detailed in the Appendix.

The empirical test also focus on MFFP versions and "classic" versions of affine Gausssian models, detailed in Appendix A.2, **and** of quadratic models, detailed in Appendix A.3. **Appendix** A.3 **explains that**, **unlike in affine MFFP models, in quadratic MFFP models $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}^*$ feature a cascade structure as in Calvet and others (2018).**

3 Empirical tests

The US yields were sourced from Thomson-Reuters Eikon, which provides discount factors implied by US Treasury bond prices. Such discount factors were downloaded for all yearly maturities from one year to twenty years for every Wednesday in the period from 3rd January 1995 to 29th November 2017. **This** provided 1188 weekly observations of yields. Also Calvet and others (2018) used Wednesday prices, but they used US interest rate swap rates instead of US Treasury bond yields. Table 1 presents descriptive statistics of the **US sample used to estimate and test all models**. **MFFP models** $\mathbb{A}_0(n)$, $\mathbb{A}_1(n)$ **and their "classic" counterparts** $\mathbb{A}_0(3)c$, $\mathbb{A}_1(3)c$ **were also estimated and tested using the AAA rated Euro area sovereign bonds yield curve provided by the European Central Bank. Since the Euro yield curve was often negative in recent years, models that rule out negative yields were not estimated for the Euro curve. The sample of Euro yields is made up of 688 Wednesday weekly observations from 8th September 2004 to 13th December 2017 for all yearly maturities from one year to twenty years.**

[Table 1 here]

All models are estimated through Kalman Filter or Extended Kalman Filter, as all yields are assumed to be observed with Gaussian errors, which are mutually and serially independent as well as independent of the latent factors. **For MFFP models the standard deviation of observation errors is as-**

sumed to be the same for all yield maturities. This assumption is in the spirit of parameters parsimony **of MFFP models**. **Instead for "classic" models the standard deviations of observation errors differ for each yield maturity. In spite of this advantage**, "classic" models underperform their **MFFP** counterparts, as shown below. On a practical level, **it is difficult to overstate how the small number of parameters of MFFP** models considerably simplifies the burdensome optimisation of the **likelihood function** of the Kalman Filters. It is much quicker to find the global optimum for models with **six or seven** parameters, than for models with thirty or forty parameters. As in Calvet and others (2018), **to** assess models predictive ability, we compute predictive variance (PV) as

$$PV_j = 1 - \frac{\sum_{t=1}^{N-1} \left(l_{j,t+1} - \widehat{l}_{j,t+1} \right)^2}{\sum_{t=1}^{N-1} \left(l_{j,t} - l_{j,t+1} \right)^2}.$$

$l_{j,t}$ is the continuously compounded yield observed in the t-th week of the sample for the j-year maturity, with $j = 1, .., 20$. N is the number of weeks in the sample. $\widehat{l}_{j,t+1}$ is the yield predicted by the Kalman Filter for week $t+1$ conditional on week t information, i.e. the one week ahead forecast according to any of the tested models. $(l_{j,t} - l_{j,t+1})$ is the one week ahead forecast error assuming that the j-year yield follows a random walk. PV compares model forecast errors with random walk-based forecast errors. For each model global PV is computed across all maturities as $PV = 1 - \frac{\sum_{j=1}^{20} \sum_{t=1}^{N-1} \left(l_{j,t+1} - \widehat{l}_{j,t+1} \right)^2}{\sum_{j=1}^{20} \sum_{t=1}^{N-1} \left(l_{j,t} - l_{j,t+1} \right)^2}$. Root mean squared errors (RMSE) for the j-year yield are computed as $RMSE_j = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(l_{j,t} - \widehat{l}_{j,t} \right)^2}$.

3.1 Starting values of the latent factors

In the Kalman Filter or Extended Kalman Filter for "classic" models the initial values of the latent factors are parameters to be estimated. This avoids arbitrary assumptions about a prior probability distribution for the latent factors at the start of the sample. In the Kalman Filter or Extended Kalman Filter for MFFP models the starting values of the latent factors in the first week of the sample are assumed to have zero variance and are computed as follows:

- first a "window" is generated of 200 artificial weekly yield curve observations, all equal to the yield curve observed on the first week of the sample;

- then the Kalman Filter or Extended Kalman Filter is run on the said window of 200 artificial observations using the given model and parameters; at the start of the 200 artificial observations the value of all factors is set to 0.01; the filtered mean value of the latent factors at the end of the 200 observations is set equal to the value of the latent factors in the first week of the sample.

The outcome of this method is similar to "inverting" the yields observed in the first week of the sample to determine the latent factors at that time, but the method has the advantages that it can be used whatever the number of factors, whose number can even exceed the number of observed yields, that it need not assume that some yields in the first week be perfectly observed, that it can be used for

quadratic models, not only for affine ones, that it needs no arbitrary assumptions about the latent factors prior distribution at the start of the sample, and that it requires the estimation of no extra parameters. Again this is in the spirit of parameters parsimony of MFFP models.

3.2 Sample split

The US sample is split into two periods of 594 weeks each: the first period is the "in sample" period used for parameter estimation and terminates on 31st May 2006, while the second period is the "out of sample" period. Both in sample and out of sample, every week the Kalman Filter produces one week ahead yield forecasts, but yield forecasts in the out of sample period are computed using the following rolling-window estimation: forecasts for week 595 to 891 are computed using parameters estimated in the window from week 1 to week 594; forecasts for week 892 to 1188 are computed using parameters estimated in the window from week 298 to week 891. The Euro sample is split into two halves: the in sample period, which is the only estimation window, is the first 344 weeks and the out of sample period is the last 344 weeks.

3.3 First results

Tables 2-5 report results for all the models presented above. Tables 2-3 concern MFFP models for the US, Table 4 MFFP models for the Euro, Table 5 "classic" models for the US and Euro. Table **5** reports the estimation results for **the four** "classic" models with many parameters and **three** factors, namely $\mathbb{A}_0(3)\,c$ which is an affine Gaussian model, $\mathbb{A}_3(3)\,c$ which is an affine model with stochastic volatility driven by all three factors, $\mathbb{A}_1(3)\,c$ which is an affine model with stochastic volatility driven by just one of the three factors, **and** $\mathbb{Q}(3)\,c$ which is a quadratic **model**.

The top panels in Tables 2-5 report the parameters estimated using the full US and Euro samples respectively. The stars highlight parameter estimates significant at the 1% **level when the corresponding asymptotic standard errors were estimated with the "Sandwich" estimator. In Table 5 the rows "**$x_{i,1}$ or $z_{i,1}$**" for** $i = 1, 2, 3$ **report the estimates of the starting values of the latent factors for the "classic" models. Tables 2-5 also show in sample and out of sample RMSE and PV for the one week ahead yield predictions of all models.**

3.4 The indications of predictive variance

Predictive variance (PV) provides early indications of various results. PV is negative for most models and maturities both in sample and out of sample, **both for the US and the Euro. Only some MFFP models with more than ten factors achieve Global PV close to zero. Such examples are**

models $\mathbb{A}_0(n)$ with $n \geq 10$ for US and Euro yields. Negative PV signals that current yields better predict future yields than the model. **Tables 2-4 show that the PV of MFFP models tends to increase, i.e. improve, with the number of factors n. Tables 2-5 also show that PV tends to be higher and RMSE lower**, i.e. PV and RMSE tend to be better, for **MFFP models** than for the corresponding "classic" **models**, which have **many** more parameters and only three factors. **Out of sample PV of "classic" models for one year yields is particularly disappointing**. **For most models, PV for the one year maturity tends to be worse, i.e. lower, than for the ten and twenty year maturities. This is often the case both in and out of sample and may be due to a degree of segmentation in the market for the shortest bonds. Random walk based one year yield forecasts appear more difficult for models to beat. Across models, PV for the ten maturity may be higher or lower than for the twenty year maturity, and that may differ in sample and out of sample. No clear pattern emerges for the term structure of PV for the longer maturities.**

3.4.1 Predictive variance for US yields during the "great recession"

From the last quarter of 2007 to the second quarter of 2009 the US economy experienced the "great recession". **The "great recession" is part of the US out of sample period.** Yet Tables **2-3 and Table 5** show that the Global PV of all models **is either similar or** improves during the "great recession".

The performance of no model seems negatively affected by the sharp decline in yields that occurred during the "great recession". **However this seems more due to a weakness of the random walk assumption than to merits of the models. Random walk based foreacasts cannot capture well the sharp and protracted drop of US yields, especially of one year yields, during the "great recession". This is shown by the positive PV of one year yields for most models in the Panels of Tables 2-3 dedicated to the US "great recession".**

3.5 Other measures of fit to observed yields

For "classic" models **Table 5 reports** $Avg\ h = \left(\sum_{i=1}^{20} h_i\right)/20$, which is the average of the estimated standard deviations of observation errors (h_i) across all maturities from one year to twenty years. For models with MFFP h is the same for all yield maturities and is often much smaller than $Avg\ h$ for "classic" models, **indicating that Kalman Filter** observation errors **are often** larger for "classic" models. For **MFFP models** h is often around two to four basis points, and sometimes even below one basis point. **Tables 2-4 show that** h **is mostly around one to two basis point for models** $\mathbb{A}_0(n)$, $\mathbb{A}_1(n)$, $\mathbb{Q}(n)$ **with** $n = 10, 15, 20$ **and slightly higher for** $\mathbb{A}_n(n)$ **with the same number of factors.** So low observation errors significantly alleviate the concerns raised by Adrian and others (2013) or by Golinski and Spencer(2017), who pointed to large and economically significant observation errors when estimating term structure models using yield levels rather than bond returns or yield changes.

"Avg RMSE" in Tables 2-5 denotes average RMSE across all yield maturities. For example for the US Avg RMSE in Tables 2-3 range between 12 and **16** basis points and, for any single model, are **larger** than h, whose estimates **do not exceed 11** basis points. RMSE reflect one week ahead prediction errors and **Kalman Filter** observation errors, while h only measures **Kalman Filter** observation errors. **For the Euro in sample Avg RMSE are about** 12 **basis points for "classic" models in Table 5 and about** 9 **to** 10 **basis points for MFFP models in Table 4.**

3.6 Out of sample RMSE and predictive variance of MFFP and of "classic" models

Out of sample Avg RMSE and out of sample Global PV in Tables 2-5 show that "classic" models $\mathbb{A}_0(3)\,c$, $\mathbb{A}_3(3)\,c$, $\mathbb{A}_1(3)\,c$, $\mathbb{Q}(3)\,c$ **predict out of sample US and Euro yields worse than their respective MFFP counterparts** $\mathbb{A}_0(n)$, $\mathbb{A}_n(n)$, $\mathbb{A}_1(n)$, $\mathbb{Q}(n)$ with $n \geq 5$. **The difference between models predictions seems greater out of sample than in sample and points to the risk that "classic" models may over-fit in sample due to their many parameters. This risk of over-fitting seems much lower for MFFP models, which only have six to seven parameters. Moreover "classic" models under-perform especially for the shortest yield maturities, as if three factors were not enough to model the whole yield curve from one year to twenty years at the same time. Instead MFFP models appear to better match short, medium and long term yields**

at the same time, thanks to their long "mean reversion chains" with many factors. Therefore MFFP models appear to predict yields better than "classic" models for two reasons. More factors entail more model flexibility to match yields of different maturities, and fewer parameters make MFFP models less prone to in sample over-fitting. These results are largely confirmed by the statistical tests below.

[Table 2 here]

[Table 3 here]

[Table 4 here]

[Table 5 here]

3.7 SBIC and Vuong tests, "classic" three-factor models and MFFP models, the number of factors in MFFP models

Table 6 reports Vuong tests that compare how pairs of models predict US and Euro in sample yields, when the entire US and Euro samples are used. Panels A and B in Table 6 are an indicative summary of models in sample performance. For each model Panel A shows the value of the maximised log-likelihood of the Kalman Filter lk**.** lk **always rises with the number of factors** n**, while the Schwartz Bayesian information criterion (SBIC) in Panel B always decreases, i.e. improves, with** n**. This is true for all MFFP models and for both**

US and Euro. For $n \geq 5$ the SBIC of all MFFP models is less than the SBIC of the respective classic models at the bottom of Panel B. These indications show that adding factors to MFFP models improves their predictions of in sample yields, while also enhancing their outperformance over their respective "classic" counterparts. According to SBIC in Panel B of Table 6 $A_1(20)$ best predicts in sample US and Euro yields among all tested models. **The Akaike information criterion gave the same insight and almost the same values as SBIC and therefore is not reported.**

Panels C, D, E of Table 6 present Vuong tests that compare models estimated using the whole samples of US and Euro yields. Using whole samples increases the power of the Vuong tests. The grey cells in Table 6 show individual Vuong tests that are significant after the Bonferroni correction. Table 6 presents a family of 258 Vuong tests. The Bonferroni correction implies that the level of significance of the whole family of Vuong tests does not exceed 1% if, for each individual Vuong test, the alternative (null) model is deemed to significantly outperforms the null (alternative) model when the test p value is lower than $0.005/258$ **(higher than** $0.995/258$**). Panels C, D, E in Table 6 report the p value of the Vuong likelihood ratio statistic for each test. A p value close to 1 supports the null model, while a p value close to 0 supports the alternative model.**

The Vuong tests in each cell of Panel C assume that the null model

is the n factor version of the model indicated in the respective column heading and the alternative model is the same model with $n-1$ factors; n ranges from 2 to 20. The Vuong tests in Panel C show that $\mathrm{Q}(n)$, $\mathrm{A}_0(n)$, $\mathrm{A}_1(n)$, $\mathrm{A}_n(n)$ respectively significantly outperform $\mathrm{Q}(n-1)$, $\mathrm{A}_0(n-1)$, $\mathrm{A}_1(n-1)$, $\mathrm{A}_{n-1}(n-1)$ for $n=2,..,20$. This confirms the insight from lk and SBIC in Panels A and B of Table 6. In sample yield predictions by MFFP models significantly improve as n increases. As for a given model n increases, the number of parameters remains the same. Therefore only the difference in the number of factors can explain why models with more factors outperform. This conclusion is later largely confirmed, as even out of sample yields predictions by MFFP models improve as n increases. These results support MFFP models with up to twenty factors and other unreported results support MFFP models with up to thirty factors, but in practice the computations become very burdensome with more than twenty factors, especially for quadratic models.

The Vuong tests in each cell of Panel D assume that the null model is the n factor version of the model indicated in the respective column heading and the alternative model is the "classic" model indicated in bold at the bottom of the same column of Panel D; n ranges from 1 to 20. The Vuong tests in Panel D confirm that in sample yields predicted by $\mathrm{A}_0(n)$, $\mathrm{A}_1(n)$, $\mathrm{Q}(n)$, $\mathrm{A}_5(n)$ with $n \geq 5$ respectively significantly outperform those predicted by $\mathrm{A}_0(3)\,c$, $\mathrm{A}_1(3)\,c$, $\mathrm{Q}(3)\,c$, $\mathrm{A}_3(3)\,c$.

In other words MFFP models with five or more factors predict in sample US and Euro yields significantly better than their respective three factor "classic" counterparts at the bottom of Panel D. This confirms the insight from SBIC in Panel B of Table 6.

The Vuong tests in each cell of Panel E assume that the null model is the model indicated in the respective column heading and the alternative model is the model indicated for the respective row. The two models in each cell of Panel E have the same number of factors n, with $n = 5, 10, 15, 20$. Among the four models $\mathbb{A}_n(n)$, $\mathbb{A}_0(n)$, $\mathbb{A}_1(n)$, $\mathbb{Q}(n)$, model $\mathbb{A}_n(n)$ fits US in sample yields significantly worse than the others for all the tested values of n. In $A_n(n)$ all the n factors are non-negative and drive the stochastic volatility of yields. Instead the ranking of in sample yield predictions by the other three models $\mathbb{A}_0(n)$, $\mathbb{A}_1(n)$, $\mathbb{Q}(n)$ varies with n.

The Vuong tests in Panel E show that $\mathbb{Q}(5)$ best predicts in sample US yields among five factor models. Among ten factor models, $\mathbb{Q}(10)$ and $\mathbb{A}_1(10)$ predict in sample US yields better than the other models, but neither significantly outperforms the other, while among fifteen and twenty factor models $\mathbb{A}_1(n)$ predicts in sample US yields significantly better than all the other models. In $\mathbb{A}_1(n)$ models $z_{i,t}$ is the only non-negative factor and drives the volatility of all factors. The Vuong tests in Panel E show that, while both $\mathbb{Q}(n)$ and $\mathbb{A}_n(n)$ rule out negative yields, the former clearly better predicts in sample yields

than the latter. As for the Euro, Panel E shows that $\mathbb{A}_1(n)$ predicts in sample yields significantly better than $\mathbb{A}_0(n)$, for $n = 5, 10, 15, 20$. Also for Euro yields $\mathbb{A}_1(n)$ seems preferable to $\mathbb{A}_0(n)$. This conclusion is largely confirmed by the tests for out of sample Euro yields in Table 7.

[Table 6 here]

3.8 Yields forecasting out of sample

Table 7 presents Giacomini-White (2006) tests, in short GW tests, of models out of sample forecast conditional densities of yields. The out of sample periods are week 595 to 1188 for US yields and week 345 to 688 for Euro yields. The loss function for each model in the GW statistics is the log of the (Extended) Kalman Filter one week ahead twenty variate (approximate) conditional density of yields forecast by the model. Such forecast conditional density of yields is Gaussian; in sample it coincides with the conditional quasi-likelihood function of the (Extended) Kalman Filter. The said **density** forecasts are out of sample **and use parameters estimated in sample for the Euro or estimated in the rolling windows described above for the US: the first window is the first 594 weeks and the second window is from week 298 to 891. GW tests are applicable to approximate forecast densities, as they, like Vuong tests, can test mis-specified models. The GW test is similar to the Diebold-Mariano test of forecast accuracy, but takes**

into account the fact that forecasts depend on estimated and therefore uncertain model parameters. The GW tests with logarithmic scores used in this paper are similar to the out of sample Vuong-type likelihood ratio test of Amisano and Giacomini (2007). The GW statistics shown in Table 7 for each pair of models employ an estimate of the variance of the weekly loss differential that is heteroschedasticity-and-autocorrelation-consistent (HAC) and that uses ten weekly lags.

Table 7 presents a family of 127 GW tests and their GW statistics. The GW statistic asymptotic distribution is the standard normal. The grey cells in Table 7 show the GW tests that are significant after a Bonferroni correction. The Bonferroni correction ensures that the level of significance of the whole family of GW tests does not exceed 1% and implies that, for an individual GW test, the null (alternative) model significantly outperforms the alternative (null) model if the test p value is lower than $0.005/127$ **(higher than** $0.995/127$**).**

The GW tests in each cell of Panel A test two versions, with different numbers of factors, of the model indicated for the respective column. For example for the $\mathbb{A}_n(n)$ **model the cell in the first row assumes that the alternative model is** $\mathbb{A}_3(3)$ **factors and the null model is** $\mathbb{A}_4(4)$**; the cell in the second row assumes that the alternative model is** $\mathbb{A}_4(4)$ **and the null model is** $\mathbb{A}_5(5)$**; and so on as indicated in Panel A for all rows and all models. The GW tests in each cell of Panel B assume that the alternative model is the model indicated in the re-**

spective column heading with $n = 3, 4, 5, 6, 10, 15, 20$ factors as indicated for the corresponding row, and the null model is the "classic" model in bold at the bottom of the corresponding column. The GW tests in each cell of Panel C assume that the alternative model is the model indicated for the respective column and the null model is the model indicated for the respective row. Panel C shows the p values of some of the GW statistics in Panel D. The GW tests in each cell of Panel D assume that both the indicated alternative model and null model have the same number of factors n. Large negative (positive) values of the GW statistic in Table 7 support the null (alternative) model.

3.8.1 US yields

The GW tests in Panel A of Table 7 show that out of sample US yields density forecasts by $\mathbb{A}_0(n), \mathbb{A}_1(n), \mathbb{A}_n(n), \mathbb{Q}(n)$ become significantly more accurate as n increases, with only two exceptions: the density forecast accuracy of $\mathbb{A}_{20}(20), \mathbb{Q}(20)$ does not significantly exceed that of $\mathbb{A}_{15}(15), \mathbb{Q}(15)$ respectively. Then the GW tests in Panel B of Table 7 show that the density forecasts by $\mathbb{A}_0(n), \mathbb{Q}(n)$ with $n \geq 4$ respectively outperform those by $\mathbb{A}_0(3), \mathbb{Q}(3)$ and the density forecasts by $\mathbb{A}_1(n), \mathbb{A}_n(n)$ with $n \geq 5$ respectively outperform those by $\mathbb{A}_1(3)$,$\mathbb{A}_3(3)$. MFFP models with five or more factors tend to forecast the conditional density of out of sample US yields significantly better their respective "classic" three factor counterparts.

The GW tests in Panel C of Table 7 show that, among five factor MFFP models, $\mathbb{Q}(5)$ best forecasts US yields out of sample conditional density, followed by $\mathbb{A}_0(5)$. Instead among ten, fifteen and twenty factor MFFP models, i.e. for $n = 10, 15, 20$, affine model $\mathbb{A}_1(n)$ forecasts the density of US yields significantly better than $\mathbb{A}_0(n)$ and better, but not significantly better, than $\mathbb{Q}(n)$. $\mathbb{A}_n(n)$ is the worst forecaster yet again. Panel D of Table 7 confirms that US yields density forecasts by $\mathbb{A}_0(n)$ significantly beat those by $\mathbb{A}_1(n)$ for $n \leq 5$, while the opposite is true for $n > 5$, and that $\mathbb{A}_n(n)$ is the worst out of sample forecaster of US yields.

3.8.2 Euro yields

Panel A of Table 7 shows that MFFP models $\mathbb{A}_1(n), \mathbb{A}_0(n)$ with more factors forecast out of sample Euro yields density significantly better than the same MFFP models with fewer factors. Then Panel B of Table 7 shows that for $n \geq 3$ models $\mathbb{A}_1(n), \mathbb{A}_0(n)$ forecast out sample Euro yields density significantly better than $\mathbb{A}_1(3)c, \mathbb{A}_0(3)c$ respectively. Again MFFP models forecast out of sample Euro yields density significantly better than corresponding "classic" models. The GW tests in Panels C and D of Table 7 show that for $n > 3$ model $\mathbb{A}_1(n)$ forecasts out sample Euro yields density significantly better than $\mathbb{A}_0(n)$, while the oppose it true for $n = 3$. In particular $\mathbb{A}_1(20)$ outperforms $\mathbb{A}_0(20)$. This out of sample evidence from Euro yields largely confirms the in

sample evidence as well as results from US yields.

3.8.3 Conclusions from out of sample forecasts of yields density

Overall the tests of out of sample forecasts of yields conditional density in Table 7 largely agree with the in sample tests in Table 6. MFFP models outperform "classic" models; adding factors tends to improve the performance of MFFP models; while $\mathbb{A}_1(n)$ is often the preferable MFFP model, especially for $n \geq 10$, for US yields it does not dominate $\mathbb{Q}(n)$; $\mathbb{Q}(5)$ seems the best among five factor MFFP models for the US. While Calvet and others (2018) showed the good performance of affine Gaussian models with MFFP that are similar to $\mathbb{A}_0(n)$, MFFP affine models with stochastic volatility $\mathbb{A}_1(n)$ and MFFP quadratic models $\mathbb{Q}(n)$ can predict in and out of sample yields even better.

[Table **7** here]

3.9 Yields volatilities and correlations

MFFP models are also more accurate than their "classic" counterparts in matching the volatilities of and the correlations between changes in **observed** yields of different maturities. **This is the case for both US and Euro yields.** Yields are observed weekly in our sample and therefore **the** yield changes **we consider** are weekly changes, **whose volatility and correlations Table 8 reports for the different maturities**. **The correlations are between changes in the one year yield and changes in yields of all**

other maturities. The two top panels in Table 8 refer to US yields and the **two bottom panels refer** to Euro yields. The first rows **in grey in each panel** show **volatilities and correlations of changes in observed yields**, while the other rows in each panel show **volatilities and correlations of changes in model predicted yields. The yield changes span the whole US and Euro samples respectively and the model yields are those predicted one week ahead. The relevant models** are those estimated using the full **samples of US and Euro yields. The first and the third panels of Table 8 show that, for both the US and Euro, changes in the one year observed yields** are little correlated with changes in long maturity yields and the said correlation decreases for the longer maturities. The correlations between changes in yields predicted by "classic" models are shown in the **grey rows at the bottom of each panel** and tend to be too high **both for US and Euro short maturity yields and too low for the longest Euro maturities**. The three factors of "classic" models entail too little independence between changes in yields of different maturities. This shortcoming seems overcome by MFFP models with **five or more** factors in the mean reversion chain, **as correlations in changes in their predicted yields match the US and Euro yields change correlations more closely**. **The second and fourth panels of Table 8 also show that MFFP models match the volatilities of changes in shorter term yields more accurately than their corresponding "classic" models. Again MFFP models appear more flexible, thanks to their many factors, than "classic" models in**

matching observed changes in yields of different maturities.

[Table **8** here]

4 Conclusion

Calvet and others (2018) showed that affine Gaussian term structure models **with up to ten factors and as few as six parameters** predict swap rates very accurately. This paper shows **that some** affine models with stochastic volatility **and quadratic models with up to twenty factors and six to seven parameters can** predict US Treasury **yields** and Euro sovereign yields **even more accurately than affine Gaussian models with the same number of factors and parameters**. **Because these models feature many factors and few parameters (MFFP), they predict** US Treasury and Euro sovereign yields better than their corresponding "classic" versions **with only** three factors and tens of parameters. **This is the case both in sample and out of sample and even if the affine MFFP models lack the cascade parameter structure of Calvet and others (2018)**. **MFFP models also fit the volatility of yields changes and the correlations between yields changes of different maturities better than "classic" models. MFFP models outperform for two reasons**. **Fewer parameters entail less in sample over-fitting and many factors linked through a mean reversion chain** give models more flexibility to match yields of different maturities. **Adding factors to a model in this way improves its yields predictions**

in sample and usually also out of sample. For US yields, quadratic models seem preferable, both in sample and out of sample, among MFFP models with five factors, while among MFFP models with more than five factors affine models with stochastic volatility of the type $\mathbb{A}_1(n)$ outperform homoschedatic models $\mathbb{A}_0(n)$ with the same number of factors; models $\mathbb{A}_1(n)$ significantly outperform quadratic models with the same number of factors only in sample and only when they feature fifteen to twenty factors. For Euro yields, affine models $\mathbb{A}_1(n)$ seem preferable, both in sample and out of sample, among MFFP models with five or more factors.

A Appendix

A.1 "Classic" models

A.1.1 "Classic" three factor model $\mathbb{A}_1(3)$

For example $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}^*$ are different in $\mathbb{A}_1(3)$. Model $\mathbb{A}_1(3)$ is such that

$$
\begin{aligned}
r_t &= z_{3,t} \\
z_{1,t+1} &= z_{1,t} + (\mu_1 - \kappa_{1,1} z_{1,t})\,\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta} + \psi_1\left(\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta}\right)^2 \\
z_{2,t+1} &= z_{2,t} + (\mu_2 - \kappa_{2,2} z_{2,t})\,\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(2)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta} \\
z_{3,t+1} &= z_{3,t} + \kappa_{3,3}\,(z_{2,t} - z_{3,t})\,\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(3)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta} \\
z_{1,t+1} &= z_{1,t} + \left(\mu^*_1 - \kappa^*_{1,1} z_{1,t}\right)\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta} + \psi_1\left(\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta}\right)^2 \\
z_{2,t+1} &= z_{2,t} + \left(\mu^*_2 - \kappa^*_{2,2} z_{2,t}\right)\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(2)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta} \\
z_{3,t+1} &= z_{3,t} + \kappa^*_{3,3}\,(z_{2,t} - z_{3,t})\,\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(3)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta}
\end{aligned}
$$

$$
\mathbf{S} = diag\,(s_1, s_2, s_3)\cdot\boldsymbol{\Upsilon}, \quad \boldsymbol{\Upsilon} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho^2_{12}} & 0 \\ \rho_{13} & \frac{\rho_{32}-\rho_{12}\cdot\rho_{13}}{\sqrt{1-\rho^2_{12}}} & \sqrt{1-\rho^2_{13}-\frac{(\rho_{32}-\rho_{12}\cdot\rho_{13})^2}{1-\rho^2_{12}}} \end{pmatrix}. \tag{4}
$$

$diag\,(s_1, s_2, s_3)$ is a 3×3 diagonal matrix whose three diagonal elements are the parameters s_1, s_2, s_3. The parameter ρ_{12} is the conditional correlation between $\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}$ and $\mathbf{S}^{(2)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}$. ρ_{13} and ρ_{32} have similar interpretation. All $\kappa_{i,i}$ and $\kappa^*_{i,i}$ are independent parameters. The conditions $k_1 = \frac{\mu_1\Delta}{1-\kappa_{1,1}\Delta}$ and $\psi_1 = \frac{1}{4(1-\kappa_{1,1}\Delta)}$ again guarantee that $z_{1,t}$ be non-negative at all times.

A.1.2 "Classic" three factor model $\mathbb{A}_3(3)$

Model $\mathbb{A}_3(3)$assumes

$$
\begin{aligned}
r_t &= z_{3,t} \\
z_{1,t+1} &= z_{1,t} + (\mu_1 - \kappa_{1,1} z_{1,t})\,\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta} + \psi_1\left(\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta}\right)^2 \\
z_{2,t+1} &= z_{2,t} + (\mu_2 + \kappa_{2,1} z_{1,t} - \kappa_{2,2} z_{2,t})\,\Delta + \sqrt{k_2 + z_{2,t}}\mathbf{S}^{(2)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta} + \psi_2\left(\mathbf{S}^{(2)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta}\right)^2 \\
z_{3,t+1} &= z_{3,t} + (\mu_3 + \kappa_{3,3}\,(z_{1,t} + z_{2,t} - z_{3,t}))\,\Delta + \sqrt{k_3 + z_{3,t}}\mathbf{S}^{(3)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta} + \psi_3\left(\mathbf{S}^{(3)}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta}\right)^2 \\
z_{1,t+1} &= z_{1,t} + \left(\mu^*_1 - \kappa^*_{1,1} z_{1,t}\right)\Delta + \sqrt{k_1 + z_{1,t}}\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta} + \psi_1\left(\mathbf{S}^{(1)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta}\right)^2 \\
z_{2,t+1} &= z_{2,t} + \left(\mu^*_2 + \kappa^*_{2,1} z_{1,t} - \kappa^*_{2,2} z_{2,t}\right)\Delta + \sqrt{k_2 + z_{2,t}}\mathbf{S}^{(2)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta} + \psi_2\left(\mathbf{S}^{(2)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta}\right)^2 \\
z_{3,t+1} &= z_{3,t} + \left(\mu^*_3 + \kappa^*_{3,3}\,(z_{1,t} + z_{2,t} - z_{3,t})\right)\Delta + \sqrt{k_3 + z_{3,t}}\mathbf{S}^{(3)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta} + \psi_3\left(\mathbf{S}^{(3)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta}\right)^2
\end{aligned}
$$

with $\mathbf{S} = diag(s_1, s_2, s_3)$. All $\kappa_{i,i}$ and $\kappa^*_{i,i}$ are independent parameters and $\kappa^*_{2,1} = \kappa_{2,1} \geq 0$. Again $\psi_i = \frac{1}{4(1-\kappa_{i,i}\Delta)}$ for $i = 1, 2, 3$ to ensure that $z_{1,t}, z_{2,t}, z_{3,t}$ are non-negative for all t.

A.2 Affine Gaussian models $\mathbb{A}_0(n)$

The empirical tests also concern affine Gaussian models $\mathbb{A}_0(n)$ with MFFP, which are special cases of the above DTASTM-SGS whereby: $\boldsymbol{\rho} = \mathbf{e}_n$, so that $r_t = x_{n,t}$; $k_i = 1$, $\mathbf{h}_i = \mathbf{0}_{n\times 1}$, $\psi_i = 0$ for all i, so that $diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right) = \mathbf{I}_n$ and $diag\left(\psi_i\left(\mathbf{S}^{(i)}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\right)^2\right) = \mathbf{0}_{n\times n}$; $\boldsymbol{\mu} = \mu_1 \cdot \mathbf{e}_1$, $\boldsymbol{\mu}^* = \mu^*_1 \cdot \mathbf{e}_1$, $\mathbf{S} = s_1 \cdot \mathbf{I}_n$, $\boldsymbol{\kappa}$ is given in 3 and with $\kappa_{i,i} = \kappa_{1,1}$ for $i = 2, .., n$ and $\kappa_{i+1,i} = -\kappa_{i+1,i+1}$ for $i = 1, .., n-1$; therefore **$\boldsymbol{\kappa}$ implies a "mean reversion chain"**. $\kappa^*_{i,j} = \kappa_{i,j} \cdot \varphi$ **for all** i, j. Then the price of a discount bond is still $P_{t,m} = e^{A_m + \mathbf{B}_m'\mathbf{x}_t}$ and the Riccati equations in Appendix A.7 determine A_m and $\mathbf{B}_m$. **The** "classic"

version of affine Gaussian models that is tested below is $\mathbb{A}_0(3)$, with $\mathbf{S}$ given by 4 and with $\kappa^*_{1,1} \neq \kappa^*_{2,2} \neq \kappa^*_{3,3} \neq \kappa_{1,1} \neq \kappa_{2,2} \neq \kappa_{3,3}$. Therefore $\mathbb{A}_0(3)$ only has three factors, but more parameters than $\mathbb{A}_0(n)$, partly because of more general risk premia and partly because of correlated shocks to factors.

A.3 Discrete time quadratic models

Quadratic Gaussian models $\mathbb{Q}(n)$ assume

$$r_t = \mathbf{x}_t' \boldsymbol{\Theta} \mathbf{x}_t$$

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, .., x_{n,t})'$$

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \Delta(\boldsymbol{\mu} - \boldsymbol{\kappa}\mathbf{x}_t) + \mathbf{S}\boldsymbol{\xi}^{\mathbb{Q}}_{t+1}\sqrt{\Delta} \tag{5}$$

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \Delta(\boldsymbol{\mu}^* - \boldsymbol{\kappa}^*\mathbf{x}_t) + \mathbf{S}\boldsymbol{\xi}^{\mathbb{P}}_{t+1}\sqrt{\Delta} \tag{6}$$

$$\boldsymbol{\xi}^{\mathbb{Q}}_{t+1} \backsim \mathbb{N}(\mathbf{0}_n, \mathbf{I}_n), \quad \boldsymbol{\xi}^{\mathbb{P}}_{t+1} \backsim \mathbb{N}(\mathbf{0}_n, \mathbf{I}_n).$$

$\mathbf{x}_t$ are n latent stochastic factors at time t that follow a Gaussian vector autoregressive process under $\mathbb{P}$ and $\mathbb{Q}$. **All** shocks are serially and mutually **independent**. $\boldsymbol{\Theta}$ is an $n \times n$ symmetric matrix**.** The conditional covariance of $\mathbf{x}_{t+1}$ is $\mathbf{S}\mathbf{S}'\Delta$. It can be shown that, according to $\mathbb{Q}(n)$, $P_{m,t} = \exp(\mathfrak{A}_m + \mathfrak{B}'_m\mathbf{x}_t + \mathbf{x}'_t\mathfrak{C}_m\mathbf{x}_t)$ and that $\mathfrak{A}_m, \mathfrak{B}_m, \mathfrak{C}_m$ **are** functions of m that satisfy Riccati difference equations shown in Appendix A.8. $\mathfrak{A}_m$ **is a** scalar. $\mathfrak{B}_m$ is an $n \times 1$ vector. $\mathfrak{C}_m$ is an $n \times n$ symmetric matrix. The process $\mathbf{x}$ is mean-reverting under both $\mathbb{Q}$ and $\mathbb{P}$ as long as all the eigenvalues of $(\mathbf{I}_n - \Delta\boldsymbol{\kappa})$ and of $(\mathbf{I}_n - \Delta\boldsymbol{\kappa}^*)$ are smaller than 1 in absolute value. We assume this condition.

The quadratic models $\mathbb{Q}(n)$ with MFFP that are tested in this paper as-

sume that: $\boldsymbol{\mu} = \mu_1 \cdot \mathbf{e}_1$, $\boldsymbol{\mu}^* = \mu_1^* \cdot \mathbf{e}_1$, $\mathbf{S} = s_1 \cdot \mathbf{I}_n$, $\boldsymbol{\Theta} = \mathbf{e}_n \cdot \mathbf{e}_n'$; **$\boldsymbol{\kappa}$ is as in 3 with** $\kappa_{i,i} = \kappa_{1,1} \cdot (1+\delta)^{i-1}$ **and** $\kappa_{i,i-1} = -\kappa_{i,i}$ **for** $i = 2, .., n$. **δ is a parameter to be estimated and** $\kappa_{i,j}^* = \kappa_{i,j}$ **for all** i, j. **In quadratic models $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}^*$ feature a cascade structure as in Calvet and others (2018) and** $r_t = x_{n,t}^2$. Quadratic models imply that $P_{t,m} = e^{\mathfrak{A}_m + \mathfrak{B}_m' \mathbf{x}_t + \mathbf{x}_t' \mathfrak{C}_m \mathbf{x}_t}$ where $\mathfrak{A}_m, \mathfrak{B}_m, \mathfrak{C}_m$ are given in Appendix A.8. **The** empirical tests concern models $\mathbb{Q}(5), \mathbb{Q}(10), \mathbb{Q}(5)\, s, \mathbb{Q}(10)\, s$. The "classic" version of quadratic Gaussian models that is tested below is $\mathbb{Q}(3)$, but with $\kappa_{i,i}^* \neq \kappa_{i,i}$ for $i = 1, .., n$, with all $\kappa_{i,i}$ and $\kappa_{i,i}^*$ independent parameters, and with $\mathbf{S}$ given by 4. Thus $\mathbb{Q}(3)$has only three factors but more parameters than $\mathbb{Q}(n)$ with **MFFP**.

A.4 The stochastic discount factor and the link between the processes of equations 1 and 2

Given the $\mathbf{z}$ process under $\mathbb{Q}$ in equation 1 we derive the $\mathbf{z}$ process under $\mathbb{P}$. The time t stochastic discount factor $\mathbb{M}_t$ is such that

$$\begin{aligned} \mathbb{M}_{t+1} = \ & \mathbb{M}_t \cdot e^{-r_t \Delta} \cdot e^{-\frac{1}{2} \boldsymbol{\Lambda}_t' \boldsymbol{\Lambda}_t \Delta - \boldsymbol{\Lambda}_t' \boldsymbol{\xi}_{t+1}^{\mathbb{P}} \sqrt{\Delta}} \\ \boldsymbol{\Lambda}_t = & \, diag\left(\left(k_i + \mathbf{h}_i' \mathbf{z}_t \right)^{-1/2} \right) \left(\mathbf{g}_2 + \mathbf{G}_3 \cdot \mathbf{z}_t \right). \end{aligned}$$

$\mathbf{g}_1, \mathbf{g}_2$ are an $n \times 1$ vectors of parameters, $\mathbf{G}_3$ is an $n \times n$ matrix of parameters and

$$\boldsymbol{\xi}_{t+1}^{\mathbb{Q}} = \boldsymbol{\xi}_{t+1}^{\mathbb{P}} + \boldsymbol{\Lambda}_t \sqrt{\Delta}. \tag{7}$$

It can be shown that, if the lower bound of the process $\mathbf{z}$ under $\mathbb{Q}$ is $\mathbf{0}_{n\times 1}$ (this is the case under conditions provided in the text), $\mathbf{0}_{n\times 1}$ is also the lower bound

of $\mathbf{z}$ under $\mathbb{P}$, irrespective of $\mathbf{\Lambda}_t$. Given the $\mathbf{z}$ process under $\mathbb{Q}$ in equation 1, substituting for $\boldsymbol{\xi}_{t+1}^{\mathbb{Q}}$ according to 7 in equation 1 implies that the $\mathbf{z}$ process under $\mathbb{P}$ is

$$\mathbf{z}_{t+1} = \mathbf{z}_t + (\boldsymbol{\mu} - \boldsymbol{\kappa}\mathbf{z}_t)\,\Delta + \mathbf{S}\cdot diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right)\mathbf{\Lambda}_t\Delta + diag\left(\mathbf{\Lambda}_t'\mathbf{S}'\mathbf{\Psi}_i\mathbf{S}\mathbf{\Lambda}_t\Delta\right)\boldsymbol{\iota}_n\Delta + \tag{8}$$

$$+diag\left(2\mathbf{\Lambda}_t'\mathbf{S}'\mathbf{\Psi}_i\mathbf{S}\boldsymbol{\xi}_{t+1}^{\mathbb{P}}\sqrt{\Delta}\right)\boldsymbol{\iota}_n\Delta + \mathbf{S}\cdot diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right)\boldsymbol{\xi}_{t+1}^{\mathbb{P}}\sqrt{\Delta} + diag\left(\boldsymbol{\xi}_{t+1}^{\mathbb{P}\prime}\mathbf{S}'\mathbf{\Psi}_i\mathbf{S}\boldsymbol{\xi}_{t+1}^{\mathbb{P}}\right)\boldsymbol{\iota}_n\Delta.$$

$\Psi_i = \mathbf{e}_i\cdot\mathbf{e}_i'\cdot\psi_i$, where $\mathbf{e}_i$ is an $n\times 1$ vector whose only non-zero element is the i-th element which is 1. Therefore all elements of Ψ_i are zero except for the i-th diagonal element which is ψ_i. It follows that $diag\left(\boldsymbol{\xi}_{t+1}^{\mathbb{Q}\prime}\mathbf{S}_i'\mathbf{\Psi}_i\mathbf{S}\boldsymbol{\xi}_{t+1}^{\mathbb{Q}}\right) = diag\left(\psi_i\left(\mathbf{S}^{(i)}\boldsymbol{\xi}_{t+1}^{\mathbb{Q}}\right)^2\right)$. The terms in 8 that are proportional to Δ^2 and $\Delta^{\frac{3}{2}}$ vanish in continuous time as $\Delta\to 0$ and, with about 52 weeks in one year and $\Delta = \frac{1}{52}$, those same terms may be omitted in estimation with little loss in accuracy, giving

$$\mathbf{z}_{t+1} \simeq \mathbf{z}_t + (\boldsymbol{\mu} - \boldsymbol{\kappa}\mathbf{z}_t)\,\Delta + \mathbf{S}\cdot diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right)\mathbf{\Lambda}_t\Delta + \mathbf{S}\cdot diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right)\boldsymbol{\xi}_{t+1}^{\mathbb{P}}\sqrt{\Delta} +$$

$$+diag\left(\boldsymbol{\xi}_{t+1}^{\mathbb{P}\prime}\mathbf{S}'\mathbf{\Psi}_i\mathbf{S}\boldsymbol{\xi}_{t+1}^{\mathbb{P}}\right)\boldsymbol{\iota}_n\Delta$$

$$E_t^{\mathbb{P}}\left[\mathbf{z}_{t+1}\right] \backsimeq \boldsymbol{\mu}\Delta + (\mathbf{I}_n - \boldsymbol{\kappa}\Delta)\,\mathbf{z}_t + \mathbf{S}\cdot diag\left(\sqrt{k_i + \mathbf{h}_i'\mathbf{z}_t}\right)\mathbf{\Lambda}_t\Delta + diag\left(\mathrm{tr}\left(\mathbf{S}'\mathbf{\Psi}_i\mathbf{S}\right)\right)\boldsymbol{\iota}_n\Delta$$

$$Cov_t^{\mathbb{P}}\left[\mathbf{z}_{t+1}\right] \backsimeq \mathbf{S}\cdot diag\left(k_i + \mathbf{h}_i'\mathbf{z}_t\right)\cdot\mathbf{S}'\cdot\Delta$$

where $E_t^{\mathbb{P}}\left[..\right]$ and $Cov_t^{\mathbb{P}}\left[..\right]$ are respectively the time t one period conditional expectation and covariance under the $\mathbb{P}$ measure, which enter the Kalman filter. According to these approximations $\mathbf{\Lambda}_t$ affects $E_t^{\mathbb{P}}\left[\mathbf{z}_{t+1}\right]$ but not $Cov_t^{\mathbb{P}}\left[\mathbf{z}_{t+1}\right]$ and the $\mathbf{z}$ process of equation 2 under $\mathbb{P}$ is consistent with the $\mathbf{z}$ process of equation 1 under $\mathbb{Q}$ **if** $\boldsymbol{\mu} + \mathbf{S}\cdot\mathbf{g}_2 = \boldsymbol{\mu}^*$ **and** $\boldsymbol{\kappa} - \mathbf{S}\mathbf{G}_3 = \boldsymbol{\kappa}^*$. This paper uses Kalman Filter quasi-maximum-likelihood estimation, which needs the first two conditional

moments of factors and yields, and such moments are known in closed form for DTATSM-SGS.

A.5 DTASTM-SGS $\mathbb{A}_n(n)$

DTASTM-SGS $\mathbb{A}_n(n)$ are such that $r_t = z_{n,t}$ and

$$A_m = A_{m-1} + \mathbf{B}'_{m-1}\boldsymbol{\mu}\Delta + \ln\left(abs\left(|\boldsymbol{\Gamma}|\right)\right) + \frac{1}{2}\Delta\sum_{i=1}^{n}\frac{B_{i,m-1}^2 s_{i,i}^2 k_i}{1-2s_{i,i}^2\Delta B_{i,m-1}\psi_i}$$

$$\mathbf{B}'_m = -\boldsymbol{\rho}'\Delta + \mathbf{B}'_{m-1}\left(\mathbf{I}_n - \boldsymbol{\kappa}\Delta\right) + \frac{1}{2}\Delta\sum_{i=1}^{n}\frac{B_{i,m-1}^2 s_{i,i}^2\cdot\mathbf{h}'_i}{1-2s_{i,i}^2\Delta B_{i,m-1}\psi_i}$$

$$\boldsymbol{\Gamma} = \left(\mathbf{I}_n - 2\cdot\mathbf{S}'\left(\sum_{i=1}^{n}B_{i,m-1}\boldsymbol{\Psi}_i\Delta\right)\mathbf{S}\right)^{-1/2}$$

$$\boldsymbol{\Psi}_i = \mathbf{e}_i\cdot\mathbf{e}'_i\cdot\psi_i$$

$$\rho_0 = 0,\ \boldsymbol{\rho}_1 = \mathbf{e}_n,\ \mathbf{h}_i = \mathbf{e}_i,\ \boldsymbol{\mu}\geq\mathbf{0}_{n\times 1},\ \kappa_{i,j\neq i}\leq 0 \text{ for } i,j = 1,..,n$$

$$\psi_i = \frac{1}{4\left(1-\kappa_{i,i}\Delta\right)},\quad k_i = \frac{\mu_i\Delta}{1-\kappa_{i,i}\Delta}.$$

A.6 DTASTM-SGS $\mathbb{A}_1(n)$

DTASTM-SGS $\mathbb{A}_1(n)$ are such that $r_t = z_{n,t}$, $\mathbf{S}$ is lower triangular, $k_i = k_1, \mathbf{h}_i = \mathbf{e}_1$ for $i = 1,..,n$ and

$$A_m = A_{m-1} + \mathbf{B}'_m\boldsymbol{\mu}\Delta + \ln\left(abs\left(|\boldsymbol{\Gamma}|\right)\right) + \frac{1}{2}\mathbf{B}'_{m-1}\mathbf{S}\left(\mathbf{I}_n - 2\cdot\mathbf{S}'\left(\sum_{i=1}^{n}B_{i,m-1}\boldsymbol{\Psi}_i\Delta\right)\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{B}_{m-1}k_1\Delta$$

$$\mathbf{B}'_m = -\boldsymbol{\rho}'\Delta + \mathbf{B}'_m\left(\mathbf{I}_n - \boldsymbol{\kappa}\Delta\right) + \frac{1}{2}\mathbf{B}'_{m-1}\mathbf{S}\left(\mathbf{I}_n - 2\cdot\mathbf{S}'\left(\sum_{i=1}^{n}B_{i,m-1}\boldsymbol{\Psi}_i\Delta\right)\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{B}'_{m-1}\mathbf{e}'_1\Delta.$$

A.7 DTASTM-SGS $\mathbb{A}_0(n)$

Affine Gaussian models $\mathbb{A}_0(n)$ are special cases of DTASTM-SGS whereby $r_t = \boldsymbol{\rho}'\mathbf{z}_t$, $k_i = 1$, $\mathbf{h}_i = \mathbf{0}_{n\times 1}$, $\psi_i = 0$ for all i, so that

$$A_m = A_{m-1} + \mathbf{B}'_m \boldsymbol{\mu}\Delta + \frac{1}{2}\mathbf{B}'_{m-1}\mathbf{S}\mathbf{S}'\mathbf{B}_{m-1}\Delta$$

$$\mathbf{B}'_m = -\boldsymbol{\rho}'\Delta + \mathbf{B}'_m\left(\mathbf{I}_n - \boldsymbol{\kappa}\Delta\right)$$

$$A_0 = 0, \quad \mathbf{B}_0 = \mathbf{0}_{n\times 1}.$$

A.8 Gaussian quadratic models

This Appendix presents the discrete time quadratic model $\mathbb{Q}(n)$ tested in the paper. According to the quadratic model $P_{m,t} = \exp\left(\mathfrak{A}_m + \mathfrak{B}'_m\mathbf{x}_t + \mathbf{x}'_t\mathfrak{C}_m\mathbf{x}_t\right)$ and

$$\mathfrak{A}_m = \mathfrak{A}_{m-1} + \mathfrak{B}'_{m-1}\boldsymbol{\mu}\Delta + \Delta^2\boldsymbol{\mu}'\mathfrak{C}_{m-1}\boldsymbol{\mu} + \ln\frac{|\boldsymbol{\gamma}|}{abs\left|\sqrt{\Delta}\mathbf{S}\right|} + \frac{1}{2}\mathbf{q}'_{m-1}\boldsymbol{\gamma}\boldsymbol{\gamma}'\mathbf{q}_{m-1}$$

$$\mathfrak{B}'_m = \mathbf{q}'_{m-1}\left(\mathbf{I}_n + 2\boldsymbol{\gamma}\boldsymbol{\gamma}'\mathfrak{C}_{m-1}\right)\left(\mathbf{I}_n - \boldsymbol{\kappa}\Delta\right)$$

$$\mathfrak{C}_m = -\boldsymbol{\Psi}\Delta + \left(\mathbf{I}_n - \boldsymbol{\kappa}\Delta\right)'\mathfrak{C}_{m-1}\left(\mathbf{I}_n + 2\boldsymbol{\gamma}\boldsymbol{\gamma}'\mathfrak{C}'_{m-1}\right)\left(\mathbf{I}_n - \boldsymbol{\kappa}\Delta\right).$$

$$\mathbf{q}'_{m-1} = \mathfrak{B}'_{m-1} + 2\Delta\boldsymbol{\mu}'\mathfrak{C}_{m-1}, \quad \boldsymbol{\gamma} = \left(\left(\Delta\mathbf{S}\mathbf{S}'\right)^{-1} - 2\mathfrak{C}_{m-1}\right)^{-1/2}$$

$$\mathfrak{A}_0 = 0, \quad \mathfrak{B}_0 = \mathbf{0}_{n\times 1}, \quad \mathfrak{C}_0 = \mathbf{0}_{n\times n}.$$

$\mathbf{0}_{n\times n}$ is an $n \times n$ matrix of zeros.

References

[1] Amisano G. and Giacomini R. 2007, "Comparing Density Forecasts via Weighted Likelihood Ratio Tests", Journal of Business & Economic Statistics 25, n. 2, 177-190.

[2] Adrian T., Crump R.K., Moench E., 2013, "Pricing the term structure with linear regressions", Journal of Financial Economics, vol. 110, n. 1, 110-138.

[3] Calvet L.E, Fisher A.J. and Wu L., 2018, "Staying on Top of the Curve: A Cascade Model of Term Structure Dynamics", Journal of Financial and Quantitative Analysis vol. 53, n. 2, 937–963.

[4] Dai Q. and Singleton K., 2000, "Specification analysis of affine term structure models", The Journal of Finance 55, n.5, 1943-1978.

[5] Dai Q. and Singleton K., 2002, "Expectation puzzles, time-varying risk premia, and affine models of the term structure", Journal of Financial Economics 63, 415–441.

[6] Duffee, G.R., 2002, "Term premia and interest rate forecasts in affine models", Journal of finance 57, 405–443.

[7] Giacomini R. and White H., 2006, "Tests of conditional predictive ability", Econometrica 74, n. 6, 1545-1578.

[8] Golinski A. and Spencer P., 2017, "The advantages of using excess returns to model the term structure", Journal of Financial Economics 125, 163–181.

[9] Le A., Singleton K. and Dai Q., 2010, "Discrete-time Affine Term Structure Models with Generalized Market Prices of Risk", Review of Financial Studies 23, 2184-227.

[10] Realdon M., 2018, "Discrete time affine term structure models with squared Gaussian shocks (DTATSM-SGS)", unpublished manuscript of Swansea University (UK).