Netspar

Network for Studies on Pensions, Aging and Retirement

Netspar DISCUSSION PAPERS

Daniela Osterrieder and Peter Schotman

The Volatility of Long-Term Bond Returns

Persistent Interest Shocks and Time-Varying Risk Premiums

DP 08/2012-060

The Volatility of Long-term Bond Returns: Persistent Interest Shocks and Time-varying Risk Premiums*

Daniela Osterrieder† and Peter C. Schotman‡

August 3, 2012

Abstract: We develop a model that can match two stylized facts of the term-structure. The first stylized fact is the predictability of excess returns on long-term bonds. Modeling this requires sufficient volatility and persistence in the price of risk. The second stylized fact is that long-term yields are dominated by a level factor, which requires persistence in the spot interest rate. We find that a fractionally integrated process for the short rate plus a fractionally integrated specification for the price of risk leads to an analytically tractable almost affine term structure model that can explain the stylized facts. In a decomposition of long-term bond returns we find that the expectations component from the level factor is more volatile than the returns themselves. It therefore takes a volatile risk premium that is negatively correlated with innovations in the level factor to explain the volatility of long-term bond returns. The model also implies that excess bond returns do not exhibit mean reversion, consistent with the empirical evidence.

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

JEL codes: C58, G12, C32

*We thank the participants Quantitative Methods in Finance Conference in Sydney for their helpful comments. We also thank Antoon Pelsser, Jean-Pierre Urbain, Rolf Tschernig, and Stefan Straetmans for their constructive feedback. This research was financially supported by a grant of the Dutch Organization for Scientific Research (NWO).

†CREATES and Aarhus University, Bartholins Allé 10, 8000 Aarhus C, Denmark.
Email: dosterrieder@creates.au.dk

‡Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
Email: p.schotman@maastrichtuniversity.nl

Electronic copy available at: http://ssrn.com/abstract=2123412

1 INTRODUCTION

Affine term structure models assume a stationary vectorautoregressive (VAR) process for the factors that drive interest rates. Time series estimates of the VAR typically imply that long-term expectations of future spot rates have very little variation. With constant risk premiums long-term yields would then have very low volatility. In the data long-term yields are, however, almost as volatile as short-term yields. To explain this volatility in the data risk premiums need to be very volatile. This implication hinges on the estimated persistence in the VAR model.[1] To explain the volatility puzzle without highly volatile risk premiums mean reversion in interest rates must be very slow, and hence the VAR models need a near unit root.

The sensitivity with respect to near unit root parameters is illustrated in Cochrane and Piazzesi (2008) and Jardet, Monfort and Pegoraro (2011). Both studies compare the long-run forecasts of the short-term interest rate from estimated VAR models with and without imposing cointegration among yields of different maturities. These long-run forecasts are very different. For the stationary VAR they are close to the constant unconditional mean of the spot rate, whereas the cointegrated model produces forecasts that are very close to the current level of the spot rate.

Affine models are usually estimated using data with maturities up to 10 years, as this is often the longest maturity for which long time series with accurately measured yield data are available. For many applications, like the valuation of long-dated liabilities of insurance companies and pension funds, or managing the risk of portfolios of mortgage loans, one needs to extrapolate the yield curve and its volatility to maturities of thirty years and more. Both extrapolations are very sensitive to the near unit-root parameter in the VAR model.

How to deal with this sensitivity remains problematical, however. Estimates of the largest autoregressive root are biased downwards due to the well-known Kendall bias,

[1] See Bauer (2011) for a recent analysis of the relation between the *excess volatility puzzle* and persistent interest rate shocks. The puzzle has a long history. In his seminal paper Shiller (1979) already conjectured a link between volatility and unit roots: '... we have no real information in small samples about possible trends or long cycles in interest rates. Indeed, [...] short-term interest rates may be unstationary' (p. 1213).

Electronic copy available at: http://ssrn.com/abstract=2123412

which is exacerbated in multivariate systems (see Abadir, Hadri and Tzavalis (1999)). Recently various bias-adjustment procedures have been proposed.[2] Instead of further refining the estimators of (Gaussian essentially) affine models, we extend the class of models to include fractionally integrated processes. Fractional integration, denoted $I(d)$, is a parsimonious and flexible means to model the long memory properties of interest rate dynamics, as it allows a smooth transition between stationary $I(0)$ processes and non-stationary unit-root $I(1)$ processes. The fractional model can generate forecasts that are in between the stationary and cointegrated models.

We see three motivations for applying fractional integration to model the term structure of interest rates. First, many studies have estimated the fractional integration parameter of interest rate time series and report that the order d is between 0.8 and 1, but significantly different from zero.[3] For our empirical data of 58 years of monthly observations we confirm the general result in the literature and estimate the order of integration as $d = 0.89$. That means that the level of interest rates is non-stationary, but less persistent than a random walk or $I(1)$ process.

Second, fractional integration models are linear and therefore analytically tractable. We show that one obtains a closed-form solution for the term structure for any Gaussian linear process for the spot rate combined with any Gaussian linear process for the price of risk, jointly driven by K shocks. For excess returns the solution is K-factor affine. The solution is very similar to the Gaussian essentially affine model of Duffee (2002). Because it leads to an affine term structure, we prefer the fractional model to other model classes that can also generate long-memory like behavior such as regime switching models. As Diebold and Inoue (2001) have shown, a fractionally integrated model provides a good approximation for long-run predictions for time series that are subject to occasional

[2] Bauer, Rudebusch and Wu (2011) develop a bootstrap adjustment for the VAR parameters under the $\mathbb{P}$ measure. Jardet *et al* (2011) suggest to take a weighted average of the stationary and cointegrated forecasts. Joslin, Priebsch and Singleton (2010) impose the condition that the largest eigenvalue under $\mathbb{Q}$ is equal to the largest eigenvalue under $\mathbb{P}$. De Wachter and Lyrio (2006) impose the unit root under the $\mathbb{P}$ measure, whereas Christensen, Diebold and Rudebusch (2011) impose the unit root under the $\mathbb{Q}$ measure. Cochrane and Piazzesi (2008) estimate persistence under the $\mathbb{Q}$ measure and infer the persistence under $\mathbb{P}$ by constraining the specification for risk prices.

[3] See, *e.g.,* Shea (1991), Gil-Alana (2004), Sun and Phillips (2004), Iacone (2009), and Gil-Alana and Moreno (2012).

breaks in the mean. Connolly, Güner and Hightower (2007) further demonstrate that a long-memory model for the short rate may describe the series more accurately than a structural change model.

Third, fractional models have been shown to fit certain characteristics of the cross section better than stationary models. Our approach has been motivated by Backus and Zin (1993), who assumed a fractional ARFIMA(p,d,q) time-series process for the spot rate together with a constant risk premium. With this model they succeed in matching the observed mean and volatility of yields. We extend their work by relaxing two of their assumptions. First, we relax their assumption that interest rates must be stationary with $d < \frac{1}{2}$. The assumption was necessary for their tests of unconditional moments of yield levels, but not required for our tests on the volatility of excess returns. Indeed, most time-series estimates of the order of integration indicate that interest rate levels are non-stationary with $d > \frac{1}{2}$. Second, we allow for time-varying risk premiums, both because there is strong empirical evidence that excess returns have a predictable component, and because these may be an important source of volatility in bond returns.[4]

A related approach is the shifting endpoints model of Kozicki and Tinsley (2001). They specify long-horizon expectations of the short-term interest rate as a (nonlinear) function of inflation and inflation expectations. The shifting endpoint serves as the 'level' factor for the term structure. Kozicki and Tinsley (2001) argue that this factor is successful in tracking long-maturity yields. The long-memory properties of the factor are due to the non-stationarity of inflation and in some specifications to regime shifts in the monetary policy target inflation rate. As an alternative, Kozicki and Tinsley (2001) also explicitly model the endpoint as the result of a unit root process for the short-term interest rate. Empirically this model performs worse than the shifting endpoints model. We therefore interpret their evidence as an indication that fractional models may provide the right amount of persistence. As with Backus and Zin (1993), an important difference

[4] See Fama (1984) and Fama and Bliss (1987) for early evidence that excess long-term bond returns have a small but significant predictable component. Campbell and Shiller (1991) extensively document that high spreads predict higher than average excess returns. More recently, Cochrane and Piazzesi (2005) capture approximately 40 percent of the variation in bond returns with a *return forecasting factor*, defined as a linear combination of forward premiums of different maturities.

between our approach and Kozicki and Tinsley (2001) is that we allow for time-varying risk premiums.

For the price of risk we also consider a fractional model. This deviates both from the affine literature and much of the cointegration literature, which generally relates the risk premium to yield spreads and forward premiums. In the literature it is mostly assumed that spreads are $I(0)$.[5] Recent time series tests reach a different conclusion. Both Chen and Hurvich (2003) and Nielsen (2010) find that spreads have a fractional order that is significantly larger than zero, but with point estimates that are less than a half. This means that spreads and risk premiums are still stationary, but more persistent than an $I(0)$ process. For our data we confirm these time-series estimates. More importantly, we obtain the same order of integration for expected excess returns when we estimate the persistence from the cross-sectional factor loadings of long-maturity bond returns. Both lead to a value of $d \approx 0.4$. Risk premiums are therefore stationary.

Our focus is on long-maturity bonds. At long maturities only the low-frequency components in the 'level' factor matter and we can work with a single-factor model as in Backus and Zin (1993) and Kozicki and Tinsley (2001). We use a two-stage estimation procedure. In the first stage we estimate the fractional order of integration d of the short-term interest rate. In the second stage we use moments related to the volatility and predictability of excess returns on discount bonds with 5- and 10-years to maturity to estimate the parameters of the risk-price process.

When we compare term-structure estimates for different orders of integration, we find a strong interaction between the persistence of the 'level' factor and the risk premium in long-maturity bond returns. With low persistence in the 'level' factor, the correlation between risk prices and the spot rate is positive. When persistence is above the threshold of $d = 0.7$, the correlation changes sign. In this case the expectations driven part of the volatility of excess returns is larger than the total volatility. This implies that the covariance between risk premiums and changes in expectations must be negative to match the volatility in the data. Our estimate of the persistence of the level factor is $d = 0.89$. This estimate implies the negative correlation and is consistent with stylized

[5] See the large literature following Campbell and Shiller (1987).

facts in the data like the regressions of excess returns on lagged spreads or on a prediction factor. With lower values of fractional integration the model is not able to replicate these regression results in the data.

2 LONG-MEMORY AFFINE TERM STRUCTURES

2.1 An essentially affine model for general linear time series processes

Our model is a generalization of a discrete-time, Gaussian, essentially affine model for the term structure of interest rates developed by Duffee (2002). Two differences are important: (*i*) the dynamics for the short-term interest rate can be more general than a VAR and (*ii*) the dynamic structure for the price of risk can be different from the lag structure of the factors.

We assume that the one-period spot rate, r_t, can be represented by the linear MA specification

$$r_t = \mu_r + \sum_{j=0}^{\infty} c_j' \epsilon_{t-j}, \tag{1}$$

where c_j and ϵ_t are vectors of length K and μ_r is a scalar constant. We assume that ϵ_t is a normal, independently and identically distributed innovation, with mean zero and covariance matrix Σ. The c_j coefficients are the impulse responses of the short rate with respect to the shocks ϵ_{t+1}. The general formulation in (1) encompasses both stationary finite order VAR models as well as fractionally integrated processes depending on the assumptions on the c_j sequence.

The logarithmic stochastic discount factor $m_{t+1} = \ln M_{t+1}$ is specified as

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

where λ_t is a K-vector of risk prices following the linear process

$$\Sigma\lambda_t = \Sigma\mu_\lambda + \sum_{j=0}^{\infty} F_j'\epsilon_{t-j}, \tag{3}$$

with μ_λ a K-vector and F_j matrices of coefficients of order $(K \times K)$.

Specification (1)-(3) reduces to an essentially Gaussian affine term-structure model if both the spot rate and the risk prices are affine in K state variables, X_t, which follow a first-order VAR with coefficient matrix $\mathcal{A}$. With $r_t = \delta_0 + \delta_1' X_t$ and $\Sigma\lambda_t = \Lambda_0 + \Lambda_1 X_t$, the MA coefficients for the spot rate will then be $c_j' = \delta_1'\mathcal{A}^j$ and the risk dynamics follow as $F_j' = \Lambda_1\mathcal{A}^j$. The coefficients c_j and F_j have the same rate of decay, determined by the VAR coefficient matrix $\mathcal{A}$. In the current model c_j and F_j can be unrelated and do not have to decline exponentially at the same rate. In particular, we will consider models in which a shock $\epsilon_{i,t}$ has a permanent effect on the spot rate, but where the same shock has only a temporary effect on the price of risk. We model the spot rate as a non-stationary fractionally integrated process, whereas the price of risk remains stationary.

Prices of discount bonds of maturity n are denoted $P_t^{(n)}$. Log bond prices are $p_t^{(n)} = \ln P_t^{(n)}$ and continuously compounded yields are $y_t^{(n)} = -p_t^{(n)}/n$. Given the pricing assumptions (1)-(3) the full term structure can be derived recursively using the basic pricing equation

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

which in logarithms becomes

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

with initial condition $p_t^{(0)} = 0$. The solution is summarized in theorem 1 (see appendix for all derivations)

Theorem 1. *The logarithmic price of a discount bond of maturity n is*

$$p_t^{(n)} = -a^{(n)} - \sum_{j=0}^{\infty} b_j^{(n)\prime}\epsilon_{t-j} \tag{6}$$

with recursively defined coefficients

$$b_j^{(n+1)} = c_j + b_{j+1}^{(n)} + F_j b_0^{(n)} \tag{7}$$

$$a^{(n+1)} = \mu_r + a^{(n)} - \tfrac{1}{2}b_0^{(n)\prime}\Sigma b_0^{(n)} + b_0^{(n)\prime}\Sigma\mu_\lambda. \tag{8}$$

and initial conditions $a^{(1)} = \mu_r$ and $b_j^{(1)} = c_j$.

Most of our analysis will be based on the excess returns

$$\begin{aligned} rx_{t+1}^{(n+1)} &= p_{t+1}^{(n)} - p_t^{(n+1)} - r_t \\ &= p_{t+1}^{(n)} - \mathsf{E}_t\left[p_{t+1}^{(n)}\right] - \mathsf{Cov}_t\left[m_{t+1}, p_{t+1}^{(n)}\right] - \tfrac{1}{2}\mathsf{Var}_t,\left[p_{t+1}^{(n)}\right] \end{aligned} \tag{9}$$

where the second line in (9) is obtained through direct substitution of (5). For an explicit solution we use (6) with the coefficients in (7). The result for $rx_{t+1}^{(n+1)}$ is in theorem 2.

Theorem 2. *Let the spot rate be generated by the linear process* (1) *and risk prices be generated by the linear process* (3)*. Then excess returns on discount bonds have the factor structure*

$$rx_{t+1}^{(n+1)} = b_0^{(n)\prime}\left(-\epsilon_{t+1} + \Sigma\lambda_t\right) - \tfrac{1}{2}b_0^{(n)\prime}\Sigma b_0^{(n)} \tag{10}$$

with factor loadings obeying $b_0^{(1)} = c_0$*, and*

$$b_0^{(n)} = C_{n-1} + \sum_{i=1}^{n-1} F_{n-1-i} b_0^{(i)}, \qquad n > 1, \tag{11}$$

where $C_n = \sum_{i=0}^{n} c_i$ *denote cumulative impulse responses of the spot rate process.*

Excess returns have three components. First, the shocks ϵ_{t+1} enter with factor loadings $b_0^{(n)}$. The recursion (11) for these loadings includes both the effects of the short rate process through the c_j terms as well as the risk price dynamics through the F_j terms. The second element is the predicted excess return, which is linear in $\Sigma\lambda_t$ with the same factor loading $b_0^{(n)}$. The last term is the Jensen-inequality adjustment $\frac{1}{2}b_0^{(n)\prime}\Sigma b_0^{(n)}$. The general structure (10) is the same as for the Gaussian essentially affine model class. What is different are the coefficients $b_0^{(n)}$.

Time-varying risk premiums have two effects on the volatility of excess returns. First, given factor loadings $b_0^{(n)}$ they add volatility through the term $\Sigma\lambda_t$ within the common factors. Since ϵ_{t+1} is by assumption orthogonal to $\Sigma\lambda_t$, this will always increase the volatility relative to a model with constant risk premiums. Second, the time series process of the risk prices affects the factor loadings. The second term in (11) depends on the persistence of shocks to the price of risk. The interaction with the expectations effect C_{n-1} crucially depends on the sign and the rate of decline of the F_i matrices. A shock $\epsilon_{i,t+1}$ that increases both the spot interest rate ($c_{i,0} > 0$) and the price of its

risk ($F_{ii,0} > 0$), will have a bigger impact on the factor loadings than a shock that has opposite effects on the two processes. In the latter case, factor loadings will be smaller than under the expectations hypothesis and the overall volatility of excess returns may be less than in a model with a constant risk premium.[6]

Because of theorem 2 it is much more convenient to work with excess returns than yield levels. With excess returns we still maintain a low-dimensional factor structure, whereas the yields (or prices) in (6) do not allow any factor structure at all for more general linear time series processes.

Theorem 2 is important for models where the factor dynamics follow a process that differs from a first order VAR. Consider, for example, a VAR with longer lags like in Ang and Piazzesi (2003),

$$X_t = \mu_X + \sum_{i=1}^{p} A_i(X_{t-i} - \mu_X) + \epsilon_t, \tag{12}$$

The normal procedure is to write (12) in companion form by stacking all lags in the augmented state vector $\mathbf{X}_t = (X_t' \; X_{t-1}' \; \ldots \; X_{t-p+1}')'$ and defining the $Kp \times Kp$ companion matrix $\mathcal{A}$ with the $(K \times K)$ matrices A_i on its first K rows. The resulting affine term structure model then has Kp factors for the yields. Theorem 2 implies that the factor structure for the excess returns has only K factors.

Theorem 2 holds for more general time series process than higher order VAR's. In the remainder of the paper we consider processes with long memory.

2.2 A fractionally integrated level factor

The dominant term in the factor loadings is the cumulative impulse response vector C_n. Elements in the $(K \times 1)$ vector C_n will increase faster in n the stronger the persistence of the corresponding shock. It is therefore at long maturities where we should expect to see the largest effects of alternative estimates of persistence. For this reason we will concentrate our analysis on long-maturity bond returns. Likewise we will focus on the

[6] The interaction between the c_j and F_j coefficients could be so strong that one or more elements of $b_0^{(n)}$ become exactly zero. This is the case of a 'hidden' factor as in Duffee (2011). In our model this will not happen for large n, since we assume that $\Sigma\lambda_t$ is stationary, whereas r_t appears to be non-stationary. The C_n term then grows faster than the convolution of $b_0^{(i)}$ and F_{i-n} in (11).

most persistent term structure factor, known as the level factor. Time-series evidence shows that long-term yields are (fractionally) cointegrated and share a single common factor for the low-frequency behavior.[7] From the results of Cochrane and Piazzesi (2005, 2008) we also know that the level factor is responsible for most of the volatility in time-varying risk premiums. In the remainder we specialize our model to the single factor case $K = 1$ with the level factor as the only factor, and test this model on discount bond prices with maturities of 5 years and longer. This is a similar assumption as in Kozicki and Tinsley (2001) and based on the fact that factors with low persistence have negligible effects at the longer end of the yield curve.[8]

With $K = 1$ the short-rate process (1) has scalar coefficients c_j. The variance of the innovations ϵ_t is denoted by σ^2. For the time series process of λ_t we likewise assume that the general process (3) can be specialized to the univariate process,

$$\sigma^2(\lambda_t - \mu_\lambda) = \xi \sum_{j=0}^{\infty} f_j \epsilon_{t-j}, \tag{13}$$

where the f_j are also scalars, normalized by $f_0 = 1$, and where ξ is a scalar parameter that determines both the volatility of the risk premium and the sign of the covariance between r_t and λ_t. If $\xi = 0$, the risk premium is constant.

To allow for long memory we assume that both r_t and λ_t can be fractionally integrated. A fractionally integrated series x_t is generated by

$$(1 - L)^d x_t = u_t, \tag{14}$$

where L denotes the usual lag operator and d is the fractional integration parameter, for which we will assume $0 \leq d \leq 1$. The process u_t is assumed to be stationary $I(0)$ with zero mean and bounded spectral density at the zero frequency. The fractional filter $(1 - L)^d$ is defined as the infinite sum $(1 - L)^d = \sum_{i=0}^{\infty} \Theta(d)_i L^i$, with coefficients

$$\Theta(d)_i = (-1)^i \binom{d}{i} = \prod_{j=0}^{i-1} \frac{j - d}{j + 1}, \qquad i > 0, \tag{15}$$

[7] See Engsted and Tanggaard (1994) for tests on regular cointegration concluding that there is a single $I(1)$ trend in yields of different maturities. Likewise Nielsen (2010) finds a single stochastic trend in yields, but with cointegrating residuals (spreads) that are still fractionally integrated.

[8] Multiple factors, with less persistence than the level factor, create substantial transitory interest rate dynamics and flexible yield curve shapes at short and intermediate maturities. Ignoring these factors of course entails a misspecification of short and intermediate maturities.

and $\Theta(d)_0 = 1$. For the MA representation, which we need for the term structure solutions, the fractional difference operator in (14) can be inverted as

$$x_t = (1-L)^{-d} u_t = \sum_{i=0}^{\infty} \Theta(-d)_i u_{t-i} \tag{16}$$

with coefficients $\Theta(-d)_i = \prod_{j=0}^{i-1} \frac{j+d}{j+1} \geq 0$ following directly from (15). The coefficients $\Theta(-d)_i$ decline hyperbolically at rate i^{d-1}.

For a pure fractional process, *i.e.* when u_t are uncorrelated, with $0 \leq d < \frac{1}{2}$ the variance $S_x(0)$ and autocovariances $S_x(i)$ are given by (see, *e.g.,* Lo (1991)):

$$S_x(0) = \sigma_u^2 \frac{\Gamma(1-2d)}{(\Gamma(1-d))^2} \tag{17}$$

$$S_x(i) = S_x(0) \prod_{j=0}^{i-1} \frac{j+d}{j+1-d}, \qquad i > 0, \tag{18}$$

with $\Gamma(\centerdot)$ the Gamma function.

If $d > \frac{1}{2}$, the solution of (14) for the process x_t is not well defined in mean-square sense. Yet, following Marinucci and Robinson (1999) one can define x_t as a Type II $I(d)$ process,

$$x_t = \sum_{i=0}^{\infty} \Theta(-d)_i u_{t-i} \mathbb{1}_{\{(t-i)>0\}}, \tag{19}$$

where $\mathbb{1}_{\{(t-i)>0\}}$ is the indicator function which is equal to zero for $i \geq t$. It truncates the process at some initial condition $x_0 = 0$. This modification has no effect on the term structure relations once we assume that the process has started far enough before the actual sample.

For both the spot rate and the price of risk we will make specific parametric assumptions to obtain the impulse response sequences c_j and f_j. We will study parametric ARFIMA(p,d,0) specifications. For the spot rate we have the general model

$$\left(1 - \sum_{i=1}^{p} \nu_i L^i\right) (1-L)^{d_r} (r_t - \mu_r) = \epsilon_t \tag{20}$$

The fractional differencing parameter d_r and the AR parameters ν_i are estimated from time series data on the spot rate. Even though the long-run properties of the spot rate depend only on the fractional differencing parameter d_r, we still include the transitory

AR dynamics to facilitate a comparison between stationary autoregressive $I(0)$ and fractional models. It allows us to fit time series models that provide almost identical short-term predictions. The fractional and autoregressive models that we consider only differ in their long-run implications based on evidence about low-frequency dynamics.

For the price of risk we need a more parsimonious model, since the parameters must be estimated from the cross section of long-maturity bond returns. We therefore consider either the pure fractional $I(d_\lambda)$ model or an $I(0)$ AR(1) specification, both special cases of the ARFIMA(1,d_λ,0) model

$$(1-\phi L)\,(1-L)^{d_\lambda}\sigma^2(\lambda_t-\mu_\lambda)=\xi\epsilon_t, \tag{21}$$

The price of risk is thus determined by either (ϕ, $d_\lambda = 0$, ξ) or ($\phi = 0$, d_λ, ξ). To impose stationarity and positive autocorrelations we restrict the fractional differencing parameter to $d_\lambda \in [0, \frac{1}{2})$ and the AR parameter to $\phi \in [0, 1)$.

3 IMPLICATIONS AND TESTABLE RESTRICTIONS

3.1 Illustration and stylized facts

The autoregressive and fractional models have widely different implications for long-maturity bond prices. To illustrate we confront the main implications of the model with a few stylized facts. The first stylized fact is the autocorrelation function of the spot rate. As spot rate we take the nominal 3-month US Treasury Bill for the period January 1954 to February 2012. Autocorrelations, shown in figure 1, decrease very slowly: the first order autocorrelation is 0.984, while the 84^{th} order autocorrelation is still 0.396. This slow decay motivates the long memory time series model. Our estimate for the fractional parameter is $d_r = 0.89$, consistent with estimates in the literature.[9]

For this illustration we compare the pure fractional model with an AR(1) model with OLS parameter estimate $\nu_1 = 0.988$. Figure 2 shows the impulse responses of both models. Initially the fractional impulse responses decline faster, but since they have slower rate of decay they are larger than the AR(1) impulses responses beyond a horizon

[9] See section 4 for estimation details.

of about 40 months. Cumulative impulse responses for the fractional model are larger than for the AR(1) model beyond lag 75.

With the impulse responses we can evaluate the implied volatility of long-term bonds under the assumption of a constant risk premium ($\xi = 0$), exactly as in Backus and Zin (1993). To confront these implied volatilities with actual volatilities we use data on excess returns on five- and ten-year discount bonds for the same sample period as the spot rate.[10] Table 1 shows summary statistics. The volatility of the ten-year excess returns is substantially larger than the volatility of the five-year bond, but less than twice this volatility, the benchmark value implied by a pure random-walk level factor. Both excess returns also exhibit small but significant autocorrelations, indicating time-variation in the price of risk. The yield levels all have about the same standard deviation. These numbers may be meaningless, however, for a fractional process, since these unconditional moments do not exist if $d > \frac{1}{2}$.

In figure 3 we plot the implied volatility of excess bond returns

$$V_{rx}^{(n+1)} = \sigma b_0^{(n)} \tag{22}$$

against the maturity n for both models under the assumption that the price of risk is constant. For both specifications we calibrate the volatility parameter σ such that the volatility of the excess returns on 5-year ($n = 60$) bonds is the same. The starting point for the two curves ($\nu = 0.988$ and $d_r = 0.89$) is thus the same by construction. For longer maturities the AR model clearly implies much lower return volatility than the fractional model. For a 10-year bond the fractional model already implies 25% more volatility, while for a 20-year bond the volatility is almost twice the AR volatility. This divergence was the motivation for Backus and Zin (1993) to propose the fractional model. They noted that the fractional model provided a close fit to the observed yield volatilities. They restricted d_r to the stationary interval $d_r \leq \frac{1}{2}$, which is much smaller than unrestricted time series estimates in the literature. With our higher value of d_r the

[10] Up to September 1997 we use the replication data from Campbell and Viceira (2001), available at `http://hdl.handle.net/1902.1/FZLJAXFHBW UNF:3:S/WHW96SHNHXv0jyZv+vMA==`. From October 1997 onwards these are supplemented by the Gürkaynak, Sack and Wright (2007) data from `www.federalreserve.gov/econresdata/researchdata.htm`.

fractional model overestimates the volatility of long-maturity bond returns. The actual volatility of the 10-year bond return is approximately halfway between the AR(1) model and the fractional model in figure 3.

For a cross-sectional fit of both the 5-year and 10-year volatilities we determine the implied ν_1 and d_r together with the baseline volatility σ that matches the volatility $V_{rx}^{(n+1)}$ in (22) for $n = (60, 120)$ simultaneously. The resulting cross-sectional parameters are $\tilde{\nu}_1 = 0.993$ and $\tilde{d}_r = 0.71$. The persistence of the fractional model has gone down, while the cross-sectional AR parameter is above its time series estimate. This leads to the other two lines in figures 2 and 3. Since both models now fit two points on the volatility curve, they are much closer to each other. Even so, the two models continue to generate different volatilities for longer maturities beyond the ten years horizon.

For the AR model it has generally been noted that the time series estimate of ν_1 is lower than the cross-sectional estimate $\tilde{\nu}_1$, see *e.g.* De Jong (2000). With the development of the essentially affine model of Duffee (2002) the difference between ν_1 and $\tilde{\nu}_1$ is explained by a time-varying risk premium of the form $\sigma^2\lambda_t = \Lambda_0 + \Lambda_1 r_t$, from which $\tilde{\nu}_1$ follows as $\tilde{\nu}_1 = \nu_1 + \Lambda_1$. In our notation this corresponds to the setting $\phi = \nu_1$ and $\xi = \Lambda_1 > 0$ for the risk parameters in (21).

A positive ξ has implications for the predictability of excess returns. From the factor structure in (10) we see that a positive shock ϵ_t has a negative impact on the excess returns, but a positive impact on λ_t due to the positive ξ. As a result the response to the shock in the next periods will be positive. A bad shock is partially compensated by higher expected excess returns in the future. The level factor in the model therefore implies mean reversion dynamics for excess bond returns.

Implications for the fractional model are more complicated. Figure 3 shows a volatility curve for $\tilde{d}_r = 0.71$, but there is no clear interpretation for this curve in models with time-varying risk premiums. Unlike in the autoregressive model, adding a pure fractional risk process to a pure fractional spot rate process does not result in a pure fractional model for the spot rate under the risk-neutral $\mathbb{Q}$ measure. Inspecting the general form of the factor loadings in theorem 2 reveals that a negative ξ is required to reduce the slope of the factor loadings $b_0^{(n)}$. It will further take substantial persistence,

high ϕ or high d_λ, to flatten the slope of $b_0^{(n)}$ enough to match the volatility curve. Most important though is the implication that we need $\xi < 0$, since this is exactly opposite to the implication for the affine model with AR(1) dynamics. With a negative ξ, a bad shock to returns is followed by a lower risk premium. This is a different implication regarding the predictability of excess returns.[11] To discriminate among the alternative models we therefore consider two sets of implications: the volatility structure and the predictability of excess returns.

This example only considered the AR(1) and a pure fractional model. In the empirical analysis we will use the more general specifications (20) and (21). We can then compare models that fit the short-term dynamics of the spot rate equally well, but still produce substantially different implications for long-term bonds. The term structure models provide a connection between the persistence of the spot rate and the dynamics of the risk premium. In models where the long-run volatility of the spot rate exceeds the volatility of long-maturity excess returns, *i.e.* if

$$\sigma C_{n-1} > V_{rx}^{(n+1)}, \tag{23}$$

the risk premium on long-maturity bonds must have a negative covariance ($\xi < 0$) with the spot rate. In these models the spot rate r_t will exhibit non-stationary or nearly non-stationary behavior. Conversely, with low spot rate persistence the observed volatility in long-term bond returns cannot be explained by volatility in discount rates alone, thus requiring time-varying risk premiums that are positively correlated with movements in the spot rate. The latter case is the classic excess volatility puzzle identified by Shiller (1979).

3.2 Moment conditions

We adopt a two-stage estimation strategy. First we estimate a time series model for the actual dynamics of the short term interest rate. This will determine the c_j coefficients. In the second step we estimate the term structure parameters, either (ϕ, ξ) or (d_λ, ξ), using

[11] In both models, the average risk premium will be positive as long as the parameter $\mu_\lambda > 0$. The difference is in how the risk premium varies with shocks to the level factor.

data on long-term bonds. We opt for this two-step method instead of joint estimation by QML for two reasons. First, semi-parametric time series estimates are preferred for robust inference on the order of fractional integration. The semi-parametric methods estimate d_r in the frequency domain. Second, since the single factor model is incomplete due to the omitted transitory factors, we only want to rely on moments that are least distorted by omitted factors.

We base the selection of our first data moment on the result of Cochrane and Piazzesi (2008) that the price of risk is a univariate time series that only loads on the level factor of the term structure. In that case the autocorrelations of excess returns identify the time series properties of the price of risk. For our model the unconditional variance of excess returns is

$$\begin{aligned} V_{rx}^{(n+1)} \equiv \mathsf{Var}\left[rx_{t+1}^{(n+1)}\right] &= \left(b_0^{(n)}\right)^2 \mathsf{Var}\left[-\epsilon_{t+1} + \sigma^2\lambda_t\right] \\ &= \left(b_0^{(n)}\right)^2 \sigma^2\left(1+\xi^2\omega^2\right), \end{aligned} \tag{24}$$

where $\sigma^2\xi^2\omega^2$ is the unconditional variance of $\sigma^2\lambda_t$, *i.e.* either $\omega^2 = \frac{\Gamma(1-2d_\lambda)}{(\Gamma(1-d_\lambda))^2}$ or $\omega^2 = \frac{1}{1-\phi^2}$. Similarly, for the first order autocovariance we have

$$\begin{aligned} \mathsf{Cov}\left[rx_{t+1}^{(n+1)}, rx_t^{(n+1)}\right] &= \left(b_0^{(n)}\right)^2 \mathsf{Cov}\left[\left(-\epsilon_{t+1} + \sigma^2\lambda_t\right), \left(-\epsilon_t + \sigma^2\lambda_{t-1}\right)\right] \\ &= \left(b_0^{(n)}\right)^2 \sigma^2\left(-\xi + \rho_1\xi^2\omega^2\right), \end{aligned} \tag{25}$$

where ρ_1 is the first order autocorrelation of the price of risk, *i.e.* either $\rho_1 = d_\lambda/(1-d_\lambda)$ or $\rho_1 = \phi$. Since both the variance and autocovariance depend on the maturity n only through $b_0^{(n)}$, the autocorrelation

$$\mathcal{M}_\rho = \frac{-\xi + \rho_1\xi^2\omega^2}{1+\xi^2\omega^2} \tag{26}$$

does not depend on maturity and does not depend on parameters c_i of the spot rate process. This is an implication of the assumption that the price of risk only loads on a single factor. Maturity independence of $\mathcal{M}_\rho$ is a testable restriction in the data and a useful overidentifying restriction to obtain precise estimates of $\mathcal{M}_\rho$.

The second moment that identifies the properties of λ_t is the relative volatility of two long-term bonds. This relative volatility measures the rate of decay of the volatility term

structure at the long end. We avoid relating the variance of long-term bonds directly to the volatility of the spot rate, since the spot rate may also be subject to additional transitory factors. The spot rate variance σ^2 is likely overestimated in a univariate time series model if expectations are formed on a broader information set than lagged short-term interest rates alone. To obtain a moment condition that does not depend on σ^2 we divide (24) by the variance of another long-term bond and use

$$\mathcal{M}_\sigma = \left(\frac{V_{rx}^{(m+1)}}{V_{rx}^{(k+1)}} \right)^{1/2} = \left(\frac{b_0^{(m)}}{b_0^{(k)}} \right) \tag{27}$$

The relative volatility is a function that exclusively depends on the factor loadings of the longer versus the shorter maturity. By taking k and m large enough even a multi-factor term structure model will have factor loadings on the level factor that are almost completely determined by the low-frequency time series properties of the factors. By taking k and m far enough apart we obtain information on how fast factor loadings increase with maturity. We therefore select five year ($k = 60$) and ten year ($m = 120$) bonds for the cross-sectional estimation of the risk parameters (ϕ, ξ) or (d_λ, ξ).

A benchmark case for the factor loadings is when the risk premium is constant ($\xi = 0$). In that case the factor loadings are the same as under the expectations hypothesis, and the volatility ratio becomes $\mathcal{M}_\sigma = C_{m-1}/C_{k-1} \equiv EH$, which only depends on the parameters of the spot rate process. We will refer to this as the volatility implied by the expectations hypothesis.

3.3 Predictive regressions

With the estimated risk parameters (ξ, d_λ) and (ξ, ϕ) we can check other implications of the term structure model. An important set of results are the predictive regressions. The standard predictive regressions in the term structure literature relate excess returns to yield spreads $s_t^{(n)} = y_t^{(n)} - r_t$ as[12]

$$rx_{t+1}^{(n+1)} = \alpha_n + \beta_n s_t^{(n)} + e_{t+1}^{(n+1)}, \tag{28}$$

[12] See, *e.g.*, Fama (1984), Campbell and Shiller (1991) and Duffee (2002).

The typical finding is that β_n is significantly positive. Given the risk parameters we can work out the implied slope coefficient β_n and the fit R_n^2 of the predictive regression (28). The results, derived in the appendix, are infinite sum expressions,

$$\beta_n = \frac{b_0^{(n)} \xi \sum_{j=0}^{\infty} f_j d_j^{(n)}}{\sum_{j=0}^{\infty} \left(d_j^{(n)}\right)^2} \tag{29}$$

$$R_n^2 = \frac{\xi^2}{1+\xi^2\omega^2} \frac{\left(\sum_{j=0}^{\infty} f_j d_j^{(n)}\right)^2}{\sum_{j=0}^{\infty} \left(d_j^{(n)}\right)^2} \tag{30}$$

where $d_j^{(n)} - c_j$ are the MA coefficients of the spread $s_t^{(n)}$. These expressions can be checked for given parameters, as in Dai and Singleton (2002). In our case they don't lend themselves for formal estimation procedures, since the convergence of both expressions is very slow for the fractional models. We have therefore chosen the somewhat unusual autocorrelation moment $\mathcal{M}_\rho$ instead of β_n and R_n^2 to estimate the risk parameters.

Theorem 2 implies that the predictable component of excess returns for all maturities should be the same $\sigma^2\lambda_t$. Such a specification has obtained empirical support from the factor regressions of Cochrane and Piazzesi (2005, 2008), who defined the prediction factor as a linear combination of yields (or forward rates) leading to predictive regressions of the form

$$rx_{t+1}^{(n+1)} = \alpha_n + \beta_n w_t + e_{t+1}^{(n+1)}, \tag{31}$$

where $w_t = \sum_{i=1}^{K} \gamma_i s_t^{(n_i)}$ is the prediction factor and n_i are the maturities included in the definition of the prediction factor. For identification we normalize $\gamma_1 = 1$.[13] For any given linear combination of spreads we can again work out the implied coefficients β_n and the fit R_n^2. The expressions are similar to (29) and (30) before (see appendix).

This form of the predictive regressions is also useful for inference on the moment condition $\mathcal{M}_\sigma$. For a single factor model the predictable component has the form (see (10)):

$$\mathsf{E}_t\left[rx_{t+1}^{(n+1)}\right] = b_0^{(n)}\sigma^2\lambda_t - \tfrac{1}{2}\sigma^2\left(b_0^{(n)}\right)^2 \tag{32}$$

[13] Cochrane and Piazzesi (2008) specify a linear combination of forward premiums, but with unrestricted coefficients this can always be rewritten as a linear combination of yield spreads.

The time-varying part $\sigma^2\lambda_t$ is the same for all maturities except for the scaling with the maturity-specific factor loadings $b_0^{(n)}$. Running the predictive regression (31) for the two long-term maturities $k = 60$ and $m = 120$ we can estimate $\mathcal{M}_\sigma$ as

$$\mathcal{M}_\sigma = \frac{\beta_m}{\beta_k} \tag{33}$$

Likewise the ratio of prediction error standard deviations $\sigma_e^{(m+1)}/\sigma_e^{(k+1)}$ in (31) should give the same volatility ratio $\mathcal{M}_\sigma$.

If the prediction factor w_t is indeed a good proxy for the theoretical $\sigma^2\lambda_t$, it should be stationary. Moreover, under the fractional model $w_t \sim I(d_\lambda)$. Time series analysis of the prediction factor should give the same order of integration d_λ that we obtain from the cross-sectional fit of the λ_t process through $\mathcal{M}_\rho$ and $\mathcal{M}_\sigma$. This is one more testable implication.

4 SHORT RATE DYNAMICS

We estimate the fractional integration parameter d_r with semiparametric techniques. That is, we estimate the long-memory behavior of the process close to frequency zero, while allowing for some unparameterized short-run dependence. A commonly used estimator is the local Whittle estimator (LW), originally developed by Künsch (1987). Shimotsu and Phillips (2006) show that the LW is consistent and asymptotically normally distributed if $d \in (-\frac{1}{2}, \frac{1}{2})$. Another well-known estimator is the exact local Whittle estimator (EW), introduced by Shimotsu and Phillips (2005). The EW estimator is consistent and asymptotically normal without restrictions on d. The EW hence seems preferable, as the range of the true fractional integration parameter does not have to be known. The LW estimator, on the other hand, has the advantage that its distribution is robust against conditional heteroskedasticity (see Robinson and Henry (1999)). For robustness, we consider both the EW and the LW estimator. The negative likelihood functions are defined as

$$Q_J^{(EW)} = \frac{1}{J}\sum_{j=1}^{J}\left(\ln\left(\tau\omega_j^{-2d}\right) + \frac{1}{\tau}I_{(1-L)^d x}\left(\omega_j\right)\right) \tag{34}$$

$$Q_J^{(LW)} = \frac{1}{J}\sum_{j=1}^{J}\left(\ln\left(\tau\omega_j^{-2d}\right) + \frac{\omega_j^{2d}}{\tau} I_x\left(\omega_j\right)\right), \tag{35}$$

which are minimized with respect to τ and d. $I_x(\omega_j)$ and $I_{(1-L)^d x}(\omega_j)$ denote the periodograms of the series x_t and $(1-L)^d x_t$, respectively; ω_j are the harmonic Fourier frequencies given by $\omega_j = \frac{2\pi j}{T}$; and J is a bandwidth parameter satisfying $\frac{J}{T} + \frac{1}{J} \to 0$ as $T \to \infty$.

Both estimators are distributed as $\sqrt{J}(\hat{d} - d) \sim N(0, \frac{1}{4})$. The choice for the bandwidth J is crucial. Setting J too low decreases the rate of convergence of the estimator and implies the risk of not capturing all the low-frequency dynamics. If J is too large, however, the estimator will be polluted by the high-frequency noise.[14] We have chosen the conventional value of $T^{0.5}$ for J.

Previous work suggests that the short rate is nonstationary, with $d_r \in [0.8, 1]$. As the LW estimator is well behaved only for stationary series, we initially prefilter the series by differencing it once and add one back to the parameter estimate later. The consistency and asymptotics of both LW and EW rely on the knowledge of the true mean of the data generating process. As the series is considered in first differences, this assumption has no consequences for the LW estimator. For the EW estimator Shimotsu (2001) finds that the estimator remains consistent and asymptotically normal for $d \geq \frac{1}{2}$, if the series is detrended by its first observation. Hence, to satisfy this requirement, the EW estimator is applied to the process $r_t - r_1$.

Table 2 summarizes our findings. In line with others we find that the nominal T-Bill is fractionally integrated of the approximate order of 0.9.[15] The estimates are not

[14] The literature is not unanimous on the choice of the optimal bandwidth. For instance, Henry and Robinson (1996) develop an optimal bandwidth choice criterion that is based on the asymptotic mean-square error of the LW estimator. Henry (2001) derives a feasible version of the optimal bandwidth rule and demonstrates its robustness to conditional error heteroskedasticity. Yet, in small sample Monte Carlo experiments such automatic selection rules for J are often outperformed by simple rules of thumb (see, *e.g.*, Schotman, Tschernig and Budek (2008)).

[15] Estimates in the literature seem to be independent of the sample period. For example, Shea (1991) reports bootstrapped estimates in the range (0.77, 0.93) for the sample period 1949-1964. Sun and Phillips (2004) study a quarterly sample 1934 - 1999 and find fractional integration of the order (0.75,1.0) for both nominal and real rates. In Gil-Alana and Moreno (2012) the estimate ranges between 0.80 and 0.89 for their sample period 1971-2009.

significantly different from the unit root $d_r = 1$; yet, parameter estimates are significantly larger than $\frac{1}{2}$ at a 5% level. Thus, the short rate is non-stationary. Estimates for long-maturity yields in table 2 are very similar. If we would define a level factor as a weighted average of these yields, we still obtain the same approximate value of $d \approx 0.9$.

As a further check on the estimates, and to obtain a parametric model for the spot rate, we estimate the ARFIMA$(p, d_r, 0)$ model in (20) by the approximate maximum likelihood method of Beran (1995). As higher order short-term dynamics appear insignificant, table 3 only displays the results for the first order case $p = 1$. The estimated d_r is equal to 0.892; hence, it is very close to the semiparametric estimates. Again, it is statistically larger than $\frac{1}{2}$, but indistinguishable from 1. The short-memory parameter is small and not statistically different from zero.

In testing the term structure implications we will consider alternative values for d_r. Conditional on several values $d_r \in [0, 1]$ we estimate the short-memory parameters in (20) by OLS. From the results in table 3 it is almost impossible to draw conclusions on the relative fit of the models. All of the estimated ARFIMA$(1, d_r, 0)$ models result in the same residual volatility. Only after two digits does it become apparent that the model for $d_r = 0.9$ has the best fit. For comparison, table 3 also shows estimates for the stationary AR(1) and AR(2) models. The fit of the AR(2) is as good as the best ARFIMA$(1, d_r, 0)$ models. For short-term prediction the fractional and autoregressive models are almost identical.

Even though the transitory dynamics are important for short-term forecasts and for explaining short- to medium-term maturities, they should have a negligible impact on the long end of the maturity spectrum. We would expect that $EH = C_{m-1}/C_{k-1}$ ($k = 60$; $m = 120$) only depends on the long memory properties. To check this, we compute the ratio EH for several values of the fractional integration parameter d_r and transitory dynamics ν. The lines in figure 4 depict the EH ratio as a function of ν keeping d_r fixed. All lines in the figure are horizontal, indicating that the ratios are fully determined by d_r and completely independent of ν. Changing d_r, however, has substantial effects on the EH ratio: the ratio increases monotonically with d_r. For comparison, the figure also shows the estimates implied by stationary AR(1) and AR(2) models. The implied

volatility ratio for these models is much lower than for any of the d_r's. The conclusion from figure 4 is that transitory dynamics of the short-term interest rate have no effect on relative volatility of excess returns of bonds with maturities five and ten years. We have chosen our k and m sufficiently large to concentrate fully on the low frequency dynamics.

5 EXCESS RETURNS ON LONG-TERM BONDS

We estimate the term structure moments $\mathcal{M}_\sigma$ and $\mathcal{M}_\rho$ using GMM exploiting the contemporaneous covariance matrix of excess returns and the first order autocovariances. Since yield data are constructed from coupon bonds using splines or other interpolation methods, there is inevitably some measurement error in long-term yield data. When yields are measured with error, so are the excess returns. Adding measurement error to the excess returns we obtain the observed data series $\hat{rx}_t^{(n+1)}$

$$\hat{rx}_t^{(n+1)} = rx_t^{(n+1)} + u_t^{(n)}, \tag{36}$$

where $u_t^{(n)}$ are measurement errors with standard deviation $\sigma_u^{(n)}$. Assuming, like Campbell and Viceira (2001), Christensen *et al* (2011) and Duffee (2011), that the size of the yield measurement error in $y_t^{(n)}$ is the same across maturities, the measurement error in excess returns $rx_t^{(n+1)} \approx -n\Delta y_t^{(n)} + s_{t-1}^{(n)}$ will be proportional to the maturity n, meaning $\sigma_u^{(n)} = n\sigma_u$. With the classical measurement error assumptions of being uncorrelated with each other and with the true excess returns, the covariance matrix of the observed excess returns $\mathbf{\hat{rx}} = (\hat{r}x^{(k+1)}\ \hat{r}x^{(m+1)})'$ can be written as

$$\hat{\mathbf{V}} = \mathsf{cov}[\mathbf{\hat{rx}_t}, \mathbf{\hat{rx}_t'}] = V_{rx}^{(k+1)} \begin{pmatrix} 1 & \mathcal{M}_\sigma \\ \mathcal{M}_\sigma & \mathcal{M}_\sigma^2 \end{pmatrix} + \sigma_u^2 \begin{pmatrix} k^2 & 0 \\ 0 & m^2 \end{pmatrix}, \tag{37}$$

where we have used that $V_{rx}^{(m+1)} = \mathcal{M}_\sigma^2 V_{rx}^{(k+1)}$.

To estimate the autocorrelation we make the additional assumption that both measurement errors have the same autocorrelation coefficient θ_1,

$$\mathsf{cov}\left[u_t^{(n)}, u_{t-1}^{(n)}\right] = \theta_1 n^2 \sigma_u^2 \qquad\qquad n = k, m \tag{38}$$

The measurement error autocorrelation will most likely be negative due to the differencing of yields in the construction of excess returns. We can write the matrix of first order autocovariances as

$$\hat{\mathbf{C}} = \mathsf{cov}\left[\mathbf{r}\hat{\mathbf{x}}_{\mathbf{t}}, \mathbf{r}\hat{\mathbf{x}}'_{\mathbf{t-1}}\right] = \mathcal{M}_\rho V_{rx}^{(k+1)} \begin{pmatrix} 1 & \mathcal{M}_\sigma \\ \mathcal{M}_\sigma & \mathcal{M}_\sigma^2 \end{pmatrix} + \theta_1 \sigma_u^2 \begin{pmatrix} k^2 & 0 \\ 0 & m^2 \end{pmatrix}, \qquad (39)$$

In total (37) and (39) comprise 7 moment conditions for 5 parameters, leaving two overidentified moments.

Table 4 contains the GMM estimates. The overidentifying moment conditions are not rejected by the data. The measurement error is small: the estimates imply that the single factor model explains 94% of the variance of 5- and 10-year excess returns. The estimated $\mathcal{M}_\rho$ is significantly different from zero. The autocorrelation in the measurement error is substantially negative and this causes the estimate of the autocorrelation moment $\mathcal{M}_\rho$ to be above the sample autocorrelation of the excess returns.

The volatility ratio is estimated very precisely. The point estimate of $\mathcal{M}_\sigma$ is five standard errors below the benchmark value of 2, which would obtain if the level factor had a unit root with $d_r = 1$. The estimate of $\mathcal{M}_\sigma$ is also about four standard errors above the value implied by the stationary AR models estimated in table 3. This is consistent with the usual result that the AR parameter under the actual probability measure is less than the risk-neutral AR parameter.

An alternative estimate of the volatility ratio is from the Cochrane-Piazzesi predictive regressions (31) relating excess returns $rx_{t+1}^{(n+1)}$ to a common prediction factor w_t. We use the two spreads $s_t^{(k)}$ and $s_t^{(m)}$ as predictor variables, leading to $w_t = s_t^{(k)} + \gamma_2 s_t^{(m)}$. Table 5 shows the estimation results. Most remarkable is the estimate of $\mathcal{M}_\sigma = \frac{\beta_m}{\beta_k}$, which is almost identical to the estimate obtained from the covariance matrix of excess returns in table 4 and the ratio of the residual standard deviations $\sigma_e^{(120)}/\sigma_e^{(60)}$. The result is even more remarkable, as the predictive regressions are known to be plagued by econometric problems. Since the predictor variables are highly autocorrelated, coefficient estimates are subject to the Stambaugh (1999) bias. This is already a problem with a single regressor, but in the present case we use two regressors that are strongly multi-collinear. Finally, measurement errors may add another source of bias, since the measurement

error in yields affects both the right-hand side spreads as well as the left-hand side excess returns.

The predictive regressions also imply testable restrictions. In unrestricted form they can be written as

$$rx_{t+1}^{(n+1)} = \alpha_n + \sum_{i=1}^{K} \gamma_{ni} s_t^{(n_i)} + e_{t+1}^{(n+1)}, \qquad n = k, m, \tag{40}$$

with $2K$ maturity specific parameters γ_{ni}. The predictive factor model, with the normalization $\gamma_1 = 1$, has only $K + 1$ parameters. In our case, with $K = 2$, this leaves one overidentified parameter. A Wald test on the restriction in the unrestricted system gives a $\chi^2(1)$-statistic equal to 1.83 (p-value 0.18). The restriction that the expected excess returns are in a one-dimensional linear space, spanned by the vector $(1 \ \mathcal{M}_\sigma)'$ is not rejected by the data.[16] Both this test and the result that the ratio of the regression coefficients β_m/β_k from the predictable component is the same as the ratio of standard deviations $\sigma_e^{(m+1)}/\sigma_e^{(k+1)}$ from the unpredictable shocks, are important testable implications of the single factor assumption.

The prediction factor has a negative coefficient γ_2 on the 10-year spread. This implies that the prediction factor has tent-shaped weights,

$$w_t = -0.46 r_t + y_t^{(60)} - 0.54 y_t^{(120)}, \tag{41}$$

that are similar to the predictive factor in Cochrane and Piazzesi (2005, 2008). The factor is plotted in figure 5. Even though we use a different forecasting horizon and only two different long-term yields, the time series looks very similar to the factor in Cochrane and Piazzesi (2008).

One of the assumptions in our term structure model is that the price of risk is stationary. For the fractional model the prediction factor should have an order of integration $d < \frac{1}{2}$. Table 6 report estimates of the fractional differencing parameter for the predictor series. For the 5-year spread and the prediction factor we find $d \approx 0.4$. The point estimate for the 10-year spread is slightly above one half, but not significantly so.

[16] We have worked with two additional predictors in the regressions: the spot rate level r_t and last year's average excess return $\frac{1}{24}\sum_{n=60,120}\sum_{\ell=1}^{12} rx_{t+1-\ell}^{(n+1)}$. Both of them turned out to be insignificant and therefore made no difference to the estimates of $\mathcal{M}_\sigma$ or on the test of the overidentifying restrictions.

These estimates indicate the range of values to expect when we estimate d_λ from the cross-section using the moments $\mathcal{M}_\sigma$ and $\mathcal{M}_\rho$.

6 THE IMPLIED PRICE OF RISK

6.1 Solving the moment conditions

We determine the parameters d_λ and ξ by inverting the moment conditions for $\mathcal{M}_\rho$ and $\mathcal{M}_\sigma$, assuming that the price of risk follows a fractional noise process. In the robustness section we do the same for the AR specification with parameters (ϕ, ξ). With two moment conditions for two parameters, the system is exactly identified. Multiple solutions may exist, however, since the moment conditions are nonlinear functions of the risk parameters.

We first discuss the autocorrelation moment separately, since the solutions for $\mathcal{M}_\rho$ are independent of the spot rate process parameters c_j. Given d_λ and $\mathcal{M}_\rho$ the moment condition (26) implies a quadratic relation in ξ. This will generally provide two solutions for ξ. In the appendix we show that the moment condition will normally have two real roots, one negative and the other positive.

Figure 6 plots the relation (26) for the estimate $\mathcal{M}_\rho = 0.115$. The bold solid line represents the solutions for ξ for given d_λ. The two dotted lines are the solutions for $\mathcal{M}_\rho$ plus and minus two standard errors. The negative solutions for ξ are fairly stable and increase slowly toward zero when d_λ gets close to the non-stationarity boundary of one half. For the positive solutions, the value of ξ increases rapidly when d_λ moves further away from the $d_\lambda = 0.5$ boundary. Both solutions converge to zero as d_λ approaches the non-stationarity boundary. For the negative solutions the estimation error in $\mathcal{M}_\rho$ does not cause much uncertainty in the solution for ξ. Effects on the positive solutions are much bigger.

The second moment describing the relation between d_λ and ξ is the relative volatility $\mathcal{M}_\sigma$. The moment condition (27) does not have a closed-form solution, as the right-hand side of the equation depends on ξ through the recursive factor loadings $b_0^{(i)}$. In addition, (27) depends on the dynamics of the short rate.

Figure 7 plots the relation $\mathcal{M}_\sigma(d_\lambda, \xi) = \hat{\mathcal{M}}_\sigma$ for all short-rate models in table 3. The colored lines represent the combinations (d_λ, ξ) that match $\hat{\mathcal{M}}_\sigma$ for different models for r_t. The black line is the relation between ξ and d_λ that solves the autocorrelation moment. The intersection points are the complete solutions (ξ, d_λ) that match both data moments ($\mathcal{M}_\rho$, $\mathcal{M}_\sigma$). Numerical values of these solutions plus standard errors are reported in table 7. Solutions depend strongly on d_r. We discuss four cases:

1. For $d_r \approx 0.7$ the moment condition for $\mathcal{M}_\sigma$ is flat at $\xi \approx 0$. In this case $EH \approx \mathcal{M}_\sigma$, *i.e.* the relative volatility is fully explained by the expectations hypothesis and there is no scope for time-varying risk premiums to alter the volatility or the factor loadings. The $d_r = 0.7$ line for $\mathcal{M}_\sigma$ intersects the $\mathcal{M}_\rho$ condition almost at the boundary $(d_\lambda, \xi) = (\frac{1}{2}, 0)$. A solution with ξ exactly equal to zero is not valid, since $\xi = 0$ implies that excess returns are not predictable and therefore it can not match the observed autocorrelation. Fitting the observed predictability is only feasible by the knife-edge solution of letting the price of risk become non-stationary.

2. For $d_r > 0.7$ all solutions that match the volatility condition $\mathcal{M}_\sigma$ have negative values for ξ. As long as $0.7 < d_r < 0.9$ the lines that fit the volatility condition intersect the predictability conditions at two points. For example, when $d_r = 0.8$ one solution has d_λ close to the nonstationarity boundary, while the other solution is obtained for a smaller value $d_\lambda = 0.32$. Only the second solution produces an estimate $\hat{\xi}$ that is statistically different from zero. For the maximum likelihood estimate $d_r = 0.89$ the two solutions are almost identical with an estimate $\hat{\xi} = -0.09$ that is statistically significant. In this case the cross-sectional estimate for $d_\lambda = 0.47$ is also consistent with the time series estimates of d_λ obtained from spreads and the prediction factor reported earlier in table 2.

   When $d_r > 0.9$, and in particular for $d_r = 1$, the two moment conditions can not be solved simultaneously. Under the expectations hypothesis, a random walk model for the short rate overshoots observed bond volatilities as noted by Backus and Zin (1993). It is not possible to find a pure long-memory price of risk process with impulse responses that decay sufficiently fast to correct this overshooting, while

still producing the right amount of predictability in excess bond returns.

3. When d_r is less than 0.7, matching the relative volatility condition requires a positive ξ. For these positive solutions ξ increases quickly as d_λ moves away from a half, just as we saw when fitting the autocorrelation moment. In this case there is always a unique solution with $\xi > 0$ and $d_\lambda \approx \frac{1}{2}$. The risk premium at these solutions is hard to distinguish from a nonstationary process regardless of the exact value of d_r. The estimates are also hardly influenced by possible estimation imprecision of the estimated moments and short-term interest rate dynamics.

4. The solution for the AR models closely reflects the solution for the ARFIMA(1,0.6,0). Implied values for ξ for this $I(0)$ process are even larger than for the non-stationary $d = 0.6$ model.

One of the implications of the parameter estimates is the maximum predictability of excess returns if we would be able to observe the true price risk $\sigma^2\lambda_t$. According to (24) the population R^2 of the hypothetical regression of $rx_{t+1}^{(n+1)}$ on λ_t is given by

$$R^2 = \frac{\xi^2\omega^2}{1+\xi^2\omega^2} \tag{42}$$

The last line in table 7 shows the maximum predictability implied by the parameters that solve the moment conditions. For the AR models predictability can be substantial. In these models the variation of risk premiums must be large enough to add sufficient volatility to the volatility of long-term expectations of the spot rate (note: adding volatility means $\xi > 0$). In the AR(2) model the risk premium must account for about 15% of the variance of long-maturity excess returns. A similar value is implied for the fractional model with low persistence $d_r = 0.6$.

Consistent with the earlier results, the predictability is very limited when $d_r = 0.7$. With larger persistence the maximum R^2 rises again, but remains much smaller than for the AR models, primarily because $\xi < 0$ means that the covariance between shocks to the spot rate and the price of risk contributes negatively to the overall volatility. For the ML estimate $d_r = 0.89$ the maximum R^2 in a predictive regression using the true λ_t is not more than 5%. This is not far above the $R^2 \approx 4\%$ obtained with the prediction factor w_t.

6.2 Overidentifying moments and implications

The previous section has presented sets of parameter values (d_r, d_λ, ξ) that all satisfy the moment conditions $\mathcal{M}_\rho$ and $\mathcal{M}_\sigma$ and provide almost equal short-term predictions for the short-term interest rate. We now check how well the different models match the results from the predictive regressions. For this we compare the regression estimates in table 5 with the implied results from equations (28) and (31).

We evaluate β_n and R_n^2 ($n = k, m$) at all solutions in table 7. For regressions on the own spread the only solution that produces an R^2 that is consistent with table 5 is when $d_r = 0.89$. For this model we find $R_k^2 = 2.3\%$, identical to the 2.3% found in the data. The implied β_k is equal to 1.86 and thus substantially lower than the value 3.31 in the data. Part of the difference could be attributed to the small-sample upward bias in the predictive regressions. From the formulas in Bekaert, Hodrick and Marshall (1997) the upward bias in the regression estimate would be around 1.7 and this would reduce the slope coefficient in the data almost to the value implied by the term structure model.[17] The results for $m = 120$ are very similar.

Most of the other models have zero implied predictability ($R^2 < 0.1\%$) and are thus inconsistent with the regression results. For the models with low persistence, *i.e.* the AR models and $d_r = 0.6$, the implied spread has very low volatility and therefore has low predictive power. For the model $d_r = 0.7$ we have already seen that the maximum implied predictability is zero. The models with $d_r = 0.8$ move in the right direction: with a nearly non-stationary risk premium the implied value for ξ is still too small to generate enough volatility in the spread, however. Similarly, the parameter combination $(d_r, d_\lambda, \xi) = (0.8, 0.318, -0.109)$ results in too little persistence in risk prices to produce sufficient spread volatililty.

Results for the regression on the prediction factor are very similar. Only the $d_r = 0.89$ model is able to match the observed pattern in the data regressions. Both the implied R^2 as well as the slope coefficients match the results from the predictive regressions.

[17] The analytical expressions in Bekaert *et al* (1997) provide the small-sample bias in β_n assuming that the spot rate is generated by either an AR(1) or first order VAR process. Results may not be fully comparable, since we assume a different, but also very persistent, process for the spot rate.

All other models do not come near to matching the predictive regressions. Closest competitor is the AR(2) model, which does generate a substantial slope coefficient, but with a tiny R^2.

The models also have very different implications for the dynamics of long-term yields. In figure 8 we compare the implications of the AR(2) model with the fractional model based on the ML estimate $d_r = 0.89$. The figure shows the impulse responses of the ten-year rate with respect to a one standard deviation shock in the 10-year rate. The initial responses are the same, since the volatility is calibrated to fit the volatility of both 5- and 10-year rates through the moment condition $\mathcal{M}_\sigma$. The AR(2) model shows decreasing impulse responses, since by construction it generates stationary long-term yields. After 5 years (60 months) about one-third of the initial shock is left. The fractional model implies much stronger persistence. After the initial shock the yield is expected to go up by another six basispoints and to remain at that level for the next five years. Both implied processes are more extreme than unrestricted time series estimates of the 10-year rate, shown as the red lines in figure 8. Both fractional and AR(2) time-series processes agree with the term structure implied model on the initial increase of the impulse responses. Both, however, are also downward sloping at longer lags, just as the AR(2) term structure implied process.[18]

Solutions that require a positive ξ imply that an increase in the spot rate is associated with an increase in the price of risk. In this case excess bond returns would exhibit mean reversion: a positive shock in the spot rate is associated with an instantaneous loss ($b_0^{(n)} > 0$), but also with higher expected future returns (increase in $\sigma^2\lambda_t$). Yet, from the first few autocorrelation of excess returns in table 1 we can already infer that excess bond returns do not exhibit mean reversion. The first order autocorrelation is positive. Although the second order autocorrelation in negative, this is not enough to produce long-run mean reversion. For instance, the quarterly variance of the 5-year

[18] Of course the two unrestricted time series estimates differ much more at longer lags than shown in the figure. The AR(2) decreases exponentially, whereas the fractional time series process has the slower hyperbolical decay. Standard errors on the parameters estimates of both time series models are large enough that we can not reject an $I(1)$ process, which would be very close to the fractional term structure.

excess returns equals 15.94, which is much larger than three times the monthly variance $3V_{rx}^{(61)} = 9.312$.[19] Campbell and Viceira (2005) provide further evidence against mean reversion in nominal bond returns.

From the results we conclude that fractional model with $d_r = 0.89$ is best in fitting the long end of the term structure. First, it deals with the near unit-root behavior of short rates that requires a fractional-integration parameter of $d_r \approx 0.9$. Second, the combination of $d_r = 0.89$ with the risk parameters (d_λ, ξ) match the predictability and volatility properties of long-maturity excess returns.

6.3 Robustness

6.3.1 Sensitivity with respect to $\mathcal{M}_\sigma$

In the previous section we argued that the parameter set $(0.89, 0.47, -0.09)$ is the best solution that is consistent with time-series as well as cross-sectional implications of the term structure. We also showed that estimates of the risk parameters depended strongly on the time-series persistence d_r, since d_r is the main determinant of the volatility ratio $\mathcal{M}_\sigma$. Although GMM yielded precise estimates of $\mathcal{M}_\sigma$, it is worthwhile to check the sensitivity of results with respect to small changes in $\mathcal{M}_\sigma$.

Figure 9 shows the implied relative volatility ratio for different fractional orders d_r. On the horizontal axis are values of the fractional order d_λ with the implied value for ξ that fit the the autocorrelation $\mathcal{M}_\rho$. Figure 9(a) considers the solutions $\xi < 0$. The figure demonstrates that variations of d_λ have a limited effect on the possible values of $\mathcal{M}_\sigma$, when the price of risk is required to be stationary. The figure also shows that a slightly higher estimate of $\mathcal{M}_\sigma$, moving toward the upper bound of the 95% confidence interval, would enable solutions for values of d_r larger than 0.9. On the other hand, if $\mathcal{M}_\sigma$ were at the lower end of the confidence interval, we would not have been able to find a solution (d_λ, ξ) for $d_r = 0.89$. Constructing dynamics of the price of risk that generate the right amount of predictability and volatility requires precise estimation of not only the time-series properties of the short rate, but also the cross-sectional term-structure

[19] The estimate of the quarterly variance has been corrected for measurement error in the same way as the monthly variance. The same inequality holds for the 10-year excess returns.

moments.

Figure 9(b) shows the same implied volatility ratios, but now for the solutions $\xi > 0$. With positive ξ it is even harder to find parameter pairs that match both data moments of excess bond returns. For any d_λ only slightly below the nonstationary boundary the parameter ξ becomes fairly large, which leads to explosive behavior of implied $\mathcal{M}_\sigma$. Volatility in the price of risk is too big to keep the volatility of excess returns within bounds. In the figure this shows up in the vertical scale of the implied volatility ratio $\mathcal{M}_\sigma$.

6.3.2 An AR(1) risk price process

As an alternative for the fractional process for the price of risk we can also solve both moment conditions of excess returns for the AR(1) process with parameters (ϕ, ξ) and impose $d_\lambda = 0$. Table 9 contains these solutions for different models for the short rate. Overall, the results are very similar to the ones previously discussed. For $d_r \leq 0.7$ we find one solution with a positive value for ξ and a ϕ parameter that is almost equal to a unit root, and for $d_r > 0.7$ there are two possible parameter pairs with $\xi < 0$, one close to the nonstationarity boundary and one with $\phi < 1$. For the near non-stationary solutions it often takes more than four digits to see the difference between ϕ and a unit root. The only plausible solutions are the ones in the lower panel of the table 9 with solutions for ϕ further away from the unit root. One notable difference with the fractional model is that the AR(1) specification also has a solution for $d_r = 1$. With an AR(1) process for the price of risk ,the limit as $\phi \to 1$ is an I(1) process, which is more persistent than the $I(\frac{1}{2})$ process at the limit $d_\lambda \to \frac{1}{2}$ for the fractional specification for λ_t. The additional limiting persistence enables the extra solution, since it allows the f_j coefficients just enough freedom to adjust the expectations driven component C_n toward the required volatility for excess returns.

The AR(1) specification for λ_t also has very similar implications for the predictive regressions. Table 10 shows results for those solutions where the spot rate is $I(d_r)$ but with ϕ away from the non-stationarity boundary.[20]

[20] For solutions with $\phi > 0.9999$ the implications for the predictive regressions are numerically

6.3.3 Average term premium

A well-known argument against non-stationary models of the short-term interest rate is that they imply negative average yields at very long maturities. Since our model is non-stationary and deals with long-maturity excess returns, we check how serious the problem is. From (10) and theorem 2 we have that for a single factor model

$$\mathsf{E}\left[rx_t^{(n+1)}\right] = b_0^{(n)}\sigma^2\left(\mu_\lambda - \tfrac{1}{2}b_0^{(n)}\right) \tag{43}$$

Since $b_0^{(n)}$ is unbounded for a fractional model with $d_r > \frac{1}{2}$, the average of log excess returns will indeed be negative for large n due to the Jensen inequality term $\frac{1}{2}b_0^{(n)}$. We calibrate positive values μ_λ and σ^2 to fit the data averages for the 5- and 10-year excess returns, and subsequently determine for which n the average excess return will turn negative. For our model with $d_r = 0.89$ this does not happen before a maturity of 40 years.

7 CONCLUSION

Predictions from term-structure models in the affine modeling class are sensitive to the persistence in factor dynamics. This paper develops a generalization of the essentially affine model of Duffee (2002) that allows general linear processes for the factors. This enables a specification with fractional long memory. While adding modeling flexibility, the model still retains an affine structure for excess bond returns and produces a closed-form solution.

The empirical analysis favors a fractionally integrated specification for the short-term interest rate, with a long-memory parameter of $d \approx 0.9$. It further suggests that the term spread is strongly persistent, with $d \approx 0.45$. Such a specification is not only consistent with the observed time-series properties of the two series, but also fits cross-sectional implications of the term structure. With a spot rate that is close to a unit root process and risk prices that possess stationary long memory, we can capture the observed volatility and predictability of long-maturity excess bond returns. An impor-

unstable, similar to the fractional model when $d_\lambda \to \frac{1}{2}$.

tant implication of our findings is that the covariance between the risk premium and the spot rate is negative. This is consistent with the absence of mean reversion in nominal bond returns.

Whereas our term-structure model allows for an arbitrary number of factors, we restrict the empirical analysis to a single factor. We focus on the most persistent factor that is commonly referred to as the 'level' factor. This is a limitation, but with respect to matching the volatility of excess returns at the long end of the maturity spectrum it is a reasonable assumption to make. We find that a single factor model explains 94% of the variation in 5-year and 10-year excess returns. Similarly, the results in Duffee (2002) imply that the 'level' factor captures 94% for the very short maturities and 96% for maturities of two years and longer.

APPENDIX: PROOFS

A1 Proof of theorem 1

To derive equation (7), we need to compute the three conditional moments in the recursive pricing equation (5) guessing that prices are given by (6). Initial conditions for $n = 1$ follow trivially from the definition $p_t^{(1)} = -r_t$. To derive the representation for maturities $n > 1$, we evaluate the conditional moments as

$$\mathsf{E}_t\left[p_{t+1}^{(n)}\right] = -a^{(n)} - \sum_{j=0}^{\infty} b_{j+1}^{(n)\prime}\epsilon_{t-j}, \tag{A1}$$

$$\mathsf{Var}_t\left[p_{t+1}^{(n)}\right] = b_0^{(n)\prime}\Sigma b_0^{(n)}, \tag{A2}$$

$$\mathsf{Cov}_t\left[m_{t+1}, p_{t+1}^{(n)}\right] = -\lambda_t'\Sigma b_0^{(n)}, \tag{A3}$$

and substitute these expression back in (5) leading to

$$\begin{aligned} p_t^{(n+1)} &= -r_t - a^{(n)} - \sum_{j=0}^{\infty} b_{j+1}^{(n)\prime}\epsilon_{t-j} + \tfrac{1}{2}b_0^{(n)\prime}\Sigma b_0^{(n)} - \lambda_t'\Sigma b_0^{(n)} \\ &= -(\mu_r + a^{(n)} - \tfrac{1}{2}b_0^{(n)\prime}\Sigma b_0^{(n)} + \mu_\lambda'\Sigma b_0^{(n)}) - \sum_{j=0}^{\infty}\left(c_j + b_{j+1}^{(n)} + F_j b_0^{(n)}\right)'\epsilon_{t-j}, \end{aligned} \tag{A4}$$

where the second line follows from substituting the dynamic specifications for r_t and λ_t.

A2 Proof of theorem 2

To derive (10) note that the innovation in the price in (9) is

$$p_{t+1}^{(n)} - \mathsf{E}_t\left[p_{t+1}^{(n)}\right] = -b_0^{(n)}\epsilon_{t+1} \tag{A5}$$

The other two terms in (9) are already given in (A2) and (A3). Direct substitution leads to (10). Starting from (7) we can simplify the expression for the factor loadings $b_0^{(n)}$ for $n > 1$, to derive equation (11). The lines below use the definition $C_n = \sum_{i=0}^{n} c_i$.

$$\begin{aligned}
b_0^{(n)} &= c_0 + b_1^{(n-1)} + F_0 b_0^{(n-1)} \\
&= c_0 + \left(c_1 + b_2^{(n-2)} + F_1 b_0^{(n-2)}\right) + F_0 b_0^{(n-1)} \\
&= C_1 + \left(c_2 + b_3^{(n-3)} + F_2 b_0^{(n-3)}\right) + F_1 b_0^{(n-2)} + F_0 b_0^{(n-1)} \\
&\vdots \\
&= C_{n-2} + b_{n-1}^{(n-(n-1))} + \sum_{i=0}^{n-2} F_i b_0^{(n-1-i)} \\
&= C_{n-1} + \sum_{i=0}^{n-2} F_i b_0^{(n-1-i)}
\end{aligned} \tag{A6}$$

A3 Roots of (26)

Note that the quantities ω^2 and ρ_1 in the moment condition (26) for $\mathcal{M}_\rho$ only depend on either d_λ or ϕ. To characterize the solution for ξ given d_λ or ϕ we consider the quadratic function

$$q(\xi) = (\rho_1 - \mathcal{M}_\rho)\,\omega^2\xi^2 - \xi - \mathcal{M}_\rho, \tag{A7}$$

which has the same roots as (26) in the text. If $\mathcal{M}_\rho = \rho_1$, the function is linear and the only solution is $\xi = -\mathcal{M}_\rho$. Otherwise, the normal case is $\rho_1 > \mathcal{M}_\rho$, meaning that we consider values for d_λ or ϕ which imply a positive autocorrelation in the price of risk, but a much lower positive autocorrelation in excess returns. This is the normal case, since excess returns are the sum of the risk price plus uncorrelated noise and therefore will generally have lower autocorrelations. The solutions for ξ can be written

$$\xi_{1,2} = \frac{1 \pm \left(1 + 4\omega^2\mathcal{M}_\rho(\rho_1 - \mathcal{M}_\rho)\right)^{1/2}}{2\omega^2(\rho_1 - \mathcal{M}_\rho)} \tag{A8}$$

If $\rho_1 > \mathcal{M}_\rho$, the discriminant is positive, so that both roots are real. Moreover, since $4\omega^2\mathcal{M}_\rho(\rho_1 - \mathcal{M}_\rho) > 0$, and the denominator in (A8) is also positive, one root will be positive and the other negative.

A4 Predictive regressions

We derive the slope coefficient and the coefficient of determination of a regression of excess bond returns on spreads, lagged by one period, as in (28). From theorem 1 and the relation $y_t^{(n)} = -(1/n)p_t^{(n)}$ the model-implied spread follows as

$$y_t^{(n)} - r_t = \frac{a^{(n)}}{n} - \mu_r + \sum_{j=0}^{\infty}\left(\frac{b_j^{(n)}}{n} - c_j\right)' \epsilon_{t-j}. \tag{A9}$$

Denote the MA coefficients of the spread by $d_j^{(n)} = \frac{b_j^{(n)}}{n} - c_j$. The slope coefficient in (29) can be derived as

$$\begin{aligned}
\beta_n &= \frac{\mathsf{Cov}\left[rx_{t+1}^{(n+1)}, y_t^{(n)} - r_t\right]}{\mathsf{Var}\left[y_t^{(n)} - r_t\right]} \\
&= \frac{b_0^{(n)'}\mathsf{E}\left[(-\epsilon_{t+1} + \Sigma(\lambda_t - \mu_\lambda))\left(\sum_{j=0}^{\infty} d_j^{(n)'}\epsilon_{t-j}\right)\right]}{\mathsf{E}\left[\sum_{i=0}^{\infty}\sum_{j=0}^{infty} d_i^{(n)'}\epsilon_{t-i}\epsilon_{t-j}' d_j^{(n)}\right]} \\
&= \frac{b_0^{(n)'}\sum_{j=0}^{\infty} F_j'\Sigma d_j^{(n)}}{\sum_{j=0}^{\infty} d_j^{(n)'}\Sigma d_j^{(n)}}.
\end{aligned} \tag{A10}$$

In the case of a univariate factor, $K = 1$, we can further simplify (A10). In this case $\Sigma = \sigma^2$, which is a scalar and therefore cancels in numerator and denominator. With $K = 1$, we also have $F_j = \xi f_j$ and $b_0^{(n)}$ as scalars, leading to (29) in the text.

The coefficient of determination, R_n^2, in (30) is given by

$$R_n^2 = \frac{\mathsf{Var}\left[\beta_n(y_t^{(n)} - r_t)\right]}{\mathsf{Var}\left[rx_{t+1}^{(n+1)}\right]} = \frac{\left(b_0^{(n)'}\sum_{j=0}^{\infty} F_j'\Sigma d_j^{(n)}\right)^2}{\left(\sum_{j=0}^{\infty} d_j^{(n)'}\Sigma d_j^{(n)}\right)\left(b_0^{(n)'}\left(\Sigma + \sum_{j=0}^{\infty} F_j'\Sigma F_j\right) b_0^{(n)}\right)}. \tag{A11}$$

Again, if $K = 1$, the equation simplifies. In that case the scalars σ^2 and $b_0^{(n)}$ cancel in numerator and denominator, while the term $\sum_{j=0}^{\infty} F_j'F_j = \xi^2\omega^2$. Direct substitution then leads to (30).

Expressions for the regression of excess returns on the prediction factor $w_t = s_t^{(k)} + \gamma s_t^{(m)}$ are very similar. Just replace $d_j^{(n)}$ by $\mathbf{d}_j = d_j^{(k)} + \gamma d_j^{(m)}$, which is independent of n. For the single factor case we find

$$\beta_n = \frac{\mathsf{Cov}\left[rx_{t+1}^{(n+1)}, w_t\right]}{\mathsf{Var}\left[w_t\right]} = b_0^{(n)}\xi \times \frac{\sum_{j=0}^{\infty} f_j \mathbf{d}_j}{\sum_{j=0}^{\infty} \mathbf{d}_j^2} \tag{A12}$$

$$R^2 = \frac{\mathsf{Var}\left[\beta_n w_t\right]}{\mathsf{Var}\left[rx_{t+1}^{(n+1)}\right]} = \frac{\xi^2}{1+\xi^2\omega^2} \times \frac{\left(\sum_{j=0}^{\infty} f_j \mathbf{d}_j\right)^2}{\sum_{j=0}^{\infty} \mathbf{d}_j^2}, \tag{A13}$$

where we have omitted the subscript n for the R^2, since it does not depend on n. In a single factor model excess returns of all maturities should have the same R^2 when regressed on a common prediction factor w_t.

REFERENCES

Abadir, K., Hadri, K. and Tzavalis, E.: 1999, The Influence of VAR Dimensions on Estimator Biases, *Econometrica* **67**, 163–181.

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. and Zin, S.: 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 2011-03.

Bauer, M., Rudebusch, G. and Wu, J.: 2011, Unbiased Estimation of Dynamic Term Structure Models. Chicago Booth Working Paper 11-29.

Bekaert, G., Hodrick, R. and Marshall, D.: 1997, On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates, *Journal of Financial Economics* **44**, 309–348.

Beran, J.: 1995, Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models, *JRSS-B* **57**, 659–672.

Campbell, J. and Shiller, R.: 1987, Cointegration and Tests of Present Value Relations, *Journal of Political Economy* **95**, 1062–1088.

Campbell, J. Y. and Shiller, R. J.: 1991, Yield Spreads and Interest Rate Movements: A Bird's Eye View, *Review of Economic Studies* **58**, 495–514.

Campbell, J. Y. and Viceira, L.: 2001, Who Should Buy Long-Term Bonds?, *American Economic Review* **91**, 99–127.

Campbell, J. Y. and Viceira, L.: 2005, The Term-Structure of the Risk-Return Trade-Off, *Financial Analysts Journal* **61**, 34–44.

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

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. and Piazzesi, M.: 2005, Bond Risk Premia, *American Economic Review* **95**, 138–160.

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

Connolly, R. A., Güner, Z. N. and Hightower, K. N.: 2007, Evidence on the Extent and Potential Sources of Long Memory in U.S. Treasury Security Returns and Yields, *Journal of Money Credit and Banking* **39**, 689–702.

Dai, Q. and Singleton, K.: 2002, Expectation Puzzles, Time-Varying Risk Premia, and Dynamic Models of the Term Structure, *Journal of Financial Economics* **63**, 415–441.

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

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

Duffee, G.: 2011, Information in (and not in) the Term Structure, *Review of Financial Studies* **24**, 2895–2934.

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

Fama, E. F.: 1984, The Information in the Term Structure, *Journal of Financial Economics* **13**, 509–528.

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

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

Gürkaynak, R. S., Sack, B. and Wright, J. H.: 2007, The U.S. Treasury Yield Curve: 1961 to the Present, *Journal of Monetary Economics* **54**, 2291–2304.

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

Henry, M. and Robinson, P.: 1996, Bandwidth Choice in Gaussian Semiparametric Estimation of Long Range Dependence, *Athens Conference on Applied Probability and Time Series*, Vol. 115 of *Lecture Notes in Statistics*, Springer, New York, pp. 220–232.

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

Jardet, C., Montfort, A. and Pegoraro, F.: 2011, No-arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth. SSRN working paper 1340303.

Joslin, S., Priebsch, M. and Singleton, K.: 2010, Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. Stanford University working paper.

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

Künsch, H.: 1987, Statistical Aspects of Self-Similar Processes, *in* Y. V. Prohorov and V. V. Sazonov (eds), *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, pp. 67–74.

Lo, A. W.: 1991, Long-term Memory in Stock Market Prices, *Econometrica* **59**, 1279–1313.

Marinucci, D. and Robinson, P. M.: 1999, Alternative Forms of Fractional Brownian Motion, *Journal of Statistical Planning and Inference* **80**, 111–122.

McCulloch, J. and Kwon, H.: 1993, U.S. Term Structure Data, 1947-1991. OSU Working Paper # 93-6.

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

Robinson, P. M. and Henry, M.: 1999, Long and Short Memory Conditional Heteroskedasticity in Estimating the Memory Parameter of Levels, *Econometric Theory* **15**, 299–336.

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.

Shiller, R.: 1979, The Volatility of Long Term Interest Rates and Expectations Models of the Term Structure, *Journal of Political Economy* **87**, 1190–1219.

Shimotsu, K.: 2010, Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend, *Econometric Theory* **26**, 501–540.

Shimotsu, K. and Phillips, P. C. B.: 2005, Exact Local Whittle Estimation of Fractional Integration, *American Statistician* **33**, 1890–1933.

Shimotsu, K. and Phillips, P. C. B.: 2006, Local Whittle Estimation of Fractional Integration and Some of its Variants, *Journal of Econometrics* **130**, 209–233.

Stambaugh, R. F.: 1999, Predictive Regressions, *Journal of Financial Economics* **54**, 375–421.

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

TABLES AND FIGURES

Table 1: Summary statistics

The monthly T-Bill rate is constructed from the daily series *DTB3* from the FRED database by taking the last observation of each month and transforming from discount basis to continuously compounded yields. Sample period is January 1954 to February 2012, a total of 698 observations. Monthly five-year and ten-year zero-coupon yields of US Treasury bonds originate from two sources. The largest part of the sample (January 1954 - September 1996) is from Campbell and Viceira (2001). From October 1996 onwards data are from Gürkaynak *et al* (2007).

			Autocorrelations		
	Average	Std. Dev.	1	2	3
Yields (% p.a.)					
3-months TBill r_t	4.85	2.977	0.984	0.965	0.946
5-year bond $y_t^{(60)}$	5.97	2.797	0.988	0.974	0.962
10-year bond $y_t^{(120)}$	6.33	2.608	0.990	0.979	0.969
Excess returns (% p.m.)					
5-year bond $rx_t^{(61)}$	0.10	1.809	0.099	-0.071	-0.027
10-year bond $rx_t^{(121)}$	0.13	3.036	0.071	-0.068	-0.016

Table 2: Time-series Estimates of d

The table reports the EW and the LW estimates (on data in first differences) for the fractional differencing parameter d of the nominal T-Bill, the 5-year bond yield, and the 10-year bond yield. The size of the spectral window is $J = T^{0.5} = 26$. The asymptotic standard error in all cases equals $\hat{\sigma}(\hat{d}) = 0.0981$.

Variable	$\hat{d}_{EW}$	$\hat{d}_{LW}$
Spot rate r_t	0.9260	0.8999
5-Year yield $y_t^{(60)}$	0.9467	0.9166
10-Year yield $y_t^{(120)}$	0.9426	0.9204

Table 3: ARFIMA estimates

The table reports estimates of alternative models for the short term interest rate of the form

$$\left(1 - \nu_1 L - \nu_2 L^2\right)(1 - L)^{d_r}(r_t - \mu_r) = \epsilon_t$$

Joint estimates of (d_r, ν_1) are ML estimates based on Beran (1995). All other estimates are OLS and conditional on d_r. All standard errors are robust to heteroskedasticity; for OLS they are conditional on d_r. σ_ϵ is the standard deviation of the residuals in percent per month.

	ARFIMA(1,d_r,0)						AR(p)	
	ML	conditional OLS					OLS	
d_r	0.892	0.6	0.7	0.8	0.9	1.0	0	0
	(0.100)							
ν_1	0.226	0.583	0.454	0.330	0.217	0.117	0.988	1.120
	(0.125)	(0.058)	(0.067)	(0.076)	(0.084)	(0.089)	(0.012)	(0.094)
ν_2								-0.134
								(0.097)
σ_ϵ	0.0395	0.0400	0.0398	0.0396	0.0395	0.0396	0.0398	0.0395

Table 4: Term structure moments

The table reports GMM parameter estimates of the autocorrelation ($\mathcal{M}_\rho$) and relative volatility ($\mathcal{M}_\sigma$) of excess returns. Auxiliary parameters are the variance 5 year excess returns ($\mathsf{Var}[rx_t^{(61)}]$), the measurement error variance ($(60\sigma_u)^2$) and measurement error autocorrelation (θ_1). Standard errors are heteroskedasticity and autocorrelation robust. The test statistic $J(2)$ evaluates the overidentified moments.

	$\mathcal{M}_\sigma$	$\mathcal{M}_\rho$	$\mathsf{Var}[rx_t^{(61)}]$	$(60\sigma_u)^2$	θ_1	$J(2)$
Estimate	1.636	0.115	3.104	0.209	-0.199	0.551
se	(0.053)	(0.043)	(0.466)	(0.030)	(0.046)	

Table 5: Predictive regressions

The first two columns report SUR estimates of the system

$$rx_{t+1}^{(61)} = \alpha_{60} + \beta_{60,60} s_t^{(60)} + \beta_{60,120} s_t^{(120)} + e_{t+1}^{(60)}$$
$$rx_{t+1}^{(121)} = \alpha_{120} + \beta_{120,60} s_t^{(60)} + \beta_{120,120} s_t^{(120)} + e_{t+1}^{(120)}$$

under the restriction $\beta_{120,n} = \mathcal{M}_\sigma \beta_{60,n}$. The next two columns report the unconstrained OLS estimates. The final two columns report OLS estimates of the predictive regression with the own spread as single predictor ($\beta_{60,120} = \beta_{120,60} = 0$). Asymptotic t-statistics are in parentheses. Standard errors for both OLS and SUR regressions are heteroskedasticity and autocorrelation robust.

	$rx_{t+1}^{(61)}$	$rx_{t+1}^{(121)}$	$rx_{t+1}^{(61)}$	$rx_{t+1}^{(121)}$	$rx_{t+1}^{(61)}$	$rx_{t+1}^{(121)}$
constant	-0.24	-0.47	-0.18	-0.46	-0.21	-0.42
	(1.8)	(2.3)	(1.3)	(2.3)	(1.5)	(2.4)
$s_t^{(60)}$	11.69	19.31	10.09	10.65	3.31	
	(3.2)	(5.8)	(2.8)	(1.7)	(2.8)	
$s_t^{(120)}$	-6.32	-10.44	-5.37	-3.24		4.47
	(2.5)	(2.5)	(2.5)	(4.7)		(3.3)
$\mathcal{M}_\sigma$	1.652					
	(12.4)					
R^2	0.032	0.027	0.033	0.034	0.023	0.026
σ_e	1.783	2.999	1.781	2.989	1.789	2.998
$\sigma_e^{(120)}/\sigma_e^{(60)}$	1.682		1.678		1.676	

Table 6: Estimates of d for spreads and predictive factor

The table reports the EW and the LW estimates (on data in levels) of the fractional differencing order for the 5-year spread, the 10-year spread, and the prediction factor from (41). The asymptotic standard error in all cases equals $\hat{\sigma}(\hat{d}) = 0.0981$.

Variable	$\hat{d}_{EW}$	$\hat{d}_{LW}$
5-year spread $s_t^{(60)}$	0.4219	0.4313
10-year spread $s_t^{(120)}$	0.5422	0.5640
Prediction Factor w_t	0.4002	0.3819

Table 7: GMM estimates for d_λ and ξ

The table reports the solutions for a fractional noise risk-price process (d_λ, ξ) that fit the moment conditions (26) and (27) using the estimated data moments in table 4. Figure 7 contains a graphical representation of the same solutions. Standard errors in parentheses are derived from the standard errors of the GMM data moments in table 4. $-.-$ means that the moment conditions have no real-valued solution. Entries > 0.499 indicate that the estimate is larger than 0.499, but still smaller than 0.5.

	AR(p)		ARFIMA(1,d_r,0)					
			$d_\lambda \to \frac{1}{2}$			$d_\lambda < \frac{1}{2}$		
Par	1	2	0.6	0.7	0.8	0.8	0.89	1
d_λ	0.499	0.498	>0.499	>0.499	0.499	0.318	0.471	-.-
	(0.001)	(0.001)	(0.000)	(0.001)	(0.003)	(0.054)	(0.039)	-.-
ξ	0.028	0.042	0.022	0.001	-0.030	-0.109	-0.089	-.-
	(0.007)	(0.007)	(0.008)	(0.010)	(0.024)	(0.041)	(0.024)	-.-
R^2	0.139	0.152	0.134	0.003	0.088	0.016	0.045	-.-

Table 8: Implied predictive regressions

The table shows the implied regression results from predicting excess returns with maturity n ($n = k, m$) by the lagged own spread $s_t^{(n)}$ or the prediction factor $w_t = s_t^{(k)} + \gamma s_t^{(m)}$. The entries report both the regression slope β_n and the fit R^2. Results are reported for different sets of parameters values (d_r, d_λ, ξ) that match the data moments. Column headings refer to the assumed value of d_r or the AR(p) model for the spot rate. ($d_\lambda \to \frac{1}{2}$) refers to solutions in table 7 with d_λ close to non-stationarity boundary; ($d_\lambda < \frac{1}{2}$) refers to solutions with d_λ further away from the boundary. The final column repeats the values from the actual regressions in table 5.

			AR(p)		ARFIMA(1,d_r,0)					data
					d_r ($d_\lambda \to \frac{1}{2}$)			d_r ($d_\lambda < \frac{1}{2}$)		
	n	Pred.	1	2	0.6	0.7	0.8	0.8	0.89	
$100\times R_n^2$	k	$s_t^{(k)}$	0.0	0.1	0.0	0.0	0.3	0.8	2.3	2.3
	m	$s_t^{(m)}$	0.0	0.1	0.0	0.0	0.4	0.8	2.2	2.6
	k	w_t	0.0	0.1	0.0	0.0	0.2	0.9	2.2	3.2
	m	w_t	0.0	0.1	0.0	0.0	0.2	0.9	2.2	2.7
β_n	k	$s_t^{(k)}$	0.14	0.80	-0.28	-0.03	1.14	1.60	1.86	3.31
	m	$s_t^{(m)}$	0.13	0.77	-0.24	-0.05	1.30	1.82	1.90	4.47
	k	w_t	2.85	6.76	-1.35	-0.09	3.91	5.86	9.75	11.69
	m	w_t	4.69	11.08	-2.21	-0.16	6.40	9.59	15.93	19.31

Table 9: GMM estimates for ϕ and ξ

The table reports the solutions (ϕ, ξ) that fit the moment conditions (26) and (27) using the estimated data moments in table 4, for a stationary AR(1) risk-price process. Standard errors in parentheses are derived from the standard errors of the GMM data moments in table 4. Entries > 0.999 indicate that the estimate is larger than 0.999, but still smaller than 1.0.

	AR(2)	ARFIMA(1,d_r,0)					
		0.8	0.8	0.89	0.89	1	1
ϕ	>0.999	>0.999	0.945	>0.999	0.968	>0.999	0.980
	(0.001)	(0.001)	(0.369)	(0.001)	(0.010)	(0.001)	(0.012)
ξ	0.004	-0.002	-0.073	-0.004	-0.062	-0.008	-0.054
	(0.001)	(0.001)	(0.185)	(0.001)	(0.044)	(0.001)	(0.058)
R^2	0.111	0.012	0.048	0.011	0.058	0.108	0.066

Table 10: Implied predictive regressions: AR(1) price of risk

The table shows the implied regression results from predicting excess returns with maturity n ($n = k, m$) by the lagged own spread $s_t^{(n)}$ or the prediction factor $w_t = s_t^{(k)} + \gamma s_t^{(m)}$. The entries report both the regression slope β_n and the fit R^2. Results are reported for different sets of parameters values (d_r, ϕ, ξ) that match the data moments. The final column repeats the values from the actual regressions in table 5.

	n	pred.	AR(2)		ARFIMA(1,d_r,0)				data
			ν_1	1.120	d_r	0.8	0.89	1	
			ν_2	-0.134	ν_1	0.330	0.226	0.117	
			ϕ	>0.999	ϕ	0.945	0.968	0.980	
			ξ	0.002	ξ	-0.073	-0.062	-0.054	
$100\times R_n^2$	k	$s_t^{(k)}$		4.7		4.7	5.8	6.6	2.3
	m	$s_t^{(m)}$		4.8		4.5	5.7	6.6	2.6
	k	w_t		2.8		4.7	5.8	6.6	3.2
	m	w_t		2.8		4.7	5.8	6.6	2.7
β_n	k	$s_t^{(k)}$		1.75		2.03	2.16	2.52	3.31
	m	$s_t^{(m)}$		1.70		2.87	3.12	3.61	4.47
	k	w_t		18.00		5.20	5.49	6.62	11.69
	m	w_t		29.67		8.50	8.97	10.83	19.31

Figure 1: Autocorrelations of the nominal 3-month T-Bill rate

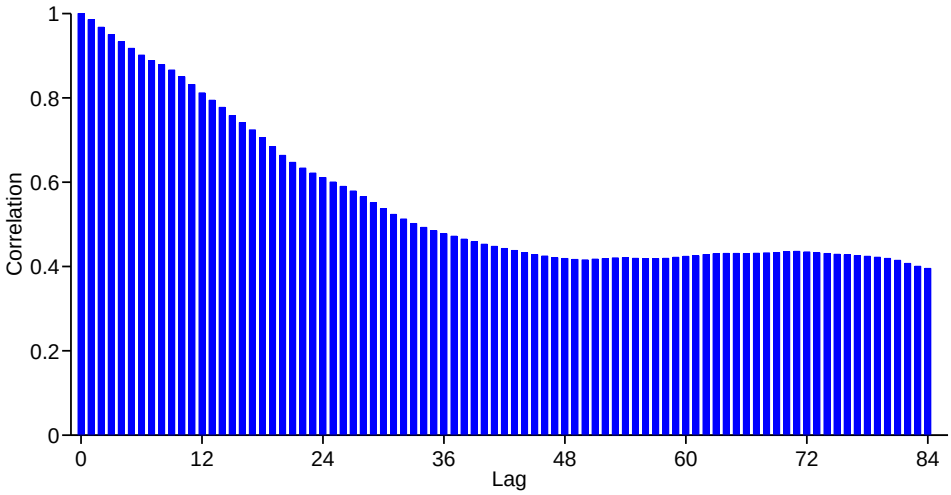


Figure 2: Spot rate impulse responses

The figure shows the spot rate impulse responses c_j for four different specifications. Two models are AR(1) with parameters $\nu_1 = 0.988$ and $\tilde{\nu}_1 = 0.993$, respectively, while the other two are pure fractional models with $d_r = 0.89$ and $\tilde{d}_r = 0.71$, respectively.

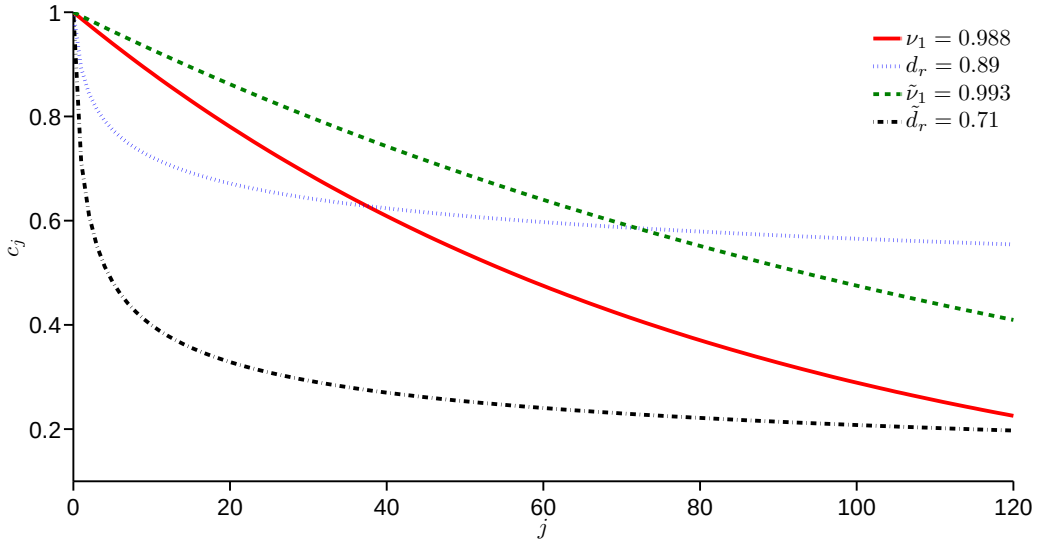


Figure 3: Volatility term structure

The figure shows the implied volatility $\sigma b_0^{(n)}$ as a function of n for four different specifications. For two models (AR(1) with $\nu_1 = 0.988$ and fractional with $d_r = 0.89$) the parameter σ has been calibrated to fit the volatility of 5-year excess returns. For the other two models (AR(1) with $\tilde{\nu}_1 = 0.993$ and fractional with $\tilde{d}_r = 0.71$) parameters $(\sigma,\ \tilde{\nu}_1)$ and $(\sigma,\ \tilde{d}_r)$ have been calibrated to fit both the 5-year and 10-years excess returns.

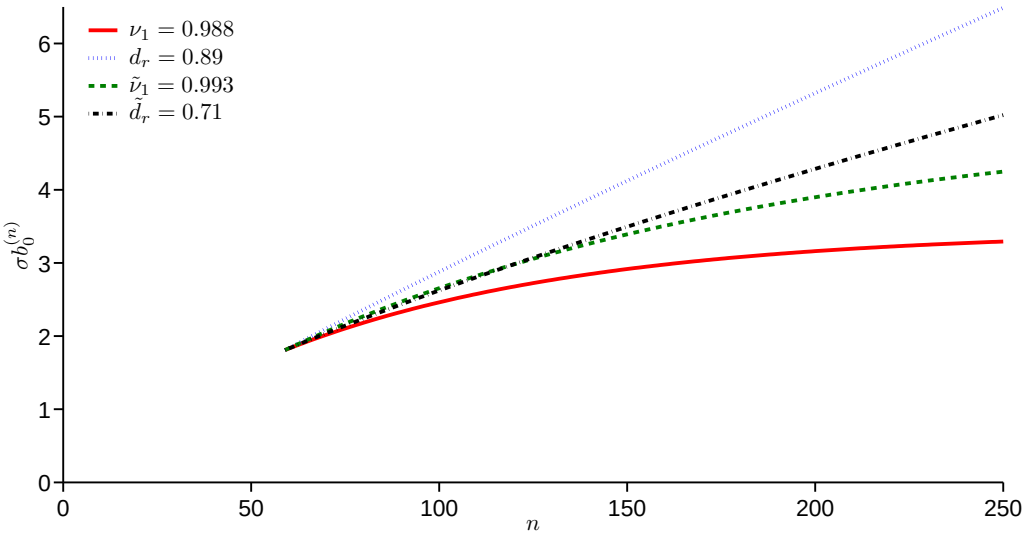


Figure 4: Relative volatility of excess bond returns

The figure shows the ratio $EH = C_{119}/C_{59}$ of the cumulative impulse responses of the short term interest rate implied by an ARFIMA$(1, d_r, 0)$ model with AR parameter ν. The lines show the ratio for fixed d_r as ν varies along the x-axis. Also shown are the EH ratios implied by stationary AR(1) and AR(2) models with the parameter estimates in table 3.

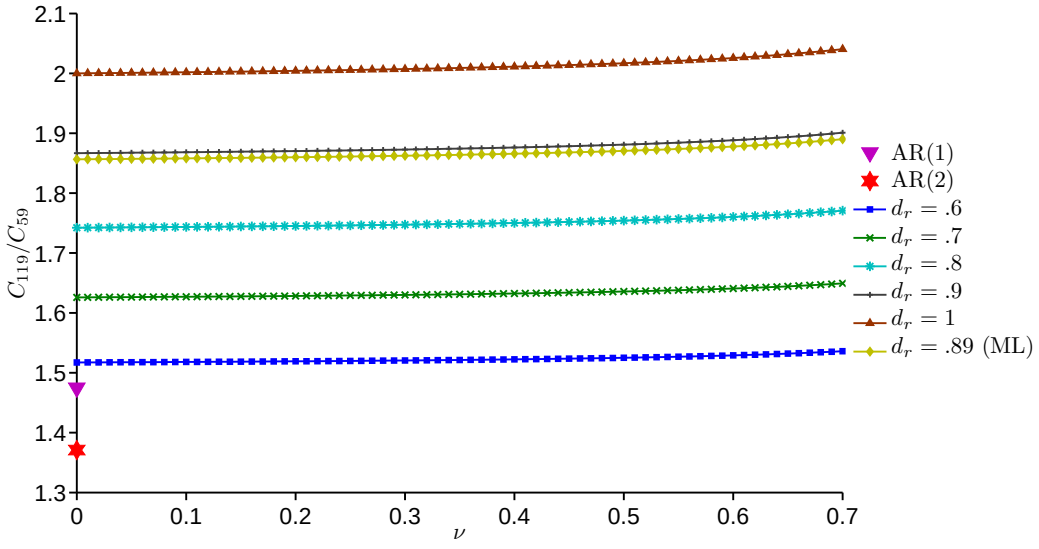


Figure 5: Prediction factor

The figure shows the prediction factor $w_t = -0.46r_t + y_t^{(60)} - 0.54y_t^{(120)}$.

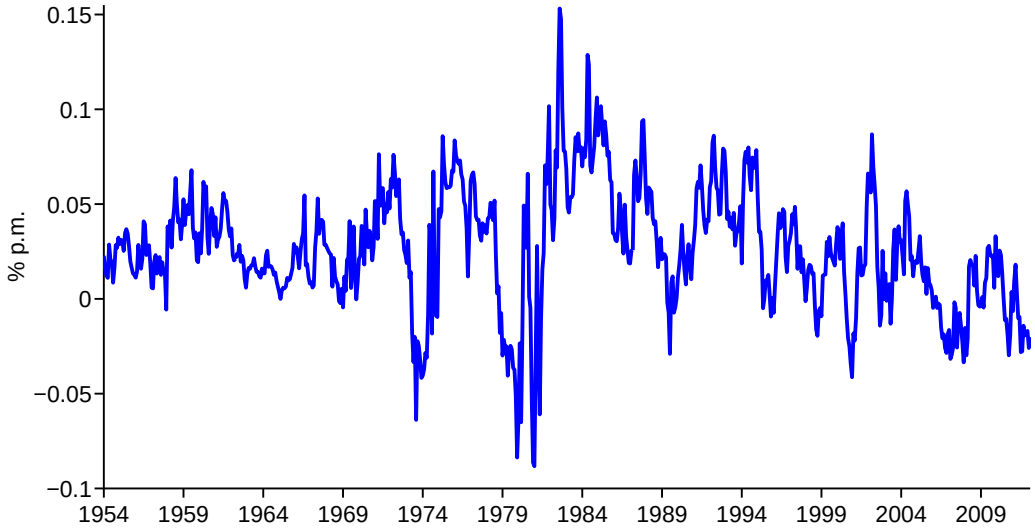


Figure 6: Matching the first order autocorrelation of excess returns

The figure shows the relation between d_λ and ξ that is consistent with the autocorrelation in excess returns. The solid line corresponds to the point estimate for $\mathcal{M}_\rho$. The two dotted lines are for $\mathcal{M}_\rho$ plus or minus two standard errors.

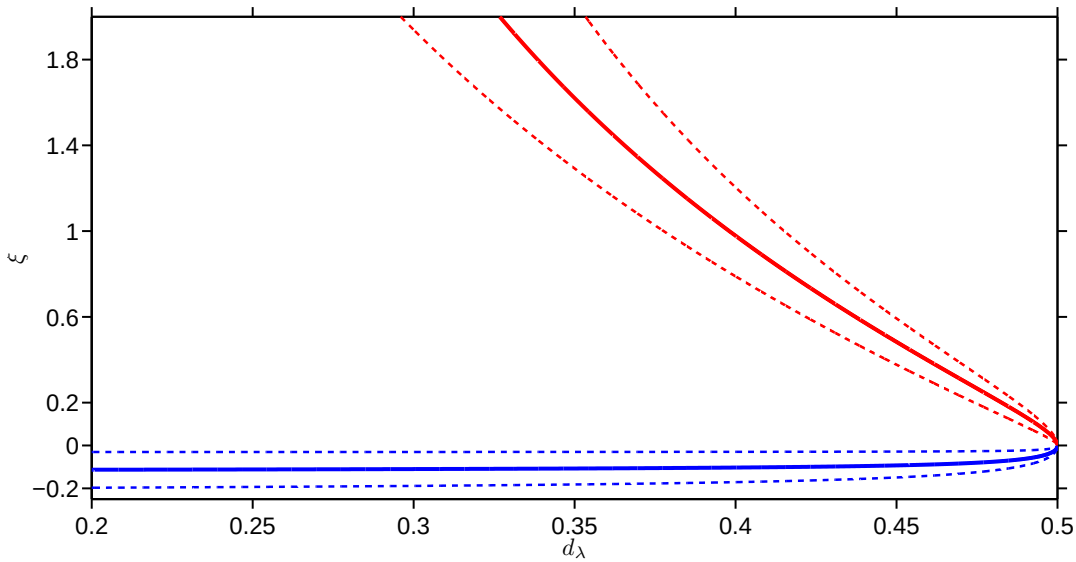


Figure 7: Matching both moments of excess bond returns

The figure shows the relation between d_λ and ξ implied by the relative variance of the 5 and 10-year excess returns for different values of the long-run persistence parameter d_r of the short-term interest rate. The vertical axis denote ξ and the horizontal axis d_λ. We match the dynamic version of the estimate for $\mathcal{M}_\sigma$, i.e. $\mathcal{M}_\sigma$=1.636. Also shown in the graph, is the relation between d_λ and ξ consistent with the respective autocorrelation, $\mathcal{M}_\rho$.

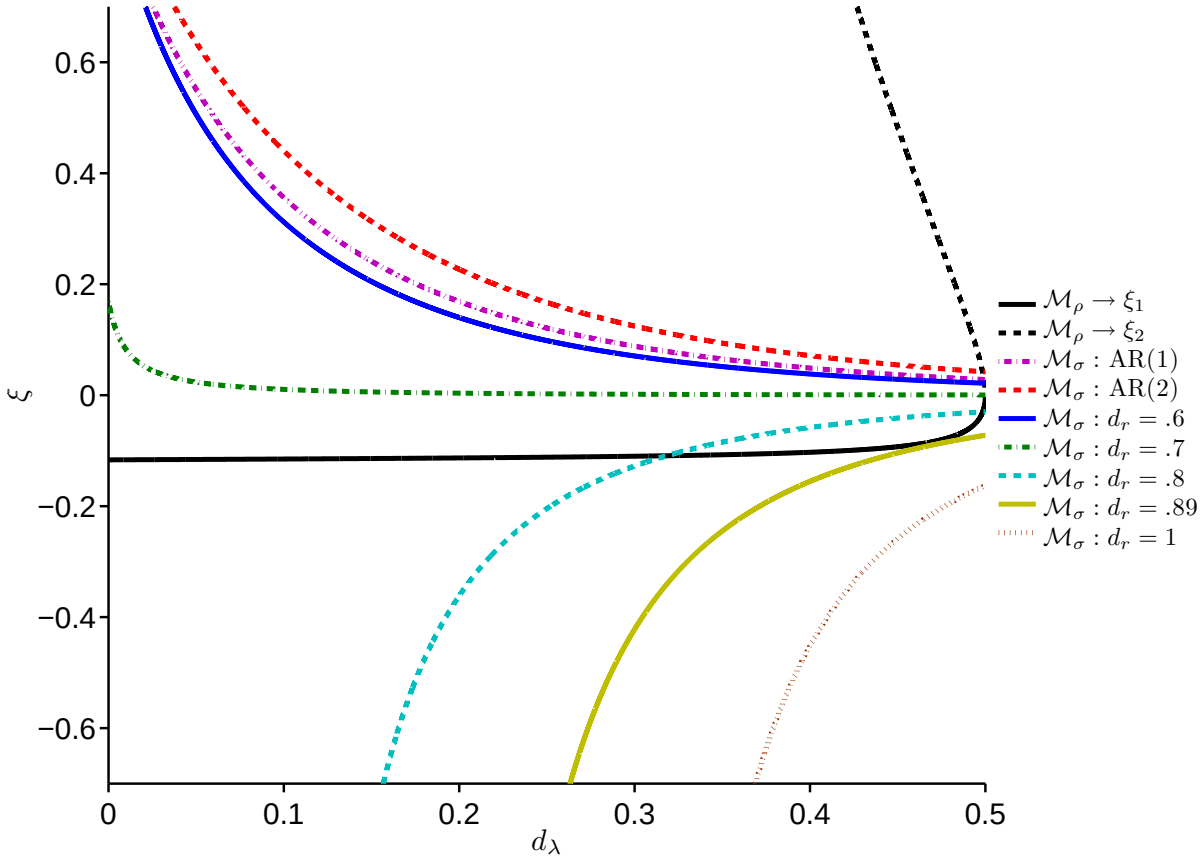


Figure 8: Impulse responses of 10-year rate

The figure shows model-implied impulse responses of the 10-year discount rate for two different models. The *fractional* model uses the parameters in the column d_r=0.892 in table 7. The *AR(2)* model uses the parameters of the AR(2) process in the same table. By construction the initial one standard deviation shock is equal for both models. The two red lines are the impulse responses of unrestricted time series estimates. The fractional model (*TS fractional*) is ARFIMA(1,0.93,0) with AR-parameter equal to 0.118; the AR(2) model (*TS AR*) has an estimated maximum root of 0.99. Units on the vertical axis are percent per annum.

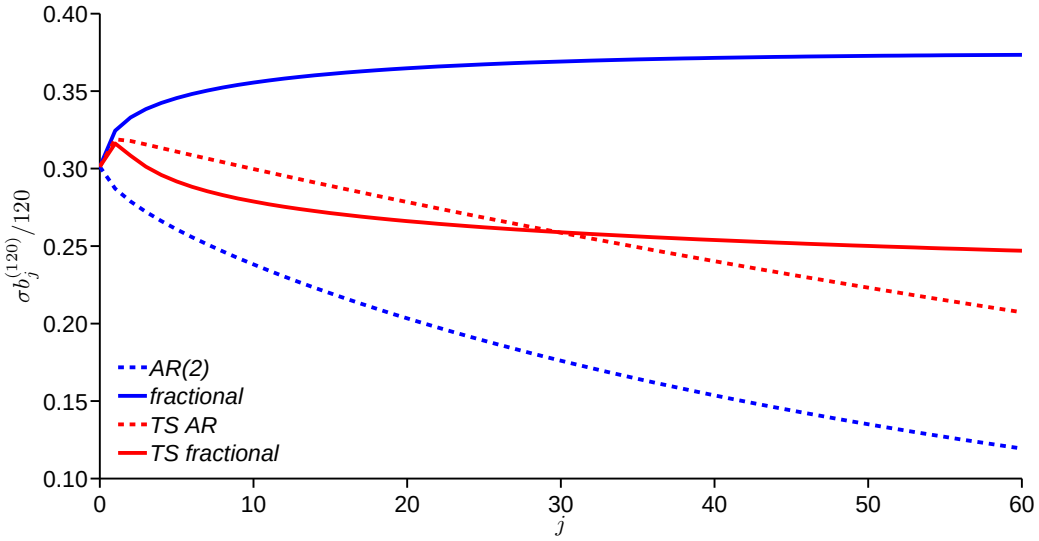
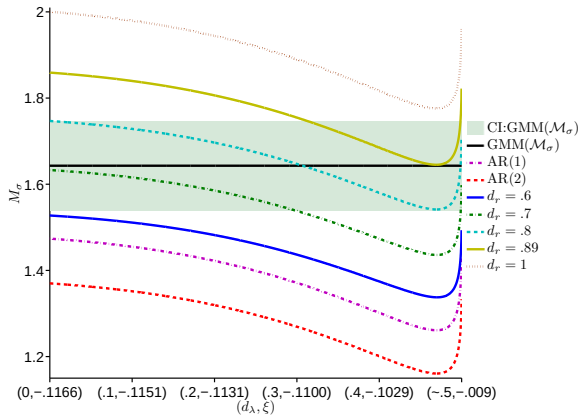
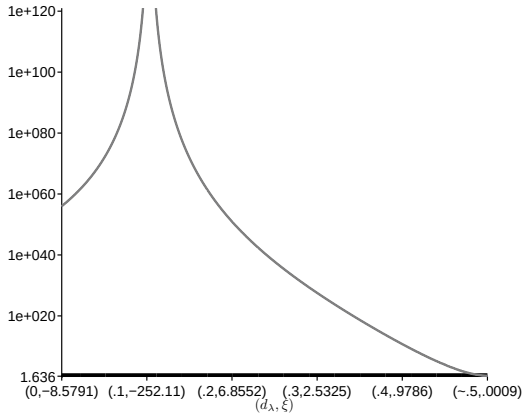


Figure 9: Model-implied relative volatility at solutions consistent with observed autocorrelations

The figure shows model-implied $\mathcal{M}_\sigma$ at values ξ and d_λ consistent with the autocorrelation, $\hat{\mathcal{M}}_\rho$, in the data. The thick black line is the dynamic data moment, $\hat{\mathcal{M}}_\sigma = 1.636$. The shaded area is the 95% confidence interval of the GMM estimate for $\mathcal{M}_\sigma$. The colored lines correspond to model-implied $\mathcal{M}_\sigma$ for different values of d_r. The horizontal axis shows the values of the pair (d_λ, ξ) that match the autocorrelation of excess returns. Panel (a) shows the negative ξ_1 root from (26); panel (b) the ξ_2 root. The vertical axis in the right panel (b) has a log-scale.



(a) Solutions ξ_1 and d_λ



(b) Solutions ξ_2 and d_λ