AARHUS
UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Interest Rates with Long Memory: A Generalized Affine Term-Structure Model

Daniela Osterrieder

CREATES Research Paper 2013-17

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Interest Rates with Long Memory: A Generalized Affine Term-Structure Model

Daniela Osterrieder[‡]

May 30, 2013

Abstract: We propose a model for the term structure of interest rates that is a generalization of the discrete-time, Gaussian, affine yield-curve model. Compared to standard affine models, our model allows for general linear dynamics in the vector of state variables. In an application to real yields of U.S. government bonds, we model the time series of the state vector by means of a co-fractional vector autoregressive model. The implication is that yields of all maturities exhibit nonstationary, yet mean-reverting, long-memory behavior of the order $d \approx 0.87$. The long-run dynamics of the state vector are driven by a level, a slope, and a curvature factor that arise naturally from the co-fractional modeling framework. We show that implied yields match the level and the variability of yields well over time. Studying the out-of-sample forecasting accuracy of our model, we find that our model results in good yield forecasts that outperform several benchmark models, especially at long forecasting horizons.

Keywords: term structure of interest rates, fractional integration and cointegration, affine models

JEL codes: G12, C32, C58

[*]I acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

[‡]CREATES and Aarhus University, School of Business and Social Sciences, Department of Economics and Business, Fuglesangs Allé 4, 8210 Aarhus V, Denmark.
Email: dosterrieder@creates.au.dk

1 Introduction

One of the salient properties of short-term interest rates is their near unit-root behavior. Figure 1 plots the U.S. one-month government bond rate rate over the past 40 years, starting in January 1970. The time series is characterized by long and persistent upward and downward trends. Typically, stationary autoregressive/moving-average (ARMA) time-series models within the $I(0)$ class cannot replicate such a dependence structure as they impose an exponential decay of autocorrelations. The slow yet steady decay of autocorrelation estimates over time (see e.g. Table 1) casts doubt that standard nonstationary $I(1)$ models such as the random walk present a more accurate modeling framework. One possibility for modeling such strongly dependent interest rates is the near-integrated approach, where a statistical uncertainty is attributed to the order of integration of the series (see, for instance, Jardet et al., 2013). Long-memory models, where the rate of decay of a shock is described by a single fractional integration parameter, are a parsimonious alternative. In this work, we estimate the order of fractional integration, d, of the real monthly short-term interest rate and find that the process is nonstationary yet mean-reverting, with $d \approx 0.87$. Using different sample periods, different data frequencies, and different estimators, many other works suggest a very similar estimator, such as Gil-Alana (2004), Gil-Alana and Moreno (2012), Iacone (2009), Shea (1991), Schotman et al. (2008), Sun and Phillips (2004).

The short-term interest rate is one of the main ingredients to most term-structure models in the literature. Under the expectations hypothesis (EH), yields can be described by the cumulative future expected short rate and potentially a constant term premium. Thus, the dynamics of the short rate directly affect the volatility of the implied yield curve. For instance Cochrane (2005) shows that a stationary AR model for the short-term interest rate produces standard deviations of yields that are always smaller than the volatility of the short-term interest rate. In fact, model-implied standard deviations decay exponentially as maturity increases. A random-walk model for the short rate, in contrast, leads to a flat volatility structure over maturities. Backus and Zin (1993) show that both models violate the empirically observed slow decay of yield variation as time to maturity increases. They provide evidence that the observed shape can instead be captured by allowing for fractional integration in the factor that drives yields. The same issue arises in affine term-structure models due to Duffie and Kan (1996), where the yield of a bond is assumed linear in a (possibly multivariate) state vector. Arbitrage-free yields are derived as expectations of the stochastic discount factor, which again depends on the short rate (see e.g. Duffee, 2002, and Gürkaynak and Wright, 2012). It follows that a stationary VAR model for the state vector, as is typically used in the literature (see, for instance, Ang and Piazzesi, 2003), will

produce too little volatility at the long end of the yield curve. An example of this is the benchmark model of Piazzesi and Schneider (2006).

The dynamic properties of the short rate and potentially other state variables are hence not only important in modeling their time series, but also in fitting the cross section of yields. The apparent incompatibility of the time-series properties of observed state variables with the dynamics required to match the cross section of yields over different maturities is a well documented stylized fact of term-structure models. According to Singleton (2006), this is one of the criticisms of yield-curve models that is quoted most widely. For example, Brown and Schaefer (1994) find that the strength of factor mean reversion that is necessary to capture the yield curve at a point in time is not consistent with the empirical dynamics of the series. We suggest that explicitly accounting for long memory can provide a solution to this conundrum. This insight makes it necessary to alter existing term-structure models in such a way that they allow for a fractionally integrated specification of the factors that drive the yield curve over time. Consequently, we define a model for zero-coupon government bond yields that is a generalization of the class of discrete-time, Gaussian, affine term-structure models of Duffie and Kan (1996) and Dai and Singleton (2000). Our model implies that bond prices are defined as the sum of an initial (possibly time-varying) value, plus the sum of current and past innovations to the state vector, scaled by term-structure loadings, plus a term premium. Our model retains the attractive property of affine models, in that we find a closed-form solution.

We model the state variables as a trivariate vector of yields, composed of the short rate, an intermediate rate, and a long rate. Our term-structure model implies that yields across all maturities inherit the persistence of the short-term interest rate. As we have evidence that the short rate has a fractional order of integration, we model the state vector by a multivariate long-memory model with low-frequency dynamics described by d. We further detect fractional cointegration between the three yields, similarly to Nielsen (2010) and Chen and Hurvich (2003a,b). As a consequence, we model the trivariate state vector by a co-fractional model due to Johansen (2008a,b). The model implies that the low-frequency dynamics of the short rate, and consequently of yields at all maturities, are driven by one variable with nonstationary long memory that has the natural interpretation of a *level* factor, and two stationary linear combinations of yields, which can be viewed as the *slope* and *curvature* factor. The seminal principal component analysis of yields up to a maturity of 30 years of Litterman and Scheinkman (1991) reveals that the first three factors are necessary to capture the comovements in U.S. treasury yields. The first principal component, or *level* factor, covers parallel shifts of the yield curve and, depending on the sampling

period and frequency, can account for up to 89% of the total explained variation. The second component, the *slope* factor, explains shifts in the slope of the yield curve, and the third component, the *curvature* (or *hump*) factor, captures the curvature. In our framework these traditional yield-curve factors arise naturally from the co-fractional relationship between yields.

We find that the co-fractional model fits the dynamics of the state vector very well, and we show that the term-structure loadings resulting from this model are very intuitive. In the time-series dimension, the *level*-factor loadings that scale state-vector shocks decay only very slowly, whereas the corresponding loadings for the *slope*-factor and *curvature*-factor decline faster. In the cross-sectional dimension, i.e. across maturities, the first factor loading is almost flat, the second one is curved, and the third loading exhibits a hump at short maturities. The resulting yields are persistent and match the observed term structure well in sample, especially for maturities above three years. More importantly, we can successfully forecast yields out of sample. Contrasting our forecasts to benchmark models such as for instance Diebold and Li (2006), we find that our model produces superior forecasts at long horizons up to ten years. Accurate long-horizon forecasts can be viewed as especially important for insurance companies and pension funds. Yet, also central banks and treasuries rely on long-run predictions.

Recently, there has been an increasing amount of research in the term-structure literature reaching the conclusion that the term premium, or risk premium, is rather stable and that the expectations component of yield-curve models captures most of the variation in bond yields. An example is Bauer (2011), who finds that imposing restrictions on risk prices that are supported by the data leads to a model that is very close to the EH. Similar evidence favoring that most volatility is generated from the expectations component and only little from the term premium, is provided by Jardet et al. (2013) and Kim and Orphanides (2007, 2012). The method of Kozicki and Tinsley (2005), based on investors' uncertainty about the long-run level of inflation, provides a further argument in favor of constant term premiums. Motivated by these findings, we model and estimate term premiums as depending only on maturity[1]. As we find a good in-sample and out-of-sample fit of yields under this modeling assumption, we view our work as additional empirical support for the conclusions above.

Strictly speaking, assuming that the term premium is time invariant may be too restrictive. When investigating the predictability of excess bond returns, this can typically not

[1]We model term premiums as constants; nevertheless the theoretical affine model that we propose could easily be extended to allow for time-varying term premiums.

be rejected, which implies that term premiums must be time varying (see e.g. Fama and Bliss, 1987, and Cochrane and Piazzesi, 2005, 2008). While this empirical evidence has also been challenged by, for instance, peso-problem arguments, or persistent-variable explanations (for an overview, see e.g. Gürkaynak and Wright, 2012)[2], some doubt regarding the time invariance of term premiums remains. In the context of this discussion, it should be noted that we estimate our model on real yields of nominal bonds, however. Inflation is hence implicitly included as an additional term-structure factor in our estimations. As we model the yield-curve factors in a (fractional) error-correction model, the long-run level of inflation and deviations from it in the short run are implicitly captured. This is very closely related to the approach of Kozicki and Tinsley (2001a,b), who argue that we can move back to EH if we account for shifts in the perception of the long-run inflation level. Our model does not allow for shifts in expectations; yet we allow for a possibly very slow adjustment to the long-run value[3]. Other evidence emphasizing the important role of inflation as a term-structure factor is provided by e.g. Rudebusch and Wu (2008) and Diebold et al. (2006).

The rest of this article is structured as follows. In Section 2, we define a generalized Gaussian affine model for the term structure of interest rates, and derive closed-form solutions for the parameters. Section 3 summarizes the time-series evidence in favor of long memory in the state vector and presents the results of the co-fractional model. In Section 4 we discuss the fit of model-implied yields in sample, and in Section 5 we show the results of the out-of-sample forecasts. Section 6 concludes.

2 Yield-curve model

We propose a term-structure model that is a generalization of the class of discrete-time affine yield curve models originally developed by Duffie and Kan (1996) and further specified by Dai and Singleton (2000). Let $y_t^{(n)}$ denote the yield to maturity of a zero-coupon bond with maturity n, where $t = 1, 2, \ldots, T$ and $n = 1, 2, \ldots, N$. Denote the one-period logarithmic short rate by r_t and note that bond yields are related to prices by $y_t^{(n)} = -(1/n)p_t^{(n)}$. The term-structure of interest rates model that we propose here has the solution for log prices

$$p_t^{(n)} = -p_{0,t}^{(n)} - \sum_{j=0}^{t-1} \left(U_j^{(n)} + V_j^{(n)'} \epsilon_{t-j} \right) - W^{(n)}, \tag{1}$$

[2]For instance Kim and Orphanides (2007) and Sack (2006) hint at the possibility that the predictability of excess returns in Cochrane and Piazzesi (2005) may be overstated.

[3]For a discussion of the indistinguishability of long-memory series and data generated by level shifts in finite sample, see e.g. Diebold and Inoue (2001).

where ϵ_t is an innovation process of size $K \times 1$ that follows and independent and identical normal distribution $(0, \Sigma)$. The term-structure parameters $U_j^{(n)}$ and $W^{(n)}$ are scalar values, and $V_j^{(n)}$ is a vector of size $K \times 1$. $p_{0,t}^{(n)}$ denotes a function of the initial values of the price of an n-period bond.

Model (1) collapses to a traditional Gaussian affine discrete-time yield-curve model of the form $y_t^{(n)} = A^{(n)} + B^{(n)'} x_t$, if the $K \times 1$-state vector x_t has a vector autoregressive representation of the order q, VAR(q). If $q = 1$ with coefficient matrix D and intercept term C, the usual affine-model parameters relate to the parameters in (1) by $p_{0,t}^{(n)} = nB^{(n)'} D^t x_0$, $U_j^{(n)} = nB^{(n)'} D^j C$, $V_j^{(n)} = nB^{(n)'} D^j$, and $W^{(n)} = nA^{(n)}$. Our model allows for more general dynamics assuming that the state vector is generated by the linear moving average (MA) specification

$$x_t = x_t^0 + \sum_{j=0}^{t-1} \Phi_j \left(\epsilon_{t-j} + \mu \right), \tag{2}$$

where Φ_j is a $K \times K$-matrix and μ a $K \times 1$ vector. x_t^0 is a function of the initial values of x_t. In what follows, we will derive (1) for different modeling assumptions about the coefficients Φ_j.

Let the short rate be the third element of the state vector such that it holds that $r_t = e3' x_t$, where $e3$ is the third unit vector, $[0,\ 0,\ 1]'$. Define the logarithmic stochastic discount factor $m_{t+1} = \ln M_{t+1}$ as

$$m_{t+1} = -r_t - \tfrac{1}{2} \lambda' \Sigma \lambda + \lambda' \epsilon_{t+1}, \tag{3}$$

where λ is the market price of risk and has size $K \times 1$.

We obtain solution (1) by deriving bond prices as the expectation of the product of future stochastic discount factors over different horizons. The basic pricing equation for zero-coupon bonds is

$$P_t^{(n+1)} = \mathsf{E}_t \left(M_{t+1} P_{t+1}^{(n)} \right), \tag{4}$$

and in logarithmic form

$$p_t^{(n+1)} = -r_t + \mathsf{E}_t \left(p_{t+1}^{(n)} \right) + \tfrac{1}{2} \mathsf{Var}_t \left(p_{t+1}^{(n)} \right) + \mathsf{Cov}_t \left(m_{t+1}, p_{t+1}^{(n)} \right). \tag{5}$$

The price of a bond with maturity zero is equal to one, $P_t^{(0)} = 1$ and $p_t^{(0)} = 0$. Then, from (5) it follows that $p_t^{(1)} = -r_t$, because $r_t = -\mathsf{E}_t(m_{t+1})$ for $n = 1$.

Equations (3)-(5) describe bonds that have a riskless payoff in real terms. $P_t^{(n)}$ and $p_t^{(n)}$ are real prices of real bonds with maturity n and m_t is the real logarithmic stochastic dis-

count factor. In this study we focus on nominal bonds and model their prices and yields in real terms. If we assume that the stochastic discount factor and inflation are independent, Campbell et al. (1997) show that the real price of a nominal bond equals the real price of a real bond times the expected future real value of money. Under the assumption that inflation risk is not priced in our model, Equations (3)-(5) hold for real prices and yields of nominal bonds.

If (1) is the solution to the pricing equation for bonds, it must hold that the elements of (5) are equal to

$$\mathsf{E}_t\left(p_{t+1}^{(n)}\right) = -p_{0,t+1}^{(n)} - \sum_{j=0}^{t-1} U_{j+1}^{(n)} - \sum_{j=0}^{t-1} V_{j+1}^{(n)\prime}\epsilon_{t-j} - \left(W^{(n)} + U_0^{(n)}\right) \tag{6}$$

$$\mathsf{Var}_t\left(p_{t+1}^{(n)}\right) = V_0^{(n)\prime}\Sigma V_0^{(n)} \tag{7}$$

$$\mathsf{Cov}_t\left(m_{t+1}, p_{t+1}^{(n)}\right) = -V_0^{(n)\prime}\Sigma\lambda. \tag{8}$$

These solutions are derived in Appendix A.

Plugging the elements (6)-(8) into the recursion for bond prices (5), and relying on the inital condition $p_t^{(1)} = -r_t$, we thus find the following expressions for the term-structure parameters.

$$p_{0,t}^{(n)} = \sum_{i=0}^{n-1} e3'x_{t+i}^0 \tag{9}$$

$$U_j^{(n)} = \sum_{i=0}^{n-1} e3'\Phi_{j+i}\mu \tag{10}$$

$$V_j^{(n)} = \sum_{i=0}^{n-1} \Phi_{j+i}' e3 \tag{11}$$

$$W^{(n)} = \sum_{i=1}^{n-1}\left(\sum_{k=0}^{i-1} e3'\Phi_k\mu - \frac{1}{2}\left[\sum_{k=0}^{i-1}\Phi_k' e3\right]'\Sigma\left[\sum_{k=0}^{i-1}\Phi_k' e3\right] + \left[\sum_{k=0}^{i-1}\Phi_k' e3\right]'\Sigma\lambda\right). \tag{12}$$

Our term-structure model (1) permits a closed-formed solution for all parameters that describe bond prices. In addition, the parameters are fully determined by the data-generating process of the state vector x_t. The only exception is $W^{(n)}$, which depends on the market price of risk.

Our term-structure model implies that the yield on a bond with maturity n is the average conditionally expected value of the sum of n successive investments in a one-period

bond plus a constant, which is a function of the market price of risk. More precisely, yields are determined by

$$y_t^{(n)} = \frac{1}{n}\sum_{i=0}^{n-1} e3' x_{t+i}^0 + e3' \sum_{j=0}^{t-1} \frac{1}{n}\sum_{i=0}^{n-1} \Phi_{j+i}\left(\epsilon_{t-j} + \mu\right) + f\left(\lambda\right), \tag{13}$$

where $f(\lambda) = \frac{1}{n} W^{(n)}$. Equation (13) suggests that the dynamics of yields reflect the time-series properties of the short-term interest rate. To see this, assume for simplicity that x_t contains only r_t, and r_t is a univariate fractional white noise process of order d. We can write coefficients $\Phi_j = (-1)^j \binom{-d}{j}$ and we know that this can be approximated by $j^{d-1}/\Gamma(d)$ as $j \to \infty$ (see, for instance, Robinson, 2003b). Hence

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} \Phi_{j+i} &\approx \frac{1}{n}\sum_{i=0}^{n-1} \frac{(j+i)^{d-1}}{\Gamma(d)} \\
&\approx \frac{j^{d-1}}{\Gamma(d)} \frac{1}{n}\sum_{i=0}^{n-1} \left(1 + \frac{i}{j}\right)^{d-1} \\
&\approx \frac{j^{d-1}}{\Gamma(d)} \text{ as } j \to \infty,
\end{aligned} \tag{14}$$

implying that $y_t^{(n)}$ and r_t have the same slow speed of decay of news impacts. Alternatively, if the short rate has short memory captured by an AR(1) model with coefficient matrix c, it holds that $\frac{1}{n}\sum_{i=0}^{n-1} \Phi_{j+i} = \frac{1}{n}\frac{1-c^n}{1-c} c^j$. As in the long-memory case, the yield process inherits the dynamic behavior of r_t and its MA coefficients decay at exponential rate. In the general case, where x_t is multivariate, the low-frequency dynamics of yields are determined by the low-frequency dynamics of r_t. That is, if the time-series of short rates is integrated of the order $I(d)$, our term-structure model implies that yields of all maturities of integrated of the same order d.

3 State-vector dynamics

Our term-structure model implies that the dynamics of yields at all maturities are almost completely determined by the expectations of the short rate, r_t. It follows that the closest approximation to the actual data-generating process of the short rate and the resulting conditional expectations can be expected to result in the best fit for yields. We assume that the dynamics of r_t can be described by its own past realizations and by the history of two other yields, one at the short end of the maturity spectrum, $y_t^{(24)}$, and one at the long end of the yield curve, $y_t^{(120)}$. Thus, the state vector $x_t = [y_t^{(24)},\ y_t^{(120)},\ r_t]'$ is trivariate with $K = 3$, where both t and n are denoted in monthly frequencies.

In estimating term-structure models, it is common to rely on Fama and Bliss (1987)-type data for yields (see e.g. Cochrane and Piazzesi, 2005, and Diebold and Li, 2006). We obtain continuously compounded U.S. government zero-coupon bond yields computed as unsmoothed Fama-Bliss rates from Hillebrand et al. (2012). The data is observed monthly from January 1970 to December 2009 and spans 50 different maturities of one month to ten years[4]. We transform the data into real yields, using CPI data from the Bureau of Labor Statistics. The short rate, r_t, is proxied by the real one-month yield, $y_t^{(24)}$ and $y_t^{(120)}$ are two-year real yields and ten-year real yields, respectively. Table 1 presents summary statistics of the elements of the state vector x_t, and for real yields with maturity n =12, 36, 60, and 84 months.

The time-series averages of yields indicate that the term-structure of interest rates has had a *normal* shape over the 40 sample years, that is the yield curve has been upward sloping. The standard deviation of yields decays very slowly for $n \geq 12$ as time to maturity increases, which is in line with Backus and Zin (1993) and provides a *cross-sectional motivation* for considering fractionally integrated models. The slowly decreasing autocorrelations of the elements of x_t in Table 1 provide a *time-series motivation* for a model that permits persistent state-vector dynamics. A shock to the short rate dies out at a very slow rate with almost 50% of the effect remaining in the system after two years. Based on these stylized facts, we allow our state variables r_t, $y_t^{(24)}$, and $y_t^{(120)}$ to have a long-memory data generating process.

3.1 Low-frequency dynamics

As a first step, we estimate the univariate low-frequency dynamics of the short rate r_t, as these determine the long-run dependence in yields by (13). We assume that the series r_t is a fractionally integrated process, defined as

$$(1-L)^d r_t = u_t, \tag{15}$$

where u_t is assumed to be stationary $I(0)$ with zero mean and spectral density $f_u(\lambda)$ satisfying $f_u(\lambda) \sim \tau$ for frequencies $\lambda \to 0+$, where $0+$ implies that the limit is approached from above. τ is assumed a positive constant. L denotes the usual lag operator and d is the fractional integration parameter. The fractional filter $(1-L)^d$ is defined as the infinite sum $(1-L)^d = \Delta^d = \sum_{i=0}^{\infty} \Theta_i L^i$, with $\Theta_i = (-1)^i \binom{d}{i}$.

[4]Hillebrand et al. (2012) do not compute higher-order yields beyond the maturity horizon of ten years as this would involve bonds that are not actively traded. Further details concerning the computation of Fama-Bliss rates can be found in their paper.

For $d = 0$, the series follows a stationary $I(0)$ process. If $d = 1$, the series is nonstationary with a unit root. In general, a process with $d \in (-\frac{1}{2}, \frac{1}{2})$ is stationary. In line with Robinson (2003a), we say that r_t is nonstationary if $d \geq \frac{1}{2}$, yet mean-reverting for $d \in [\frac{1}{2}, 1)$[5].

It is common to rely on semiparametric techniques for the estimation of the long-memory parameter, d, given that the short-memory structure of the data is usually not known a priori. The exact local Whittle estimator (ELW) of Shimotsu and Phillips (2005) is well suited for our estimation as it is generally robust and efficient (Henry and Zaffaroni, 2002) and asymptotically normally distributed for any value of d, with $\sqrt{m}(\hat{d} - d) \sim N(0, \frac{1}{4})$. m is the size of the spectral window and its value determines the speed of convergence. Some academic work has been devoted to the determination of an optimal bandwidth in spectral analysis (see e.g. Henry and Robinson, 1996, and Henry, 2001), but a unanimous conclusion is still absent. In line with Schotman et al. (2008), we therefore rely on a simple rule of thumb and let $m = T^o$, where $o = 0.45$.

The resulting point estimate for the fractional differencing parameter of r_t, d_r, is equal to 0.8620. The short rate exhibits nonstationary long memory behavior, yet remains mean reverting. This is in line with previous findings by Shea (1991), Sun and Phillips (2004), and Schotman et al. (2008), among others. According to our term-structure model (13), the substantial persistence in the short rate implies that long-term bond yields asymptotically exhibit the same slow decay of shocks to the process and are, thus, also nonstationary. This coincides with the existing empirical evidence, as discussed by Diebold and Li (2006), for instance. The standard error of d_r is 0.1213. A standard t-test reveals that we clearly reject the null hypothesis that d is equal to zero. Using a one-sided t-test, we can further reject that $d_r = \frac{1}{2}$ in favor of the alternative $d_r > \frac{1}{2}$. Given the fairly narrow bandwidth that we choose for the estimation, the t-test for the hypothesis that $d_r = 1$ against $d_r < 1$ cannot be rejected at the 5% level, however.

Our theoretical term-structure of interest rates model implies that yields of all maturities have equal integration order. We assess this implication for the elements of the state vector x_t by a Wald test. Testing the equality of $d_{y^{(24)}} = d_{y^{(120)}}$ results in a Wald statistic of 0.1356, which is smaller than the 5% $\chi^2(1)$ critical value of 3.841. Similarly, the Wald statistic for $H_0 : d_{y^{(24)}} = d_r$ is 0.0845, and for $H_0 : d_{y^{(120)}} = d_r$ is 0.0060. Hence, we fail to reject the equality of all fractional integration orders, which provides evidence in favor of

[5]The term *mean reversion* has to be interpreted with care, here. In case of $d > \frac{1}{2}$, the unconditional mean is not defined, yet the conditional mean (such as $\mathsf{E}_0(r_t)$) does exist.

the restrictions imposed by the term-structure model (1)-(13).

3.2 Co-fractional VAR estimation

The state vector x_t is fractionally integrated with $d_x \approx 0.8620$, according to the univariate estimates. In addition, $y_t^{(120)}$, $y_t^{(24)}$, and r_t can be expected to cointegrate fractionally[6]. Our term-structure model implies that yields of all maturities are driven by the same process r_t. This implies that yields with different maturities cannot be independent stochastic trends. Rather, we expect yields to cointegrate with one common stochastic trend. Therefore, we model the trivariate process x_t by a co-fractional VAR due to Johansen (2008a,b). The co-fractional VAR, $\text{CFVAR}_d(q)$, is defined as

$$\Delta^d x_t = \gamma \left[\delta' \left(1 - \Delta^d \right) x_t + \rho' \right] + \sum_{i=1}^{q} \Gamma_i \Delta^d \left(1 - \Delta^d \right)^i x_t + \epsilon_t, \tag{16}$$

where ϵ_t is independently and identically distributed with $(0, \Sigma)$. γ is the adjustment vector of size $3 \times s$, where $s = 2$ is the cointegrating rank, and δ is the cointegrating vector of same size. ρ is a restricted constant of size $1 \times s$. Johansen and Nielsen (2012) show that we can estimate model (16) by maximum likelihood (ML)[7].

Before estimating model (16), we have to determine the number of lags q. To that end, we estimate the model for several different lags lengths under the hypothesis of full rank $s = K$ as is commonly done in the traditional $I(0)/I(1)$ cointegration literature. The Bayesian-Schwartz information criterion (BIC) is minimized for $q = 1$. We condition the estimation of the trivariate process on the first ten initial values of x_t to avoid inaccuracies due to the application of the truncated fractional filter. Table 2 summarizes our ML estimation results. The corresponding 95% confidence intervals are obtained from a wild bootstrap as we cannot exclude heteroskedasticity in the residuals[8]. The multivariate estimate for the long-run persistence is $d = 0.8871$. Hence, this parametric estimate is very close to the univariate semiparametric estimate from the previous section. The estimate is statistically larger than $\frac{1}{2}$ and smaller than one, again suggesting that yields are nonstationary long-memory processes. The adjustment parameters for the short-rate equation, $e3'\gamma$, are both positive, whereas in the equation for the long-term yields, $y_t^{(120)}$, they are both negative. $y_t^{(24)}$ adjusts positively to the first cointegrating vector and negatively to the sec-

[6] For a derivation of term-structure cointegration implications in a traditional $I(0)/I(1)$ framework, see e.g. Hall et al. (1992) and Engsted and Tanggaard (1994). An extension to fractional and co-fractional processes is discussed in Nielsen (2010).

[7] The corresponding Matlab code has been provided by Nielsen and Morin (2012).

[8] For a treatment of the wild bootstrap, see e.g. Davidson and MacKinnon (2007), Davidson and Flachaire (2008), and Cavaliere et al. (2010a,b).

ond. The reported R^2 values suggest that the model captures the dynamics of x_t very well[9].

The normalized the cointegrating vector is given by

$$\delta = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ \delta_{(31)} & \delta_{(32)} \end{pmatrix}. \tag{17}$$

Let $a_{(ij)}$ denote the element in row i and column j of any vector or matrix a. We can write the equation for the short rate that is implied by the CFVAR$_d(1)$ as

$$\begin{aligned} r_t &= \left(1-\Delta^d\right)\left\{r_t + \gamma_{(31)}\left(2y_t^{(24)} - y_t^{(120)} + \delta_{(31)}r_t + \rho_{(11)}\right) + \gamma_{(32)}\left(y_t^{(120)} + \delta_{(32)}r_t + \rho_{(12)}\right)\right\} \\ &\quad + e3'\Gamma_1\Delta^d\left(1-\Delta^d\right)x_t + e3'\epsilon_t. \end{aligned} \tag{18}$$

Equation (18) implies that the long-run dynamics of the short rate are driven by three terms. Firstly, the infinite history of the variable itself, since $(1-\Delta^d) = -\sum_{i=1}^{\infty}\Theta_i L^i$. This component is very persistent, i.e. it is integrated of the order $\hat{d} = 0.8871$. Thus, it can be interpreted as a *level* factor, with a factor loading of 1. Secondly, the short rate is determined by the entire history an $I(0)$ linear combination of the three elements of the state vector $2y_t^{(24)} - y_t^{(120)} + \hat{\delta}_{(31)}r_t$, plus a constant. If $\delta_{(31)}$ were equal to -1, the interpretation of this second driver of r_t would be that of a *curvature* factor, as in Diebold and Li (2006), with a factor loading of $\gamma_{(31)}$. We estimate $\delta_{(31)}$ to be equal to -1.2369. Hence, the value is close to -1 and it is statistically indistinguishable from -1. The final driver of low-frequency dynamics of the short rate is a zero-mean stationary linear combination of the long rate, $y_t^{(120)}$, and r_t. Our estimate for $\delta_{(32)}$ is equal to -1.0076, and we cannot reject that it is statistically equal to -1 at a 95% confidence level. We conclude that this third driver resembles the traditional *slope* factor with a factor loading of $\gamma_{(32)}$.

It is well known that the *level* factor is highly persistent. For instance, Ang and Piazzesi (2003) find that the first yield-curve factor has a VAR(1) coefficient estimate of above 0.99. In our model, it exhibits nonstationary long memory. The *slope* factor is persistent, as well. Among others, Diebold et al. (2006) find that the impact of shocks on the second term-structure factor declines only slowly over time. Considering the autocorrelations in Table 1, we find the same results for the *slope* vector resulting from the co-fractional VAR. The first three autocorrelation estimates are fairly high; yet after the first three lags they decay rapidly as expected for an $I(0)$ process. The *curvature* factor is the least persistent of the three as is typically found in the literature.

[9]Note that R^2 values have to be interpreted with care here, however, since x_t is nonstationary. Thus, as $T \to \infty$, it holds that $R^2 \to 1$.

In estimating model (16), we made two a priori assumptions. Firstly, we assumed that the three elements of x_t share one common trend, which implies a cointegration rank of $s = 2$. We can check this assumption by computing the likelihood ratio (LR) statistic of Johansen and Nielsen (2012) that provides a test for the null hypothesis that $s \leq 2$ against the alternative that $s \leq p$. The resulting estimate of the LR statistic is 1.1475. The asymptotic distribution of the LR statistic is non-standard and depends on fractional type-II Brownian motions. Yet, MacKinnon and Nielsen (2012) tabulate the corresponding numerical critical value, which is 8.7616 at a 5% significance level. We strongly fail to reject that the cointegrating rank is two or smaller.

Secondly, we imposed the restriction that the strength of the cointegrating relation between the elements of x_t, b, is equal to the order of fractional integration d in the co-fractional model. Stating this differently, we restricted the cointegrating relations, $\delta' x_t$, in (16) to $I(0)$ processes. We can test this assumption in several ways. First, we re-estimate a generalized version of the co-fractional VAR model (16) by maximum likelihood without imposing the restriction that $d = b$ [10]. The resulting estimates are $\hat{d}$ =0.8769 and $\hat{b}$ =0.8769. As they are identical, this lends support to the imposed restriction. A second test relies on the fact that if we estimated a model with $d = b$, but the true b were unequal to d, the remaining low-frequency dynamics would appear in the residuals of the regression. We estimate the integration order of our estimated residuals from (16) and check whether they are $I(0)$. The ELW estimates (using the same bandwidth size as above) are 0.1492, -0.1066, and -0.1234 for the residuals of the equations for $y_t^{(24)}$, $y_t^{(120)}$, and r_t respectively. The corresponding standard error is 0.1213, which implies that we fail to reject that the residuals are integrated of the order zero. Finally, a more general misspecification test is examining the residuals for serial correlation. A standard Lagrange-Muliplier (LM) test results in p-values 0.1163, 0.6428, and 0.7373 for the three residual series. We have no evidence of serial correlation in ϵ_t up to the eighth lag. However, we discover heteroskedasticity in the residuals of the CFVAR$_d$(1); this motivates the application of the wild bootstrap.

3.3 Impulse responses of the co-fractional VAR model

To compute yields according to our term-structure model (1)-(13), we need the initial values x_t^0 of the state vector as well as the impulse responses Φ_j. We invert the co-fractional model (16) to find a MA solution for x_t given by

$$x_t = -\Xi_+^{-1}(L)\Xi_-(L)x_t + \Xi_+^{-1}(L)\left(\epsilon_t + \gamma\rho'\right), \tag{19}$$

[10] Note that we cannot estimate ρ, if $d \neq b$, as the parameter is not identified.

where $\Xi_+(L) = \sum_{j=0}^{t-1} \Xi_j L^j$ and $\Xi_-(L) = \sum_{j=t}^{\infty} \Xi_j L^j$, as shown by Johansen and Nielsen (2012). For the CFVAR$_d$(1) model, the elements of the AR polynomial $\Xi(L)$ are given by $\Xi_0 = I_{K\times K}$, and $\Xi_j = (I_{K\times K} + \gamma\delta')\,\Theta_j + \Gamma_1 \sum_{k=0}^{j-1} \Theta_k \Theta_{j-k}\ \forall j \geq 1$.

Solution (19) is only defined, if the conditions for inversions are satisfied. Johansen and Nielsen (2012) show that it must hold that $|\gamma_\perp'\,(I_{K\times K} - \Gamma_1)\,\delta_\perp| \neq 0$. In our estimation, this value is equal to 3.6145, and thus different from zero. Secondly, the roots z of the characteristic polynomial $|(1-z)I_{K\times K} - \gamma\delta' z - \Gamma_1(1-z)z| = 0$ must be either equal to one or outside a complex disk $\mathbb{C}_{max(d,1)}$. Figure 2 shows that all roots fulfill that condition, and hence our model has solution (19).

As a final step, we re-write the solution (19) in the form of the state-vector representation (2) in the previous section. μ is equal to $\gamma\rho'$. We find that the MA coefficients are given by the following recursion

$$\Phi_j = -\sum_{k=0}^{j-1} \Phi_k \Xi_{j-k}\ \forall\ j \geq 1, \tag{20}$$

with initial condition $\Phi_0 = I_{K\times K}$. The initial values, x_t^0, are given by

$$x_t^0 = \sum_{j=0}^{\infty} \sum_{k=0}^{t-1} \Phi_k \Xi_{t+j-k} L^{t+j} x_t. \tag{21}$$

All solutions are derived in Appendix B. The expression for the initial values, (21), involves an infinite sum. In computing initial values, we truncate this initial sum after 10 observations using the first 10 observations on x_t that we conditioned on in the ML estimation to compute x_t^0.

Solution (19) allows us to analyze the conditional correlations between the three factors that we identified as drivers of the low-frequency dynamics of the short-term interest rate, r_t. We can denote the *level* factor as $e3'x_t$, and the *slope* and *curvature* factor as $\delta' x_t + \rho'$ (see Equation (18)). The conditional (co-)variance between the factors that describe long-run short-rate movements, that is $\mathsf{Var}_t([x_{t+1}' e3,\ x_{t+1}'\delta + \rho]')$, is given by

$$\begin{aligned}
&\mathsf{E}_t\left[\begin{pmatrix} e3'x_{t+1} \\ \delta' x_{t+1} + \rho' \end{pmatrix} - \mathsf{E}_t\begin{pmatrix} e3'x_{t+1} \\ \delta' x_{t+1} + \rho' \end{pmatrix}\right]\left[\begin{pmatrix} e3'x_{t+1} \\ \delta' x_{t+1} + \rho' \end{pmatrix} - \mathsf{E}_t\begin{pmatrix} e3'x_{t+1} \\ \delta' x_{t+1} + \rho' \end{pmatrix}\right]' \\
&= \begin{pmatrix} e3'\Sigma e3 & e3'\Sigma\delta \\ \delta'\Sigma e3 & \delta'\Sigma\delta \end{pmatrix}.
\end{aligned} \tag{22}$$

The conditional correlation between the short rate and the *slope* factor is equal to -0.8633, and the corresponding value for short-rate and *curvature*-factor correlation is -0.5827. The second and third factor are strongly negatively correlated with the first factor, that is with the long-run *level* factor. Hence, the term-structure factors in our model are related in a way that is consistent with the results in Dai and Singleton (2000), who find that conditional factor correlations should be negative. The conditional correlation among the second and the third factor is positive, however.

4 Results

4.1 Term-structure factor loadings

The term-structure model (1) suggests bond prices are a function of the residuals of the CFVAR$_d$(1) model scaled by loadings $-V_j^{(n)}$. By the inverse relation between prices and yields it holds that the loading for yields are equal to $(1/n)V_j^{(n)}$. The raw residuals capture shocks to the state variables $y_t^{(24)}$, $y_t^{(120)}$, and r_t, respectively. To interpret the shocks as shocks to the *level*, *slope*, and *curvature* factor instead, we rely on the permanent-transitory shock decomposition of Gonzalo and Ng (2001). That is, we re-parametrize shocks ϵ_t from the CFVAR$_d$(1) model by multiplying with a vector G, given by

$$G = \begin{pmatrix} \gamma_\perp' \\ \delta' \end{pmatrix}, \tag{23}$$

where $\gamma_\perp$ has size $(K \times (K-s))$ and is defined by $\gamma_\perp'\gamma = 0$. The re-scaled shocks are $\tilde{\epsilon}_t = G\epsilon_t$. The interpretation of $\tilde{\epsilon}_t$ is that the first element is a permanent[11] *level* shock, whereas the second and third element are transitory shocks to the *curvature* factor and the *slope* factor. The corresponding impulse responses are $\tilde{\Phi}_j = \Phi_j G^{-1}$. This decomposition is not unique as it depends on the normalization of $\gamma_\perp$. Yet, we can compute unique orthogonalized shocks with covariance matrix $I_{K\times K}$. Let H be the Cholesky decomposition of $\tilde{\epsilon}_t$. Then $\bar{\epsilon}_t = H^{-1}\tilde{\epsilon}_t$ are orthogonalized shocks with MA parameters $\bar{\Phi}_j = \tilde{\Phi}_j H$. For these, we compute the corresponding yield loadings as $\bar{V}_j^{(n)} = (1/n)\sum_{i=0}^{n-1} \bar{\Phi}_{j+i}' e3$.

Figures 3, 4, and 5 plot these orthogonalized loadings on yield-curve factor shocks. Figure 3 shows that the loadings on the *level*-factor shock decay only very slowly, both over time and over maturity. It also demonstrates that the further we go back in the history of the process, the slower is the rate of decay over maturities n. The rate of decay of the loadings over lags j does not change much over different maturities n, however. As expected, shocks

[11] Strictly speaking, the first shock $\gamma_\perp'\epsilon$ in our setup is not permanent. Its impact decays very slowly over time, however, exhibiting nonstationary long-memory behavior.

to the *level* factor have a very persistent effect on yields.

Figure 4 plots the effect of a *slope* shock on yields. By definition, the contemporary effect of a shock to the *slope*, or to the term spread, on the short-term interest rate (with maturity $n = 1$) is negative. Beyond $n = 1$, the effect is slowly upward sloping for increasing maturities, such that the shape of $\bar{V}_0^{(n)}$ very much resembles the shape of the average yield curve. That is, the effect marginally diminishes (in absolute value), as n increases. Just like for *level* shocks, we observed that the impacts of *slope* shocks exhibit a slower decay rate over n, as the lags j increase. In contrast to shocks to the *level* factor, we observe that the effect decays relatively quickly over time, however. The same is true for the *curvature*-shock effect, plotted in Figure 5. Again, the impact is negative for $j = 0$ and $n = 1$, by definition. The usual initial 'hump' that the curvature factor is known to possess, is well visible as lags j increase. Less visible, but still present, is the 'hump' as n increases. For small values of j, the effect of the *curvature* factor first increase for low values of n and decreases thereafter. All in all, the term-structure factors that arise naturally from the co-fractional modeling framework exhibit the typical time-series and cross-sectional characteristics of a *level*, *slope*, and *curvature* that have been established in the literature.

4.2 Implied yield curve

We calibrate the prices of risk, λ, such that average yields from model (13) match observed yields at maturities equal to 12 months, five years, and ten years. More precisely, the market price of risk is estimated at the short end, the intermediate end, and the long end of the yield curve by the generalized method of moments (GMM). Using the resulting estimate as well as the estimates for the dynamics of the state vector, x_t, from Section 3, we compute the term-structure parameters from Section 2 and model-implied yields.

Figure 6 plots fitted versus observed yields for six different maturities, $n = \{1, 12, 24, 36, 60, 120\}$, over time. Overall, yields implied by our affine term-structure model together with co-fractional state-vector dynamics track the time series of observed yields very closely. At maturity $n = 1$, model-implied and observed yields are virtually indistinguishable. As the time to maturity increases, the fit gets marginally worse. Yet, even at maturity $n = 120$, the match is exemplary. Over all of the 50 different maturities, for which we have observed data, the R^2 measure for the in-sample model fit fluctuates between 0.8881 and 1.0000. Our model captures the lowest proportion of yield variation at maturity $n = 96$ and the largest percentage at $n = 1$. Since we are working under the maintained assumption that the short rate and hence yields are nonstationary long-memory processes, caution is required in conducting inference on the basis of R^2. As an alternative, we suggest to evaluate the

fit of the term-structure model by regressing observed yields of different maturities on the yields implied by Model (13). That is, we regress

$$y_{OBS,t}^{(n)} = \alpha^{(n)} + \beta^{(n)} y_{MODEL,t}^{(n)} + u_t^{(n)}, \tag{24}$$

where the subscripts 'OBS' and '$MODEL$' denote observed and modeled values, respectively. We estimate $\beta^{(n)}$ in (24) by narrow-band frequency domain least squares (FDLS) due to Robinson (1994), which is more efficient than OLS in this setting. As before, the size of the spectral window is limited to $m = T^o$, with $o = 0.45$[12]. As the asymptotic distribution of the FDLS estimate is non-standard (see, for instance, Robinson and Marinucci, 2001), we again resort to the application of the wild bootstrap to construct 95% confidence intervals.

Figure 8(i) depicts the resulting estimates for $\beta^{(n)}$ and corresponding confidence intervals. Across all maturities, the estimated $\beta^{(n)}$ is close to one. At the initial maturities, we somewhat underestimate the variability of yields, yet for maturities $n \geq 33$ we always find that $\beta^{(n)} = 1$ statistically. At higher maturities, the most persistence factor will dominate the dynamics of yields. In our estimations, this is a nonstationary long-memory factor. The results in Figure 8(i) therefore confirm that a long-memory model can generate the correct amount of variation in yields at the long end, which is in line with the findings of Backus and Zin (1993). The coefficient estimate for $\alpha^{(n)}$ measures how well our model captures the level of yields over time. The estimated value is plotted in Figure 7(i). We slightly overestimate the level of yields at initial maturities, but for $n \geq 22$, we cannot reject that the level of observed and modeled yields is identical. We conclude that yields that are modeled by our affine term-structure model in (1)-(13) together with a CFVAR$_d$(1) model for x_t fits the observed term structure well and for maturities $n \geq 33$ the level and variability of yields are matched exactly.

4.3 Comparison with other models

To obtain an understanding of the comparative fit of our term-structure model, when the state vector is modeled by a CFVAR$_d$(1), we contrast our results to a VAR(q) model for the state vector x_t. In the existing literature, it is very common to fit a VAR model (see, for instance, Ang and Piazzesi, 2003). In that case, the term-structure model in (1)-(13) collapses to the usual affine model for yields. According to the BIC, we select $q = 2$ for the trivariate state vector. Any VAR(q) model can be inverted to find solution (2) and

[12]The estimate of $\alpha^{(n)}$ is obtained in the second step. As its value should capture the intercept over the entire frequency spectrum, we estimate α by setting $m = T - 1$, which is equivalent to OLS.

we can obtain the term-structure parameters for (13) from this representation. If we compare model-implied yields with observed yields in terms of R^2, we find a maximal value of 1.000 at maturity $n = 1$ and a minimum value of 0.7004 at $n = 120$. As pointed out before, the R^2 value may be misleading as an absolute measure of fit, yet it can be used to infer the relative fit of models. Comparing the R^2 of the VAR(2) with the one from the CFVAR$_d$(1), we observe that the former always produces a worse yield fit than the latter, with one exception at $n = 2$. The fit also becomes relatively worse compared to the co-fractional model, as maturity increases. Figure 8(ii) clarifies why; it plots the estimates of $\beta^{(n)}$ of Equation (24) for the VAR(2) model. The estimated value is well above one and the degree of underestimation of the variability in yields increases as maturity increases. The underestimation of yield variability with a stationary VAR model[13], especially at the long end of the yield curve, is a well-known result. The coefficients of the MA solution of the (stationary) VAR(q) model decay at exponential rate over time; by (13) this implies a fast decay of the term-structure of yield volatilities over maturities, as well[14]. The coefficient $\beta^{(n)}$ is only statistically indistinguishable from one at very low maturities $n \in [1, 5]$. The 95% parameter confidence interval is increasingly wider than the corresponding interval for the co-fractional model. It follows that modeling yields from a state vector with VAR dynamics also implies relatively more parameter uncertainty. Figure 7(ii) shows that the stationary VAR model statistically significantly overestimates the level of yields, in addition to underestimating the variability in yields. The resulting negative $\alpha^{(n)}$ estimate is only statistically equal to zero at $n = \{1, 2\}$. As for $\beta^{(n)}$, the 95% confidence interval for $\alpha^{(n)}$ is increasingly wider than the one of the CFVAR$_d$(1), with the exception of $n < 5$.

We further compare our implied yields from (1)-(13) with a co-fractional model for x_t, with a unit-root model for x_t. Many empirical studies conclude that yields are $I(1)$ series (see e.g. Engsted and Tanggaard, 1994, and Nielsen, 2010). As a consequence, it is not uncommon to impose a unit root on the dynamics of the state vector, as for instance done by Dewachter and Lyrio (2006) and Christensen et al. (2011). To capture such very persistent long-run dynamics in x_t, while correcting for potential short-run dependence, we estimate a VAR(1) in first differences. The fitted yields from this model compare to observed yields with a maximal R^2 of 1.000 at $n = 1$. All three models for x_t thus result in an equally well in-sample fit for r_t. The lowest R^2 that results from the VAR(1) in first differences is below zero, however. Overall, the fit of yields from this model is always worse than from

[13]Note that the largest root of the VAR(2) model for x_t has modulus 0.9858. The model is therefore stationary.

[14]In the context of this discussion, it should be noted that one can always model a restricted VAR model for x_t that will perfectly fit the variability of yields. Yet, all tests in this section are based on unrestricted VAR models, estimated from the time series of observed x_t.

the CFVAR$_d$(1). Figures 7(iii) and 8(iii) plot the estimates of $\alpha^{(n)}$ and $\beta^{(n)}$ in Equation (24), respectively. 7(iii) shows that the unit-root model strongly overestimates the level of yields at initial maturities. The estimate for $\alpha^{(n)}$ is only statistically equal to zero at maturities $n = \{12, 60, 63, 66, 72, 108\}$. The 95% confidence interval for $\alpha^{(n)}$ for the unit root model is at almost all marturities substantially wider than the correspoding interval for the co-fractional model, again suggesting more uncertainty within this modeling framework. In contrast to the level, the slope of yields over time is captured fairly well by the VAR(1) model in first differences. Only at $n = 120$ do we statistically significantly overestimate the variablility of yields; yet, again confidence intervals are substantially wider at all maturities compared to the CFVAR$_d$(1).

To summarize, the co-fractional model for x_t results in yields that mostly have the correct level and dynamic properties to match observed yields. The stationary VAR(2) model overestimates the level and underestimates the slope of yields over time. The unit-root model does not come close to matching the level of yields, yet it introduces almost the right amount of yield dynamics. The disappointing results from the stationary and unit-root model are one of the reasons why previous studies called for highly volatile term premiums. If term premiums have a lot of variability, they can move to correct these over- and underestimations of the level and slope of yields over time. With a co-fractional model we show that we can obtain a good fit even when term premiums are constant.

The term-structure model of Diebold and Li (2006), henceforth DL model, has become a benchmark in the yield-curve literature as it is very successful in capturing the dynamics of yields over time as well as over maturity. As a final comparison, we contrast our implied yields to yields resulting from the DL model. At each time t, we evaluate a cross-sectional regression over 50 observed maturities given by

$$y_t^{(n)} = b_{1,t} + \left(\frac{1 - e^{-\kappa n}}{\kappa n}\right) b_{2,t} + \left(\frac{1 - e^{-\kappa n}}{\kappa n} - e^{-\kappa n}\right) b_{3,t} + \eta_t^{(n)}. \tag{25}$$

This results in a time-series of factors $b_{i,t}$, with corresponding loadings $F^{(n)}$, where

$$F^{(n)} = \left[1,\ \frac{1 - e^{-\kappa n}}{\kappa n},\ \frac{1 - e^{-\kappa n}\,(1 + \kappa n)}{\kappa n}\right]'.$$

We estimate κ by GMM, such that average estimated and observed yields at maturities 12, 60, and 120 months match. The resulting value for R^2 varies between 0.9870 and 0.9989 for different maturities. We observe the lowest R^2 value at maturity $n = 1$ and the highest at $n = 25$. Hence, the in-sample fit of the DL model is exceptionally good and the resulting

R^2 is always larger than the corresponding figure from the term-structure model (1)-(13) together with a CFVAR$_d$(1) for x_t, except at maturity $n = 1$. Figures 7(iv) and 8(iv) demonstrate that the estimates for $\alpha^{(n)}$ and $\beta^{(n)}$ in Equation (24) are never statistically different from zero and one, respectively. In addition, the confidence intervals of the DL model are somewhat narrower than the corresponding ones from the affine term-structure model together with CFVAR$_d$(1) state-vector dynamics.

What explains the outperformance of the DL model in sample? The approach of Diebold and Li (2006) to yield curve modeling is very distinct from the method that this paper relies on. In this work, we first model the dynamics of an observed state vector and from that we infer yields. The DL approach reverses this methodology; the first step consists in fitting the cross section of yields, from which one can then infer the dynamics of a latent state vector. In that respect, it may not come as a complete surprise that DL-fitted yields are closer to observed yields, as at every point in time t one obtains the best cross-maturity yield fit. Diebold and Li (2006) assume their yield curve factors, b_t, follow a VAR(1) process. This implies that we can write b_t in the form of x_t in (2). Our affine term-structure model (1) then implies the following factor specifications: $p_{0,t}^{(n)} = nF^{(n)\prime}b_t^0$, $U_j^{(n)} = nF^{(n)\prime}\Phi_j\mu$, $V_j^{(n)} = n\Phi_j' F^{(n)}$, and $W^{(n)} = 0$. The term-structure parameters in the DL model are exogenously derived from a Nelsen-Siegel exponential components model, together with dynamics of three factors that match the cross section of yields. In contrast, the yield-curve parameters in our model are derived from a time-series model for the short rate together with equilibrium dynamics for the stochastic discount factor. The affine model (1) that we propose is arbitrage free, whereas the DL model is not guaranteed to be arbitrage free[15]. We expect this difference between the models to become an important factor, when forecasting the yield curve out of sample. Especially when forecasting over long horizons, we expect our arbitrage-free framework together with explicitly modeling the low-frequency dynamics of the state vector, as we do in the CFVAR$_d$(1), to be well suited.

5 Forecasting Yields

Our term-structure model together with a CFVAR$_d$(1) model for the state vector x_t results in a good fit of the yield curve in sample. This fit was investigated by modeling the dynamics of x_t by accounting for the fractional integration and cointegration in the state variables, using 40 years of observed data. From these estimates, we inferred yield dynamics for the same sample period. Potentially more important than fitting yields in sample is the ability

[15] In an extension of the model, Christensen et al. (2009) introduce a maturity-dependent constant term into yield dynamics as well, i.e. $W^{(n)} \neq 0$, which renders the term-structure model arbitrage free.

of a term-structure model to provide accurate yield-curve forecasts, when the dynamics of the state vector are estimated in an earlier separate sample. Such a forecasting ability, especially over long horizons, could be very beneficial to asset managers of portfolios with long-dated investment horizons that are typically faced by pension funds and (life) insurances. In this section, we investigate the out-of-sample forecast accuracy of term-structure model (1)-(13).

If prices of zero-coupon bonds follow our term-structure specification in Section 2, we can compute k-step ahead forecasts by the following recursion.

$$\mathsf{E}_t p_{t+k}^{(n)} = \mathsf{E}_t p_{t+k-1}^{(n+1)} + e3' \mathsf{E}_t x_{t+k-1} - \frac{1}{2} V_0^{(n)'} \Sigma V_0^{(n)} + V_0^{(n)'} \Sigma \lambda, \tag{26}$$

where the initial price forecast and the forecasts of the state vector are given by

$$\mathsf{E}_t p_{t+1}^{(n)} = p_t^{(n+1)} + e3' x_t - \frac{1}{2} V_0^{(n)'} \Sigma V_0^{(n)} + V_0^{(n)'} \Sigma \lambda \tag{27}$$

$$\mathsf{E}_t x_{t+k-1} = x_{t+k-1}^0 + \sum_{j=0}^{t-1} \Phi_{j+k-1} \epsilon_{t-j} + \sum_{j=0}^{t+k-2} \Phi_j \mu. \tag{28}$$

From Equation (26), we can obtain forecasts for yields from the identity $y_t^{(n)} = -(1/n) p_t^{(n)}$. We compare yield-curve forecasts from our term-structure model together with a CFVAR$_d$(1) model for x_t, with the three competing models from the previous section. To that end we rely on the mean-squared forecast error (MSE). Forecasts for the DL model are obtained by assuming that b_t follows a VAR(1) model, as in Diebold and Li (2006)[16]. We consider forecasting horizons from one month to ten years, $k = 1, 2, \ldots, 120$, and employ a rolling-window forecasting framework. We fit a model to x_t (respectively b_t) on the basis of 260 months initial observations, and forecast x_t as well as yields for all observed 50 maturities between one month and ten years beyond that. We repeat this 221 times. Thus, for $k = 1$ we obtain 221 time-series observations of the yield curve to compare to the target, and for $k = 120$ the time-series of out-of-sample forecasts has length 102.

We re-estimate the model parameters in each on of the 221 in-sample periods. The only exception is the fractional differencing parameter d, which remains equal to its full-sample estimated value $\hat{d} = 0.8871$. The reason is that we view long memory as a long-run statistical property of the respective series. The fact that others such as Shea (1991), Sun and Phillips (2004), and Schotman et al. (2008) find comparable estimates for d for the

[16] Diebold and Li (2006) argue that uncorrelated AR(1) models for each element of b_t, $b_{i,t}$, produce better forecasts than a VAR(1) model. For our data, we cannot confirm this conclusion. The resulting MSEs from both specification are virtually indistinguishable.

short-term interest rate in different samples, supports this assumption. In addition, e.g. Orphanides and Wei (2012) find that the persistence of the short rate is stable throughout their sample. Finally, roughly 22 years of data, which corresponds to the in-sample period, may not be sufficient to correctly infer low-frequency dynamics.

Figures 9 and 10 plot the resulting MSE for all but two ($n = 2, 4$) maturities in our sample. We start the discussion of the results by comparing the DL model to the affine model with co-fractional dynamics. Figures 9 and 10 show that forecasts from the latter model are always more accurate than the corresponding DL-model forecasts at long horizons. The only deviations from that rule occur at very short maturities. For $n = 1$ the predictions from the CFVAR$_d$(1) yield model are better than DL forecasts at any forecasting horizon k. Thus, the short rate is predicted more precisely by a co-fractional model. At short maturities $n \in [2, 5]$, the DL model provides superior forecasts, even slightly so at very large k. Beyond that, that is for $n \geq 6$, long-run forecasts from the CFVAR$_d$(1) are more accurate than DL forecasts. For maturities $n \in [9, 57]$, the co-fractional model consistently produces lower MSE values than the DL model for approximately $k \geq 30$. At all other maturities the CFVAR$_d$(1) outperforms the DL model in terms of MSE for horizons larger than $k \in [40, 50]$. Thus, for forecasting horizons longer than approximately four years, the forecast error is minimized by the affine term-structure model based on equilibrium yield dynamics with co-fractional dynamics for x_t, relative to the DL model (except for $n = \{3, 5\}$).

How do yield forecasts from the CFVAR$_d$(1) model compare to the other two specifications for the state-vector presented in Section 4? If we forecast x_t with a VAR(2) model in levels, we find that the co-fractional model produces more accurate yield forecasts, especially at long forecasting horizons. As for the DL model, the only exceptions are yields with very low maturities, $n \in [2, 5]$, where the VAR(2) forecasts yields more precisely. For maturities $n \in [6, 10]$, the co-fractional model outperforms the VAR(2) only in terms of long-run forecasts. For all remaining maturities, $n > 10$, the latter model always produces larger MSE values for any k. These results are very much in line with the findings from the in-sample analysis, where we showed that the VAR(2) did not fit the observed data very well, except at low maturities. We attribute this evidence to the fact that the stationary VAR(2) model does not introduce sufficient persistence into the model, which is especially important for fitting the long end of the yield curve, and for forecasting over long horizons. Although less severely, the DL model seems to suffer partly from the same problem. In fact, Figures 9 and 10 demonstrate that the mean-square forecast errors of the VAR(2) and the DL model converge as k grows.

Yield forecasts following from a VAR(1) specification for Δx_t are generally substantially worse than forecasts from the other models. Noting that Figures 9 and 10 have a logarithmic scale, they demonstrate that this model produces mostly very high MSE values. Yet, there is one exception. At maturities, where the level of yields is explicitly modeled in the in-sample estimation period, long-run forecasts from the model that imposes a unit root on the state vector are superior. More precisely, this occurs at maturity $n = 1$, where the level and dynamics of r_t are estimated by a CFVAR$_d$(1) model, and at $n \approx \{12, 60, 120\}$, where the level of yields is matched by the market price of risk. We conclude that the unit-root model for x_t introduces too much volatility into model-implied yields. This results in a continuous over-shooting and under-shooting of the level of yields, except at those points on the yield curve, where the level is explicitly matched. One possibility to remedy this short-coming is to introduce volatile term premiums that correct for the strong deviations from the level of yields. The alternative that is suggested in this paper, does not require the introduction of more (or less) volatility through term premiums. The previous discussion shows that one can obtain good long-run yield forecasts from an equilibrium model with constant term premiums, when modeling state vector dynamics as a co-fractional process. Our term-structure model is hence much closer to the expectations hypothesis, as recommended by e.g. Bauer (2011).

5.1 Model-confidence sets

Graphs 9 and 10 suggest that the forecasting performance of our term-structure model together with co-fractional dynamics is superior at long horizons for most maturities. However, the figures do not indicate whether this outperformance is statistically significant. We remedy this shortcoming by computing forecast model confidence sets (MCS) due to Hansen et al. (2003, 2011). For several different forecast horizons, $k = \{1, 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120\}$, we compute the set of models that contains the *best* forecasting model for a yield with an average maturity with a confidence level of 95%. The *best* model is the one for which MSE is minimized.

In addition to the forecasting models considered above, we add some more candidates. The first addition consists of yield forecasts from the affine term-structure model (1)-(13) together with CFVAR$_d$(1) dynamics for the state vector, where d is re-estimated in each rolling in-sample period, as opposed to keeping its value fixed as above. Secondly, we add DL-forecasts, where the vector b_t is assumed to have three univariate AR(1) specifications, as Diebold and Li (2006) find this model to produce superior forecasts compared to a VAR(1). Furthermore, we consider random-walk forecasts, where $\mathsf{E}_t y_{t+k}^{(n)} = y_t^{(n)}$. The

random walk is an obvious candidate, since it is well known that it produces yield forecasts that are hard to beat by equilibrium affine models (see e.g. Duffee, 2002). Finally, we consider a historical mean model, that is $\mathsf{E}_t y_{t+1} = \frac{1}{t}\sum_{i=1}^{t} y_{t+1-i}$ and for $k > 1$ forecasts can be obtained recursively, which is a standard benchmark in the equity-premium literature (see e.g. Campbell and Thompson, 2008, and Goyal and Welch, 2008).

Table 3 summarizes our results. As indicated by the findings in the previous section, the affine term-structure model with CFVAR$_d$(1) dynamics for x_t provides superior forecasts at long horizons. When the forecast target lies five years or more in the future, we conclude that our model is within the set of best predictor models. At these horizons, $k \geq 60$, we can discard all other models as having significantly worse predictive ability, with one exception. The MCS at long horizons suggest that we cannot reject that our model with co-fractional dynamics and the random walk are equally good prediction models.

For forecasting horizons from one to four years, our co-fractional interest rate model is also within the set of best forecasting models. For these values of k, the MCS contains many other models as well, however. That is, both specifications of the DL model and the random walk are in the MCS, as well. The MCS methodology is known to acknowledge the limitations of the data (Hansen et al., 2003). Thus, for these horizons where the MCS has many elements, the data are simply not informative enough to select a smaller set of best models, and it cannot be ruled out that several models possess equal predictive ability in population. For very short forecast horizons, $k \leq 6$, the model of Diebold and Li (2006) together with the random walk provide superior forecasts, again suggesting that accounting for long memory in the way it is done in this paper, is especially important for long-horizon forecasts. Finally, we note that a yield with average maturity cannot be forecasted well by our term-structure model with VAR-dynamics for x_t in levels or first differences, nor by the historical average. These models are found to have statistically inferior forecasting ability relative to the competing models, at all forecast horizons.

6 Conclusion

Affine models for the term-structure of interest rates are subject to the criticism that the dynamic properties required to match the time-series properties of state variables seem incompatible with the persistence required to match the cross-section of yields over different maturities. In this article, we tackle this problem by allowing for a more flexible specification in the state-vector dynamics. The resulting yield curve model is a generalization of discrete time, Gaussian, affine term-structure models. Even though more flexible, the

yield-curve model maintains the attractive property of affine models in that it produces a closed-form solution for the term-structure parameters.

We model the state-vector dynamics by a co-fractional VAR model due to Johansen (2008a,b). This model generates yields that are fractionally integrated. More precisely, implied yields exhibit long-memory yet mean-reverting behavior with a fractional integration coefficient of $d \approx 0.87$. Hence, implied yields are very persistent, which matches their empirical properties (see e.g. Diebold and Li, 2006). The co-fractional VAR specification fits the dynamics of the state vector well. Three low-frequency factors arise from the specification, that have the natural interpretation as *level*, *slope*, and *curvature* factors.

Predicting yields from the generalized affine model and co-fractional dynamics for the state vector results in a good term-structure fit, both in sample as well as out of sample. In sample, the model captures the level and the variability of yields well, especially at longer maturities. The out-of-sample forecasting exercise reveals that the model produces good long-horizon forecasts that outperform other well-established benchmark models and are equally accurate as random-walk forecasts.

Appendix

A Derivation of the term-structure model

We derive the generalized Gaussian affine term-structure model in Equations (1)-(12). We assume that logarithmic state-vector dynamics x_t, logarithmic prices of zero-coupon bond with maturity n, $p_t^{(n)}$, the log short rate $r_t = -p_t^{(1)}$, and the logarithmic stochastic discount factor, m_t, have to following specifications.

$$x_t = x_t^0 + \sum_{j=0}^{t-1} \Phi_j \left(\epsilon_{t-j} + \mu\right) \tag{A1}$$

$$p_t^{(n)} = -p_{0,t}^{(n)} - \sum_{j=0}^{t-1} \left(U_j^{(n)} + V_j^{(n)'} \epsilon_{t-j}\right) - W^{(n)} \tag{A2}$$

$$r_t = e3' x_t \tag{A3}$$

$$m_{t+1} = -r_t - \tfrac{1}{2}\lambda'\Sigma\lambda + \lambda'\epsilon_{t+1}. \tag{A4}$$

From (A2), we obtain conditional moments of prices by

$$\begin{aligned}
\mathsf{E}_t\left(p_{t+1}^{(n)}\right) &= -p_{0,t+1}^{(n)} - \sum_{j=1}^{t}\left(U_j^{(n)} + V_j^{(n)'}\epsilon_{t+1-j}\right) - \left(U_0^{(n)} + W^{(n)}\right) && \text{(A5)}\\
\mathsf{Var}_t\left(p_{t+1}^{(n)}\right) &= \mathsf{E}_t\left\{\left[p_{t+1}^{(n)} - \mathsf{E}_t\left(p_{t+1}^{(n)}\right)\right]\left[p_{t+1}^{(n)} - \mathsf{E}_t\left(p_{t+1}^{(n)}\right)\right]'\right\}\\
&= \mathsf{E}_t\left\{\left[-V_0^{(n)'}\epsilon_{t+1}\right]\left[-V_0^{(n)'}\epsilon_{t+1}\right]'\right\}\\
&= V_0^{(n)'}\Sigma V_0^{(n)} && \text{(A6)}\\
\mathsf{Cov}_t\left(p_{t+1}^{(n)}, m_{t+1}\right) &= \mathsf{E}_t\left\{\left[p_{t+1}^{(n)} - \mathsf{E}_t\left(p_{t+1}^{(n)}\right)\right]\left[m_{t+1} - \mathsf{E}_t\left(m_{t+1}\right)\right]'\right\}\\
&= \mathsf{E}_t\left\{\left[-V_0^{(n)'}\epsilon_{t+1}\right]\left[\lambda'\epsilon_{t+1}\right]'\right\}\\
&= -V_0^{(n)'}\Sigma\lambda. && \text{(A7)}
\end{aligned}$$

The basic recursive pricing equation for a zero-coupon bond in levels is $P_t^{(n+1)} = \mathsf{E}_t(M_{t+1}P_{t+1}^{(n)})$, where capital letters are the level analogue of the logarithmic counterparts. Imposing Gaussianity, we can solve for log bond prices recursively.

$$\begin{aligned}
p_t^{(n+1)} &= -r_t + \mathsf{E}_t\left(p_{t+1}^{(n)}\right) + \tfrac{1}{2}\mathsf{Var}_t\left(p_{t+1}^{(n)}\right) + \mathsf{Cov}_t\left(p_{t+1}^{(n)}, m_{t+1}\right)\\
&= -e3'x_t^0 - \sum_{j=0}^{t-1} e3'\Phi_j\left(\epsilon_{t-j} + \mu\right) - p_{0,t+1}^{(n)} - \sum_{j=0}^{t-1}\left(U_{j+1}^{(n)} + V_{j+1}^{(n)'}\epsilon_{t-j}\right)\\
&\quad - \left(U_0^{(n)} + W^{(n)}\right) + \tfrac{1}{2}V_0^{(n)'}\Sigma V_0^{(n)} - V_0^{(n)'}\Sigma\lambda\\
&= -\left(e3'x_t^0 + p_{0,t+1}^{(n)}\right) - \sum_{j=0}^{t-1}\left(\Phi_j'e3 + V_{j+1}^{(n)}\right)'\epsilon_{t-j}\\
&\quad - \sum_{j=0}^{t-1}\left(e3'\Phi_j\mu + U_{j+1}^{(n)}\right) - \left(U_0^{(n)} + W^{(n)} - \tfrac{1}{2}V_0^{(n)'}\Sigma V_0^{(n)} + V_0^{(n)'}\Sigma\lambda\right) && \text{(A8)}
\end{aligned}$$

Equation (A8) shows that bond prices with maturity $n+1$ can be expressed in the same form as (A2) with corresponding parameters

$$\begin{aligned}
p_{0,t}^{(n+1)} &= e3'x_t^0 + p_{0,t+1}^{(n)} && \text{(A9)}\\
V_j^{(n+1)} &= \Phi_j'e3 + V_{j+1}^{(n)} && \text{(A10)}\\
U_j^{(n+1)} &= e3'\Phi_j\mu + U_{j+1}^{(n)} && \text{(A11)}\\
W^{(n+1)} &= U_0^{(n)} + W^{(n)} - \tfrac{1}{2}V_0^{(n)'}\Sigma V_0^{(n)} + V_0^{(n)'}\Sigma\lambda. && \text{(A12)}
\end{aligned}$$

From the equality of the price of a zero-coupon bond with maturity $n=1$ with the negative short rate, i.e. from $p_t^{(0)} = -r_t = -e3'x_t^0 - \sum_{j=0}^{t-1} e3'\Phi_j\left(\epsilon_{t-j} + \mu\right)$, we can derive the initial

condition for the bond pricing recursion in (A9)-(A12) as

$$\begin{aligned} p_{0,t}^{(1)} &= e3' x_t^0 && \text{(A13)} \\ V_j^{(1)} &= \Phi_j' e3 && \text{(A14)} \\ U_j^{(1)} &= e3' \Phi_j \mu && \text{(A15)} \\ W^{(1)} &= 0. && \text{(A16)} \end{aligned}$$

Combining the recursive parameter expressions (A9)-(A12) with the initial values (A13)-(A16), we obtain closed form solutions for $p_{0,t}^{(n)}$ by

$$\begin{aligned} p_{0,t}^{(2)} = &\quad e3' x_t^0 + p_{0,t+1}^{(1)} &&= e3' x_t^0 + e3' x_{t+1}^0 \\ p_{0,t}^{(3)} = &\quad e3' x_t^0 + p_{0,t+1}^{(2)} &&= e3' x_t^0 + e3' x_{t+1}^0 + e3' x_{t+2}^0 \\ &\qquad \vdots \\ p_{0,t}^{(n)} = &\quad e3' x_t^0 + p_{0,t+1}^{(n-1)} &&= \sum_{i}^{n-1} e3' x_{t+i}^0. \end{aligned} \tag{A17}$$

The closed-form expression for $V_j^{(n)}$ is defined by

$$\begin{aligned} V_j^{(2)} = &\quad \Phi_j' e3 + V_{j+1}^{(1)} &&= \Phi_j' e3 + \Phi_{j+1}' e3 \\ V_j^{(3)} = &\quad \Phi_j' e3 + V_{j+1}^{(2)} &&= \Phi_j' e3 + \Phi_{j+1}' e3 + \Phi_{j+2}' e3 \\ &\qquad \vdots \\ V_j^{(n)} = &\quad \Phi_{j+1}' e3 + V_{j+1}^{(n-1)} &&= \sum_{i=0}^{n-1} \Phi_{j+1}' e3, \end{aligned} \tag{A18}$$

and for $U_j^{(n)}$ we find

$$\begin{aligned} U_j^{(2)} = &\quad e3' \Phi_j \mu + U_{j+1}^{(1)} &&= e3' \Phi_j \mu + e3' \Phi_{j+1} \mu \\ U_j^{(3)} = &\quad e3' \Phi_j \mu + U_{j+1}^{(2)} &&= e3' \Phi_j \mu + e3' \Phi_{j+1} \mu + e3' \Phi_{j+2} \mu \\ &\qquad \vdots \\ U_j^{(n)} = &\quad e3' \Phi_j \mu + U_{j+1}^{(n-1)} &&= \sum_{i=0}^{n-1} e3' \Phi_{j+i} \mu. \end{aligned} \tag{A19}$$

The closed-form solution for the parameter $W^{(n)}$ is obtained by

$$\begin{aligned}
W^{(2)} &= U_0^{(1)} + W^{(1)} - \tfrac{1}{2} V_0^{(1)\prime} \Sigma V_0^{(1)} + V_0^{(1)\prime} \Sigma \lambda \\
&= U_0^{(1)} - \tfrac{1}{2} V_0^{(1)\prime} \Sigma V_0^{(1)} + V_0^{(1)\prime} \Sigma \lambda \\
W^{(3)} &= U_0^{(2)} + W^{(2)} - \tfrac{1}{2} V_0^{(2)\prime} \Sigma V_0^{(2)} + V_0^{(2)\prime} \Sigma \lambda \\
&= \sum_{i=1}^{2} \left(U_0^{(i)} - \tfrac{1}{2} V_0^{(i)\prime} \Sigma V_0^{(i)} + V_0^{(i)\prime} \Sigma \lambda \right) \\
&\vdots \\
W^{(n)} &= U_0^{(n-1)} + W^{(n-1)} - \tfrac{1}{2} V_0^{(n-1)\prime} \Sigma V_0^{(n-1)} + V_0^{(n-1)\prime} \Sigma \lambda \\
&= \sum_{i=1}^{n-1} \left(U_0^{(i)} - \tfrac{1}{2} V_0^{(i)\prime} \Sigma V_0^{(i)} + V_0^{(i)\prime} \Sigma \lambda \right).
\end{aligned}$$

Replacing $V_0^{(n)}$ and $V_0^{(n)}$ by their solutions from (A18) and (A19), we find

$$W^{(n)} = \sum_{i=1}^{n-1} \left(\sum_{k=0}^{i-1} e3' \Phi_k \mu - \frac{1}{2} \left[\sum_{k=0}^{i-1} \Phi_k' e3 \right]' \Sigma \left[\sum_{k=0}^{i-1} \Phi_k' e3 \right] + \left[\sum_{k=0}^{i-1} \Phi_k' e3 \right]' \Sigma \lambda \right). \tag{A20}$$

B Inversion of the CFVAR$_d(1)$ model

The co-fractional CFVAR$_{d,b}(q)$ model of Johansen (2008a,b) with one lag, $q = 1$, and with the strength of the co-fractionality equal to the order of fractional integration, $d = b$, has the representation

$$\Delta^d x_t = \gamma \left[\delta' \left(1 - \Delta^d \right) x_t + \rho' \right] + \Gamma_1 \Delta^d \left(1 - \Delta^d \right) x_t + \epsilon_t. \tag{B1}$$

We start the derivation of the inversion by re-writing Equation (B1) as follows.

$$\begin{aligned}
x_t &= \left[I + \gamma \delta' \right] \left(1 - \Delta^d \right) x_t + \gamma \rho' + \Gamma_1 \Delta^d \left(1 - \Delta^d \right) x_t + \epsilon_t \\
&= -\left[I + \gamma \delta' \right] \sum_{j=1}^{\infty} \Theta_j L^j x_t + \gamma \rho' - \Gamma_1 \left(\sum_{j=0}^{\infty} \Theta_j L^j \right) \left(\sum_{j=1}^{\infty} \Theta_j L^j \right) x_t + \epsilon_t \\
&= -\sum_{j=1}^{\infty} \left[\left(I + \gamma \delta' \right) \Theta_j + \Gamma_1 \sum_{k=0}^{j-1} \Theta_k \Theta_{j-k} \right] x_{t-j} + \gamma \rho' + \epsilon_t.
\end{aligned} \tag{B2}$$

From (B2), the autoregressive representation of the process x_t is given by

$$\begin{aligned} x_t + \sum_{j=1}^{\infty} \left[\left(I + \gamma\delta' \right) \Theta_j + \Gamma_1 \sum_{k=0}^{j-1} \Theta_k \Theta_{j-k} \right] x_{t-j} &= \gamma\rho' + \epsilon_t \\ \sum_{j=0}^{\infty} \Xi_j x_{t-j} &= \gamma\rho' + \epsilon_t \\ \Xi(L) x_t &= \gamma\rho' + \epsilon_t, \end{aligned} \tag{B3}$$

where $\Xi_0 = I$. Let $\Xi_+(L) = \sum_{j=0}^{t-1} \Xi_j L^j$ and $\Xi_-(L) = \sum_{j=t}^{\infty} \Xi_j L^j$. It follows that (B3) can be expressed as

$$\Xi_+(L) x_t + \Xi_-(L) x_t = \gamma\rho' + \epsilon_t \tag{B4}$$

$$\Xi_+(L)^{-1} \Xi_+(L) x_t + \Xi_+(L)^{-1} \Xi_-(L) x_t = \Xi_+(L)^{-1} \left(\gamma\rho' + \epsilon_t \right). \tag{B5}$$

We want to find a moving-average solution for x_t, such that the right-hand side of (B5) can be written as

$$\Xi_+(L)^{-1} \left(\gamma\rho' + \epsilon_t \right) = \sum_{j=0}^{t-1} \Phi_j \left(\gamma\rho' + \epsilon_{t-j} \right). \tag{B6}$$

Note that it has to hold that $\Xi_+(L)^{-1}\Xi_+(L) + \Xi_+(L)^{-1}\Xi_-(L) = \Xi_+(L)^{-1}\Xi(L)$. Expanding these polynomials, we obtain

$$\begin{aligned} \Xi_+(L)^{-1}\Xi(L) &= \left(\sum_{j=0}^{t-1} \Phi_j L^j \right) \left(\sum_{j=0}^{\infty} \Xi_j L^j \right) \\ &= \sum_{j=0}^{t-1} \sum_{k=0}^{j} \Phi_k \Xi_{j-k} L^j + \sum_{j=0}^{\infty} \sum_{k=0}^{t-1} \Phi_k \Xi_{t+j-k} L^{t+j}. \end{aligned} \tag{B7}$$

Equation (B7) consist of two terms. The first polynomial involves only in-sample values for $t = 1, 2, \ldots, T$, whereas the second sum involves pre-sample values, that is initial values of the process x_t. We therefore conclude that

$$\Xi_+(L)^{-1}\Xi_+(L) x_t = \sum_{j=0}^{t-1} \sum_{k=0}^{j} \Phi_k \Xi_{j-k} x_{t-j} \tag{B8}$$

$$\Xi_+(L)^{-1}\Xi_-(L) x_t = \sum_{j=0}^{\infty} \sum_{k=0}^{t-1} \Phi_k \Xi_{t+j-k} x_{0-j}. \tag{B9}$$

Imposing the condition that $\Xi_+(L)^{-1}\Xi_+(L) = I$, we can find solutions for the moving-average coefficients Φ_j. For identification, and without loss of generality, we impose that

$\Phi_0 = I$. It follows that

$$\Phi_j = -\sum_{k=0}^{j-1} \Phi_k \Xi_{j-k} \quad \forall j \in [1, t-1]. \tag{B10}$$

From recursion (B10), we can further compute the parameters $\Xi_+(L)^{-1}\Xi_-(L)x_t$ and obtain a solution/inversion for x_t equal to

$$x_t = x_t^0 + \sum_{j=0}^{t-1} \Phi_j \left(\gamma\rho' + \epsilon_{t-j}\right), \tag{B11}$$

where $x_t^0 = -\sum_{j=0}^{\infty} \sum_{k=0}^{t-1} \Phi_k \Xi_{t+j-k} x_{0-j}$.

References

Ang, A. and Piazzesi, M. (2003). A No-arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, *Journal of Monetary Economics* **50**: 745–787.

Backus, D. K. and Zin, S. E. (1993). Long-memory Inflation Uncertainty: Evidence from the Term Structure of Interest Rates, *Journal of Money, Credit and Banking* **25**: 681–700.

Bauer, M. D. (2011). Term Premia and the News, *Federal Reserve Bank of San Francisco: Working Paper Series* **WP 2011-03**.

Brown, R. H. and Schaefer, S. M. (1994). The Term Structure of Real Interest Rates and the Cox, Ingersoll and Ross Model, *Journal of Financial Economics* **35**: 3–42.

Campbell, J. Y., Lo, A. W. and MacKinlay, C. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey, USA.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *Review of Financial St* **21**: 1509–1531.

Cavaliere, G., Rahbek, A. and Taylor, A. M. R. (2010a). Cointegration Rank Testing Under Conditional Heteroskedasticity, *Econometric Theory* **26**: 1719–1760.

Cavaliere, G., Rahbek, A. and Taylor, A. M. R. (2010b). Testing for Cointegration in Vector autoregressions with Non-stationary Volatility, *Journal of Econometrics* **158**(1): 7–24.

Chen, W. W. and Hurvich, C. M. (2003a). Estimating Fractional Cointegration in the Presence of Polynomial Trends, *Journal of Econometrics* **117**: 95–121.

Chen, W. W. and Hurvich, C. M. (2003b). Semiparametric Estimation of Multivariate Fractional Cointegration, *Journal of the American Statistical Association* **98**: 629–642.

Christensen, J. H. E., Diebold, F. X. and Rudebusch, G. D. (2009). An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model, *The Econometrics Journal* **12**: 33–64.

Christensen, J. H. E., Diebold, F. X. and Rudebusch, G. D. (2011). The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models, *Journal of Econometrics* **164**: 4–20.

Cochrane, J. H. (2005). *Asset Pricing*, revised edn, Princeton University Press, Princeton, New Jersey, USA.

Cochrane, J. H. and Piazzesi, M. (2005). Bond Risk Premia, *The American Economic Review* **95**: 138–160.

Cochrane, J. H. and Piazzesi, M. (2008). Decomposing the Yield Curve. SSRN working paper 1333274.

Dai, Q. and Singleton, K. J. (2000). Specification Analysis of Affine Term Structure Models, *Journal of Finance* **55**: 1943–1978.

Davidson, R. and Flachaire, E. (2008). The Wild Bootstrap, Tamed at Last, *Journal of Econometrics* **146**: 162–169.

Davidson, R. and MacKinnon, J. G. (2007). Bootstrap Methods in Econometrics, *in* K. Patterson and T. C. Mills (eds), *Palgrave Handbooks of Econometrics*, Vol. 1: Econometric Theory, Palgrave Macmillan, Hampshire, England.

Dewachter, H. and Lyrio, M. (2006). Macro Factors and the Term Structure of Interest Rates, *Journal of Money, Credit, and Banking* **38**: 119–140.

Diebold, F. X. and Inoue, A. (2001). Long Memory and Regime Switching, *Journal of Econometrics* **105**: 131–159.

Diebold, F. X. and Li, C. (2006). Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics* **130**: 337–364.

Diebold, F. X., Rudebusch, G. D. and Aruoba, S. B. (2006). The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach, *Journal of Econometrics* **131**: 309–338.

Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance* **57**: 405–443.

Duffie, D. and Kan, R. (1996). A Yield-factor Model of Interest Rates, *Mathematical Finance* **6**: 379–406.

Engsted, T. and Tanggaard, C. (1994). Cointegration and the US Term Structure, *Journal of Banking and Finance* **18**: 167–181.

Fama, E. F. and Bliss, R. R. (1987). The Information in Long-maturity Forward Rates, *The American Economic Review* **77**: 680–692.

Gil-Alana, L. A. (2004). Modelling the U.S. Interest Rate in Terms of I(d) Statistical Models, *The Quarterly Review of Economics and Finance* **44**(4): 475–486.

Gil-Alana, L. and Moreno, A. (2012). Uncovering the US Term Premium: An Alternative Route, *Journal of Banking and Finance* **36**: 1181–1193.

Gonzalo, J. and Ng, S. (2001). A Systematic Framework for Analyzing the Dynamic Effects of Permanent and Transitory Shocks, *Journal of Economic Dynamics and Control* **25**: 1527–1546.

Goyal, A. and Welch, I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* **21**: 1455–1508.

Gürkaynak, R. S. and Wright, J. H. (2012). Macroeconomics and the Term Structure, *Journal of Economic Literature* **50**: 331–367.

Hall, A. D., Anderson, H. M. and Granger, C. W. J. (1992). A cointegration Analysis of Treasury Bill Yields, *Review of Economics and Statistics* **74**: 116–126.

Hansen, P. R., Lunde, A. and Nason, J. M. (2003). Choosing the Best Volatility Models: The Model Confidence Set Approach, *Oxford Bulletin of Economics and Statistics* **65**: 839–861.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The Model Confidence Set, *Econometrica* **79**: 453–497.

Henry, M. (2001). Robust Automatic Bandwidth for Long Memory, *Journal of Time Series Analysis* **22**: 293–316.

Henry, M. and Robinson, P. M. (1996). Bandwidth Choice in Gaussian Semiparametric Estimation of Long Range Dependence, *in* P. M. Robinson and M. Rosenblatt (eds), *Athens Conference on Applied Probability and Tme Series Analysis, Vol. II: Time Series Analysis, In Memory of E. J. Hannan*, Springer, New York, pp. 220–232.

Henry, M. and Zaffaroni, P. (2002). The Long Range Dependence Paradigm for Macroeconomics and Finance, *Columbia University, Department of Economics, Discussion Paper Series* **0102-19**.

Hillebrand, E., Huang, H., Lee, T. H. and Li, C. (2012). Using the Yield Curve in Forecasting Output Growth and Inflation, *CREATES Research Paper* **17**.

Iacone, F. (2009). A Semiparametric Analysis of the Term Structure of the US Interest Rates, *Oxford Bulletin of Economics and Statistics* **71**(4): 475–490.

Jardet, C., Monfort, A. and Pegoraro, F. (2013). No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth, *Journal of Banking and Finance* **37**: 389–402.

Johansen, S. (2008a). A Representation Theory for a Class of Vector Autoregressive Models for Fractional Processes, *Econometric Theory* **24**: 651–676.

Johansen, S. (2008b). Representation of Cointegrated Autoregressive Processes with Application to Fractional Processes, *Econometric Reviews* **28**: 121–145.

Johansen, S. and Nielsen, M. O. (2012). Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model, *Econometrica* **80**: 2667–2732.

Kim, D. H. and Orphanides, A. (2007). The Bond Market Term Premium: What Is It, and How Can We Measure It?, *BIS Quarterly Review* **June 2007**: 27–40.

Kim, D. H. and Orphanides, A. (2012). Term Structure Estimation with Survey Data on Interest Rate Forecasts, *Journal Financial and Quantitative Analysis* **47**: 241–272.

Kozicki, S. and Tinsley, P. A. (2001a). Shifting Endpoints in the Term Structure of Interest Rates, *Journal of Monetary Economics* **47**: 613–652.

Kozicki, S. and Tinsley, P. A. (2001b). Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations, *Journal of Economic Dynamics and Control* **25**: 149–184.

Kozicki, S. and Tinsley, P. A. (2005). What Do You Expect? Imperfect Policy Credibility and Tests of the Expectations Hypothesis, *Journal of Monetary Economics* **52**: 421–447.

Litterman, R. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns, *Journal of Fixed Income* **1**: 54–61.

MacKinnon, J. G. and Nielsen, M. O. (2012). Numerical Distribution Functions of Fractional Unit Root and Cointegration Tests, *forthcoming in Journal of Applied Econometrics* . `http://www.econ.queensu.ca/faculty/mon/software`.

Nielsen, M. O. (2010). Nonparametric Cointegration Analysis of Fractional Systems with Unknown Integration Orders, *Journal of Econometrics* **155**: 170–187.

Nielsen, M. O. and Morin, L. (2012). FCVARmodel.m: A Matlab Software Package for Estimation and Testing in the Fractionally Cointegrated VAR Model, *Queens Economics Department Working Paper* **1273**. `http://www.econ.queensu.ca/faculty/mon/software`.

Orphanides, A. and Wei, M. (2012). Evolving Macroeconomic Perceptions and the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control* **36**: 239–254.

Piazzesi, M. and Schneider, M. (2006). Equilibrium Yield Curves, *in* D. Acemoglu, K. Rogoff and M. Woodford (eds), *NBER Macroeconomics Annual*, MIT Press, chapter 6, pp. 389–442.

Robinson, P. M. (1994). Semiparametric Analysis of Long-Memory Time Series, *Annals of Statistics* **22**: 515–539.

Robinson, P. M. (2003a). Long-Memory Time Series, *in* P. M. Robinson (ed.), *Time Series with Long Memory*, Oxford University Press, New York, USA, pp. 4–32.

Robinson, P. M. (2003b). *Time Series with Long Memory*, Oxford University Press, New York, USA.

Robinson, P. M. and Marinucci, D. (2001). Narrow-Band Analysis of Nonstationary Processes, *Annals of Statistics* **29**: 947–986.

Rudebusch, G. D. and Wu, T. (2008). A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy, *The Economic Journal* **118**: 906–926.

Sack, B. (2006). Comment on Kuttners Can Central Banks Target Bond Prices?, *Proceedings of the Bank of Korea International Conference, Bank of Korea, Seoul*.

Schotman, P. C., Tschernig, J. and Budek, J. (2008). Long Memory and the Term Structure of Risk, *Journal of Financial Econometrics* **6**: 459–495.

Shea, G. S. (1991). Uncertainty and Implied Variance Bounds in Long-Memory Models of the Interest Rate Term Structure, *Empirical Economics* **16**: 287–312.

Shimotsu, K. and Phillips, P. C. B. (2005). Exact Local Whittle Estimation of Fractional Integration, *The Annals of Statistics* **33**: 1890–1933.

Singleton, K. (2006). *Empirical Dynamic Asset Pricing*, Princeton University Press, Princeton, New Jersey, USA.

Sun, Y. and Phillips, P. C. B. (2004). Understanding the Fisher Equation, *Journal of Applied Econometrics* **19**: 869–886.

Tables

Table 1: Summary statistics

The table reports summary statistics for monthly real yields with maturities $n = \{1, 12, 24, 36, 60, 84, 120\}$. All yields are quoted in % per month and are continously compounded. The lower part of the table contains summary statistics for the *slope* and *curvature* factors, implied by our co-fractional model (16). The former is defined as *slope*: $y_t^{(120)} + \hat{\delta}_{(32)} r_t + \hat{\rho}_{(12)}$ and the latter as *curvature*: $2y_t^{(24)} - y_t^{(120)} + \hat{\delta}_{(31)} r_t + \hat{\rho}_{(11)}$, where a hat represents an estimate.

			Autocorrelations					
	Ave.	StDev.	1	2	3	12	24	36
1-m TBond r_t	0.470	0.249	0.960	0.931	0.908	0.729	0.483	0.355
1-y TBond $y_t^{(12)}$	0.520	0.260	0.974	0.950	0.928	0.775	0.555	0.439
2-y TBond $y_t^{(24)}$	0.542	0.253	0.981	0.959	0.939	0.796	0.608	0.510
3-y TBond $y_t^{(36)}$	0.559	0.243	0.983	0.963	0.947	0.816	0.645	0.551
5-y TBond $y_t^{(60)}$	0.581	0.230	0.985	0.967	0.951	0.830	0.687	0.604
7-y TBond $y_t^{(84)}$	0.597	0.222	0.985	0.967	0.951	0.834	0.704	0.618
10-y TBond $y_t^{(120)}$	0.604	0.214	0.987	0.971	0.956	0.830	0.708	0.631
Slope	-0.019	0.121	0.859	0.794	0.742	0.354	0.008	-0.212
Curvature	0.004	0.086	0.599	0.529	0.480	0.175	0.013	-0.129

Table 2: CFVAR$_d$(1) estimates

The table reports estimates of the CFVAR$_d$(1) model

$$\Delta^d x_t = \gamma \left[\delta' \left(1 - \Delta^d \right) x_t + \rho' \right] + \sum_{i=1}^{q} \Gamma_i \Delta^d \left(1 - \Delta^d \right)^i x_t + \epsilon_t.$$

Entries denote maximum parameter estimates, following Nielsen and Morin (2012). In addition, we provide heteroskedasitcity robust 95% confidence intervals for the parameter estimates, obtained from 20,000 repetitions of the wild bootstrap.

	estimate	95% confidence interval
d	0.8871	$[0.5072, 0.9874]$
γ	$\begin{pmatrix} -0.0434 & 0.0034 \\ -0.0312 & -0.0249 \\ 0.1487 & 0.0267 \end{pmatrix}$	$\begin{pmatrix} \text{[-0.1502,0.0398]} & \text{[-0.0568,0.0751]} \\ \text{[-0.0952,0.0328]} & \text{[-0.0909,0.0092]} \\ \text{[0.0292,0.2896]} & \text{[-0.0701,0.1273]} \end{pmatrix}$
δ	$\begin{pmatrix} 2 & 0 \\ -1 & 1 \\ -1.2369 & -1.0076 \end{pmatrix}$	$\begin{pmatrix} - & - \\ - & - \\ \text{[-1.4570,-0.9474]} & \text{[-1.9407,-0.3368]} \end{pmatrix}$
ρ'	$\begin{pmatrix} 0.1053 \\ -0.1490 \end{pmatrix}$	$\begin{pmatrix} \text{[0.0073,0.1881]} \\ \text{[-0.4184,0.1630]} \end{pmatrix}$
Γ_1	$\begin{pmatrix} 0.0981 & 0.2012 & 0.0382 \\ 0.2312 & -0.0019 & -0.0207 \\ 0.5240 & -0.0611 & -0.1946 \end{pmatrix}$	$\begin{pmatrix} \text{[-0.1972,0.7501]} & \text{[-0.1286,0.4824]} & \text{[-0.1022,0.1734]} \\ \text{[0.0321,0.4292]} & \text{[-0.2320,0.7517]} & \text{[-0.1178,0.0694]} \\ \text{[0.1041,1.0584]} & \text{[-0.4524,0.2973]} & \text{[-0.3604,0.1058]} \end{pmatrix}$
Σ	$\begin{pmatrix} 0.0019 & 0.0010 & 0.0015 \\ 0.0010 & 0.0010 & 0.0008 \\ 0.0015 & 0.0008 & 0.0038 \end{pmatrix}$	
$R^2_{y^{(24)}}$	0.9712	$[0.9120, 0.9951]$
$R^2_{y^{(120)}}$	0.9789	$[0.9237, 0.9963]$
R^2_r	0.9400	$[0.8358, 0.9882]$

Table 3: Model confidence sets for yield forecasts

The table reports the members of a 95% model confidence set for out-of-sample yield forecasts, indicated by a tick. The target is a time series of yields that are averaged accross 50 maturities. The length of the block bootstrap window is set to six, i.e. half a year, which corresponds to Newey-West's rule of thumb. The results are based on 10,000 resampled values. The loss function is the mean-sqaured error and the test statistic is $maxT$ (see, Hansen et al. (2003) and Hansen et al. (2011)). Forecasts are produced from eight different candidate models, indiciated by numbers (1)-(8), which correspond to:

(1) Affine term-structure model (1)-(13), where x_t is modeled by CFVAR$_d$(1) and d is re-estimated in each rolling sample

(2) Affine term-structure model (1)-(13), where x_t is modeled by CFVAR$_d$(1) and d is fixed at its full-sample value

(3) Affine term-structure model (1)-(13), where x_t is modeled by VAR(2)

(4) Affine term-structure model (1)-(13), where Δx_t is modeled by VAR(1)

(5) Diebold and Li (2006) model, where b_t is modeled by VAR(1)

(6) Diebold and Li (2006) model, where $b_{i,t}$ is modeled by AR(1)

(7) Random Walk

(8) Historical Mean

Forecast Horizon	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 month					✓		✓	
3 months					✓		✓	
6 months					✓		✓	
1 year	✓				✓	✓	✓	
2 years	✓	✓			✓	✓	✓	
3 years	✓	✓			✓	✓	✓	
4 years	✓	✓				✓	✓	
5 years	✓	✓					✓	
6 years	✓	✓					✓	
7 years	✓	✓					✓	
8 years	✓	✓					✓	
9 years	✓	✓					✓	
10 years	✓	✓					✓	

Figures

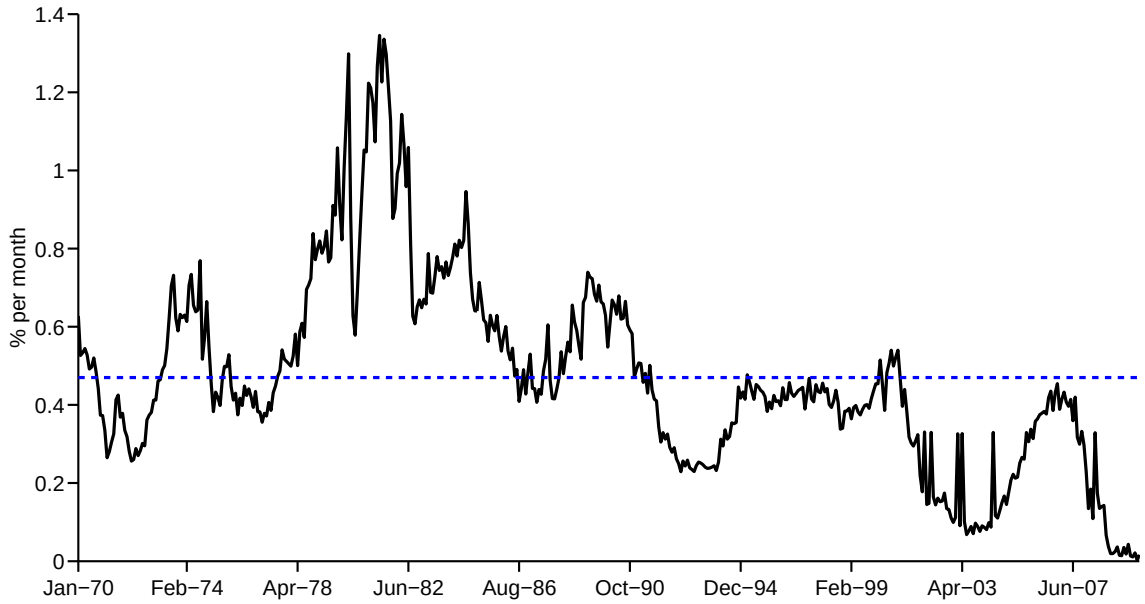


Figure 1: **Evolution of the US 1-month zero-coupon U.S. government bond yields** - The figure plots the continuously compounded monthly real yield of a the nominal U.S. government zero-coupon bond yield with maturity one month for the period 01.1970 - 12.2009 (black line). The blue dashed line is the time-series average of the series, computed over the same sample period.

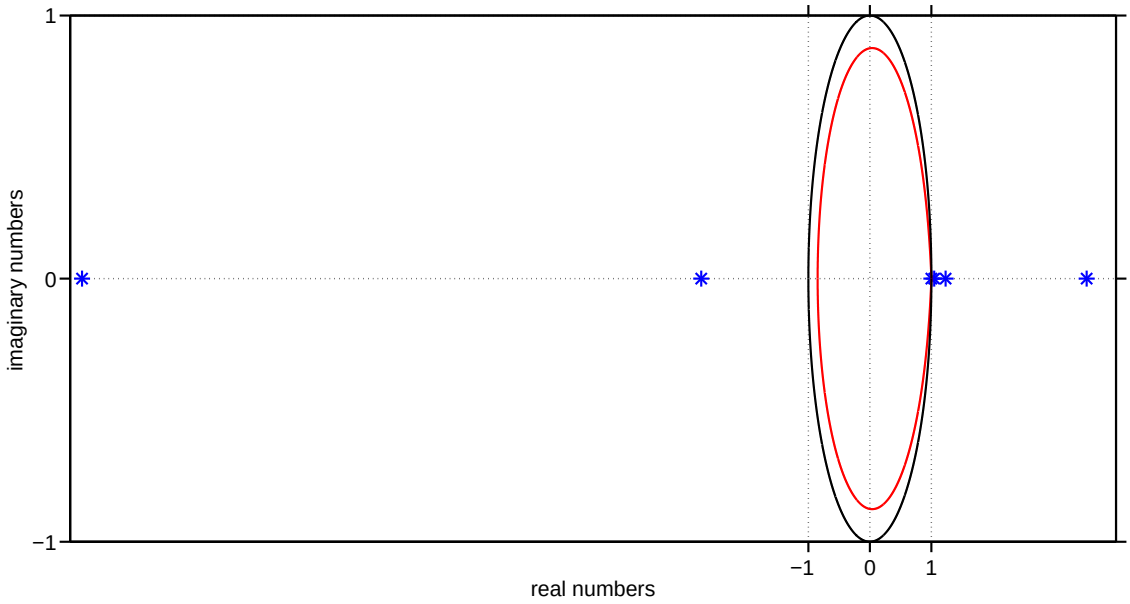


Figure 2: **Roots of the characteristic polynomial** - The figure plots the roots of the characteristic equation $|(1-z)I_{K\times K} - \gamma\delta' z - \Gamma_1(1-z)z| = 0$, indicated by the blue stars. The red line is the image of the complex disk $\mathbb{C}_d$, for $d = 0.8871$. The black line represents the image of the unit disk. For invertibility of the co-fractional VAR, all roots must be equal to one or lie out side the disk $\mathbb{C}_{max(d,1)}$.

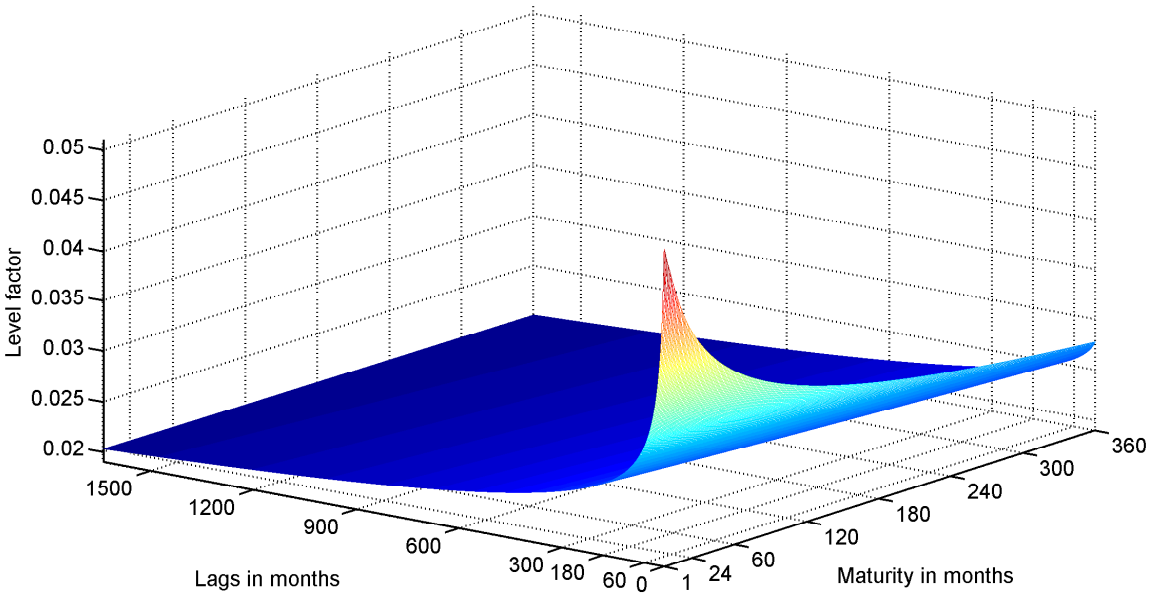


Figure 3: **Level factor** - The figure plots the impact of an orthogonalized permanent shock to the state vector on yields. The effect is computed as the first element of the vector $\bar{V}_j^{(n)} = (1/n)\sum_{i=0}^{n-1} \bar{\Phi}'_{j+i} e3$, where $\bar{\Phi}_j$ are the MA coefficients orthogonal shocks that have a permanent-transitory decomposition, as defined in Section 4.1.

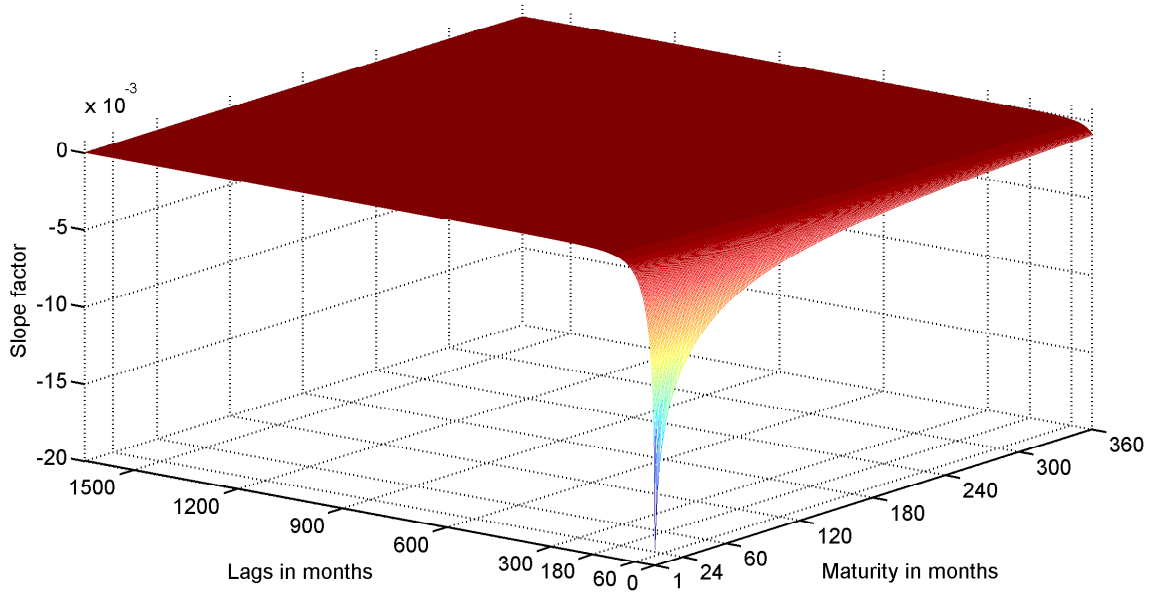


Figure 4: **Slope factor** - The figure plots the impact of an orthogonalized transitory shock to the state vector on yields. The effect is computed as the third element of the vector $\bar{V}_j^{(n)} = (1/n)\sum_{i=0}^{n-1} \bar{\Phi}'_{j+i} e3$, where $\bar{\Phi}_j$ are the MA coefficients orthogonal shocks that have a permanent-transitory decomposition, as defined in Section 4.1.

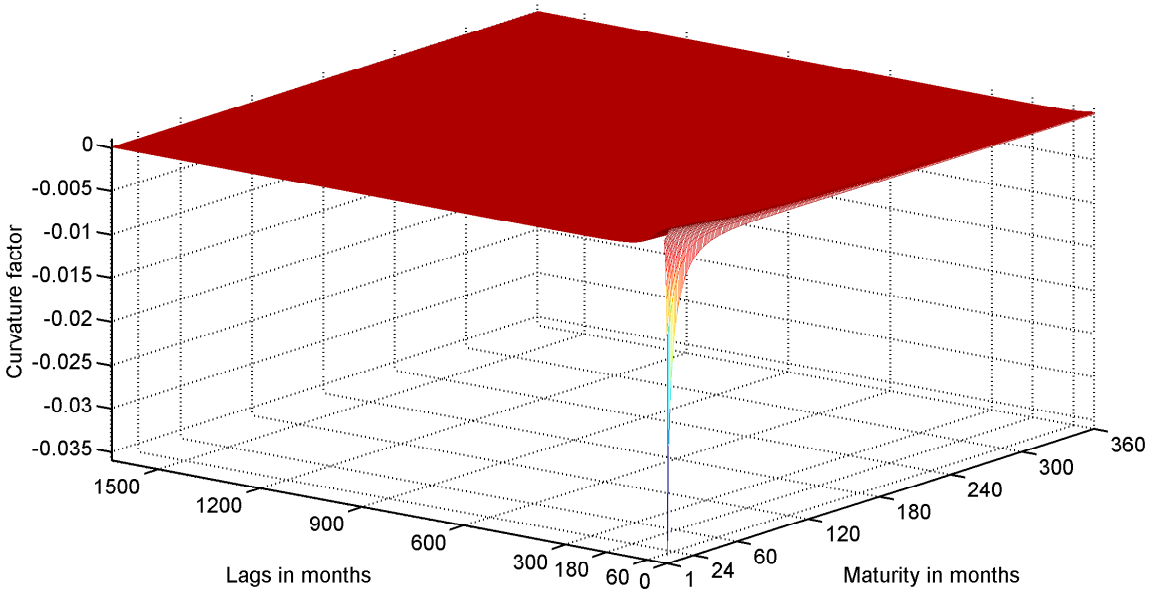


Figure 5: **Curvature factor** - The figure plots the impact of an orthogonalized transitory shock to the state vector on yields. The effect is computed as the second element of the vector $\bar{V}_j^{(n)} = (1/n)\sum_{i=0}^{n-1} \bar{\Phi}'_{j+i} e3$, where $\bar{\Phi}_j$ are the MA coefficients orthogonal shocks that have a permanent-transitory decomposition, as defined in Section 4.1.

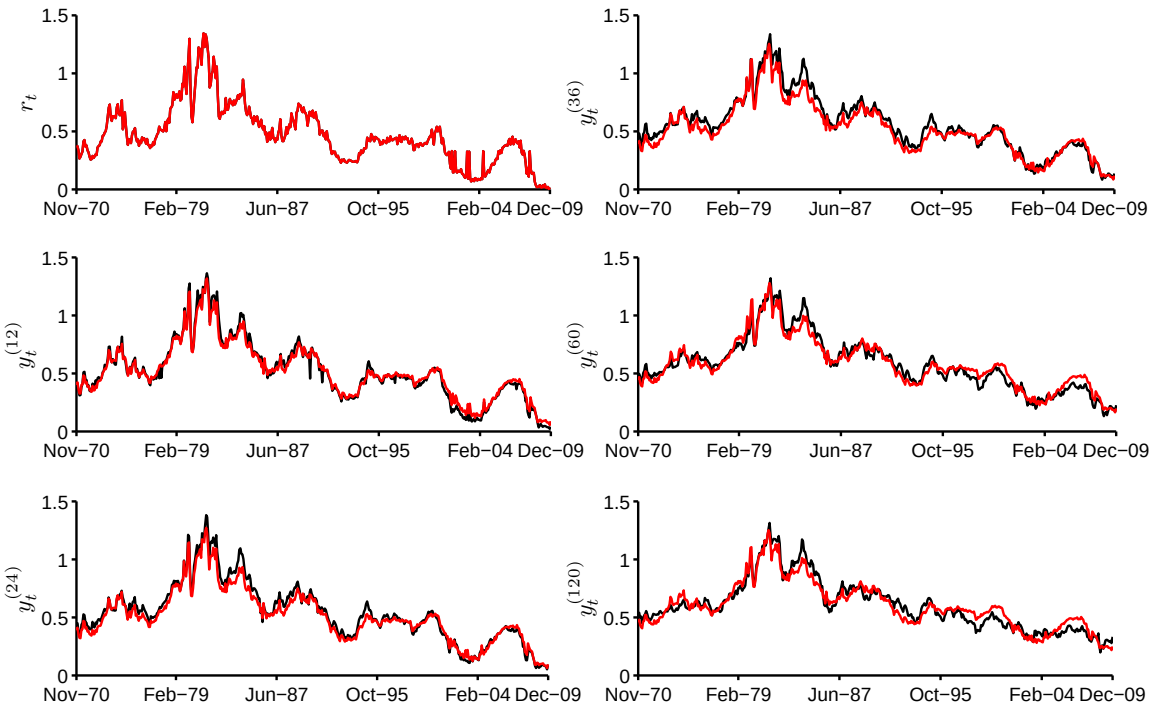
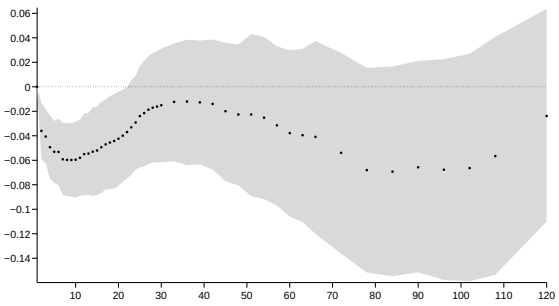
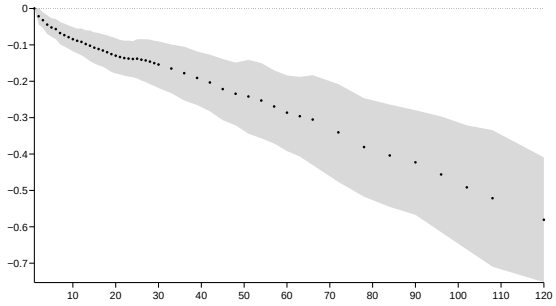


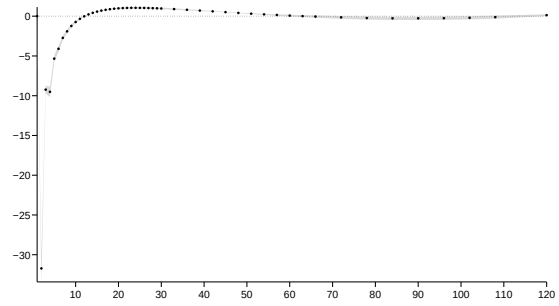
Figure 6: **Observed and model-implied yields** - The figure plots real yields of observed continuously compounded zero-coupon bonds against their model-implied counterparts. The figure depicts yields over the sample period Jan-1970 to Dec-2009 at maturities $n = \{1, 12, 24, 36, 60, 120\}$ months. A black line represents the observed yield, and a red line corresponds to the modeled yield. Modeled yields result from a CFVAR$_d(1)$ model for the state vector, and are computed as in Equations (1)-(13).



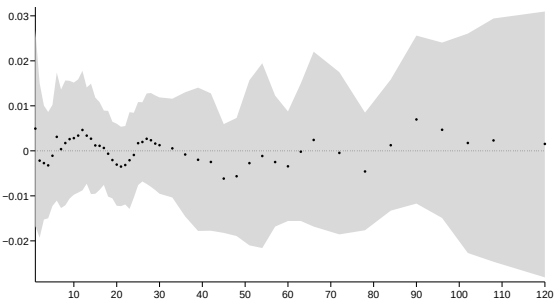
(i) x_t: CFVAR$_d$(1) and $y^{(n)}_{MODEL,t}$: Model (13)



(ii) x_t: VAR(2) and $y^{(n)}_{MODEL,t}$: Model (13)

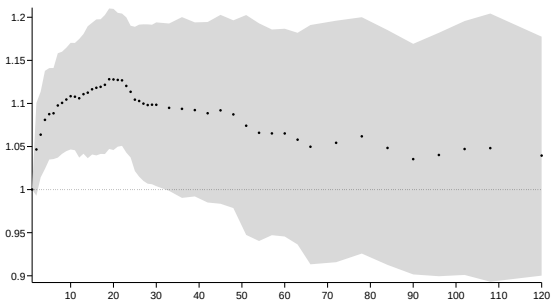


(iii) Δx_t: VAR(1) and $y^{(n)}_{MODEL,t}$: Model (13)

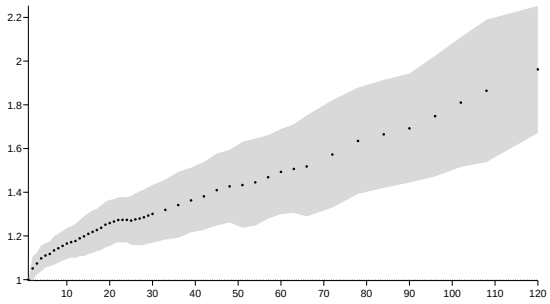


(iv) b_t: VAR(1) and $y^{(n)}_{MODEL,t}$: DL model

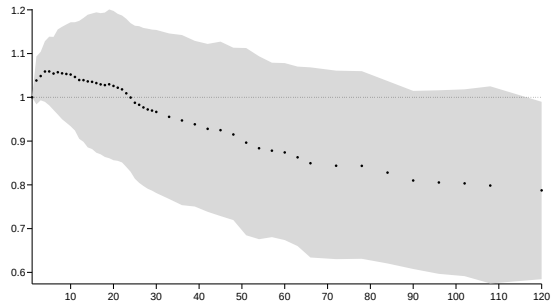
Figure 7: **Estimates for $\alpha^{(n)}$ in** (24) - The figure plots the estimate for the intercept in Regression (24) and the corresponding 95% confidence interval. The estimates are depicted by black dots, and the confidence intervals by gray shaded areas. The x-axis represents maturities n in months, and the y-axis resembles the size of the estimate for $\alpha^{(n)}$. In Figures (i), (ii), (iii), yields are estimated from the term-structure model in Section 2, whereas in Figure (iv) they result from a Diebold and Li (2006) model (DL). The state vector x_t is estimated by a CFVAR$_d$(1) in (i), a VAR(2) in (ii), and a VAR(1) in first differences in (iii). In Figure (iv), factors b_t follow a VAR(1).



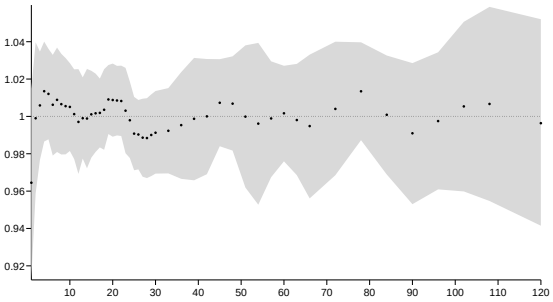
(i) x_t: CFVAR$_d$(1) and $y^{(n)}_{MODEL,t}$: Model (13)



(ii) x_t: VAR(2) and $y^{(n)}_{MODEL,t}$: Model (13)



(iii) Δx_t: VAR(1) and $y^{(n)}_{MODEL,t}$: Model (13)



(iv) b_t: VAR(1) and $y^{(n)}_{MODEL,t}$: DL model

Figure 8: **Estimates for $\beta^{(n)}$ in** (24) - The figure plots the estimate for the slope in Regression (24) and the corresponding 95% confidence interval. The estimates are depicted by black dots, and the confidence intervals by gray shaded areas. The x-axis represents maturities n in months, and the y-axis resembles the size of the estimate for $\beta^{(n)}$. In Figures (i), (ii), (iii), yields are estimated from the term-structure model in Section 2, whereas in Figure (iv) they result from a Diebold and Li (2006) model (DL). The state vector x_t is estimated by a CFVAR$_d$(1) in (i), a VAR(2) in (ii), and a VAR(1) in first differences in (iii). In Figure (iv), factors b_t follow a VAR(1).

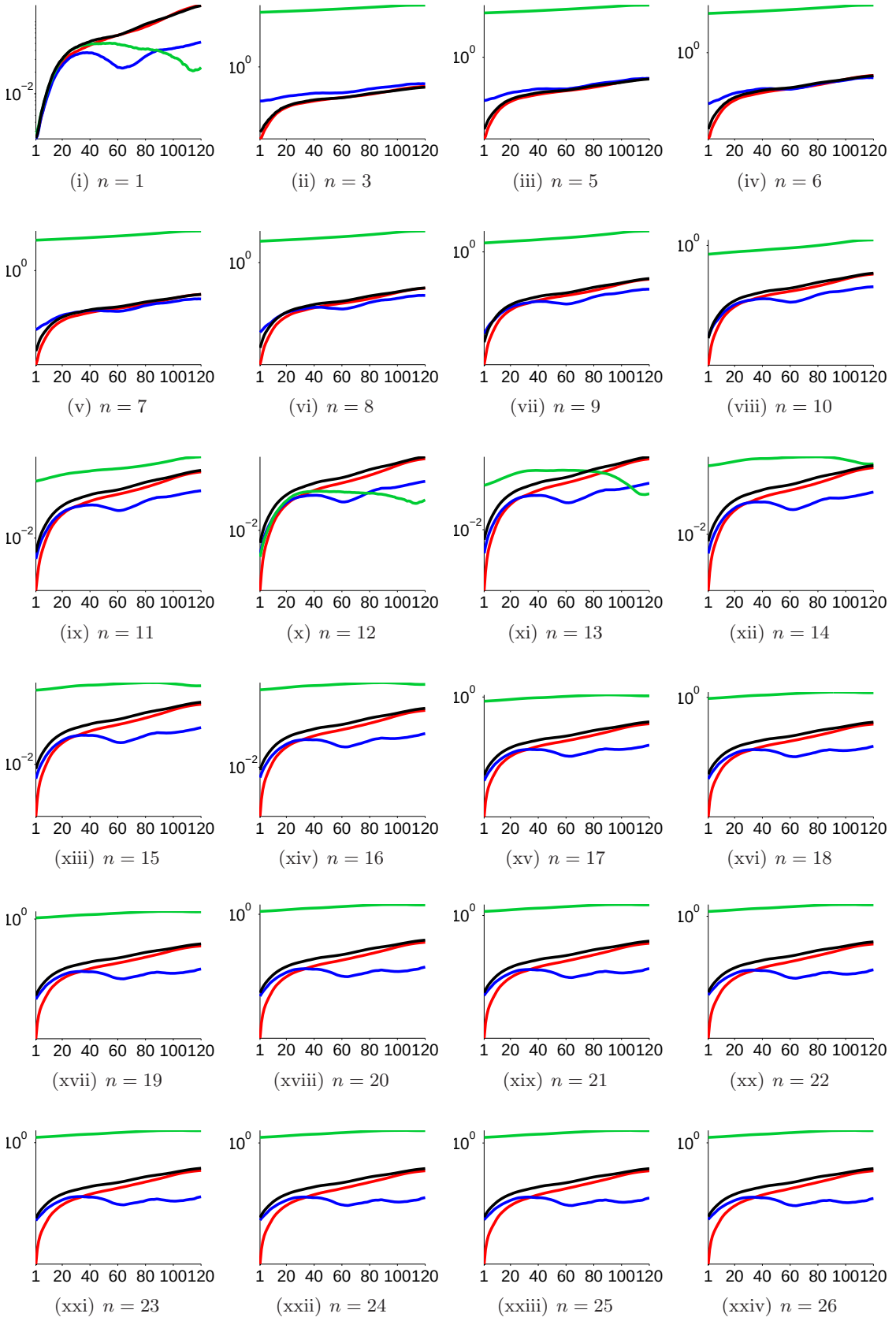


Figure 9: **Out-of-Sample Forecast MSE** - The figure plots MSEs from yield curve forecasts for maturities $n \in [1, 26]$. The red line corresponds to yield forecasts from the DL model, where factors b_t follow a VAR(1). The other lines result from yield forecasts from the term-structure model in Section 2. x_t is estimated by a CFVAR$_d$(1), indicated by a blue line; a VAR(2), represented by a black line; and a VAR(1) in first differences, indicated by a green line. The x-axis represents the forecast horizon k in months. The y-axis has log scale.

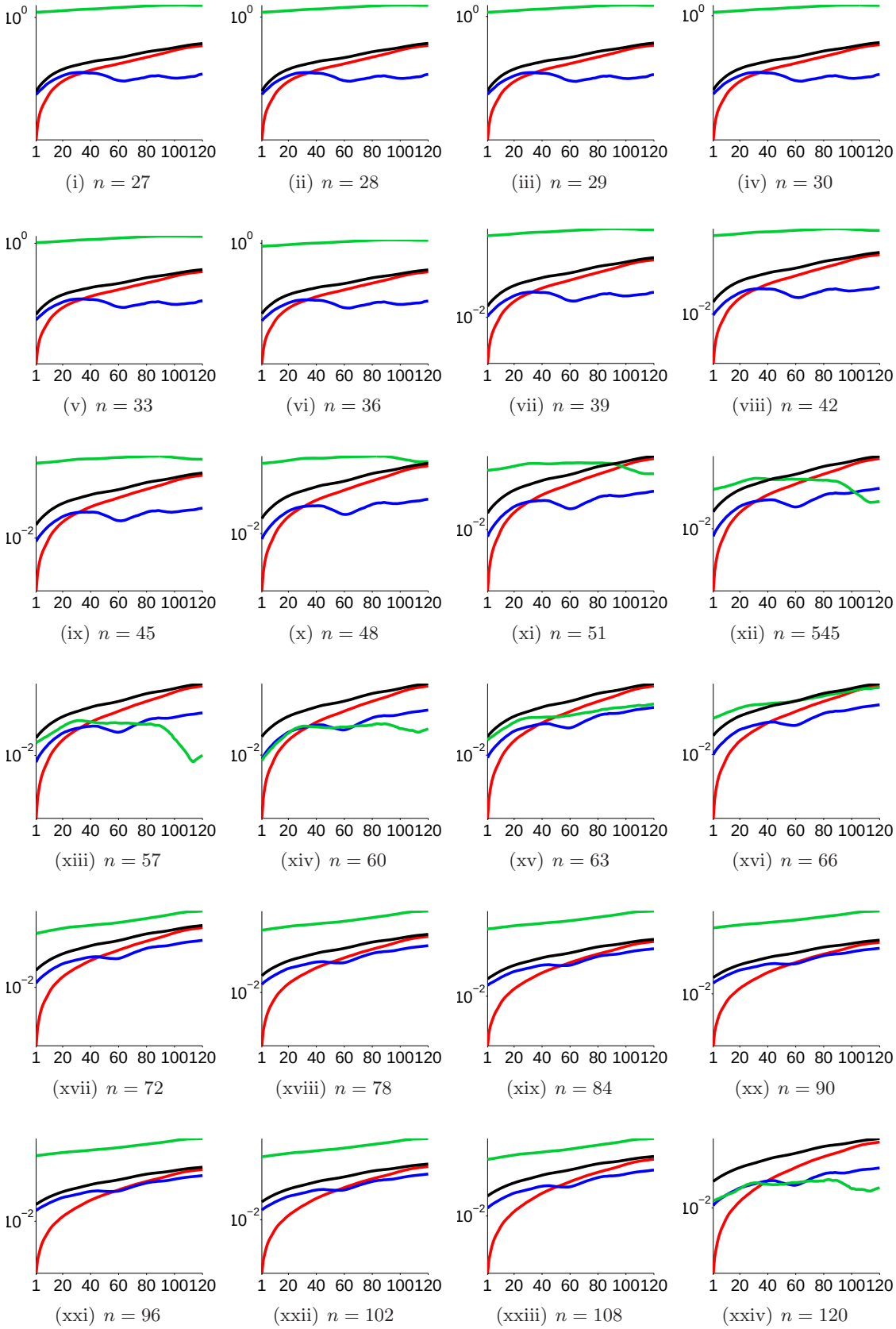


Figure 10: **Out-of-Sample Forecast MSE (cont'd)** - The figure plots MSEs from yield curve forecasts for maturities $n \in [27, 120]$. The red line corresponds to yield forecasts from the DL model, where factors b_t follow a VAR(1). The other lines result from yield forecasts from the term-structure model in Section 2. x_t is estimated by a CFVAR$_d$(1), indicated by a blue line; a VAR(2), represented by a black line; and a VAR(1) in first differences, indicated by a green line. The x-axis represents the forecast horizon k in months. The y-axis has log scale.

Research Papers 2013

2012-58: Tom Engsted and Thomas Q. Pedersen: Predicting returns and rent growth in the housing market using the rent-to-price ratio: Evidence from the OECD countries

2013-01: Mikko S. Pakkanen: Limit theorems for power variations of ambit fields driven by white noise

2013-02: Almut E. D. Veraart and Luitgard A. M. Veraart: Risk premia in energy markets

2013-03: Stefano Grassi and Paolo Santucci de Magistris: It's all about volatility (of volatility): evidence from a two-factor stochastic volatility model

2013-04: Tom Engsted and Thomas Q. Pedersen: Housing market volatility in the OECD area: Evidence from VAR based return decompositions

2013-05: Søren Johansen and Bent Nielsen: Asymptotic analysis of the Forward Search

2013-06: Debopam Bhattacharya, Pascaline Dupasand Shin Kanaya: Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets

2013-07: Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns

2013-08: Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation

2013-09: Roberto Casarin, Stefano Grassi, Francesco Ravazzolo and Herman K. van Dijk: Parallel Sequential Monte Carlo for Efficient Density Combination: The Deco Matlab Toolbox

2013-10: Hendrik Kaufmann and Robinson Kruse: Bias-corrected estimation in potentially mildly explosive autoregressive models

2013-11: Robinson Kruse, Daniel Ventosa-Santaulària and Antonio E. Noriega: Changes in persistence, spurious regressions and the Fisher hypothesis

2013-12: Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications

2013-13: Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions

2013-14: Charlotte Christiansen, Jonas Nygaard Eriksen and Stig V. Møller: Forecasting US Recessions: The Role of Sentiments

2013-15: Ole E. Barndorff-Nielsen, Mikko S. Pakkanen and Jürgen Schmiegel: Assessing Relative Volatility/Intermittency/Energy Dissipation

2013-16: Peter Exterkate, Patrick J.F. Groenen, Christiaan Heij and Dick van Dijk: Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression

2013-17: Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model