Long Memory Affine Term Structure Models*

Adam Golinski **Paolo Zaffaroni**

This draft: 13th January 2014

Abstract

We develop a Gaussian discrete time essentially affine term structure model which allows for long memory. This feature reconciles the strong persistence observed in nominal yields and inflation with the theoretical implications of affine models, especially for long maturities. We characterise the dynamic and cross-sectional implications, in particular in terms of volatility, of long memory for our model. We explain how long memory can naturally arise within the term structure of interest rates, providing a theoretical underpinning for our model. Despite the infinite-dimensional structure that long memory implies, we show how to cast the model in state space and estimate it by maximum likelihood. We present an empirical example where we estimate a two-factor version of the model whereby the unobserved factors have a clear economic interpretation as the real short rate and expected inflation.

*Golinski: University of York, Department of Economics and Related Studies, Heslington, York, YO10 5DD, UK. E-mail: `adam.golinski@york.ac.uk`.
Zaffaroni: Imperial College Business School, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK. E-mail: `p.zaffaroni@imperial.ac.uk`.

1. Introduction

Modelling the term structure of interest rates is a relevant from many different perspectives, both academic and practical. For instance, central bankers would be interested in extracting inflation expectations and future movements of short rates embedded in nominal yields. From a macroeconomics angle, deriving the term structure of real interest rates allows to measure the cost of investment and its implication for economic growth. From a finance perspective, it is crucial to price accurately nominal and inflation-indexed bonds and to quantify the associated term premia.

The main challenge is that nominal observed yields are extremely persistent, in fact hardly distinguishable from a nonstationary series. A routine test would hardly reject the hypothesis of a unit root. Although explicitly assumed in early work of term structure modelling (see Dothan (1978)), accepting the possibility of a unit root in the physical measure appears troublesome in terms of its economics implications and econometric estimation. In fact, the unit root paradigm rules out any degree of mean-reversion, namely the possibility that shocks are eventually absorbed as time goes by. Lack of mean-reversion bears implausible cross-sectional predictions, in particular in terms of the volatility term structure of yields, forward rates and holding period returns. In terms of estimation, the possibility of a unit root affects the finite sample as well as the asymptotic properties of conventional estimators of term structure models, making inference more difficult. Recognising that the notion of long memory permits to obtain a substantial degree of persistence, in fact even non stationarity, together with dynamic mean-reversion, this paper develops a class of discrete time no-arbitrage affine term structure models with long memory state variables. The idea of long memory has been postulated as a suitable description of nominal yields by Backus and Zin (1993), which can be seen as a very special case of our general theory[1].

Our long memory model belongs to the class of essentially affine (in the sense of Constantinides (1992), Duffee (2002) and Dai and Singleton (2002)) Gaussian term structure model with multiple factors. We establish the closed-form solution of the model and, relying on its state space representation, show how to carry out estimation by maxi- mum likelihood and Kalman filtering of the latent state variables. These achievements are non trivial because a critical feature of long memory models is to be non-Markov implying, in our affine term structure context, infinite-dimensional state variables.

Our approach shares the many virtues of the powerful class of affine models, formally defined by Duffie and Kan (1996) and pioneered by Vasicek (1977) and Cox et al (1985) highly influential models. First, closed-form solution for bond prices and yields can be easily obtained as affine functions of a set of state variables. Second, nominal yields can be decomposed into inflation expectations, real yields and inflation risk premia with minimal, no-arbitrage, assumptions. Third, conditional moments, in particular term premia, can be easily computed. Fourth, the model can be naturally cast in state-space implying that parameters estimation and inference can be obtained by maximum likelihood estimation. Filtered values of the latent state variables, which typically include expected inflation and the short-term real interest rate, follow by the Kalman recursion.

[1]Related work is also Comte and Renault (1996) who analyse a continuous time long memory model and the equilibrium approach of Duan and Jacobs (1996) where long memory enters through the volatility of the state variables.

To better understand the analogies, and differences, of our model with the conventional affine models, it is useful to consider the unified framework represented by the class $DA^{\mathbb{Q}}_{M}(N)$ of discrete-time affine models spelled out by Le et al (2010)[2], where M of the N factors (here $0 \leq M \leq N$) drive stochastic volatility. Gaussian affine models, whereby the unconditional distribution of the state vectors is normal, feature $M = 0$ (no stochastic volatility) and makes the $DA^{\mathbb{Q}}_{0}(N)$ class. A crucial feature of the $DA^{\mathbb{Q}}_{0}(N)$ class is that the N state variables form a Markov system, possibly of higher yet finite order, under the risk-neutral (hereafter $\mathbb{Q}$) measure, such as a vector autoregression (hereafter VAR). It is well known that the Markov property, together with stationarity under the physical measure, implies a weak form of temporal dependence for model-implied yields, as expressed by the fast decay toward zero of their autocorrelation function. At the same time, a stationary VAR under the $\mathbb{Q}$ measure implies that the theoretical volatility, both conditional and unconditional, of long yields and forward rates diminishes fast toward zero as maturity increases. Instead, the model-implied volatility of holding period returns stays bounded for large maturities. These features are completely at odds with the empirical evidence. However, if one relaxes the assumption of stationarity under the $\mathbb{Q}$ measure, within this $DA^{\mathbb{Q}}_{0}(N)$ class, a unit or even an explosive root emerges the consequences of which are also at odds with the empirical evidence: as discussed above, the theoretical (conditional) volatility of yields and forward rates is either flat (in the unit root case) or explosive across maturity whereas it will always increase sharply for returns.

In contrast, due to the long memory specification of our model, we are able to match the strong degree of persistence together with the dynamic mean-reversion observed in nominal yields. When looking at the characteristics across maturity, our model- driven term structure of volatility for yields and forward rates can be slowly decaying for intermediate maturities yet flattening out or even (slowly) increasing for long maturities. At the same time, the model-driven volatility term structure will diverge for returns. These implications are now compatible with stationarity. More importantly, these are the features observed in the data. As we shall see, long memory can be obtained by allowing the number of state variables, N, to become infinite, spanning the $DA^{\mathbb{Q}}_{0}(\infty)$ class of term structure models, with respect to the Le et al (2010) notation. Besides infinite-dimensionality of the state variables, a suitable long lags characterization of the state variables impulse response is required in order to induce long memory.

Obviously these various issues raised by the persistence in nominal bond data have attracted a great deal of interest and different approaches have been developed. These are reviewed in Appendix C, where we discuss their analogies with our long memory framework.

Although our theory is completely general, we then present a model that includes realised inflation within the set of observables, and thus expected inflation as one of the state variables. This makes our model akin to terms structure models that merge yields and macroeconomic data, such as the $DA^{\mathbb{Q}}_{0}(N)$-type models of Ang and Piazzesi (2003), Rudebush and Wu (2008) and Hordhalh et al (2008) among others. Including inflation is instrumental for recovering the canonical decomposition of nominal yields into the term structure

[2]This class nests all the exact discrete-time representation of the general class of continuous-time models of Dai and Singleton (2000). Under the physical measure this class of models might feature nonlinearity but are characterised by a closed-form expression of the exact likelihood.

of real yields, inflation expectation and inflation risk premia[3]. It is also asked for by the data. In fact long memory appears to be a robust description of realised inflation dynamics. Altissimo et al (2009) analyse how the consumer price index (hereafter CPI) construction protocol gives rise naturally to long memory in CPI inflation and provide empirical evidence for the inflation rate of the euro area. As a consequence, inflation appears to be one of the main channels that naturally leads to long memory in observed nominal yields, as argued below.

Since Rogers (1997), it is well known that assuming long memory for a tradable asset might lead to existence of arbitrage opportunities. This would undermine the possibility to identify the pricing kernel and thus, in our case, to determine model-implied (bond) prices. However, it is now understood that the conditions required to violate no-arbitrage are much more stringent in a discrete time setting (see Cheridito (2003)) such as ours. Moreover, arbitrage opportunities are ruled out whenever transaction costs, no matter how minimal, are allowed for, ensuring existence and uniqueness of the pricing kernel (see Guasoni et al (2010)). Therefore, as discussed below, in practice no pricing consequence for our model appears to arise despite its long memory feature.

The paper proceeds as follows. Section 2 describes the data for nominal yields and inflation used for estimation of the model. We highlight some features of the yields data, namely their dynamic persistence and the shape of their volatility term structure, especially for long maturities. Section 3 explores the extent to which these features can be accounted for by Vasicek-type model, spelling out the theoretical implication for long term yields, forward rates and returns. This paves the ground for the model presented in Section 4: a discrete time essentially affine non-Markov Gaussian term structure model with long memory. With no loss of generality, we focus on the case of two latent factors and establish closed-form solution of the model for a general parameterization of the state variables dynamics, in terms of the nominal and real term structures. Section 4.4. provides analytical characterization of the time series and cross-sectional properties, in terms of volatility term structure, for model-implied yields, forward rates and holding period returns, under various forms of the market prices of risk. Section 5 discusses theoretical underpinnings of long memory in real and nominal yields, leaving some formal details to Appendix A. Estimation results are described in Section 6 which presents an empirical example. A technical description of the Kalman filter and an approximate maximum likelihood estimator for long memory processes is relegated to Appendix D. Having estimated a simple two-factor version of the long memory model, we verify in Section 6.2 that the above described features of the empirical distribution of zero coupon bonds are extremely well matched by the model. The estimated two-factor model is rich enough to decompose nominal yields into the real interest rate, inflation expectation and the inflation risk premia term structures, as exemplified in Section 6.3. Some indications on the statistical performance of the estimated long memory model are examined in Section 6.4 where we presents some out-of-sample forecasting performance results, comparing our affine long memory model with well-established models such as Ang and Piazzesi (2003) macro

[3]Alternative methods for recovering the real term structure and inflation expectation uses inflation-indexed bonds (see Barr and Campbell (1997) and Evans (1998) among others), Treasury inflation-protected securities (see D'Amico et al (2008) and Christensen et al (2010) and among others), survey forecasts of inflation (see Pennacchi (1991) and Chernov and Mueller (2012) among others) and inflation-based derivatives (see Haubrich et al (2012) and Kitsul and Wright (2012)).

model, Ang et al (2008) regime switching model and Diebold and Li (2006) cross-sectional model. Final remarks make Section 7. Appendix A explains how long memory can be induced within the class of affine term structure models. Appendix B discusses the pricing implications of long memory for our model. A review of the different approaches to tackle the high persistence of observed nominal yields, and their analogies with our long memory approach, are discussed in Appendix C. Particular emphasis is given to regime switching term structure models. Appendix E contains two technical lemmas and the proofs of the main theorems.

2. Some stylised facts of nominal bonds and inflation

We now highlight the strong, well established, degree of dynamic persistence that characterises certain specific aspects of the empirical distribution of nominal bonds and consumer price index (CPI) inflation. Regarding nominal bonds, we consider the term structure of nominal yields, forward rates and holding period returns. Noticeably, we wish to emphasise how the extremely strong degree of time series persistence appears to influence certain cross-sectional aspects of the yields distribution, namely the term structure of volatility, of nominal yields, forward rates and returns. In particular, this strong persistence appears to be the main channel through which the negligible volatility of bond returns at very short maturities becomes magnified by several orders of magnitude as we move along the term structure. Similarly, the riskiness of long term yields and forward rates appear only slowly declining along the term structure, far from vanishing for very long maturities. At first glance, these stylised facts can be qualitatively rationalised by means of a simple Markov term structure model, as exemplified in Section 3. However, anticipating matters, when looking more carefully, both the time series and the cross-sectional evidence appear at odd with the quantitative predictions of such term structure model built around both stationary and non-stationary Markov state variables.

2.1. Nominal bonds

This section uses a data set comprised of monthly observations of nominal yields $r^{\$}_{n,t}$ on zero coupon bonds with maturities n equal to 1 and 3 month, $1, 3, 5, 10, 15, 20-$ and 30 year. The source for the 1 and 3 month yields is the Fama's Treasury bills term structure files, while for maturities up to 5 year are the Fama-Bliss discount bond files. The 10 to 30 year yields are extracted from the data of Gurkaynak et al (2007) corresponding to the period January 1986 to December 2011. Yields

$$r^{\$}_{n_i,t} = -\frac{1}{n_i} log P_{n_i t},$$

are continuously compounded, annualised and percent, where $P_{n_i t}$ denotes nominal zero coupon bond prices with maturity n_i. We also consider (nominal) forward rates

$$f^{\$}_{n_i,n_{i+1},t} = (n_{i+1} r^{\$}_{n_{i+1},t} - n_i r^{\$}_{n_i,t})/(n_{i+1} - n_i) \text{ with maturities } n_i < n_{i+1},$$

and holding period returns

$$y^{\$}_{n_{i-1},n_i,t} = (n_i r^{\$}_{n_i,t-(n_i-n_{i-1})} - n_{i-1} r^{\$}_{n_{i-1},t})/(n_i - n_{i-1}) \text{ with maturities } n_{i-1} < n_i,$$

referring to them as $f^{\$}_{n_i,t}$ and $y^{\$}_{n_i,t}$ when $n_{i+1} - n_i = 1$. Summary statistics are presented in Table 1.

[**Insert Table 1 near here**]

Average yields are increasing with maturity whereas their volatility, expressed in terms of standard deviation, shows a hump at about one year maturity and then flattens out. A similar pattern is obtained in terms of forward rates, the main difference being that for forwards their volatility term structure raises sharply again for long maturitities after declining from the one-year hump. Holding period returns exhibit a monotonically increasing volatility curve.

It has been known for a long time that nominal yields display a substantial degree of persistence[4]. This is evident when performing unit root tests, as illustrated Table 2 where we present the results for the standard Augmented Dickey-Fuller (ADF) unit root test.

[**Insert Table 2 near here**]

The null hypothesis of unit root is not rejected for nominal yields across all maturities. However, the unit root paradigm rules out mean-reversion, which appears implausible from an empirical and theoretical viewpoint. Moreover, as exemplified below, unit root dynamics raises implausible cross-sectional predictions within affine term structure models.

We propose to assess the persistence of nominal bonds characteristics using a somewhat more sophisticated approach that does not suffer the limits of the unit root framework. In particular, we need to use a measure that allows to disentangle the notion of nonstationarity from the one of mean-reversion.

Figure 1(a) plots the periodogram ordinates near the zero frequency for yields, forwards and returns, averaged across maturity[5] where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is

$$I_w(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} w_t e^{\imath \lambda t} \right|^2, \quad -\pi < \lambda \leq \pi,$$

where $\imath$ defines the complex unit. Data have been standardised so that the sample variance is unity.

[**Insert Figure 1 near here**]

The strengths of using the periodogram come essentially from the fact that it is a non-parametric measure and a function of the entire strings of sample autocorrelation of the data. More in general, it gives neat insights on both the low, medium and high frequency dynamics of the data, which in turn are linked to the long run persistence, mean-reversion,

[4]See for example Ball and Torous (1996) and Kim and Orphanides (2012) among many others.

[5]The same pattern is observed for the single maturities with little variation.

and cycles of the data. For instance, the periodogram near zero frequency is proportional to the sum of the entire set of sample autocorrelations corresponding to a given sample and, as such, is a clearcut measure of long run persistence[6]. Instead, the local behaviour of the periodogram, as one moves away from the zero frequency, provides indications on the degree of mean-reversion.

Given the substantial mass of the periodogram of the data near zero frequency, it could be more insightful to examine the log periodogram where large values of the periodogram are mitigated. Figure 1(b) clearly shows that yields, forwards and returns all display a negatively sloped log-periodogram near the origin, for at least the first twenty or thirty frequencies.

To provide a benchmark, any stationary autoregressive moving average (ARMA) process implies a zero-sloped logarithm of the spectral density near the zero frequency. We plot the spectral density for AR(1) model with unit variance with autoregressive parameter equal to $0.80, 0.98, 0.99999$, represented by the the blue, red and green line, respectively, in Figure 2(a) together with the periodogram of yields, forward rates and returns. Figure 2(b) reports the same quantities in the logarithmic scale.

[Insert Figure 2 near here]

The comparison is striking: even a value as large as 0.98 does not induce a sufficiently large degree of persistence able to match the peak found in the periodogram of the data near the zero frequency. The mean-reversion implied by stationary ARMA is also too strong. The case of autoregressive parameter equal to 0.99999 appears more akin to the data at zero frequency. However, for this case it would be hard to deny the existence of a unit root. Moreover, although dense near zero, the behaviour of the log-spectrum with an autoregressive parameter of 0.99999 does not match too well the other higher frequencies, in particular the slow, hyperbolic, decay of the log periodogram of the data as the frequency increases. We interpret this as the limit of the unit root paradigm, able to induce persistence but at the cost of giving up stationarity and, in particular, mean-reversion. This provides implausible predictions for the volatility cross-section of nominal bond characteristics across maturities.

We summarize this finding as follows.

Stylized Fact 1. Nominal yields, forwards and holding period returns are highly persistent across time yet mean reverting. In particular they all display a negatively sloped log periodogram near the origin, slowly decaying as the frequency increases.

Figure 3(a) displays the term structure of the sample standard deviation of yields and forward rates. As observed in Table 1, for yields, the curve is decaying yet with a hump at about two year maturity, flattening for longer maturities around a level well above zero. Forward rates have a similar pattern, although they show a more substantial increase toward the end of the term structure, clearly non vanishing with maturity.

[Insert Figure 3 near here]

[6]The periodogram can be rewritten as $I_w(\lambda) = (1/2\pi)\sum_{k=-T+1}^{T-1} \hat{cov}_w(k)e^{\imath k\lambda}$ for $\lambda \neq 0$ where $\hat{cov}_w(k) = T^{-1}\sum_{t=1}^{T-|k|}(w_t - \bar{w})(w_{t+|k|} - \bar{w})$, namely the sample autocovariance at lag k (see Brockwell and Davis (1991), Proposition 10.1.2).

Figure 3(b) reproduces the term structure of the sample standard deviation of holding period returns. Differently from yields and forward rates, the volatility of returns raises steeply with maturity. These observations lead to:

Stylized Fact 2. The term structure of the sample standard deviation of nominal yields and forward rates slowly declines up to mid maturities and then flattens out or even slowly increases for long maturities. The term structure of the sample standard deviation of nominal holding period returns rises with maturity without flattening out.

These facts are well documented in the term structure literature. Note that although Stylized Fact 1 is a time series characteristic, Stylized Fact 2 features cross-sectional aspects of the bond data. However, these are intimately related and can be rationalised within an affine framework. The approach proposed in this paper tries to explain these features.

2.2. Inflation

Inflation data are taken from Bureau of Labor Statistics of U.S. Department of Labor, where monthly observations are available from June 1947 to December 2011. We calculate (non seasonally adjusted) inflation based on the all urban consumer price index CPI_t as:

$$\pi_t = \log(CPI_t) - \log(CPI_{t-1}). \tag{1}$$

In Table 1 we report the basic summary statistics where inflation is annualised and in percent. The average inflation is 2.80% with a standard deviation of 3.57%. The last entry of Table 2 reports the Augmented Dickey-Fuller test for inflation. The unit root hypothesis now cannot be rejected at 5% significance level. However, as for yields, we look at the more revealing plot of the log periodogram, reported in Figure 4. The periodogram of (normalized) inflation is plotted (blue line) together with the spectral density of autoregressive processes of order one with parameters 0.80, 0.98 and 0.99999 (green, red and light blue line respectively), all of which appear inadequate. The strong degree of persistence is evident yet with some feature of mean-reversion, as expressed by the slow decay as the frequency increases.

[**Insert Figure 4 near here**]

Persistence of CPI inflation, in particular in the form of long memory, has been documented (for the euro area) in detail by Altissimo et al (2009) who argue that this is an unavoidable consequence of the way in which price indexes are constructed. This is revisited in Section 5 below. Persistence of observed inflation clearly reflects the persistence of expected inflation, the key variable in term structure modelling. This fact also suggests that the magnitude of the volatility of unexpected inflation (the difference between observed and expected inflation), which by construction is serially uncorrelated across time, cannot be too large for otherwise it would mask the persistence found in the data.

3. Implications for Markov affine models

We now revisit the theoretical implications of the persistence of yields, found in the data, for Gaussian Markov affine models. Consider the discrete time version of the Vasicek (1977) model, a one-factor Gaussian model. The price of a real zero-coupon bond issued at time t which expires n periods ahead, here denoted $Q_{n,t}$, satisfies the no-arbitrage condition

$$Q_{n,t} = E_t\left(e^{m_{t+1}} Q_{n-1,t+1}\right), \tag{2}$$

where $E_t(\cdot)$ is the expectation operator conditional on the information available up to time t, based on the physical measure. It is well-known that assuming no-arbitrage implies existence of the real pricing kernel $e^{m_{t+1}}$ the exponent of which, for this model, has the simple form

$$-m_{t+1} = \mu_r + \frac{1}{2}\lambda^2\sigma_x^2 + x_t + \lambda\epsilon_{x,t+1} \tag{3}$$

where the (single) factor follows an AR(1) process

$$x_t = \psi_x x_{t-1} + \epsilon_{x,t}, \;\; \epsilon_{x,t} \sim NID(0, \sigma_x^2), \tag{4}$$

and $\lambda, \mu_r, \psi_x, \sigma_x^2$ are constant parameters with $\mid \psi_x \mid< 1$.

By the standard recursive method one obtains that bond yields $r_{t,n} = -(1/n)\log Q_{n,t}$, forward rates $f_{n,t}$ and holding one-period returns satisfy, respectively,

$$\begin{aligned} r_{t,n} &= n^{-1}(A_n + B_n x_t), && (5)\\ f_{t,n} &= A_{n+1} - A_n + (B_{n+1} - B_n)x_t, && (6)\\ y_{t,n} &= A_n - A_{n-1} + B_n x_{t-1} - B_{n-1}x_t, && (7) \end{aligned}$$

where, in turn, the n-varying coefficients satisfy the well-established Riccati difference equations

$$A_n = A_{n-1} + \mu_r - \lambda\sigma_x B_{n-1} - \frac{1}{2}B_{n-1}^2\sigma_x^2, \quad B_n = 1 + \psi_x B_{n-1}, \tag{8}$$

with initial conditions $A_0 = B_0 = 0$.

Consider first the stationary case $|\psi_x| < 1$ giving $B_n = (1-\psi_x^n)/(1-\psi_x)$. Clearly yields, forward rates and returns are elementary (linear) transformation of the AR(1) process x_t, and their temporal dependence, under the physical measure, is determined by the magnitude of ψ_x. Analytically, the spectral densities for yields, forward rates and returns[7] are, for $-\pi \leq \lambda < \pi$,

$$\begin{aligned} s_{r_n}(\lambda) &= (B_n/n)^2\, s_x(\lambda),\\ s_{f_n}(\lambda) &= (B_n - B_{n-1})^2\, s_x(\lambda),\\ s_{y_n}(\lambda) &= s_x(\lambda) + B_{n-1}^2\frac{\sigma_x^2}{2\pi} + 2B_{n-1}\frac{\sigma_x^2}{2\pi}\Re\left(\frac{e^{\imath\lambda}}{1-\psi_x e^{\imath\lambda}}\right), \end{aligned}$$

where $\Re(.)$ denotes the real part of a complex number, $\imath$ is the complex unit and $s_x(\lambda)$ indicates the spectral density of the AR(1) state variable (4), equal to $\sigma_x^2/(2\pi|1-\psi_x e^{\imath\lambda}|^2)$.

[7] For returns $y_{t,n}$ the additional terms in the spectral density are due to the fact $y_{t,n}$ can be represented as $A_n - A_{n-1} + x_{t-1} - B_{n-1}\epsilon_{x,t}$. However the behaviour of the first and third term in $s_{y_n}(\lambda)$ are identical near the zero frequency.

By easy derivations, the slope of the log spectra, for $\lambda \to 0^+$, will then satisfy

$$\begin{aligned}
\frac{dlog\, s_{r_n}(\lambda)}{d\log\lambda} &\sim \frac{2\psi_x}{(1-\psi_x)^2}\lambda^2, \\
\frac{dlog\, s_{f_n}(\lambda)}{d\log\lambda} &\sim \frac{2\psi_x}{(1-\psi_x)^2}\lambda^2, \\
\frac{dlog\, s_{f_n}(\lambda)}{d\log\lambda} &\sim \left(\frac{2\psi_x}{(1-\psi_x)^2} + \frac{2B_{n-1}(1-B_{n-1}\psi_x)}{(1+B_{n-1}(1-\psi_x))^2}\right)\lambda^2.
\end{aligned}$$

In all cases the slope becomes null at zero frequency and its magnitude, near zero frequency, is larger the closer ψ_x is to unity. This was already illustrated in Figure 2(b), which shows how the log-spectral density becomes flat near the zero frequency approaching a positive finite value[8] for ψ_x equal to 0.90 and 0.98, represented by the blue and green line respectively.

The term structures of conditional and unconditional volatility for yields, forward rates and returns are

$$\begin{aligned}
\text{var}_{t-1}(r_{t,n}) &= \left(\frac{B_n}{n}\right)^2\sigma_x^2, \quad \text{var}(r_{t,n}) = \left(\frac{B_n}{n}\right)^2\frac{\sigma_x^2}{1-\psi_x^2}, \\
\text{var}_{t-1}(f_{t,n}) &= \psi_x^{2n}\sigma_x^2, \quad \text{var}(f_{t,n}) = \psi_x^{2n}\frac{\sigma_x^2}{1-\psi_x^2}, \\
\text{var}_{t-1}(y_{t,n}) &= B_n^2\sigma_x^2, \quad \text{var}(y_{t,n}) = B_n^2\sigma_x^2 + \frac{\sigma_x^2}{1-\psi_x^2},
\end{aligned}$$

where $\text{var}_t(\cdot)$ is the variance operator conditional on the information available up to time t, based on the physical measure. Since $B_n \sim 1/(1-\psi_x)$ for large n when $|\psi_x| < 1$, it follows that, as $n \to \infty$,

$$\text{var}_{t-1}(r_{t,n}) \sim \left(\frac{\sigma_x^2}{(1-\psi_x)^2}\right)\frac{1}{n^2}, \quad \text{var}_{t-1}(f_{t,n}) \sim \psi_x^{2n}\sigma_x^2, \quad \text{var}_{t-1}(y_{t,n}) \sim \frac{\sigma_x^2}{(1-\psi_x)^2}, \tag{9}$$

where $\sim$ indicates asymptotic equivalence[9]. An identical pattern is obtained for the unconditional variances. When $|\psi_x| > 1$ the conditional variances will all rapidly explode as ψ_x^{2n} (approximately so for yields) for large n. These results are obtained under the implicit assumption, embedded in (3), of constant market price of risk. If one replaces λ by $\lambda_t = \lambda_0 + \lambda_1 x_t$ into (3), for some constant non-zero parameters λ_0, λ_1, the same formulae for the conditional variance term structures apply but with ψ_x replaced by $\tilde{\psi}_x \equiv \psi_x - \lambda_1\sigma_x^2$, the $\mathbb{Q}$ measure autoregressive coefficient.

Consider now the unit root case $\psi_x = 1$ giving $B_n = n$. Now the model is nonstationary under the physical measure so the variance is not finite and the spectral density is not defined, technically speaking. The quasi-unit root case $\psi_x = 0.99999$ reported in Figure 2(b), red line, shows how the log-spectrum will exhibit a sharp peak near zero frequency. The term

[8] Equivalently, the autocorrelation function is summable and, in particular, proportional to $\psi_x^{|u|}$ at lag u.
[9] We say that $a_n \sim b_n$, where $b_n \neq 0$, when $a_n/b_n \to 1$ as $n \to \infty$.

structure of conditional volatility for yields, forward rates and returns will now be

$$\text{var}_{t-1}(r_{t,n}) = \sigma_x^2, \quad \text{var}_{t-1}(f_{t,n}) = \sigma_x^2, \quad \text{var}_{t-1}(y_{t,n}) = n^2\sigma_x^2. \tag{10}$$

The model is purposely extremely stylized, but it shares the main implications in terms of persistence and of long maturity behaviour of the volatility term structures with more sophisticated discrete affine models with ARMA state variables. In particular, the cases of both stationary and non-stationary ARMA state variables are at odd with the empirical evidence surveyed in Section 2. The stationary case generates a stronger than needed degree of mean-reversion whereas nonstationarity, either a unit or explosive root, rules out mean-reversion altogether. Moreover, postulating a unit root makes invalid the evaluation of impulse responses and variance decomposition. There appears the need for a model able to generate an intermediate degree mean-reversion between these two cases, without imposing nonstationarity. This is accomplished by the long memory affine term structure model, which we formalise in the next section.

4. Long memory affine term structure models: representation

Long memory models, in particular autoregressive fractionally integrated moving average (ARFIMA) models, bridge the gap between stationary ARMA and ARIMA (when a unit root is allowed for). In fact, not only long memory models can describe the dynamics of stationary yet highly persistent time series but can also account for non-stationary yet mean reverting series, whereby the impulse response function will eventually die out with time[10]. There is another, less known, feature of linear long memory models that makes them particularly useful with respect to affine models, namely the fact that they admit a state-space representation although with infinite-dimensional state variables. This result has been established by Chan and Palma (1998) and summarized in Appendix D. More importantly, it turns out that, despite the presence of an infinite number of transition equations, the likelihood can be computed in a finite number of steps. Therefore parameter estimates can be obtained and the Kalman filter delivers optimal out-of-sample forecasts and filtered values of the latent factors.

These considerations suggest to consider Gaussian affine models with long memory state variables. This model is described in the following subsections. We first show how to solve the model imposing the no-arbitrage condition, yielding the real and nominal term structure. This can be obtained for a general specification of the model, yet providing a closed-form solution. We then consider specific parameterizations, such as ARFIMA, which are required in order to carry our estimation.

4.1. Real term structure

We extend the basic model of Section 3 in two directions. First, we consider a two-factor model, with latent factors here denoted by x_t and z_t. Two is the minimal number

[10] In contrast, in the unit root case the impulse response function does not vanish and persists for ever.

of factors that permits to derive both the nominal and the real term structure of interest rates. Extension to multi factors is straightforward and possibly desirable from an empirical point of view. However the main theoretical features of the model would not differ from the ones of the present model. More importantly, the second departure from the basic model of Section 3 is to allow for long memory factors. Assume that the real stochastic discount factor m_t is an affine function of the two latent factors with zero mean, x_t and z_t:

$$-m_{t+1} = \mu_r + x_t + \delta_z z_t + \frac{1}{2}\lambda_t' \mathbf{\Sigma} \lambda_t + \lambda_t' \varepsilon_{t+1} \tag{11}$$

with i.i.d. innovation

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{z,t} \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_z^2 \end{pmatrix} \right) = NID\left(\mathbf{0}, \mathbf{\Sigma}\right). \tag{12}$$

The price of risk is affine in the state variables

$$\lambda_t = \begin{pmatrix} \lambda_{x,t} \\ \lambda_{z,t} \end{pmatrix} = \lambda_0 + \lambda_1 \begin{pmatrix} x_t \\ z_t \end{pmatrix} \tag{13}$$

for a 2×1 vector $\lambda_0 = (\lambda_{x,0},\, \lambda_{z,0})'$ and a $2{\times}2$ matrix $\lambda_1 = \begin{pmatrix} \lambda_{x,1} & \lambda_{z,1} \\ \lambda_{x,2} & \lambda_{z,2} \end{pmatrix}$ of parameters. Formulation (13) qualifies the model as 'essentially' affine. Expression (11) follows by specifying the one-period real interest rate to be an affine function of the factors, that is

$$r_{1,t} = \mu_r + \delta_x x_t + \delta_z z_t = \mu_r + x_t + \delta_z z_t, \tag{14}$$

where we set $\delta_x = 1$, and assuming the existence of a conditionally log-normal stochastic process $\alpha_t = \alpha_{t-1}\exp(-0.5\lambda_{t-1}' \mathbf{\Sigma} \lambda_{t-1} - \lambda_{t-1}' \varepsilon_t)$ such that $E_t^{\mathbb{Q}}(X_{t+1}) = \alpha_t^{-1} E_t(X_{t+1}\alpha_{t+1})$ for any stochastic process X_{t+1}, where $E_t^{\mathbb{Q}}(\cdot)$ defines the conditional expectation operator under the $\mathbb{Q}$ (see Harrison and Kreps (1979)). Hereafter, we shall specify all model equations and parameters in terms of the physical measure, unless stated otherwise.

The parameters μ_r and δ_z represent the unconditional mean of the one-period real interest rate and the loading of the factor z_t, respectively. In particular, as we shall see below, factor z_t represents (demeaned) expected inflation. Thus leaving δ_z unrestricted will allow for a possible transmission channel of money non-neutrality. Finally, (2) follows since the price of any asset that does not pay dividends is a martingale under $\mathbb{Q}$ (once adjusted by $e^{-r_{1,t}}$), that is for zero-coupon bonds $Q_{n,t} = E_t^{\mathbb{Q}}[e^{-r_{1,t}} Q_{n-1,t+1}] = E_t[e^{-r_{1,t}} Q_{n-1,t+1}\alpha_{t+1}/\alpha_t] = E_t[Q_{n-1,t+1} e^{m_{t+1}}]$.

To close the model one needs to specify the dynamics of the latent factors under the physical measure. In order to introduce long memory, we need to make a distinction between latent factors and state variables. The factors, x_t and z_t, bear a precise economic interpretation but their dynamics are more conveniently represented by the infinite-dimensional state vectors $\mathbf{C}_t = (\mathbf{C}_{x,t}',\, \mathbf{C}_{z,t}')'$, which obey an infinite-dimensional VAR(1) model

$$\mathbf{C}_{x,t+1} = \mathbf{F}\mathbf{C}_{x,t} + \mathbf{h}_x \varepsilon_{x,t+1}, \tag{15}$$

$$\mathbf{C}_{z,t+1} = \mathbf{F}\mathbf{C}_{z,t} + \mathbf{h}_z (\varepsilon_{z,t+1} + \beta_{zx}\varepsilon_{x,t+1}), \tag{16}$$

for a constant β_{zx}, infinite-dimensional vectors $\mathbf{h}_x, \mathbf{h}_z$ and a double-infinite dimensional matrix $\mathbf{F}$. Notice that the innovations in (5)-(5) are the same as in (12). Equations (5)-(5) represent the transition equation of the state-space of the model used for the Kalman filter recursion. Obviously we could have written (5)-(5) as $\mathbf{C}_{t+1} = \mathbf{F}^*\mathbf{C}_t + \mathbf{h}^*\varepsilon_{t+1}$ for certain matrices $\mathbf{F}^*, \mathbf{h}^*$ suitably restricted, but it is more convenient to rely on (5-5). The relationship between factors and state variables is simply

$$x_t = \mathbf{G}'\mathbf{C}_{x,t}, \;\; z_t = \mathbf{G}'\mathbf{C}_{z,t}, \tag{17}$$

for an infinite dimensional vector $\mathbf{G} = (1,\, 0,\, 0\cdots)'$ with all zeros from the second row and below. The two factors will be uncorrelated, and thus independent by (12), when $\beta_{zx} = 0$ but can nevertheless both potentially influence the price of real bonds and real yields through the price of risk λ_t.

Despite the infinite dimension of the state variables, it turns out that the model can be solved much in the same way used for the basic model of Section 3. We report the following result without proof.

Theorem 4.1. *For the pricing kernel (11), the market price of risk (13), the state variable dynamics (5)-(5) with innovations (12) and the real interest rate (14), the no-arbitrage zero coupon prices $Q_{n,t}$ satisfy*

$$q_{n,t} = -A_n - \mathbf{B}'_{x,n}\mathbf{C}_{x,t} - \mathbf{B}'_{z,n}\mathbf{C}_{z,t}$$

where $q_{n,t} = \ln Q_{n,t}$ and the coefficients satisfy the Riccati recursions

$$\begin{aligned} A_n &= \mu_r + A_{n-1} - \lambda_{x,0}\sigma_x^2(\mathbf{B}'_{x,n-1}\mathbf{h}_x + \beta_{zx}\mathbf{B}'_{z,n-1}\mathbf{h}_z) - \lambda_{z,0}\sigma_z^2\mathbf{B}'_{z,n-1}\mathbf{h}_z \\ &\quad -\frac{1}{2}\sigma_x^2(\mathbf{B}'_{x,n-1}\mathbf{h}_x + \beta_{zx}\mathbf{B}'_{z,n-1}\mathbf{h}_z)^2 - \frac{1}{2}\sigma_z^2(\mathbf{B}'_{z,n-1}\mathbf{h}_z)^2 \end{aligned} \tag{18}$$

and

$$\mathbf{B}_{x,n} = \Big(1 - \lambda_{x,1}\sigma_x^2(\mathbf{B}'_{x,n-1}\mathbf{h}_x + \beta_{zx}\mathbf{B}'_{z,n-1}\mathbf{h}_z) - \lambda_{x,2}\sigma_z^2(\mathbf{B}'_{z,n-1}\mathbf{h}_z)\Big)\mathbf{G} + \mathbf{F}'\mathbf{B}_{x,n-1} \tag{19}$$

$$\mathbf{B}_{z,n} = \Big(\delta_z - \lambda_{z,1}\sigma_x^2(\mathbf{B}'_{x,n-1}\mathbf{h}_x + \beta_{zx}\mathbf{B}'_{z,n-1}\mathbf{h}_z) - \lambda_{z,2}\sigma_z^2(\mathbf{B}'_{z,n-1}\mathbf{h}_z)\Big)\mathbf{G} + \mathbf{F}'\mathbf{B}_{z,n-1}. \tag{20}$$

Note that A_n is scalar while $\mathbf{B}_{x,n}, \mathbf{B}_{z,n}$ are infinite-dimensional vectors in general. These coefficients are interpreted as evaluated under the $\mathbb{Q}$-measure unless in (13) one sets $\lambda_0 = 0$ for A_n or $\lambda_1 = 0$ for $\mathbf{B}_{x,n}$ and $\mathbf{B}_{z,n}$. For these cases, the corresponding coefficients are interpreted to be evaluated under the $\mathbb{P}$ measure. The distribution of 'observed' bond prices and yields, viz. the physical measure, is of course function of both the $\mathbb{P}$- and $\mathbb{Q}$-measure parameters. The dynamic properties of the latent factors x_t and z_t depend on the chosen parameterization for $\mathbf{F}, \mathbf{h}_x, \mathbf{h}_z$ which, in turn, determines the degree of persistence and mean-reversion of the model.

Real yields would then be obtained as

$$r_{n,t} = -n^{-1}q_{n,t} = \widetilde{A}_n + \widetilde{\mathbf{B}}'_{x,n}\mathbf{C}_{x,t} + \widetilde{\mathbf{B}}'_{z,n}\mathbf{C}_{z,t}, \tag{21}$$

with $\widetilde{A}_n = n^{-1}A_n, \widetilde{\mathbf{B}}_{x,n} = n^{-1}\mathbf{B}_{x,n}, \widetilde{\mathbf{B}}_{z,n} = n^{-1}\mathbf{B}_{z,n}$. Similarly, forward rates and holding period return are given by $f_{t,n} = A_{n+1} - A_n + (\mathbf{B}_{x,n+1} - \mathbf{B}_{x,n})'\mathbf{C}_{x,t} + (\mathbf{B}_{z,n+1} - \mathbf{B}_{z,n})'\mathbf{C}_{z,t}$ and $y_{t,n} = A_n - A_{n-1} + (\mathbf{B}'_{x,n}\mathbf{C}_{x,t-1} - \mathbf{B}'_{x,n-1}\mathbf{C}_{x,t}) + (\mathbf{B}'_{z,n}\mathbf{C}_{z,t-1} - \mathbf{B}'_{z,n-1}\mathbf{C}_{z,t})$ respectively.

4.2. Nominal term structure

Let us define the nominal price index at time t as Π_t with the price of nominal and real bonds satisfying $Q_{n,t} = P_{n,t}/\Pi_t = E_t\left[P_{n-1,t+1}e^{m_{t+1}}/\Pi_{t+1}\right]$ or, equivalently,

$$P_{n,t} = E_t\left[P_{n-1,t+1}\frac{\Pi_t}{\Pi_{t+1}}e^{m_{t+1}}\right] = E_t\left[P_{n-1,t+1}e^{-\pi_{t+1}}e^{m_{t+1}}\right], \tag{22}$$

where the (one-period) rate of inflation $\pi_t = \ln(\Pi_t/\Pi_{t-1})$ satisfies

$$\pi_t = E_{t-1}[\pi_t] + \varepsilon_{\pi,t} = (\mu_\pi + z_{t-1}) + \varepsilon_{\pi,t}, \tag{23}$$

where

$$\varepsilon_{\pi,t} \sim NID(0, \sigma_\pi^2) \text{ mutually independent from } \varepsilon_{x,t}, \varepsilon_{z,t}. \tag{24}$$

Following the same steps used for the real term structure one obtain the following.

Theorem 4.2. *For the pricing kernel (11), the market price of risk (13), the state variable dynamics (5)-(5) with innovations (12) and the real interest rate (14) and the inflation dynamics (23)-(24) the no-arbitrage zero coupon nominal prices $P_{n,t}$ satisfy*

$$-p_{n,t} = A_n^\$ + \mathbf{B}_{x,n}^{\$\prime}\mathbf{C}_{x,t} + \mathbf{B}_{z,n}^{\$\prime}\mathbf{C}_{z,t}$$

where $p_{n,t} = \ln P_{n,t}$ and the coefficients satisfy the Riccati recursions

$$\begin{aligned} A_n^\$ &= \mu_r + \mu_\pi + A_{n-1}^\$ \\ &\quad -\frac{1}{2}\sigma_x^2(\mathbf{B}_{x,n-1}^{\$\prime}\mathbf{h}_x + \beta_{zx}\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z)^2 - \frac{1}{2}\sigma_z^2(\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z)^2 - \frac{1}{2}\sigma_\pi^2 \\ &\quad -\lambda_{x,0}\sigma_x^2(\mathbf{B}_{x,n-1}^{\$\prime}\mathbf{h}_x + \beta_{zx}\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z) - \lambda_{z,0}\sigma_z^2(\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z) \end{aligned} \tag{25}$$

and

$$\mathbf{B}_{x,n}^\$ = \Big(1 - \lambda_{x,1}\sigma_x^2(\mathbf{B}_{x,n-1}^{\$\prime}\mathbf{h}_x + \beta_{zx}\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z) - \lambda_{x,2}\sigma_z^2(\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z)\Big)\mathbf{G} + \mathbf{F}'\mathbf{B}_{x,n-1}^\$ \tag{26}$$

$$\mathbf{B}_{z,n}^\$ = \Big(1 + \delta_z - \lambda_{z,1}\sigma_x^2(\mathbf{B}_{x,n-1}^{\$\prime}\mathbf{h}_x + \beta_{zx}\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z) - \lambda_{z,2}\sigma_z^2(\mathbf{B}_{z,n-1}^{\$\prime}\mathbf{h}_z)\Big)\mathbf{G} + \mathbf{F}'\mathbf{B}_{z,n-1}^\$ \tag{27}$$

Nominal yields with maturity n are then given by $r_{n,t}^\$ = \widetilde{A}_n^\$ + \widetilde{\mathbf{B}}_{x,n}^{\$\prime}\mathbf{C}_{x,t} + \widetilde{\mathbf{B}}_{z,n}^{\$\prime}\mathbf{C}_{z,t}$ where $\widetilde{A}_n^\$ = n^{-1}A_n^\$, \widetilde{\mathbf{B}}_{x,n}^\$ = n^{-1}\mathbf{B}_{x,n}^\$$ and $\widetilde{\mathbf{B}}_{z,n}^\$ = n^{-1}\mathbf{B}_{z,n}^\$$. Nominal forward rates and holding period returns are defined accordingly.

4.3. Persistence characterization

Solution of the model, both in terms of nominal and real term structure, was obtained without the need to specify whether the state variables are stationary or not. Indeed, only linearity of the state variables dynamics is necessary. This is due to the fact that conditional moments, rather than unconditional moments, are required to solve the model for any given maturity. We now discuss possible choices for $\mathbf{F}$ and, in particular, for $\mathbf{h}_x, \mathbf{h}_z$ which define both the degree of memory, and possibly of stationarity, of the model factors x_t, z_t through (5)-(5). These choices define the time series and cross-sectional properties of yields, forward rates and holding period returns implied by the term structure model.

Throughout the paper we will maintain the assumption that the matrix $\mathbf{F}$ satisfies (see Appendix D)

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & & \cdots \\ 0 & 0 & 1 & 0 & \\ \vdots & \vdots & & \ddots & \ddots \end{bmatrix}, \tag{28}$$

By Gaussianity the factors can be expressed as linear processes in the i.i.d. innovations $\varepsilon_{x,t}, \varepsilon_{z,t}$ of (12):

$$x_t = \sum_{i=0}^{\infty} \phi_{x,i}\varepsilon_{x,t-i}, \quad z_t = \sum_{i=0}^{\infty} \phi_{z,i}(\varepsilon_{z,t-i} + \beta_{zx}\varepsilon_{x,t-i}). \tag{29}$$

Stacking together the coefficients $\phi_{x,i}, \phi_{z,i}$ gives the infinite dimensional vector

$$\mathbf{h}_x = (1 \;\; \phi_{x,1} \;\; \phi_{x,2} \;\; \phi_{x,3}...)', \mathbf{h}_z = (1 \;\; \phi_{z,1} \;\; \phi_{z,2} \;\; \phi_{z,3}...)'. \tag{30}$$

Stationarity follows if

$$\sum_{i=0}^{\infty} \phi_{x,i}^2 < \infty, \;\; \sum_{i=0}^{\infty} \phi_{z,i}^2 < \infty. \tag{31}$$

As explained below, the stationarity condition (31) includes a wide range of possibilities in terms of the degree of persistence, in turn expressed by the rate at which the coefficients $\phi_{x,i}, \phi_{z,i}$ go to zero. We briefly summarise such possibilities including the case when the stationarity condition (31) is violated. Given (29), the factors x_t, z_t will be defined short memory if

$$\sum_{i=0}^{\infty} | \, \phi_{x,i} \, | < \infty, \;\; \sum_{i=0}^{\infty} | \, \phi_{z,i} \, | < \infty. \tag{32}$$

Alternatively, the factors are said to be long memory if

$$\sum_{i=0}^{j} | \, \phi_{x,i} \, | \to \infty, \;\; \sum_{i=0}^{j} | \, \phi_{z,i} \, | \to \infty \text{ as } j \to \infty. \tag{33}$$

Note that short memory (32) implies stationarity (31) since summability is stronger than square summability. However, long memory (33) does not necessarily implies stationarity. In this case we will distinguish between stationary long memory processes and non-stationary long memory processes. The latter case (non-stationary long memory) can be separated into the mean reverting case, namely when (31) is violated and yet

$$\phi_{x,i} \to 0, \;\; \phi_{z,i} \to 0 \text{ as } i \to \infty, \tag{34}$$

and the case when even mean-reversion (34) does not occur. A simple example of this last, extreme, circumstance is given by the basic model of Section 3 when the single factor x_t is a random walk, namely $\phi_{x,i} = 1$ for all i.

4.3.1. Short Memory

We now check that the simple model of Section 3 is nested within the general solution of Section 4. To achieve this, set $\mathbf{h}_z = \mathbf{0}$ since now only the latent factor x_t is required. Also, the simple model is based on a constant market price of risk, that is $\lambda_1 = \mathbf{0}$. Now the infinite dimensional vector (30) equals

$$\mathbf{h}_x = (1 \;\; \psi_x \;\; \psi_x^2 \;\; \psi_x^3 ...)', \tag{35}$$

where ψ_x is the autoregressive parameter in (4). By standard arguments model (4) can be re-written as

$$x_t = \sum_{i=0}^{\infty} \psi_x^i \varepsilon_{x,t-i}, \tag{36}$$

implying that, obviously, the AR(1) satisfies the linearity assumption (29) with coefficients $\phi_{x,i} = \psi_x^i$. When $\mid \psi_x \mid < 1$ then the short memory condition (32) is satisfied, and thus both the stationarity and the mean-reversion conditions apply. Instead, when $\psi_x = 1$ the AR(1) becomes a random walk and even (34) fails.

One just needs to find the scalar sequence A_n and infinite dimensional sequences $\mathbf{B}_{x,n}$, solution of the recurrence equations (18)-(19), and verify that indeed the basic affine model (5) is re-obtained. By (28) and $\lambda_1 = 0$ (note that since there is one factor only $\lambda_1 = \lambda_{x,1}$ and $\lambda_0 = \lambda_{x,0}$) recursion (19) becomes

$$\mathbf{B}_{x,n} = \mathbf{G} + \mathbf{F}'\mathbf{B}_{x,n-1}$$

with initial condition $\mathbf{B}_{x,0} = \mathbf{0}$ yielding

$$\mathbf{B}_{x,n} = (\underbrace{1....1}_{n\,terms}0...)' \quad \text{for every n } \geq 1. \tag{37}$$

This implies $\mathbf{B}'_{x,n}\mathbf{h}_x = 1 + \psi_x + ... + \psi_x^{n-1} = (1 - \psi_x^n)/(1 - \psi_x)$ for every n ≥ 1 which in turn gives $A_n = A_{n-1} + \mu_r - \lambda_{x,0}\sigma_x^2(\frac{1-\psi_x^n}{1-\psi_x}) - \frac{1}{2}\sigma_x^2(\frac{1-\psi_x^n}{1-\psi_x})^2$, which coincides exactly with (8). Notice that $\mathbf{C}_{x,t}$ can be expressed as

$$\mathbf{C}_{x,t} = (E_t(x_t), E_t(x_{t+1}), E_t(x_{t+2}), ...)'$$

where $E_t(x_{t+i}) = \sum_{j=i}^{\infty} \psi_x^j \varepsilon_{x,t+i-j}$ for all $i = 0, 1, ...$ (see Appendix D). In turn, this implies

$$\mathbf{B}'_{x,n}\mathbf{C}_{x,t} = \sum_{i=0}^{n-1} E_t(x_{t+i}) = \sum_{i=0}^{n-1} \left(\sum_{j=i}^{\infty} \psi_x^j \varepsilon_{x,t+i-j} \right) = \frac{1-\psi_x^n}{1-\psi_x} x_t$$

which coincides with $B_n x_t$ re-obtaining the solution of Section 3. This shows that the general solution (21) and the particular one (5) coincide.

4.3.2. Long Memory

Consider again the representation with the single factor x_t. A particularly convenient long memory parameterization, that nests both stationary ARMA as well as the non-stationary (and non-mean reverting) random walk is the ARFIMA model. In particular, the factor x_t follows a stationary ARFIMA$(1, d, 1)$ model (see Brockwell and Davis (1991), Definition 12.4.2) when

$$(1 - \psi_x L)(1 - L)^{d_x} x_t = (1 + \theta_x L)\varepsilon_{x,t}, \tag{38}$$

where the autoregressive and moving average coefficients ψ_x, θ_x satisfy the usual stationarity and invertibility conditions

$$| \psi_x |< 1, \ \ | \theta_x |< 1, \text{ with } \psi_x \neq \theta_x, \tag{39}$$

and d_x is a real number such that

$$-1/2 < d_x < 1/2. \tag{40}$$

When (39) and (40) hold, it can be shown (see Brockwell and Davis (1991), Theorem 12.4.2) that x_t admits the linear representation (29) with coefficients $\phi_{x,i} = \phi_{x,i}(\ \xi_x)$ satisfying

$$\sum_{i=0}^{\infty} \phi_{x,i} L^i = (1 + \theta_x L)(1 - \psi_x L)^{-1}(1 - L)^{-d_x}, \tag{41}$$

function of the 3×1 vector $\xi_x = (\psi_x, \theta_x, d_x)'$. To discuss the stationarity and memory properties of the factor x_t, we use the property

$$\phi_{x,i} \sim c\, i^{d_x - 1} \quad \text{as } i \to \infty, \tag{42}$$

which stems from (41) for any $d_x < 1$, for a constant c, function of the parameters ξ_x. Stationarity (31) then follows when (40) holds. Short memory (32) requires $d_x \leq 0$ and long memory $d_x > 0$. As a particular case of short memory, stationary ARMA is obtained for $d_x = 0$. Although stationarity implies mean-reversion, the opposite is not necessarily true since mean-reversion (34) simply requires $d_x < 1$. Finally, when $d_x = 1$ one obtains the non stationary ARIMA process, a special case of which is the random walk (when $\psi_x = \theta_x = 0$).

Alternative definitions of long memory when $0 < d_x < 1/2$, equivalent to (42) for linear stationary processes, are in terms of autocovariance function

$$cov(x_t, x_{t+u}) \sim c\, u^{2d_x - 1} \text{ as } u \to \infty, \tag{43}$$

and spectral density

$$s_x(\lambda) \sim c\, \lambda^{-2d_x} \text{ as } \lambda \to 0. \tag{44}$$

4.4. $\mathbb{P}$ and $\mathbb{Q}$ measure implications of long memory

We now provide a quasi-closed form characterization of the general solution for real bond prices as from Theorem 1. This permits to explore the implications of the long memory model both in terms of dynamic persistence of yields, forwards and returns as well of the cross-sectional behaviour of their volatility.

Our interest is on the characterization of the physical measure, namely the 'true' distribution, of observed bond prices and transformation of such as yields, forward rates and holding period returns, as can be obtained by an ideal historical observation of these quantities in the market. Assuming that the model is correctly specified, the physical measure will be, generally speaking, a function of both the $\mathbb{P}$ and the $\mathbb{Q}$ measure's parameters. By this we mean that observed (log) bond prices are function of the loadings coefficients, namely the A_n and the $\mathbf{B}_{x,n}, \mathbf{B}_{z,n}$, are evaluated under the $\mathbb{Q}$ measure, and of the state variables $\mathbf{C}_{x,t}, \mathbf{C}_{z,t}$, which are evaluated under the $\mathbb{P}$ measure[11].

The results below indicate a clear dichotomy, namely that the $\mathbb{P}$ measure's parameters determine the 'long-run' dynamic properties of the physical measure whereas the $\mathbb{Q}$ measure's parameters contribute to the 'long maturity' cross-sectional properties of the physical measure. We will refer to these results, with a somewhat abuse of notation, as holding 'under the $\mathbb{P}$' and 'under the $\mathbb{Q}$ measure' respectively. In other words, the dynamic persistence induced by the model does not depend on the form of the market prices of risk or, generally speaking, on the $\mathbb{Q}$ measure. Instead, the combination of the essentially affine specification of the market price of risk together with the long memory parameterization of the factors shape the volatility term structure for yields, forwards and returns. Precisely the same results apply to nominal bond characteristics.

To get the result, a key observation is that when the matrix $\mathbf{F}$ satisfies (28) (see Appendix D), which we assume for both short and long memory parameterizations, then the recursions (19) and (20) in the infinite-dimensional loadings $\mathbf{B}_{x,n}, \mathbf{B}_{z,n}$ can in fact be reduced into a recursion of a scalar sequence. In particular, by direct evaluation the real loadings

[11] We are not interested in deriving the loading coefficients A_n and the $\mathbf{B}_{x,n}, \mathbf{B}_{z,n}$ under the $\mathbb{P}$ measure nor the distribution of the state variables $\mathbf{C}_{x,t}, \mathbf{C}_{z,t}$ under the $\mathbb{Q}$ measure. For instance, the former permits straightforward evaluation of term premia.

will satisfy the recursion:

$$
\begin{aligned}
&\mathbf{B}_{x,n} = (b_{x,n}\, b_{x,n-1}\, \ldots\, b_{x,1}\, 0\, \ldots)' \,,\, \mathbf{B}_{z,n} = (b_{z,n}\, b_{z,n-1}\, \ldots\, b_{z,1}\, 0\, \ldots)' \text{ with} \\
&b_{x,1} = 1, \\
&b_{x,k} = 1 - \kappa_{x1}(\sum_{j=0}^{k-2} b_{x,k-j-1}\phi_{x,j} + \beta_{zx}\sum_{j=0}^{k-2} b_{z,k-j-1}\phi_{z,j}) - \kappa_{x2}(\sum_{j=0}^{k-2} b_{z,k-j-1}\phi_{z,j}), \quad k \geq 2, \\
&\text{ and} \\
&b_{z,1} = \delta_z, \\
&b_{z,k} = \delta_z - \kappa_{z1}(\sum_{j=0}^{k-2} b_{x,k-j-1}\phi_{x,j} + \beta_{zx}\sum_{j=0}^{k-2} b_{z,k-j-1}\phi_{z,j}) - \kappa_{z2}(\sum_{j=0}^{k-2} b_{z,k-j-1}\phi_{z,j}), \quad k \geq 2,
\end{aligned}
$$

where we set

$$\kappa_{x1} \equiv \lambda_{x,1}\sigma_x^2, \kappa_{x2} \equiv \lambda_{x,2}\sigma_z^2, \kappa_{z1} \equiv \lambda_{z,1}\sigma_x^2, \kappa_{z2} \equiv \lambda_{z,2}\sigma_z^2. \tag{45}$$

Similarly, nominal loadings can be expressed as

$$\mathbf{B}^{\$}_{x,n} = \left(b^{\$}_{x,n}\, b^{\$}_{x,n-1}\, \ldots\, b^{\$}_{x,1}\, 0\, \ldots\right)' ,\, \mathbf{B}^{\$}_{z,n} = \left(b^{\$}_{z,n}\, b^{\$}_{z,n-1}\, \ldots\, b^{\$}_{z,1}\, 0\, \ldots\right)'$$

where the $b^{\$}_{x,i}$ and $b^{\$}_{z,i}$ satisfy a recursion analogue to the ones above based on (26) and (27).

Useful insights can be obtained by looking at the simpler one-factor case, $b_{z,k} = 0$. By recursive substitution one gets

$$
\begin{aligned}
b_{x,1} &= 1, \\
b_{x,2} &= 1 - \kappa_{x,1}\phi_{x,0}, \\
b_{x,3} &= 1 - \kappa_{x,1}(\phi_{x,0} + \phi_{x,1}) + \kappa_{x,1}^2\phi_{x,0}^2, \\
b_{x,4} &= 1 - \kappa_{x,1}(\phi_{x,0} + \phi_{x,1} + \phi_{x,2}) + \kappa_{x,1}^2(\phi_{x,0}^2 + 2\phi_{x,0}\phi_{x,1}) - \kappa_{x,1}^3\phi_{x,0}^3, \\
b_{x,5} &= 1 - \kappa_{x,1}(\phi_{x,0} + \phi_{x,1} + \phi_{x,2} + \phi_{x,3}) + \kappa_{x,1}^2(\phi_{x,0}^2 + \phi_{x,1}^2 + 2\phi_{x,0}\phi_{x,1} + 2\phi_{x,0}\phi_{x,2}) - \kappa_{x,1}^3(\phi_{x,0}^3 + 3\phi_{x,0}^2\phi_{x,1}) + \kappa_{x,1}^4\phi_{x,0}^4, \\
b_{x,6} &= \ldots.
\end{aligned} \tag{46}
$$

We need to distinguish between evaluation of the $b_{x,k}$ under the $\mathbb{P}$ and the $\mathbb{Q}$ measures. The first case is obtained when $\kappa_{x,1} = 0$, which in turn follows when $\lambda_1 = 0$ in (13), namely for a constant market price. This does not, of course, imply that bond prices are evaluated under the $\mathbb{P}$ measure[12]. In this case $b_{x,k} = 1$ for every k and one obtains a (quasi) closed-form solution to bond prices, as formalized below. When instead $\kappa_{x,1} \neq 0$ then the $b_{x,k}$, now evaluated under the $\mathbb{Q}$ measure, have a more cumbersome expression. Important implications can nevertheless be derived: by looking at the recursion above, it is evident that the behaviour of the $b_{x,k}$ as k increases, depends on the interaction between powers of the slow (hyperbolic) increase of the partial sum terms $\sum_{j=0}^{k}\phi_{x,j}$ and the fast (exponential) decay of powers of the term $\kappa_{x,1}$. For instance, whereas the latter term can dominate for small and

[12]By Gaussianity of the model, the distribution of bond prices only depend on the first two moments. The mean is evaluated under the $\mathbb{P}$ measure when $\lambda_0 = 0$ whereas the variance requires $\lambda_1 = 0$. Therefore both parameters are required to be zero, implying null market prices of risk, for observed bond prices to be expressed under the $\mathbb{P}$ measure.

intermediate maturities, the former can dominate for long maturities since $\sum_{j=0}^{k} \phi_{x,j} \sim ck^{d_x}$ as k increases when (42) holds. See Lemma 2 in Appendix E. This gives rise to a remarkable degree of flexibility of our long memory affine model in fitting the volatility term structures of yields, forward rates and returns.

With the $b_{x,k}$ and the $b_{z,k}$ at hand, the general quasi-closed solution of the model under the $\mathbb{Q}$ measure follows. In fact, setting

$$\mathbf{h}_x = (\phi_{x,0}\, \phi_{x,1}\, \phi_{x,2} ...)', \quad \mathbf{h}_x = (\phi_{z,0}\, \phi_{z,1}\, \phi_{z,2} ...)', \tag{47}$$

where $\phi_{x,i}, \phi_{z,i}$ are the linear representation coefficients of the factors x_t, z_t in (29), one gets $\mathbf{B}'_{x,n}\mathbf{h}_x = \sum_{i=0}^{n-1} b_{x,n-i}\phi_{x,i} = \Phi_{x,n,0}$, $\mathbf{B}'_{z,n}\mathbf{h}_z = \delta_z \sum_{i=0}^{n-1} b_{z,n-i}\phi_{z,i} = \delta_z \Phi_{z,n,0}$, where

$$\Phi_{x,n,j} \equiv \sum_{i=0}^{n-1} b_{x,n-i}\phi_{x,i+j}, \quad \Phi_{z,n,j} \equiv \sum_{i=0}^{n-1} b_{z,n-i}\phi_{z,i+j} \text{ for every } j \geq 0. \tag{48}$$

Plugging $\Phi_{x,n,0}$ and $\Phi_{z,n,0}$ into (18) provides the A_n, the first moment of the (log) bond prices. Next, since $E_t(x_{t+i}) = \sum_{j=0}^{\infty} \phi_{x,j+i}\varepsilon_{x,t-j}$, $E_t(z_{t+i}) = \sum_{j=0}^{\infty} \phi_{z,j+i}(\varepsilon_{z,t-j} + \beta_{zx}\varepsilon_{x,t-j})$ for all $i = 0, 1, ...$ then

$$\mathbf{B}'_{x,n}\mathbf{C}_{x,t} = \sum_{i=0}^{n-1} b_{x,n-i}E_t(x_{t+i}) = \sum_{i=0}^{n-1} b_{x,n-i}\left(\sum_{j=0}^{\infty} \phi_{x,j+i}\varepsilon_{x,t-j}\right) = \sum_{j=0}^{\infty} \Phi_{x,n,j}\varepsilon_{x,t-j},$$

$$\mathbf{B}'_{z,n}\mathbf{C}_{z,t} = \delta_z \sum_{i=0}^{n-1} b_{z,n-i}E_t(z_{t+i}) = \delta_z \sum_{i=0}^{n-1} b_{z,n-i}\left(\sum_{j=0}^{\infty} \phi_{z,j+i}(\varepsilon_{z,t-j} + \beta_{zx}\varepsilon_{x,t-j})\right) = \delta_z \sum_{j=0}^{\infty} \Phi_{z,n,j}(\varepsilon_{z,t-j} + \beta_{zx}\varepsilon_{x,t-j}),$$

the variance of which provide the second moment of (log) bond prices. Combining terms, the term structure of real yields (21) can be expressed as:

$$r_{n,t} = n^{-1}A_n + \sum_{j=0}^{\infty}(n^{-1}\Phi_{x,n,j} + \beta_{zx}n^{-1}\Phi_{z,n,j})\varepsilon_{x,t-j} + \delta_z \sum_{j=0}^{\infty}(n^{-1}\Phi_{z,n,j})\varepsilon_{z,t-j}, \tag{49}$$

since the n-period (log) bond price of real bonds is given by $q_{n,t} = -nr_{n,t} = -A_n - \sum_{j=0}^{\infty}(\Phi_{x,n,j} + \beta_{zx}\Phi_{z,n,j})\varepsilon_{x,t-j} - \delta_z \sum_{j=0}^{\infty} \Phi_{z,n,j}\varepsilon_{z,t-j}$. Simple calculations give forward rates

$$f_{n,t} = q_{n,t} - q_{n+1,t} = A_{n+1} - A_n + \sum_{j=0}^{\infty}(\Delta^f_{x,n,j} + \beta_{zx}\Delta^f_{z,n,j})\varepsilon_{x,t-j} + \delta_z \sum_{j=0}^{\infty} \Delta^f_{z,n,j}\varepsilon_{z,t-j}, \tag{50}$$

and holding one-period returns

$$y_{n,t} = q_{n-1,t} - q_{n,t-1} = A_n - A_{n-1} + \sum_{j=0}^{\infty}(\Delta^y_{x,n,j} + \beta_{zx}\Delta^y_{z,n,j})\varepsilon_{x,t-j} + \delta_z \sum_{j=0}^{\infty} \Delta^y_{z,n,j}\varepsilon_{z,t-j}, \tag{51}$$

setting

$$\begin{aligned}
\Delta^f_{x,n,j} &\equiv \Phi_{x,n+1,j} - \Phi_{x,n,j}, j \geq 0, \\
\Delta^y_{x,n,0} &\equiv -\Phi_{x,n-1,0}, \;\; \Delta^y_{x,n,j} \equiv \Phi_{x,n,j-1} - \Phi_{x,n-1,j}, j \geq 1,
\end{aligned}$$

with $\Delta^f_{z,n,j}$ and $\Delta^y_{z,n,j}$ defined accordingly.

Nominal yields $r^{\$}_{n,t}$, nominal forwards $f^{\$}_{n,t}$ and nominal returns $y^{\$}_{n,t}$ also satisfy (49), (50) and (51) respectively but one needs to replace the $\Phi_{x,n,j}, \Phi_{z,n,j}$ with

$$\Phi^{\$}_{x,n,j} \equiv \sum_{i=0}^{n-1} b^{\$}_{x,n-i}\phi_{x,i+j}, \;\; \Phi^{\$}_{z,n,j} \equiv \sum_{i=0}^{n-1} b^{\$}_{z,n-i}\phi_{z,i+j} \text{ for every } j \geq 0.$$

Consider now the restricted case of constant market prices of risk, that is $\lambda_1 = 0$. As indicated above, this implies that the recursions (19)-(20) for $\mathbf{B}_{x,n}, \mathbf{B}_{z,n}$ are evaluated under the $\mathbb{P}$ measure and are, in fact, parameters-free. Under this circumstance

$$\mathbf{B}_{x,n} = (\underbrace{1....1}_{n\,terms}0...)', \mathbf{B}_{z,n} = \delta_z(\underbrace{1....1}_{n\,terms}0...)' \quad \text{for every n } \geq 1.$$

Model-implied yields, forward rates and returns will still satisfy (49), (50) and (51), respectively, but now (48) simplifies to

$$\Phi_{x,n,j} \equiv \sum_{i=0}^{n-1} \phi_{x,i+j}, \;\; \Phi_{z,n,j} \equiv \sum_{i=0}^{n-1} \phi_{z,i+j} \text{ for every } j \geq 0. \tag{52}$$

implying $\Delta^f_{x,n,j} = \phi_{x,i+j}, \Delta^f_{z,n,j} = \phi_{z,i+j}$ $j \geq 0$ and $\Delta^y_{x,n,j} = \phi_{x,j}, \Delta^y_{z,n,j} = \phi_{z,j}$, $j \geq 1, \Delta^y_{x,n,0} = -\sum_{i=0}^{n-1} \phi_{x,i}, \Delta^y_{z,n,0} = -\sum_{i=0}^{n-1} \phi_{z,i}$.

Irrespective of the assumptions made on the market prices of risk, in general yields $r_{n,t}$, forwards $f_{n,t}$ and returns $y_{n,t}$ have a linear process representation in i.i.d. innovations by (49)-(50)-(51). Hence derivation of their theoretical spectral density and variances follows easily.

Theorem 4.3. *Under the assumptions of Theorem 4.1 yields have spectral density*

$$s_{r_n}(\lambda) = \frac{\sigma_x^2}{2\pi}\Big|\sum_{j=0}^{\infty}(n^{-1}\Phi_{x,n,j} + \beta_{zx}n^{-1}\Phi_{z,n,j})e^{\imath\lambda j}\Big|^2 + \frac{\sigma_z^2}{2\pi}\Big|\sum_{j=0}^{\infty}(n^{-1}\Phi_{z,n,j})e^{\imath\lambda j}\Big|^2, \tag{53}$$

conditional variance

$$\text{var}_{t-1}(r_{t,n}) = \sigma_x^2(n^{-1}\Phi_{x,n,0} + \beta_{zx}n^{-1}\Phi_{z,n,0})^2 + \sigma_z^2(n^{-1}\Phi_{z,n,0})^2, \tag{54}$$

and unconditional variance

$$\text{var}(r_{t,n}) = \sigma_x^2\sum_{j=0}^{\infty}(n^{-1}\Phi_{x,n,j} + \beta_{zx}n^{-1}\Phi_{z,n,j})^2 + \sigma_z^2\sum_{j=0}^{\infty}(n^{-1}\Phi_{z,n,j})^2. \tag{55}$$

The same formulae applies to forward rates and returns by substituting $n^{-1}\Phi_{x,n,j}, n^{-1}\Phi_{z,n,j}$ *with* $\Delta^f_{x,n,j}, \Delta^f_{z,n,j}$ *and* $\Delta^y_{x,n,j}, \Delta^y_{z,n,j}$ *respectively.*

Moreover, the same formulae apply to nominal yields, forwards and returns by replacing the $\Phi_{x,n,j}, \Phi_{z,n,j}$ *with the* $\Phi^{\$}_{x,n,j}, \Phi^{\$}_{z,n,j}$.

Noticeably, these formulae are extremely general since derived for generic specifications of the coefficients $\phi_{x,j}$ and $\phi_{z,j}$.

We can now fully characterise the persistence of yields, forward rates and returns when long memory is allowed for. Stationary ARFIMA x_t and z_t are included as a special, parametric, case.

Theorem 4.4. *Assume*

$$\phi_{x,j} \sim cj^{d_x-1}, \;\; \phi_{z,j} \sim cj^{d_z-1} \quad \textit{as } j \to \infty \textit{ with } 0 < d_x, d_z < 1/2 \tag{56}$$

and

$$|\phi_{x,j+1} - \phi_{x,j}| \leq cj^{-1}\phi_{x,j}, \;\; |\phi_{z,j+1} - \phi_{z,j}| \leq cj^{-1}\phi_{z,j}, \; \textit{for any } j \geq J, \; \textit{some finite } J. \tag{57}$$

Under either the $\mathbb{P}$ *and* $\mathbb{Q}$ *measure, the spectral densities of yields* $r_{t,n}$*, forward rates* $f_{t,n}$ *and returns* $y_{n,t}$ *satisfy:*

$$s_{r_n}(\lambda) \sim c\lambda^{\min(-2d_x,-2d_z)}, \; s_{f_n}(\lambda) \sim c\lambda^{\min(-2d_x,-2d_z)}, \; s_{y_n}(\lambda) \sim c\lambda^{\min(-2d_x,-2d_z)} \;\textit{as } \lambda \to 0^+.$$

Precisely the same formulae apply to nominal yields, forwards and returns.

Alternatively, taking logarithm, it follows that $log\, s_{r_n}(\lambda) \sim \min(-2d_x, -2d_z)\log\lambda, log\, s_{f_n}(\lambda) \sim \min(-2d_x, -2d_z)\log\lambda$ and $log\, s_{y_n}(\lambda) \sim \min(-2d_x, -2d_z)\log\lambda$ for $\lambda \to 0^+$. This shows that the model spectral density are all negatively sloped near the zero frequency, the more the larger the long memory parameters d_x, d_z. We are matching[13] *Stylized Fact* 1. Figure 5 reproduces the log-periodogram of the data of Section 2 superimposing $c - 2\,d\log(\lambda)$ for d equal to 0.2 (blue line), 0.3 (green line) and 0.4 (red line). The degree of memory will not depend on n although away from zero frequency the spectral densities of $r_{n,t},\; f_{n,t}$ and $y_{n,t}$ will all be affected as n varies. Alternatively, the usual characterization of long memory in terms of long lags behaviour can also be obtained (cf (43)).

[Insert Figure 5 near here]

The degree of memory or, alternatively, of nonstationarity implied by the physical measure for yields, forwards and rates does not depend on the form of the $\mathbb{Q}$ measure since the parameters, λ_0 and λ_1 , governing the market price of risk do not affect these aspects of the dynamic properties of the model although, of course, contributing to the physical measure. This result does not depend on the long memory assumption but holds true for any specification of the essentialy affine model. Instead, the cross-sectional properties of the model-implied physical measure differ markedly depending on whether the $\mathbb{P}$ or the $\mathbb{Q}$

[13]It is easy to see that Theorem 4.4 equally applies to nominal yields, forwards and returns.

measure is considered. The next theorem illustrates the long maturity behaviour of both the conditional and unconditional variance for yields, forwards and rates under the $\mathbb{P}$ measure. The corresponding $\mathbb{Q}$-measure term structure properties are presented subsequently.

Theorem 4.5. *Assume (56) and (57). Under the $\mathbb{P}$ measure, as $n \to \infty$:*
(i) the conditional variances of yields $r_{t,n}$, forward rates $f_{t,n}$ and returns $y_{t,n}$ satisfy

$$\mathrm{var}_{t-1}(r_{t,n}) = O(n^{2d_x-2}+n^{2d_z-2}),\ \mathrm{var}_{t-1}(f_{t,n}) = O(n^{2d_x-2}+n^{2d_z-2}),\ \mathrm{var}_{t-1}(y_{t,n}) = O(n^{2d_x}+n^{2d_z});$$

(ii) the unconditional variances of yields $r_{t,n}$, forward rates $f_{t,n}$ and returns $y_{t,n}$ satisfy

$$\mathrm{var}(r_{t,n}) = O(n^{2d_x-1} + n^{2d_z-1}),\ \mathrm{var}(f_{t,n}) = O(n^{2d_x-1} + n^{2d_z-1}),\ \mathrm{var}(y_{t,n}) = O(n^{2d_x} + n^{2d_z})..$$

Under the $\mathbb{P}$ measure and long memory, the term structure of volatility for yields and forwards declines to zero at the same rate when mean-reversion holds[14], namely for $d_x, d_x < 1$.

Under the same conditions, the term structure diverges, with maturity, for returns as long as long memory is manifested (either d_x or d_z greater than zero). These features are obtained in terms of both conditional and unconditional variances although different rates are obtained in these two cases. Comparing this with the short memory case (9) where the volatility term structure for yields and forwards also declines with maturity under stationarity, long memory implies a much slower rate of convergence towards zero. For returns, short memory ruled out divergence altogether (under stationarity). When $d_x = d_z = 1$ the unit root results of (10) are re-obtained as a special case of long memory. In our long memory case the speed of convergence (towards zero) or divergence is smoothly modulated by the magnitude of d_x and d_z, in contrast to the discontinuous behaviour of the simple model of Section 3. We now present the $\mathbb{Q}$ measure results.

Theorem 4.6. *Assume (56) and (57). Under the $\mathbb{Q}$ measure, when*

$$b_{x,j} \sim -\kappa_{x,1}(\sum_{i=0}^{j} \phi_{x,i}),\ \ b_{z,j} \sim -\kappa_{z,1}(\sum_{i=0}^{j} \phi_{z,i}) \text{ as } j \to \infty, \tag{58}$$

(i) the conditional variance of yields $r_{t,n}$, forward rates $f_{t,n}$ and returns $y_{t,n}$ as $n \to \infty$ satisfy:

$$\mathrm{var}_{t-1}(r_{t,n}) = O(n^{4d_x-2}+n^{4d_z-2}),\ \mathrm{var}_{t-1}(f_{t,n}) = O(n^{4d_x-2}+n^{4d_z-2}),\ \mathrm{var}_{t-1}(y_{t,n}) = O(n^{4d_x}+n^{4d_z});$$

(ii) the unconditional variances of yields $r_{t,n}$, forward rates $f_{t,n}$ and returns $y_{t,n}$ as $n \to \infty$ satisfy:

$$\mathrm{var}(r_{t,n}) = O(n^{2d_x} + n^{2d_z}),\ \mathrm{var}(f_{t,n}) = O(n^{2d_x} + n^{2d_z}),\ \mathrm{var}(y_{t,n}) = O(n^{2d_x+2} + n^{2d_z+2}).$$

Precisely the same formulae apply to nominal yields, forwards and returns by replacing the $b_{x,j}, b_{z,j}$ with the $b^{\$}_{x,j}, b^{\$}_{z,j}$.

[14] Comte and Renault (1996) derive (see their Proposition 12) the analogue result to Theorem 4.5-(i) for $r_{n,t}$ in a continuous time setting.

The picture changes markedly under the $\mathbb{Q}$ measure. Under long memory, the unconditional variance term structures are now all turning positively sloped for large n, more prominently so for returns, although could initially decline for short and intermediate maturities depending on the other parameters' value. In term of conditional variances, the term structures tend to be negatively sloped when stationarity ($0 < d_x, d_z < 1/2$) holds but diverging otherwise, including the mean-reversion case ($1/2 < d_x, d_z < 1$). Therefore, under the $\mathbb{Q}$ measure the long memory model achieve a great deal of flexibility for the volatility term structure of yields, forwards and returns. Those closed-form results rely on condition (58) which can be easily verified numerically. In turn, the latter appears to require a sufficiently small κ_x, κ_z by (46).

Summarizing, the long memory affine model is able to generate predictions more adequately aligned with the characteristics observed of the bond data, as spelled out in Stylized Facts 1 and 2.

5. Inducing long memory in affine term structure models

To allow for the possibility that long memory arises within the affine class of models, it is useful to consider the conventional decomposition of nominal yields on zero-coupon bonds into real yields, expected inflation and inflation risk premium:

$$r^{\$}_{n,t} = c_n + r_{n,t} + \frac{1}{n}\mathrm{E}_t \ln(\frac{\Pi_{t+n}}{\Pi_t}) + IP_{n,t}, \tag{59}$$

where $IP_{n,t}$ denotes the inflation risk premium and c_n is the Jensen's inequality term, constant since the model assumes conditional homoskedasticity. We consider two different sources of long memory, which in turn can be thought of as related to the expected inflation term $n^{-1}\mathrm{E}_t \ln(\Pi_{t+n}/\Pi_t)$ and to the real interest rate term $r_{n,t}$, respectively. Both channels are able to induce the form of long memory observed empirically in the data.

5.1. *Inflation channel*

Recent research (see Altissimo et al (2009))suggests that the CPI inflation in large, mature, economies is very likely to exhibit long memory, being less persistent than a unit-root process but at the same time more persistent than a stationary ARMA. Although this result is illustrated for euro area, we argue that a similar result will apply to US inflation. In particular, Altissimo et al (2009) document that sub-sectorial inflation rates for the euro area, comprised by $J = 404$ sectors, are well described by an ARMA structure with a single common factor, a simple case of which is the autoregressive structure

$$\pi_{i,t} = \mu_{\pi i} + \psi_{\pi,i}\pi_{i,t-1} + \gamma_i u_t + \epsilon_{i,t}, \;\; i = 1, ..., J,$$

where $\pi_{i,t}$ is the ith sector inflation rate, u_t is the i.i.d. common dynamic shock and $\epsilon_{i,t}$ is the i.i.d. idiosyncratic component, assumed independent from u_t at any leads and lags[15]. The autoregressive coefficients $\psi_{\pi,i}$ are assumed i.i.d., in particular random with a common distribution over the stationary region ensuring that $-1 < \psi_{\pi,i} < 1$ for any sub-sector i. Although the $\epsilon_{i,t}$ appear to dominate the variance of the individual $\pi_{i,t}$, the common factor appear to explain a large part of the aggregate CPI inflation dynamics. In fact $var(J^{-1}\sum_{i=1}^{J}\epsilon_{i,t})$ is estimated to be much smaller than, about one fourth of the average variance of the idiosyncratic components[16] $\epsilon_{i,t}$. At the same time, by well-known aggregation results (see Granger (1980) and the generalisations by Zaffaroni (2004)) under mild conditions

$$N^{-1}\sum_{i=1}^{N}\pi_{i,t} \rightarrow_2 \mu_\pi + \sum_{k=0}^{\infty}\phi_{\pi,k}u_{t-k}, \quad \text{as } N \rightarrow \infty, \tag{60}$$

where μ_π and $\phi_{\psi,k}$, $k = 0, 1, ...$ are the limit (cross-sectional) averages of the $\mu_{\pi,i}$ and $\psi_{\pi,i}^k$, $k = 0, 1, ...$ respectively, and $\rightarrow_2$ denotes convergence in mean square. The crucial result here is that under some weak conditions, in particular regarding the behaviour of the (cross-sectional) distribution of the autoregressive roots $\psi_{\pi,i}$ near unity (see Figure 3 and Table 3 in Altissimo et al (2009)), (60) occurs and the estimated impulse response of the common shock u_t to CPI inflation satisfies $\phi_{\pi,k} \sim c\, k^{d_\pi - 1}$ as $k \rightarrow \infty$, which, recalling (42), is coherent with π_t exhibiting long memory with memory parameter d_π:

$$cov(\pi_t, \pi_{t+k}) \sim c\, k^{2d_\pi - 1} \text{ as } k \rightarrow \infty. \tag{61}$$

Note that the expected inflation term in (59) consists of an average of n terms, namely $n^{-1}\mathrm{E}_t \ln(\frac{\Pi_{t+n}}{\Pi_t}) = n^{-1}E_t(\pi_{t+1} + ...\pi_{t+n})$ where $\pi_t = \ln(\Pi_t/\Pi_{t-1})$ is the one-period inflation based on the CPI index Π_t. This average turns out to have the same memory properties, for any given n, as the individual components $E_t\pi_{t+j}$, $j = 1, ..., n$ (see Chambers (1998)).

5.2. Real rate channel

Consider a multi-factor version of the Vasicek-type model of Section 3 with J independent latent factors, each following a first order stationary autoregressive process:

$$x_{j,t} = \psi_{x,j}x_{j,t-1} + \gamma_j u_t + \epsilon_{j,t}, \;\; j = 1, ..., J,$$

where $u_t \sim NID(0, 1)$, $\epsilon_{j,t} \sim NID(0, \sigma_j^2)$ mutually independent one of another and $-1 < \psi_{x,j} < 1$ for all $j = 1, ..., J$. Under suitable assumptions on the pricing kernel akin to (3),

[15] In Altissimo, Mojon and Zaffaroni (2009) $\epsilon_{i,t}$ are modelled as ARMA, mutually independent from u_s for any t, s but the same aggregation result carries through.

[16] Note that CPI inflation π_t is constructed as a weighted average of the sub-sectoral inflation rates $\pi_{i,t}$ but turns out to be strongly positively correlated with the equally weighted average $J^{-1}\sum_{i=1}^{J}\pi_{i,t}$ based on $J = 404$ sectors in France, Germany, Italy only, with a correlation above 0.8.

real bond yields satisfy the affine relationship

$$r_{t,n} = a_n + \sum_{j=1}^{J} n^{-1} B_{j,n} x_{j,t}, \tag{62}$$

where, in particular, the n-varying coefficients $B_{j,n}$, $j = 1, .., J$ satisfy $B_{j,n} = (1 - \psi_{x,j}^n)/(1 - \psi_{x,j})$. Since Litterman and Sheinkman (1991), the large majority of estimated affine models considers up to three factors[17], that is $1 \leq J \leq 3$. This approach is essentially dictated by statistical consideration since the number of parameters to be estimated increases rapidly with J. On the other hand, a small J induces spurious cross-correlation between estimated yields at different maturities, not observed in the data (Dai and Singleton (2000)), and it is often advocated as causing a modest out-of-sample performance (Duffee (2002)). We argue that this curse of dimensionality can be mitigated, by allowing for a suitable form of heterogeneity of the AR(1) coefficients $\psi_{x,j}$ and then applying the aggregation results of Granger (1980) as J increases to infinity. In particular, as illustrated in Appendix A, letting $J \to \infty$ leads to a semiparametric specification of an affine term structure model with long memory yields $r_{n,t}$:

$$\text{cov}(r_{n,t}, r_{n,t+k}) \sim c\, k^{2d-1} \text{ as } k \to \infty. \tag{63}$$

with memory parameter d satisfying $0 < d < 1/2$. This semiparametric specification is characterised by an infinite number of coefficients, akin to the $\phi_{x,j}$ and $\phi_{z,j}$ of (29), unrestricted except for the long memory property (42) (see Appendix A). A natural parameterization is then represented by the ARFIMA model with coefficients (41), so that estimation and inference on a finite, small, number of parameters can be carried out.

In conclusion, both (61) and (63) imply long memory in the nominal yields $r_{n,t}^{\$}$ through (59). Moreover, the inflation channel suggests that inflation data should be certainly included when estimating the long memory affine models since these would help pin down the dynamic persistence of the data.

6. Long memory affine term structure models: empirical example

An empirical application is presented, to describe the potential of the model in capturing the dynamic persistence of the data and the shape of their volatility term structures. Given the illustrative scope of the exercise, the simplest possible, two factor, specification is adopted that permits to disentangle the term structure of real yields and inflation expectation.

[17] Among the few exceptions, the multifrequency affine model of Calvet, Fisher and Wu (2010) where, by means of an ingenious representation, the number of parameters does not increase with J. An empirical application with $J = 15$ is presented.

6.1. Estimation results

This section uses the monthly data on nominal yields and inflation data of Section 2. To motivate further the long memory parameterization of our model, Table 3 reports the long memory parameter estimates obtained by fitting an ARFIMA$(1, d, 1)$ model to yields and inflation. It turns out that the memory parameter for yields and inflation are positive and significant, well into the stationary region (40).

[Insert Table 3 near here]

We estimate the model by means of the approximate maximum likelihood estimator based on the Kalman recursions[18]. See Appendix D for details. The data sample goes from November 1985 to December 2011. The model is cast in state space with measurement equations

$$\begin{bmatrix} r^{\$}_{n_1,t} \\ r^{\$}_{n_2,t} \\ \vdots \\ r^{\$}_{n_k,t} \\ \pi_t \end{bmatrix} = \begin{bmatrix} \widetilde{A}^{\$}_{n_1} \\ \widetilde{A}^{\$}_{n_2} \\ \vdots \\ \widetilde{A}^{\$}_{n_k} \\ \mu_\pi \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{x,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{x,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{x,n_k} \\ 0 \end{bmatrix} \mathbf{C}_{x,t-1} + \begin{bmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{z,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{z,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{z,n_k} \\ \mathbf{G}' \end{bmatrix} \mathbf{C}_{z,t-1} + \begin{bmatrix} \epsilon_{n_1,t} \\ \epsilon_{n_2,t} \\ \vdots \\ \epsilon_{n_k,t} \\ \varepsilon_{\pi,t} \end{bmatrix},$$

with transition equations (5)-(5) and the innovations to yields satisfy

$$\epsilon_{n,t} \;=\; (\widetilde{\mathbf{B}}^{\$\prime}_{x,n}\mathbf{h}_x + \beta_{zx}\widetilde{\mathbf{B}}^{\$\prime}_{z,n}\mathbf{h}_z)\varepsilon_{x,t} + \widetilde{\mathbf{B}}^{\$\prime}_{z,n}\mathbf{h}_z\varepsilon_{z,t} + \nu_{n,t}, \tag{64}$$

where we refer to Theorem 4.2 for the definition of $\widetilde{A}^{\$}_n, \widetilde{\mathbf{B}}^{\$}_{x,n}, \widetilde{\mathbf{B}}^{\$}_{z,n}$. Here the $\nu_{n,t} \sim NID(0, \sigma^2_n)$ are measurement errors introduced to enhance the flexibility of the model. If the model fits the data well at a given maturity n, one expects the estimate of σ^2_n to be small. The covariance matrix of the measurement equations innovations $(\epsilon_{n_1,t}, ..., \epsilon_{n_k,t}, \epsilon_{\pi,t})$ will be also a function of the $\mathbf{B}_{x,n}, \mathbf{B}_{x,n}, \mathbf{h}_x, \mathbf{h}_x$, a feature which must be taken into account during the optimisation. The long memory feature of the model is driven by $\mathbf{h}_x = (1\, \phi_{x,1}\, \phi_{x,2}...)'$ and $\mathbf{h}_z = (1\, \phi_{z,1}\, \phi_{x,2}...)'$ where $\phi_{x,j} = \phi_{x,j}(\xi_x)$ and $\phi_{z,j}(\xi_z)$ are the linear representation coefficients of the $ARFIMA(1, d, 1)$ factors x_t, z_t (cf. (38)) with parameters $\xi_x = (\psi_x, \theta_x, d_x), \xi_z = (\psi_z, \theta_z, d_z)'$, respectively. Recalling that the factors and the (infinite-dimensional) state variables are related by $x_t = \mathbf{G}'\mathbf{C}_{x,t},\;\; z_t = \mathbf{G}'\mathbf{C}_{z,t}$, then the last of the measurement equations sets z_t to be the inflation factor, namely expected inflation. As a consequence, we interpret x_t as the real factor.

[Insert Table 4 near here]

[18]Using ordinary least squares at the first-stage for full maximum likelihood estimation, shown by Joslin et al (2011) and Hamilton and Wu (2012) to be computationally efficient when the state variables follow an autoregressive process, is not applicable here since we consider latent factors as well as long memory. This two-stage approach is ruled out here even when observed factors are considered. In fact, although the long memory parameterization implies an (infinite order) autoregressive structure, the corresponding autoregressive coefficients are not unconstrained but satisfy a condition like (42) with the exponent $-(d_x+1)$ replacing $d_x - 1$.

Table 4 presents the estimates of the model parameters when long memory is allowed for both x_t and z_t. Standard errors, obtained by numerical evaluation of the Hessian matrix, are reported in small font. Both the real factor x_t and the expected inflation factor z_t display a significant positive long memory parameter well in the middle of the stationary region[19] . The AR and MA parameters, driving the short run dynamics, are also large and significant, although significantly below the unit root level. The real factor x_t appears more volatile than z_t in terms of the one-step ahead conditional variance since $\sigma_x = 0.55\%$ while this equals 0.030% for z_t, the latter obtained as the square root of $\sigma_z^2 + \beta_{zx}^2 \sigma_x^2$. The non-neutrality parameter δ_z is negative and significant. Regarding the mean parameters, the unconditional mean of the one-period real rate μ_r equals 0.59% whereas the mean of realised inflation μ_π equals 2.80%, the latter being set equal to the sample mean of observed inflation π_t. The last two lines of Table 4 report the estimated variances of the idiosyncratic errors. The model appears to fit well the yield curve especially for maturities between 3-year and 20-year. The fit deteriorates at the 30-year and especially for very short maturities such as at 1-month.

Regarding the estimates of the price of risk parameters, both the intercepts vector λ_0 and the slopes matrix λ_1 are significant. This suggests that the data reject the statement by which the $\mathbb{P}$ and the $\mathbb{Q}$ measures coincide. By Gaussianity of the model, this implies that both the $\mathbb{Q}$ measure mean (through λ_0) and variance (through λ_1) for yields, forwards and returns differ from the corresponding moments under the $\mathbb{P}$ measure. As illustrated below, the combination of non zero λ_1 parameters together with the long memory feature of the model drastically increases the goodness of fit of the model in terms of volatility term structures.

[Insert Table 5 near here]

It is interesting to compare these results with the parameters' estimates obtained by estimating the short memory version of the model, namely setting $d_x = d_z = 0$. These are reported in Table 5. The AR coefficients are now much larger, in fact close to the unit root bound. Noticeably, the fit of the model deteriorated across all maturities, as indicated by the estimated variances of the measurement errors. A formal test of adequacy between long and short memory will be presented below.

[Insert Figure 6 near here]

Figures 6(a) and 6(b) plot the filtered factors x_t and z_t obtained with the Kalman recursion. These factors appear to be a rotation of the 'level' and 'slope' factors as expressed by the conventional static principal components, in particular the first and the second one, extracted from the nominal yields. This is evident in Table 6 which reports the regression R^2 from projecting each of the first four principal components on the filtered values of x_t and z_t, either individually or jointly. The goodness of fit is virtually zero when projecting either the third or the fourth principal component onto the filtered factors. This is to be expected within a two-factor model.

[Insert Table 6 near here]

[19]Interestingly, Altissimo, Mojon and Zaffaroni (2009) document an estimate of the long memory parameter for the euro area CPI inflation equal to 0.13.

6.2. *Revisiting the Stylized Facts*

We evaluate the extent to which our long memory model captures the dynamic persistence found in the data, as formalised in *Stylized Fact 1.* Figure 7 plots the periodogram (averaged across maturities) of (standardised) nominal yields, forwards and returns together with a theoretical spectral density equal to

$$s_{LM}(\lambda) = c\lambda^{min(-2d_x,-2d_z)}, \quad -\pi \leq \lambda < \pi.$$

with $d_x = 0.2862, d_z = 0.1878$ as from Table 4. The constant c is set such that $s_{LM}(\lambda)$ integrates to one, viz. one obtains unit variance. This simple specification is equivalent to the model-implied spectral densities near the zero frequency for yields, forwards and returns, as indicated in Theorem 4.4, although the other parameters, beyond d_x, d_z, will be important to achieve a good fit of the model across all frequencies. Both Table 4 and Figure 7 confirm that long memory is an important feature of the yields data, inducing a degree of persistence that well agrees with *Stylized Fact 1* .

[**Insert Figure 7 near here**]

We now investigate the capabilities of the long memory model to reproduce the observed volatility term structure of yields, forwards and returns. In particular, we aim to establish whether we can capture *Stylized Fact 2.* Figure 8 reports the term structures of the sample standard deviation (blue line) together with both the long memory (green line) and short memory (red line) model-implied standard deviation for nominal yields (left panel), forward rates (centre panel) and nominal returns (right panel). The closed form formulae are reported in Theorem 4.3. The long memory and short memory term structures use the estimated parameters of Table 4 and Table 5 respectively.

[**Insert Figure 8 near here**]

The long maturity shape of the volatility term structures depend on the magnitude of the coefficients in λ_1 which ensure that the $b^{\$}_{x,i}, b^{\$}_{z,k}$ are well-behaved for large k and, moreover, satisfy condition (58).

The difference between the long and short memory is striking: the long memory model is able to capture *Stylized Fact 2*, namely a declining volatility term structure for intermediate maturities then flattening out or even raising again for long maturities for yields and especially for forward rates. Instead the short memory model implies declining curves for long maturities, in agreement with estimated autoregressive coefficients close yet smaller than unity. For nominal returns, the long memory model is able to produce a monotically increasing term structure without violating stationarity. On the contrary, for the short memory model the volatility term structure appears to flatten out due to the mean-reversion. These results are particularly insightful and not an artefact of overfitting, In fact these are obtained by using the maximum likelihood estimator which does not necessarily deliver a perfect fit of the volatility term structures.

6.3. Real yields, inflation expectation and inflation risk premia

Our simple estimated specification of the long memory model allows to decompose the nominal yields term structure ($r^{\$}_{nt} = -p_{n,t}/n$) into the real ($r_{n,t}$), the inflation expectation ($n^{-1}E_t(\ln(\Pi_{t+n}/\Pi_t))$ and the inflation risk premia ($\mathrm{IP}_{n,t}$) term structures according to

$$r^{\$}_{n,t} = c_n + r_{n,t} + n^{-1}E_t \ln(\frac{\Pi_{t+n}}{\Pi_t}) + \mathrm{IP}_{n,t}. \tag{65}$$

Concerning the term structure of real yields, the estimated model implies an upward sloping real yield curve, as presented in Figure 9(a) (green line) where real yields are defined as $r_{n,t} = \widetilde{A}_n + \widetilde{\mathbf{B}}'_{x,n}\mathbf{C}_{x,t} + \widetilde{\mathbf{B}}'_{z,n}\mathbf{C}_{z,t}$. We also report in Figure 9(a) the real yields volatility term structure (blue line) which appears hump curved, declining across maturities yet slightly increasing at the long end of the curve. Descriptive statistics of real yields for all maturities are reported in Table 7.

[**Insert Table 7 near here**]

In Figure 9(b) we plot the time series of real yields for all maturities. Wheres the short real rates become negative, the long real rates (above 10 year maturity) remain positive throughout the sample period.

[**Insert Figure 9 near here**]

Figure 10 presents the inflation risk premia across maturities. The term structure is hump shaped with a peak at 5 year. The average inflation risk premia here obtained appear generally similar to the ones reported in the literature[20].

[**Insert Figure 10 near here**]

6.4. Statistical performance

6.4.1. In-sample

Having estimated both the long memory and short memory model, we now present some specification and goodness of fit analysis. Table 8 reports log-likelihood values and likelihood ratio test statistic under the null hypotheses H_0: $d_1 = d_2 = 0$ (short memory model). The short memory model is rejected at 1% confidence level in favor of the long memory model. The different performance is also manifested when comparing the estimates of the idiosyncratic errors, much larger for the short memory model.

[**Insert Table 8 near here**]

[20] Ang et al (2008) find that the unconditional mean of inflation risk premia are equal to 31 and 114 basis points, respectively, for 1 and 5 year bonds. For the same maturities we find 29 and 89 basis points, respectively.

We have also evaluated the in-sample forecasting power of the models. This is an in-sample analysis since we use estimated parameter values based on the entire sample. We consider four different forecast horizons (1−, 3−, 6− and 12− months) for each yield, over the period January 2002 to December 2011, where we evaluate the root mean square error (RMSE) statistic. The results, presented in Table 9 show that the long memory model provides a superior goodness of fit across yields and forecasting horizon.

[Insert Table 9 near here]

6.4.2. Out-of-sample

We first compare the out of sample forecasting performance of the long memory and the short memory version of our model. The RMFE of the forecasts are reported in the top two panels of Table 10. The evaluation period is January 2002 - December 2011. Each month we re-estimate the model, using a rolling window of 595 months, and use the results for the new forecast at 1−, 3−, 6− and 12− month horizon. The predictions are obtained as the last recursion of the the Kalman filter. The results shows that the long memory model dominates the short memory one in all cases except in few instances at 1 month horizon.

[Insert Table 10 near here]

We also provide a comparison with other, non-nested, classes of term structure models. The results are reported in the last three panels of Table 10. We consider a version of Ang and Piazzesi (2003) model[21] , of the regime switching model[22] of Ang et al (2008) and of the Diebold and Li (2006) model. Details on the adopted specification of these models are not presented for sake of simplicity but are available upon request. All the considered models contain a relatively similar number of parameters, slightly larger for the Ang and Piazzesi (2003) and for the Ang et al (2008) models.

Our long memory model appears[23] to outperform the Ang and Piazzesi macro model, the regime switching model of Ang et al and the Diebold and Li model across all forecasting horizon for short yields. Instead, the other models appear marginally superior for medium term yields, especially for the 1 and 3 year maturity. The results are only illustrative since we considered one of the possibly simplest specification of the long memory model without having undertaken an extensive model specification.

7. Final remarks

In this paper we introduce the long memory affine model of the term structure, a discrete time essentially affine Gaussian factor term structure model with long memory factors,

[21]We considered a model with 3 latent factors and inflation, where inflation is uncorrelated with latent factors. The joint dynamics of the system is described by a VAR model of order 1.

[22]We exactly replicated the Ang et al (2008) model with their preferred specification, model C.

[23]Ideally the differences in forecasting performance should be assessed using a formal testing procedure such as the Diebold and Mariano (1995) test. However, no asymptotic theory exists that is valid when long memory processes are considered.

designed to account for the strong persistence in observed yields and inflation. We provide the closed-form solution of the model, both in terms of the real and nominal term structure. A detailed characterisation of the long memory implications in terms of the $\mathbb{P}$ and $\mathbb{Q}$ measures' parameters is presented. Despite the infinite dimensional state variables, we show how estimation of the model can be still carried out by maximum likelihood using the Kalman filter recursions. Closed-form expressions for term premia, and other quantities of economic significance, are easy to obtain. We present an empirical application of a stylized two factor version of the model which illustrates how extension of the model from short memory to long memory factors gives a substantial improvement in terms of fit of the model both dynamically as well as across maturity, in particular for the volatility term structure of yields, forwards and holding period returns. The model can provide superior out of sample forecasting performance over many competitive models of the term structure.

Several generalizations are of interest. Given the capability of the long memory model to induce non-negligible volatility of long term yields, its theoretical and empirical implications in terms of term premia dynamics could be substantial. Second, in view of the strong evidence of dynamic conditional heteroskedasticity in observed yields, one should relax the assumption of unconditional Gaussianity and allow for time-varying conditional volatility. We leave this and other extensions to further research.

Appendix A. Aggregation and long memory in affine term structure models

Consider the J factor affine term structure model

$$r_{t,n} = a_n + \sum_{j=1}^{J} n^{-1} B_{j,n} x_{j,t}, \tag{66}$$

where each state variable follows a stationary AR(1) model with a one-factor structure innovation:

$$x_{j,t} = \psi_{x,j} x_{j,t-1} + \gamma_j u_t + \sigma_j \epsilon^*_{j,t}, \;\; j = 1, ..., J,$$

where $-1 < \psi_{x,j} < 1$, $u_t \sim NID(0,1)$, $\epsilon^*_{j,t} \sim NID(0,1)$ mutually independent one another. Here γ_j and σ_j are parameters. If one wants to exclude the idiosyncratic component of the factor structure it suffices to set $\sigma^*_j = 0$ for $j = 1, ..., J$. No-arbitrage implies the J cross-equation restrictions

$$B_{j,n} = \frac{(1 - \psi^n_{x,j})}{(1 - \psi_{x,j})}, \;\; j = 1, ..., J,$$

We wish to evaluate the limiting behaviour of $r_{t,n}$ as $J \to \infty$ and in particular its memory properties. To formalize this, it is useful to assume that the parameters $\theta_j = (\psi_{x,j}, \gamma_j, \sigma_j)'$ are random i.i.d. draws across $j = 1, ..., J$, mutually independent one another. Note that by letting $J \to \infty$ the parameters γ_j and σ^2_j must both be $O_p(J^{-1})$, a simple form of which consists of:

$$\gamma_j = \frac{\gamma_j^*}{J}, \;\; \sigma_j = \frac{\sigma_j^*}{J^{\frac{1}{2}}}, \tag{67}$$

where γ_j^*, σ_j^* are i.i.d. random parameters such that $0 <\mid E\gamma_j^* \mid< \infty$ and $0 < E\sigma_j^{*2} < \infty$. To see why (67) is required, note that the variance of $r_{n,t}$ conditional on parameters $\theta = (\theta_1, ..., \theta_J)$ satisfies

$$var(r_{n,t}) = \sum_{k=0}^{\infty} \left(\sum_{j=1}^{J} n^{-1} B_{j,n} \gamma_j \psi_{x,j}^k \right)^2 + \sum_{k=0}^{\infty} \left(\sum_{j=1}^{J} (n^{-1} B_{j,n} \sigma_j \psi_{x,j}^k)^2 \right) < \infty$$

and (67) ensures that $var(r_{n,t})$ would not increase just because a larger number J of factors $x_{j,t}$ is considered. In other words, the larger is J, the smaller necessarily the loadings γ_j and the variances σ_j^2 must be.

Therefore the second term on the right hand side of (66) involves, through (67), averaging across $j = 1, ..., J$ and it can be decomposed as the sum of two components, one function of the common innovation u_t and the other function of the idiosyncratic innovations $\epsilon_{j,t}$:

$$\begin{aligned} \sum_{j=1}^{J} n^{-1} B_{j,n} x_{j,t} &= \sum_{k=0}^{\infty} \frac{1}{J} \sum_{j=1}^{J} \left(n^{-1} B_{j,n} \gamma_j^* \psi_{x,j}^k \right) u_{t-k} + \sum_{k=0}^{\infty} \left(J^{-1/2} \sum_{j=1}^{J} n^{-1} B_{j,n} \psi_{x,j}^k \epsilon_{j,t-k} \right) \\ &= U_{J,n,t} + E_{J,n,t}. \end{aligned}$$

To close the model assume that the $\psi_{x,j}$ are i.i.d. with density $f(\psi)$ over the interval $[0, 1)$. This ensures stationarity of the model. Instead, no distributional assumptions are required for the other parameters. However, we can leave $f(\psi)$ unspecified except for its behaviour in proximity of 1 (see Assumption II of Zaffaroni (2004)) such as[24]:

$$f(\psi) \sim c(1 - \psi)^b \text{ as } \psi \to 1^-, \tag{68}$$

for some constants b, c where $0 < c < \infty$ and $b > -1$ to ensure integrability of $f(\psi)$.

For the common component, $U_{J,n,t}$, one can show (see Theorem 5 of Zaffaroni (2004)) that for $b > -1/2$

$$U_{J,n,t} = \sum_{k=0}^{\infty} \hat{\phi}_{n,k} u_{t-k} \to_2 U_{n,t} = \sum_{k=0}^{\infty} \phi_{n,k} u_{t-k} \text{ as } J \to \infty,$$

where

$$\hat{\phi}_{n,k} = \left(\frac{1}{J} \sum_{j=1}^{J} n^{-1} B_{j,n} \gamma_j^* \psi_{x,j}^k \right) \to_p \phi_{n,k} = E(\gamma_j^*) E(n^{-1} B_{j,n} \psi_{x,j}^k) \text{ for } k = 0, 1, ... \tag{69}$$

and $\to_p$ denotes convergence in probability. Moreover, by (16) of Zaffaroni (2004) for finite

[24] Particular important cases of (68) are the uniform distribution, for $b = 0$, and the Beta (p, q) distribution, for $q = b + 1$.

n

$$\phi_{n,k} \sim c_n k^{-(b+1)} \text{ as } k \to \infty.$$

for some constant c_n. In fact, notice that the term $n^{-1}B_{j,n}$ does not interfere into the limit behaviour of $\hat{\phi}_{n,k}$ which, in turn, behaves as $E\psi_{x,j}^k$ for large k since $n^{-1}(1-\psi^n)/(1-\psi) \sim 1$ as $\psi \to 1^-$, hence not affecting the way in which (68) leads to the result.

Similarly, the idiosyncratic component $E_{J,nt}$ satisfies (see Theorem 3 of Zaffaroni (2004)) for $b > 0$

$$E_{J,n,t} \to_d E_{n,t} = \sum_{k=0}^{\infty} \upsilon_{n,k}\eta_{t-k} \text{ as } J \to \infty,$$

with $\eta_t \sim NID(0,1)$ and where, for finite n,

$$\upsilon_{n,k} \sim c_n k^{-(b+1)/2} \text{ as } k \to \infty,$$

where $\to_d$ denotes convergence in distribution.

Therefore, for large J, real yields $r_{n,t}$ can be expressed, net of constant terms, as the sum of $U_{n,t}$ and $E_{n,t}$, with coefficients satisfying (42) and hence implying

$$cov(U_{n,t}, U_{n,t+k}) \sim c_n\, k^{-2b-1} \text{ and } cov(E_{n,t}, E_{n,t+k}) \sim c_n\, k^{-b} \text{ as } k \to \infty.$$

Long memory is obtained for b not too large, in particular when $-1/2 < b < 0$ for U_{nt} and $0 < b < 1$ for $E_{n,t}$, respectively. Therefore (63), namely

$$cov(r_{n,t}, r_{n,t+k}) \sim c\, k^{2d-1} \text{ as } k \to \infty$$

holds for some $0 < d < 1/2$ under the above conditions. Note that the limit of $r_{n,t}$ can be viewed as a semiparametric affine model since $U_{n,t}$ and $E_{n,t}$ are function of the infinite sequences of coefficients $\phi_{n,k}, \upsilon_{n,k}$, $k = 0, ...$ which are unspecified except for their long lag behaviour as $k \to \infty$, as indicated above. For practical estimation of the model, as indicated in the main body of the paper, a suitable parameterization of the $\phi_{n,k}$ and $\upsilon_{n,k}$ is necessary such as the ARFIMA.

Appendix B. Pricing implications of long memory

We summarize here the pricing implications of allowing a tradeable asset to have long memory. Following Rogers (1997), assume that the log price of a generic asset, here denoted p_t, follows a fractional Brownian motion which can be represented as

$$p_t = k \int_{-\infty}^{\infty} \left((t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \right) dB_s,\ t \in R, \tag{70}$$

where $x_+ = x1(x \geq 0)$, for a positive constant k where B_t denotes a Brownian motion (set $B_0 = 0$) and $H \in (0,1)$ is a scalar parameter. It is well known that the one-period rate of return $r_t = p_t - p_{t-1}$ is a stationary, mean zero, stochastic process with long memory whenever $H \neq 1/2$ since

$$cov(r_t, r_{t+u}) \sim c\, u^{2H-2}, \quad \text{as } u \to \infty. \tag{71}$$

Expression (71) is analogue[25] to (43) by setting $H = d + 1/2$. Generally speaking, representation (70) implies some predictability so that one can obtain gains with an arbitrarily small variance over a finite period, based on a combination of 'buy-and-hold' strategies. More formally, Rogers (1997) shows that r_t is not a semi-martingale for $H \neq 1/2$ and thus, by the fundamental theorem of asset pricing (Delbaen and Schachermayer (1994)), a mild form of arbitrage exists called 'free lunch with vanishing risk'. An essential condition for this is, however, to observe the entire history of log prices. Instead, $r_t = B_t$ when $H = 1/2$, r_t is i.i.d. and therefore not predictable. Hence no-arbitrage holds.

Cheridito (2003) shows that profitable 'buy-and-hold' strategies, with a vanishing risk, can still be constructed when observing the asset price over a finite interval. However, it is essential to be able to trade over any arbitrarily small interval of time, a condition ruled out when observing data in discrete time. Therefore, observing a finite number of observations over a finite time interval rules out[26] mild forms of arbitrage such as 'free lunch with vanishing risk'. The previous results assumed a friction-less market. Guasoni et al (2010) show that even a minimal amount of transaction costs is enough to rule out arbitrage opportunities when asset (log) pricing follow a fractional Brownian motion, ensuring the existence of an analogue concept to equivalent martingale measure.

Therefore, although long memory in asset prices can have potentially dramatic consequences ruling out existence of pricing functionals, it turns out that very stringent conditions are required for this to be verified. These conditions are extremely unlikely to hold in practice.

Appendix C. Alternative approaches to model the persistence of nominal bonds

The persistence of nominal yields represents an important challenge to models of the term structure. Although solution of affine models, as exemplified in the previous sections, does not require stationarity since it is based on evaluation of conditional moments, the possibility of unit root state variables is troublesome.

Two main approaches have emerged in the literature to tackle this problem. One strand maintains the assumption that the state variables' dynamics is described by a parametric linear process such as a finite order VAR[27]. Stationarity is typically imposed in the estimation. It is well known that the ordinary least squares estimates of the maximal autoregressive

[25] It can be shown that the discrete-time process r_τ, $\tau = 0, \pm 1, ...$ admits a representation (29) with coefficients satisfying (33) and (42).

[26] Rogers (1997) notes that it is not the long memory feature (71) of the model that could lead to arbitrage opportunities. In fact he shows how to construct a Gaussian process satisfying (71) and yet with the semimartingale property (see his Section 5).

[27] Among this vast literature, see for instance Dai and Singleton (2000), Duffee (2002) and Ang and Piazzesi (2003).

root are plagued by a downward bias, the more intense the closer the root is to unity (see Kendall (1954)), suggesting a spuriously low degree of mean-reversion found in the data. The various approaches of this line of research differ for the way used to mitigate this bias. The aim here is to afford a very precise estimate of the maximal autoregressive root which, when below unity, justifies the stationarity paradigm. For instance, it has been proposed that adding further information into the state space could mitigate the bias problem, such as including both short and long term yields (see Ball and Torous (1996)) or long horizon survey forecasts of short yields (see Kim and Orphanides (2012)). Others rely on identification assumptions such as Joslin et al (2010), who impose the same degree of persistence under the $\mathbb{P}$ and $\mathbb{Q}$ measures, making the model more parsimonious and thus, as a by product, affording more precise estimation. Prompted by the recent findings of Joslin et al (2011) and Hamilton and Wu (2012), who show that ordinary least squares provides a computationally efficient first-stage method for full maximum likelihood estimation, Bauer et al (2012) realize that bias-corrected estimators could then be easily afforded in such first-stage part of the estimation procedure. An alternative bias-correction method is proposed in Jardet et al (2013) by blending stationarity-imposed estimates and unit root by means of model averaging techniques.

A second strand of the literature departs from linearity altogether and instead explores different, possibly nonlinear, models for yields dynamics. This would permit to capture a strong degree of mean-reversion for extreme values of the data, together with no or limited mean-reversion when the data are observed in the centre of their distribution. Nonparametric approaches, hence allowing for an unspecified form of nonlinearity, include Ait-Sahalia (1996), Stanton (1997) and Conley et al (1997) among others. Another attractive, parametric, nonlinear alternative is obtained by means of allowing regime switching state variables, as illustrated below, which is very effective in capturing persistence.

Our long memory model lies somewhere in between these two approaches. We are postulating a stationary Gaussian, hence linear, model, retaining the possibility of estimating the model with maximum likelihood and the Kalman recursions. In fact the factors have a Gaussian VAR(∞) representation, departing from the finite-dimensional $DA_0^{\mathbb{Q}}(N)$ class. However, in our case the autoregressive coefficients or, equivalently, the impulse response function, cannot be left unconstrained but instead must satisfy a suitably defined long lags behaviour in order to induce long memory. Nonlinear estimation cannot be avoided. In a time series context, it has been widely established that a degree of persistence similar to long memory can also be induced by regime switching models when the transition matrix has most of its mass on the diagonal terms (see Diebold and Inoue (2001)). This prompts us to investigate the extent to which regime switching term structure and long memory models capture similar features of the data. Insights can be obtained from a simple, discrete time, regime switching term structure model for real yields such as

$$r_{n,t}(i) = c_n(i) + d_n x_t, \tag{72}$$

with one factor that follows the regime switching process:

$$x_t = \mu(i) + \psi_x x_{t-1} + \sigma(i)\epsilon_{x,t}, \tag{73}$$

where the regime variable $s_t = i \in \{1, ..., K\} = \mathcal{K}$ follows a K-state Markov chain with constant (under $\mathbb{Q}$) transition probabilities $p_{i,j} = Pr(s_t = i | s_{t-1} = j)$, $i, j \in \mathcal{K}$. Model (72)-(74)-(75) is a stylized version of Ang et al (2008). The constant autoregressive parameter satisfies $|\psi_x| < 1$ whereas drift and volatility parameters are assumed to depend on the regime variable, $\mu(i), \sigma(i), i \in \mathcal{K}$. The affine function coefficients satisfy $c_n(i) = C_n(i)/n$, $d_n = D_n/n$, $i \in \mathcal{K}$ with

$$
\begin{aligned}
C_{n+1}(i) &= g\left(C_n(j), B_n, \mu(j), \lambda(j), j \in \mathcal{K}; \theta\right), && (74)\\
D_{n+1} &= -1 + \psi_x D_n, && (75)
\end{aligned}
$$

where θ are constant parameters. A particular element of this class of models will be characterised by a specific choice for the function $g(\cdot)$. Importantly, closed-form solution of a general regime switching terms structure models requires regime-invariance of the coefficients D_n (see Dai and Singleton (2003)).

From (72) it follows

$$
\begin{aligned}
&cov(r_{n,t}, r_{n,t+k}) = \\
&cov(c_n(s_t), c_n(s_{t+k})) + d_n^2 cov(x_t, x_{t+k}) + d_n cov(c_n(s_t), x_{t+k}) + d_n cov(c_n(s_{t+k}), x_t),
\end{aligned}
$$

and

$$
var(r_{n,t}) = var(c_n(s_t)) + d_n^2 var(x_t) + 2 d_n cov(c_n(s_t), x_t).
$$

The regime switching mechanism influences the second moments of the yields $r_{n,t}$ across maturity primarily through the second moments of $c_n(s_t)$. Instead, the effect of x_t is of second-order importance for large maturity n since the regime-invariant affine coefficients d_n converge rapidly toward zero with n when $| \psi_x |< 1$.

Diebold and Inoue (2001) illustrate, with a detailed Monte Carlo experiment that when the $p_{1,1} = p_{2,2} = 0.95$ (they consider $K = 2$) and for samples between 200 and 400 observations, that long memory is manifested with estimates of the memory parameter well in the stationary region (40). Such values for the transition probabilities are not too dissimilar from estimated probabilities found in the term structure literature, especially when a small number of states K is considered[28].

Therefore, regime switching and long memory models are both able to account for the persistence of observed yields[29], implying an asymptotic behaviour of the autocovariances such as (43). The two models can instead differ with respect to the term structure of volatility. In fact, the regime-invariant coefficients d_n will either decrease rapidly towards zero or explode for large maturity depending on whether ψ_x is smaller or larger than unity, much in the same way as for the basic model of Section 3. A slowly decaying volatility is not warranted and requires a suitable parameterization of the sequence $c_n(i)$. In any case, a closed-form expression of the volatility term structure does not follow in general. Instead, our

[28] Among others see Bansal and Zhou [Table 4](2002), Evans [Table 2] (2003), Ang et al [Table 3] (2008), Dai et al [eq (34)] (2007) and Bikbov and Chernov (2013)). Note that Ang et al (2008) consider four states and the transition probability matrix is less concentrated, than others, around the diagonal. Dai et al (2007) consider constant transition probabilities under the $\mathbb{Q}$ measure.

[29] The multifrequency term structure model of Calvet, Fisher and Wu (2010) might also be able to describe this feature of the data, given its strong analogies with regime switching models.

long memory model preserves a closed-form solution yet providing a volatility term structure which can be either (mildly) negatively or positively sloped for long maturities.

Appendix D. State space representation and estimation of linear long memory processes

Chan and Palma (1998) clarify that ARFIMA admit an infinite-dimensional state space representation. In particular, setting $\phi_i = \phi_i(\xi_0)'$ for the $p+q+2$ parameter $\xi = (\theta_1, ..., \theta_q, \psi_1, ..., \psi_p, d, \sigma^2)'$ where ξ_0 denotes the true value, the ARFIMA(p, d, q) process

$$y_t = \frac{\Theta(L)}{\Psi(L)}(1-L)^{-d}\epsilon_t = \sum_{i=0}^{\infty} \varphi_i \epsilon_{t-i},$$

is shown to be equivalent to the state space system (see Chan and Palma (1998), p. 723)

$$\begin{aligned} \mathbf{X}_{t+1} &= \mathbf{F}\mathbf{X}_t + \mathbf{H}\epsilon_t, \\ y_t &= \mathbf{G}\mathbf{X}_t + \epsilon_t, \end{aligned} \tag{76}$$

where $\mathbf{X}_t$ is an infinite dimensional vector defined as

$$X_t = \begin{bmatrix} E[y_t \mid y_{t-1}, y_{t-1}, ...] \\ E[y_{t+1} \mid y_{t-1}, y_{t-1}, ...] \\ E[y_{t+2} \mid y_{t-1}, y_{t-1}, ...] \\ \vdots \end{bmatrix},$$

with coefficients

$$\begin{aligned} \mathbf{F} &= \begin{bmatrix} 0 & 1 & 0 & & \cdots \\ 0 & 0 & 1 & 0 & \\ \vdots & \vdots & & \ddots & \ddots \end{bmatrix}, \\ \mathbf{H} &= [\varphi_1\, \varphi_2\, \ldots]' \text{ and} \\ \mathbf{G} &= [1\,0\,0\ldots]. \end{aligned}$$

Despite the infinite dimensionality of the system, Chan and Palma (1998) show that based on a sample of T observations $(y_1, ...y_T)$ the exact Gaussian likelihood function can be obtained through the usual Kalman recursion, based on the first T components of the Kalman equations (76). Although the exact likelihood can be computed in a finite number of steps, $O(T^3)$ evaluations are required. Therefore, Chan and Palma (1998) propose an approximate maximum likelihood approach which can be computed in a smaller number of steps, yet maintaining the same asymptotic properties. This is obtained by recognising that the first-difference $y_t - y_{t-1}$ satisfies

$$y_t - y_{t-1} = \sum_{i=0}^{\infty} \psi_i^* \epsilon_t, \;\; \psi_i^* = \psi_i - \psi_{i-1}, \;\; \psi_0^* = 1.$$

Consider its m-truncation, for an arbitrarily chosen $m > 1$,

$$z_t = \sum_{i=0}^{m} \psi_i^* \epsilon_t.$$

Then z_t is a finite-order, in fact m-order, moving average and its state space representation can be easily obtained:

$$\mathbf{X}_{t+1} = \begin{bmatrix} 0 & & \mathbf{I}_m \\ 0 & \dots & 0 \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} \psi_1^* \\ \vdots \\ \psi_m^* \end{bmatrix} \epsilon_t, \tag{77}$$

$$z_t = [\,1\,0 \cdots 0\,]\,\mathbf{X}_t + \epsilon_t. \tag{78}$$

Now the algorithm requires $O(m^2 T)$ iterations, where typically one sets $m < T$. The truncation implies an approximation error which, nevertheless, is mitigated by having taken the first-differences since the ψ_i^* decay to zero faster than the ψ_i. The asymptotic theory developed by Chan and Palma (1998) requires m to diverge to infinity with T although at a smaller rate such as $m = T^\beta, \beta \geq 1/2$. Note that the approximation is better the larger is m. The approximate maximum likelihood estimator for ξ is then

$$\hat{\xi} = \text{argmax}_\xi \;\; l_T(\xi)$$

where the approximate Gaussian log likelihood is

$$l_T(\xi) = -\frac{1}{2} \log\, det[\mathbf{M}(\xi)] - \frac{1}{2} \mathbf{z}_T' \mathbf{M}(\xi) \mathbf{z},$$

and where $\mathbf{M}(\xi_0)$ is the population covariance matrix corresponding to $\mathbf{z}_T = (z_1, ..., z_T)'$.

We rely on the above set up although we find more convenient to define the transition equations as (note the time index of the state variable):

$$\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{H}\epsilon_t.$$

Following the Monte Carlo results in Chan and Palma (1998), we set the truncation at $m = 60$ lags.

Appendix E. Proof of Theorem 4.4, 4.5 and 4.6

We first establish two preliminary lemmas.

Lemma E.1. *For a finite d assume*

$$\phi_i \sim c i^{d-1} \text{ as } i \to \infty.$$

Setting

$$\Phi_{n,0} = \sum_{j=0}^{n-1} \phi_j \tag{79}$$

then

$$\sum_{i=1}^{n} \Phi_{i,0} \sim \begin{cases} cn \log n, & d = 0, \\ cn^{d+1}, & d > 0, \\ cn, & d < 0, \end{cases} \quad \text{as } n \to \infty,$$

and

$$\sum_{i=1}^{n} \Phi_{i,0}^2 \sim \begin{cases} cn \log^2 n, & d = 0, \\ cn^{2d+1}, & d > 0, \\ cn, & d < 0, \end{cases} \quad \text{as } n \to \infty,$$

where c denotes an arbitrary constant, not always the same.

Proof. Assume with no loss of generality that $\phi_i \neq 0$ for all $i < \infty$. Consider case $d > 0$. Since ϕ_i is (asymptotically) monotone in i

$$\sum_{j=1}^{i} \phi_j \sim c \int_1^i j^{d-1} = c i^d \text{ as } i \to \infty,$$

then

$$\sum_{j=1}^{n} \Phi_{j,0} \sim cn^{d+1} \text{ as } n \to \infty.$$

When $d = 0$ instead

$$\sum_{j=1}^{i} \phi_j \sim c \int_1^i j^{-1} = c \log(i) \text{ as } i \to \infty,$$

and

$$\sum_{j=1}^{n} \Phi_{j,0} \sim cn \log(n) \text{ as } n \to \infty.$$

Finally, when $d < 0$

$$\sum_{j=1}^{i} \phi_j \sim c \int_1^i j^{d-1} = c \text{ as } i \to \infty,$$

yielding

$$\sum_{j=1}^{n} \Phi_{j,0} \sim cn \text{ as } n \to \infty.$$

The results for $\sum_{j=0}^{n} \Phi_{j,0}^2$ follow along the same lines. QED

Lemma E.2. *For a finite d assume*

$$\phi_i \sim c i^{d-1} \text{ as } i \to \infty.$$

Setting

$$\Phi_{n,i} = \sum_{j=0}^{n-1} \phi_{j+i}$$

then

$$\Phi_{n,i} = \begin{cases} O((\log(n)), & d = 0, \\ O(n^d + i^d), & d \neq 0, \end{cases} \quad \text{for } i \leq n,$$

and

$$\Phi_{n,i} = \begin{cases} O(n/i), & d = 0, \\ O(ni^{d-1}), & d \neq 0, \end{cases} \quad \text{for } i > n.$$

Proof. Consider $d = 0$. Then for $0 < i \leq n$, $\sum_{j=0}^{n-1} 1/(i+j) \leq 1/i + \sum_{j=1}^{n-1} 1/j \sim c\log(n)$. When instead $i > n$ then for some $0 < \tilde{n} < n$, by the mean value theorem,

$$\sum_{j=0}^{n-1} \frac{1}{(i+j)} \sim c(\log(n+i) - \log(i)) = c\frac{n}{\tilde{n}+i} \leq c\frac{n}{i}.$$

For $d > 0$, when $i \leq n$ then $\sum_{j=0}^{n-1} (i+j)^{d-1} \sim c((n+i)^d - i^d) \sim cn^d$ since $n^d \leq ((n+i)^d - i^d) \leq n^d(2^d - 1)$. When $d < 0$ then $\sum_{j=0}^{n-1}(i+j)^{d-1} \sim c(i^d - (n+i)^d) \sim ci^d$ whereas if $i \sim cn$ then $(i^d - (n+i)^d) \sim cn^d$. For $i > n$ by the mean value theorem, for some $0 < \tilde{n} < n$,

$$\sum_{j=1}^{n} (i+j)^{d-1} \sim c((n+i)^d - i^d) = cn(\tilde{n}+i)^{d-1} \leq cni^{d-1}.$$

Similar reasonings apply to the case $d < 0$. QED

Proof of Theorem 4.4.

We first characterise the log lags behaviour of the autocovariances and the subsequently the local behaviour of the spectra near the zero frequency. For given n, Lemma E.2 can be strengthen to

$$\Phi_{x,n,j} \sim cj^{d_x-1}, \quad \Phi_{z,n,j} \sim cj^{d_z-1} \text{ as } j \to \infty.$$

Note that, since n is fixed, this result applies for any values for $b_{x,k}, b_{z,k}$, irrespective of whether λ_1 is zero or not, that is under either the $\mathbb{P}$ or the $\mathbb{Q}$ measure.

The autocovariance of $r_{n,t}$ satisfies

$$\begin{aligned} \operatorname{cov}(r_{n,t}, r_{n,t+u}) \quad &= \sigma_x^2 \left(\sum_{j=0}^{\infty} \Pi_{x,j} \Pi_{x,j+u} \right) + \delta_z^2 \sigma_z^2 \left(\sum_{j=0}^{\infty} (n^{-1}\Phi_{z,n,j})(n^{-1}\Phi_{z,n,j+u}) \right) \\ &\sim cu^{2d_x-1} + cu^{2d_z-1} \text{ as } u \to \infty, \end{aligned}$$

setting $\Pi_{x,j} \equiv n^{-1}\Phi_{x,n,j} + \beta_{zx} n^{-1}\Phi_{z,n,j}$. To show this, we use a truncation argument as follows. From $\sum_{j=0}^{\infty} \Pi_{x,j}\Pi_{x,j+u} = \sum_{j=0}^{u} \Pi_{x,j}\Pi_{x,j+u} + \sum_{j=u+1}^{\infty} \Pi_{x,j}\Pi_{x,j+u}$ one gets, as $u \to \infty$,

$$\sum_{j=0}^{u} \Pi_{x,j}\Pi_{x,j+u} \sim c\Pi_{x,u} \sum_{j=0}^{u} \Pi_{x,j} \sim cu^{2d_x-1}$$

and, likewise,

$$\sum_{j=u+1}^{\infty} \Pi_{x,j}\Pi_{x,j+u} \sim c \sum_{j=u+1}^{\infty} \Pi_{x,j}^2 \sim cu^{2d_x-1}.$$

The same applies for the second term of $\mathrm{cov}(r_{n,t}, r_{n,t+u})$ in $\Phi_{z,n,j+u}$. Moreover, by (57), $\mathrm{cov}(r_{n,t}, r_{n,t+u})$ satisfies the quasi-monotonic convergence condition $|\mathrm{cov}(r_{n,t}, r_{n,t+u}) - \mathrm{cov}(r_{n,t}, r_{n,t+u+1})| = O(u^{-1}|\mathrm{cov}(r_{n,t}, r_{n,t+u})|)$ and the bounded variation condition $\sum_{k=u}^{\infty} |\mathrm{cov}(r_{n,t}, r_{n,t+k}) - \mathrm{cov}(r_{n,t}, r_{n,t+k+1})| = O(|\ \mathrm{cov}(r_{n,t}, r_{n,t+u}\ |)$ as $u \to \infty$. In fact, since $|\Pi_{x,j} - \Pi_{x,j+1}| \le cj^{-1}\Pi_{x,j}$ by elementary calculations,

$$\sum_{j=0}^{\infty} |\Pi_{x,j}||\Pi_{x,j+u} - \Pi_{x,j+u+1}| = \sum_{j=0}^{u} |\Pi_{x,j}||\Pi_{x,j+u} - \Pi_{x,j+u+1}| + \sum_{j=u+1}^{\infty} |\Pi_{x,j}||\Pi_{x,j+u} - \Pi_{x,j+u+1}|$$

$$\le c|\Pi_{x,u} - \Pi_{x,u+1}| \sum_{j=0}^{u} |\Pi_{x,j}| + c \sum_{j=u+1}^{\infty} |\Pi_{x,j}||\Pi_{x,j} - \Pi_{x,j+1}| \le cu^{-1}|\Pi_{x,u}| \sum_{j=0}^{u} |\Pi_{x,j}| + c \sum_{j=u+1}^{\infty} j^{-1}|\Pi_{x,j}|^2$$

$$\le cu^{-1}|\Pi_{x,u}||u^{d_x}| + c \sum_{j=u+1}^{\infty} j^{-2d_x-3} \le cu^{-1}|u^{2d_x-1}| \le cu^{-1} \sum_{j=0}^{\infty} |\Pi_{x,j}\Pi_{x,j+u}|,$$

the same holding for the term in $\Phi_{z,n,j+u}$. Therefore, the conditions of Young (1974), Lemma III-12, hold concluding the proof. The same proof apply to the spectral density of $f_{n,t}$ and $y_{n,t}$. QED

Proof of Theorem 4.5. The results easily follow by applying Lemma E.1 to the conditional variances formulae (54). For the unconditional variances, use Lemma E.2 together with a truncation argument. For example, for $\mathrm{var}(r_{t,n})$

$$\sum_{j=0}^{\infty}(n^{-1}\Phi_{x,n,j})^2 = \sum_{j=0}^{n}(n^{-1}\Phi_{x,n,j})^2 + \sum_{j=n+1}^{\infty} (n^{-1}\Phi_{x,n,j})^2,$$

and

$$\sum_{j=0}^{n}(n^{-1}\Phi_{x,n,j})^2 = O(n^{-2}n^{2d_x+1} + n^{-2}\sum_{i=1}^{n} i^{2d_x}) = O(n^{2d_x-1}),$$

$$\sum_{j=n+1}^{\infty} (n^{-1}\Phi_{x,n,j})^2 = O(\sum_{j=n+1}^{\infty} i^{2(d_x-1)}) = O(n^{2d_x-1}).$$

A similar reasoning applies to $\mathrm{var}(f_{t,n})$. For $\mathrm{var}(y_{t,n})$ notice that since $d_x, d_z < 1/2$ then $\sum_{j=0}^{\infty} \phi_{x,j}^2$ and $\sum_{j=0}^{\infty} \phi_{z,j}^2$ are bounded. QED

Proof of Theorem 4.6. Focus on $\Phi_{x,n,j}$, the same results applying to $\Phi_{z,n,j}$. Since $b_{x,j} \sim$

cj^{d_x} as $j \to \infty$, re-writing

$$\Phi_{x,n,j} = \sum_{j=0}^{n-1} b_{x,n-i}\phi_{x,i+j} = \sum_{j=0}^{[n/2]} b_{x,n-i}\phi_{x,i+j} + \sum_{j=[n/2]+1}^{n-1} b_{x,n-i}\phi_{x,i+j} \equiv I + II$$

one gets

$$I = \sum_{j=0}^{[n/2]} b_{x,n-i}\phi_{x,i+j} \sim cn^{d_x} \sum_{j=0}^{[n/2]} (i+j)^{d_x-1} \sim cn^{d_x}(([n/2]+j)^{d_x} - j^{d_x}),$$

and

$$II \sim c(j+n)^{d_x-1} n^{d_x+1}.$$

For I one obtains

$$I \sim \begin{cases} n^{2d_x} & j/n \to 0 \\ n^{d_x+1} j^{d_x-1} & n/j \to 0, \end{cases}$$

where the two cases coincide when $j/n \sim c$.

For the conditional variance of yields $r_{n,t}$ the result follows simply by substituting the above results into $\Phi_{x,n,0}/n$ and $\Phi_{z,n,0}/n$ and squaring terms. An easy truncation argument leads to the unconditional variance expression, whereby

$$\sum_{j=0}^{\infty} \Phi_{x,n,j}/n = \sum_{j=0}^{n} \Phi_{x,n,j}/n + \sum_{j=n+1}^{\infty} \Phi_{x,n,j}/n$$

and we apply the results obtained above to the two cases $0 \leq j \leq n$ and $j > n$, then squaring terms. With respect to forward rates $f_{n,t}$ the conditional variance result follows from

$$\Phi_{x,n+1,0} - \Phi_{x,n,0} \sim c((n+1)^{2d_x} - n^{2d_x}) \sim cn^{2d_x-1},$$

whereas for their unconditional variance, by a truncation argument,

$$\sum_{j=0}^{\infty} (\Phi_{x,n+1,j} - \Phi_{x,n,j})^2 \sim c \sum_{j=0}^{\infty} j^{2d_x-2}((n+1)^{d_x+1} - n^{d_x+1})^2 \sim cn^{2d_x}.$$

Finally, for returns the result follows straightforwardly substituting I and II into the conditional variance expression. In terms of the returns' unconditional variance

$$
\begin{aligned}
var(y_{n,t}) \quad &= O\Big(\sum_{j=0}^{\infty}(n^{d_x+1}j^{d_x-1} - (n-1)^{d_x+1}(j+1)^{d_x-1})^2\Big) \\
&= O\Big(\sum_{j=0}^{\infty}(n^{d_x+1}(j^{d_x-1} - (j+1)^{d_x-1}) + (n^{d_x+1} - (n-1)^{d_x+1})(j+1)^{d_x-1})^2\Big) \\
&= O\Big(2\sum_{j=0}^{\infty} n^{2d_x+2}(j^{d_x-1} - (j+1)^{d_x-1})^2 + 2\sum_{j=0}^{\infty}(n^{d_x+1} - (n-1)^{d_x+1})^2(j+1)^{2d_x-2}\Big) \\
&= O\Big(n^{2d_x+2}\sum_{j=0}^{\infty} j^{2d_x-4} + n^{2d_x}\sum_{j=0}^{\infty} j^{2d_x-2}\Big) = O(n^{2d_x+2}),
\end{aligned}
$$

since $(j+1)^{d_x-1} - j^{d_x-1} = (d_x-1)(j+\epsilon)^{d_x-2} \sim cj^{d_x-2}$ for some $0 < \epsilon < 1$. QED

References

[1] Ait-Sahalia Y., (1996), "Testing continuous-time models of the spot interest rate", Review of Financial Studies, vol 9, 385-426.

[2] Altissimo F., Mojon B. and Zaffaroni P., 2009, "Can aggregation explain the persistence of inflation?" Journal of Monetary Economics, vol 56, 231-241.

[3] Ang A., Bekaert G. and Wei M., 2008, "The term structure of real rates and expected inflation", Journal of Finance, vol 63, 797-849.

[4] Ang A. and Piazzesi M., 2003, "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables", Journal of Monetary Economics, vol 50, 745–787.

[5] Backus D.K. and Zin S.E., 1993, "Long-memory inflation uncertainty: evidence from the term structure of interest rates", Journal of Money, Credit and Banking, vol 25, 681-700.

[6] Ball C.A. and Torous W., 1996, "On unit roots and the estimation of interest rate dynamics", vol 3, Journal of Empirical Finance, vol , 215-238.

[7] Bansal R. and Zhou H. , 2002, "Term structure of interest rates with regime shifts", Journal of Finance, vol 57, 1997-2043.

[8] Bauer M.D., Rudebusch G.,D. and Wu J.C., 2012, "Correcting estimation bias in dynamic term structure models", Journal of Business & Economic Statistics, vol 30, 454-467.

[9] Bikbov R. and Chernov M., 2013, "Monetary policy regimes and the term structure of interest rates", Journal of Econometrics, 174, 2743.

[10] Boivin J., Giannoni M.P and Mihov I., 2009, "Sticky prices and monetary policy: evidence from disaggregated US data", American Economic Review, vol 99, 350-384.

[11] Brockwell P.J. and Davis R.A., 1991, "Time series: theory and methods", Springer-Verlag, New York, Inc., second edition.

[12] Campbell J.Y., Lo A.W. and MacKinlay A.C., 1997, "The econometrics of financial markets", Princeton University Press, Princeton, NJ.

[13] Calvet L.E. , Fisher A.J. and Wu L., 2010, "Dimension-invariant dynamic term structures", working paper.

[14] Chambers M. J., 1998, "Long memory and aggregation in macroeconomic time series", International Economic Review, vol 39, 1053-1072.

[15] Chan N. H and Palma W., 1998, "State space modeling of long-memory processes", Annals of Statistics, vol 26, 719-740.

[16] Cheridito P., 2003, "Arbitrage in fractional Brownian motion models", Finance and Stochastics, vol 7, 533-553.

[17] Conley T. G., Hansen L.P. and Luttmer E.C.J. and Scheinkman J.A. , 1997, Short term interest rates as subordinated diffusions, Review of Financial Studies, vol 10, 525-577.

[18] Comte F. and Renault E., 1996, "Long memory continuous time models", Journal of Econometrics, vol 73, 101-149.

[19] Cox J.C., Ingersoll J.E. and Ross S.A., 1985, "A theory of the term structure of interest rates", Econometrica, vol 53, 385-407.

[20] Dai Q. and Singleton K.J., 2000, "Specification analysis of affine term structure models", Journal of Finance, vol 55, 1943-1978.

[21] Dai Q. and Singleton K.J., 2002, "Expectation analysis of affine term structure models", Journal of Financial Economics, vol 63, 415-441.

[22] Dai Q. and Singleton K.J., 2003, "Term structure dynamics in theory and reality", Review of Financial Studies, vol 16, 631-678.

[23] Dai Q., Singleton K.J. and Yang W., 2007, "Regime shifts in a dynamic term structure model of U.S. Treasury bond yields", Review of Financial Studies, vol 20, 1669-1706.

[24] Delbaen F. and Schachermayer W., 1994, "A general version of the fundamental theorem of asset pricing", Mathematische Annalen, vol 300, 463–520.

[25] Diebold F.X. and Inoue A., 2001, "Long memory and regime switching", Journal of Econometrics, vol 105, 131-159.

[26] Diebold F.X. and Li C., 2006, "Forecasting the term structure of government bond yields", Journal of Econometrics, vol 120, 337-364.

[27] Diebold F.X. and Mariano R.S., 1995, "Comparing predictive accuracy", Journal of Business and Economic Statistics, vol 13, 253-263.

[28] Dothan L.U. , 1978, "On the term structure of interest rates", Journal of Financial Economics, vol 6, 59-69.

[29] Duan, J.C. and Jacobs, K., 1996, "A simple long memory equilibrium interest rate model", Economic Letters, vol 53, 317-321.

[30] Duffie D. and Kan R., 1996, "A yield-factor model of interest rates", Mathematical Finance, vol 6, 379 - 406.

[31] Duffee G., 2002, "Term premia and interest rate forecast in affine models", Journal of Finance, vol 57, 405-443.

[32] Duarte J., 2004, "Evaluating an alternative risk preference in affine term structure models", Review of Financial Studies, vol. 17, 379-404.

[33] Evans M.D., 2003, "Real risk, inflation risk and the term structure", The Economic Journal, vol 114, 345-389.

[34] Granger C. W. J., 1980, "Long memory relationships and the aggregation of dynamic models", Journal of Econometrics, vol 14, 227-238.

[35] Guasoni P., Rasonyi M. and Schachermayer W., 2010, "The fundamental theorem of asset pricing for continuous processes under small transaction costs" , Annals of Finance, vol 6, 157-191.

[36] Gurkaynak R. S., Sack B. and Wright J. H., 2007, "The U.S. Treasury yield curve: 1961 to the present", Journal of Monetary Economics, vol 54, 2291-2304.

[37] Hamilton J.D. and Wu J.C. 2012, "Identification and estimation of Gaussian affine term structure models", Journal of Econometrics, vol. 168, 315331.

[38] Harrison J.M. and Kreps D.M., 1979, "Martingales and arbitrage in multiperiod securities markets", Journal of Economic Theory, vol 20, 381-408.

[39] Hördahl P., Tristani O. and Vestin D., 2008, "The yield curve and macroeconomic dynamics," Economic Journal, vol. 118, 1937-1970.

[40] Jardet C., Monfort A. and Pegoraro F., 2013, "No-arbitrage Near-Cointegrated VAR(p) term structure models, term premia and GDP growth", Journal of Banking & Finance, vol 37, 389-402.

[41] Joslin S., Priebsch M. and Singleton K. J., "Risk premiums in dynamic term structure models with unspanned macro risks", Journal of Finance, forthcoming.

[42] Joslin S., Singleton K.J. and Zhu H. 2011, "A new perspective on gaussian dynamic term structure models", Review of Financial Studies, vol 24, 926-970.

[43] Kendall M.G., 1954, ”A note on bias in the estimation of autocorrelation”, Biometrika, vol 41, 403- 404.

[44] Kim D. H. and Orphanides A., 2012, ”Term structure estimation with survey data on interest rate forecasts”, Journal of Financial and Quantitative Analysis, vol 47, 241-272.

[45] Kitsul Y. and Wright J.H., 2012, ”The Economics of options-implied inflation probability density functions” , NBER Working Paper No. 18195.

[46] Le A., Singleton K.J. and Dai Q., 2010, ”Discrete-time dynamic term structure models with generalized market prices of risk”, Review of Financial Studies, vol. 23, 2184-2227.

[47] Litterman R. and Scheinkman J., 1991, ”Common factors affecting bond returns”, Journal of Fixed Income, vol 1, 62-74.

[48] Rogers L. C. G., 1997, ”Arbitrage with fractional Brownian motion”, Mathematical Finance, vol 7, 95-105.

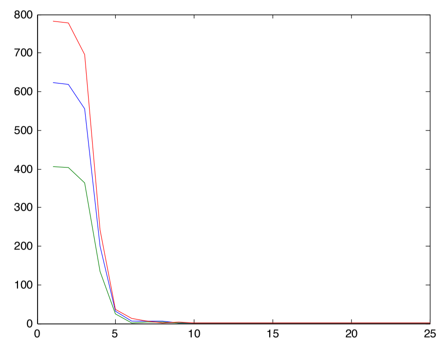
[49] Rudebusch G.W. and Wu T., 2008, ”A macro-finance model of the term structure, monetary policy, and the economy”, Economic Journal, vol 118, , 906-926.

[50] Stanton R. (1997), ”A nonparametric model of term structure dynamics and the market price of interest rate risk”, Journal of Finance vol 52, 1973-2002.

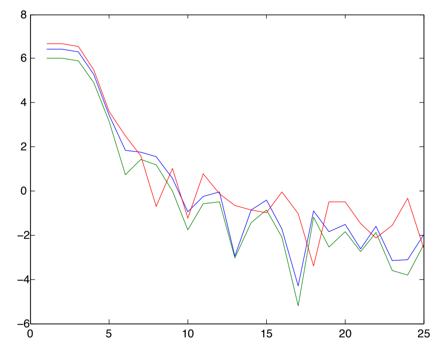
[51] Vasicek O., 1977, ”An equilibrium characterization of the term structure”, Journal of Financial Economics, vol 5, 177-188.

[52] Yared F.B. and Veronesi P., 1999, ”Short and long horizon term and inflation risk premia in the US term structure: Evidence from an integrated model for nominal and real bond prices under regime shifts”, Center for Research in Security Prices (CRSP) Working Paper No 508.

[53] Zaffaroni P., 2004, ”Contemporaneous aggregation of linear dynamic models in large economies”, Journal of Econometrics, vol 120, 75-102.



(a) Normal scale

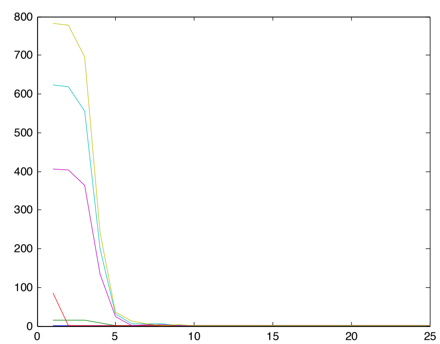


(b) Logarithmic scale

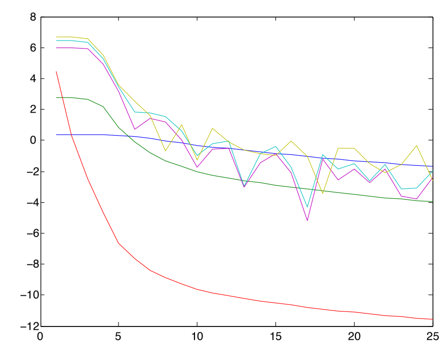
Fig. 1. We plot the periodogram ordinates (normal scale (a) and logarithmic scale (b)) near the zero frequency for yields (blue line), forwards (red line) and returns (green line), averaged across maturity, where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is:

$$I_w(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} w_t e^{\imath\lambda t}\right|^2, \quad -\pi < \lambda \leq \pi.$$

Data are standardized. On the horizontal axis the numbers $1 \leq j \leq 25$ refer to the first 25 frequencies $\lambda_j = 2\pi j/T$.



(a) Periodogram and AR(1) spectral density: normal scale

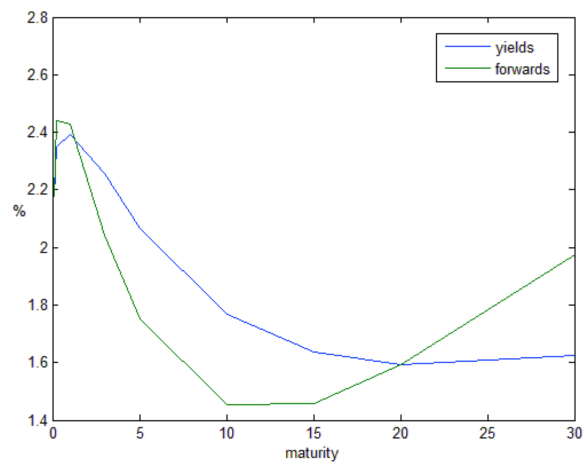


(b) Periodogram and AR(1) spectral density: logarithmic scale

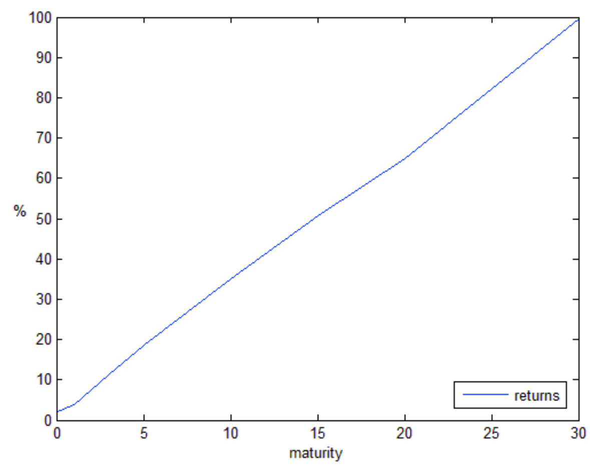
Fig. 2. We plot the periodogram ordinates (normal scale (a) and logarithmic scale (b)) near the zero frequency for nominal yields (yellow line), forwards (light blue line) and returns (purple line), averaged across maturity, where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is $I_w(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} w_t e^{\imath\lambda t} \right|^2, \ -\pi < \lambda \leq \pi$. Data are standardized. We also report the theoretical spectral density (normal scale (a) and logarithmic scale (b)) for an AR(1) process process with unit variance, equal to

$$s_{AR(1)}(\lambda) = \frac{(1-\phi^2)}{2\pi} |1 - \phi e^{\imath\lambda}|^2, -\pi < \lambda \leq \pi,$$

and AR coefficient ϕ equal to 0.80 (blue line), 0.98 (green line) and 0.99999 (red line). On the horizontal axis the numbers $1 \leq j \leq 25$ refer to the first 25 frequencies $\lambda_j = 2\pi j/T$.



(a) Sample standard deviation term structure: yields (blue) and forwards (green).



(b) Sample standard deviation term structure: holding period returns (red)

Fig. 3. We plot the sample standard deviation across maturity for nominal yields and forwards (left panel) and holding period return (right panel), where for a sample of generic observables $(w_1, ...w_T)$ the sample standard deviation is defined as

$$\left(\frac{1}{T}\sum_{t=1}^{T}(w_t - \bar{w})^2\right)^{\frac{1}{2}}$$

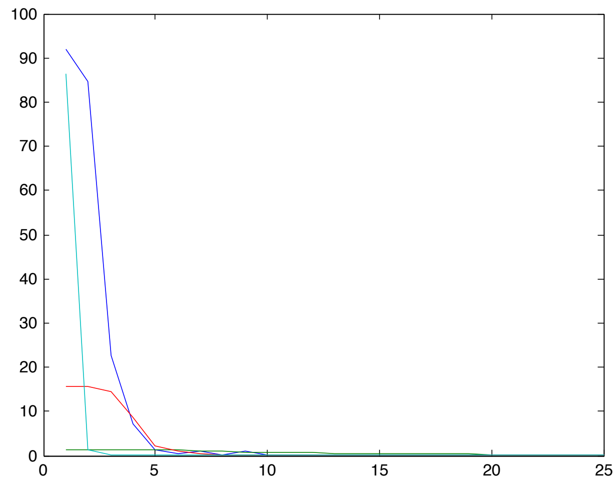


Fig. 4. We plot the periodogram ordinates near the zero frequency for inflation (blue line) where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is:

$$I_w(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} w_t e^{\imath\lambda t}\right|^2, \ \ -\pi < \lambda \leq \pi.$$

Data are standardized. We also report the theoretical spectral density for an AR(1) process process with unit variance, equal to $s_{AR(1)}(\lambda) = \frac{(1-\phi^2)}{2\pi}|1 - \phi e^{\imath\lambda}|^2$ with $-\pi < \lambda \leq \pi$ and AR coefficient ϕ equal to 0.80 (green line), 0.98 (red line) and 0.99999 (light blue line). On the horizontal axis the numbers $1 \leq j \leq 25$ refer to the first 25 frequencies $\lambda_j = 2\pi j/T$.

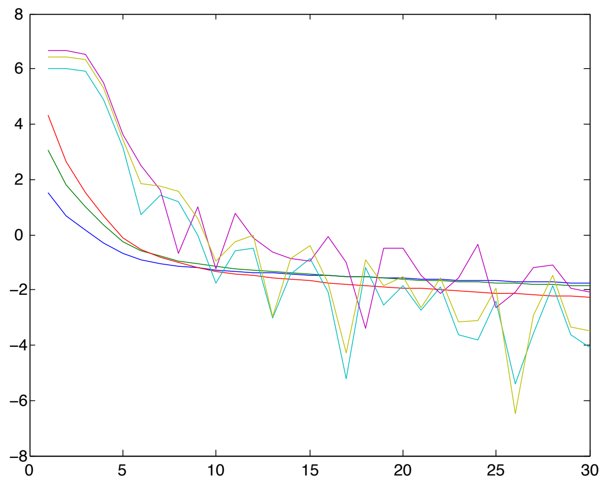
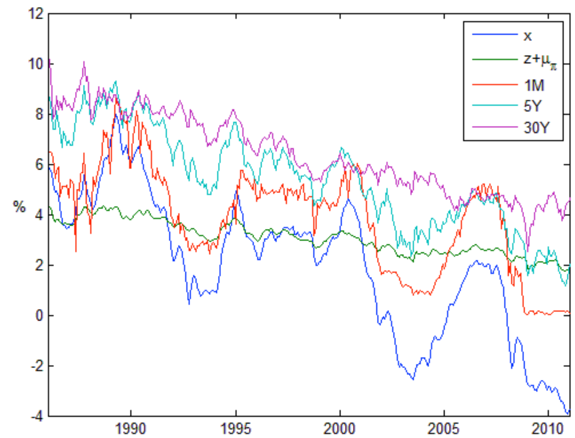


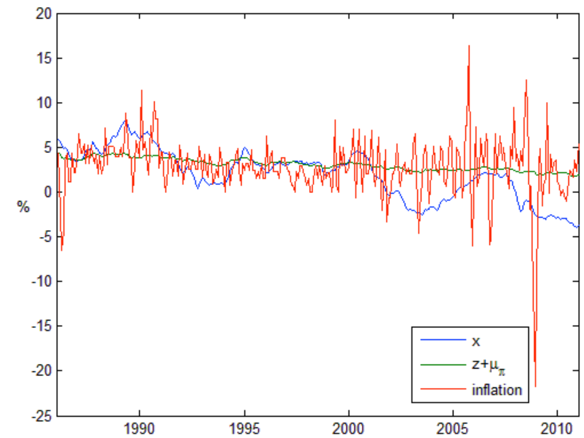
Fig. 5. We plot the logarithm of the periodogram ordinates near the zero frequency for nominal yields (light blue line), forwards (yellow blue line) and returns (purple line), averaged across maturity, where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is $I_w(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} w_t e^{\imath\lambda t}\right|^2, \ -\pi < \lambda \leq \pi$ together with the spectral density

$$s_{LM}(\lambda) = c\lambda^{-2d}, \ -\pi < \lambda \leq \pi,$$

setting $c = \frac{1-2d}{2\pi^{1-2d}}$ to ensure unit variance, with long memory parameter d equal to 0.20 (green line), 0.30 (green line) and 0.40 (red line). Data are standardized. On the horizontal axis the numbers $1 \leq j \leq 25$ refer to the first 25 frequencies $\lambda_j = 2\pi j/T$.



(a) Nominal yields $r^{\$}_{n,t}$ and filtered factors x_t (blue line) and $\mu_\pi + z_t$ (green line).



(b) Realized inflation π_t and factors x_t (blue line) and $\mu_\pi + z_t$ (green line)

Fig. 6. We plot the filtered values of the latent factors x_t (the real factor, blue line) and z_t (the expected inflation factor, green line), where the latter has been centred at the estimate of μ_π. We report the filtered factors together with nominal yields (left panel) and with realized inflation (right panel).

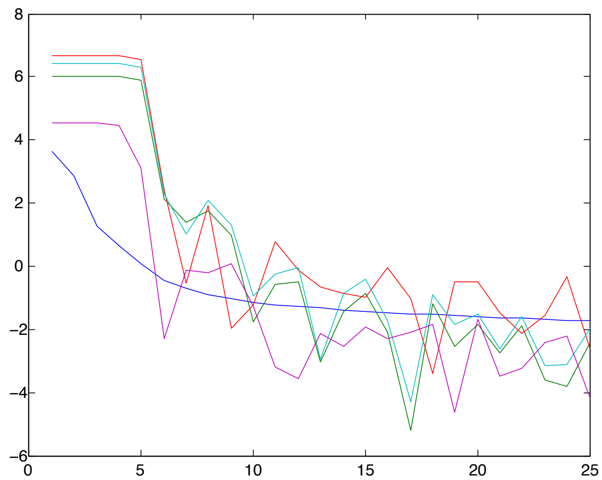


Fig. 7. We plot the logarithm of the periodogram ordinates near the zero frequency for nominal yields (green line), forwards (light blue line) and returns (red line), averaged across maturity, and inflation (purple) where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is $I_w(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} w_t e^{\imath\lambda t}\right|^2, \ -\pi < \lambda \leq \pi$ together with the spectral density (blue line)

$$s_{LM}(\lambda) = c\lambda^{-2d}, \ -\pi < \lambda \leq \pi,$$

setting $c = (1-2d)/2\pi^{1-2d}$ to ensure unit variance, with long memory parameter $d = 0.2862$. On the horizontal axis the numbers $1 \leq j \leq 25$ refer to the first 25 frequencies $\lambda_j = 2\pi j/T$.

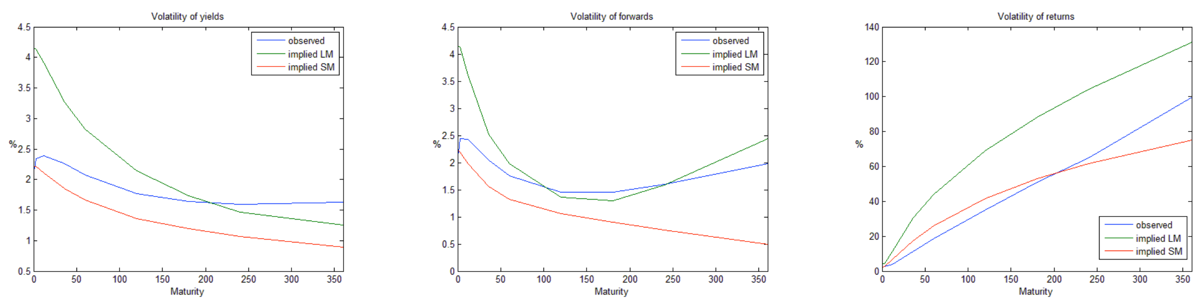
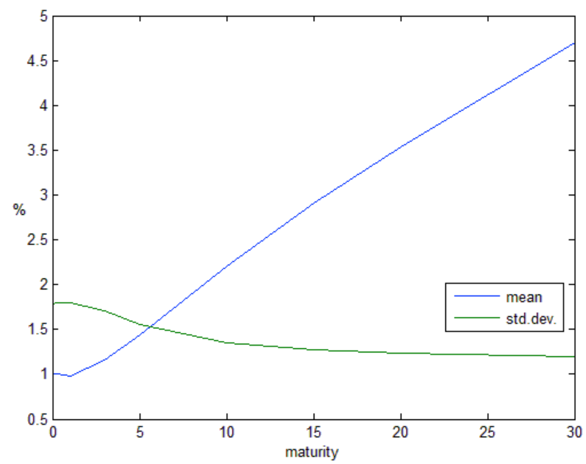
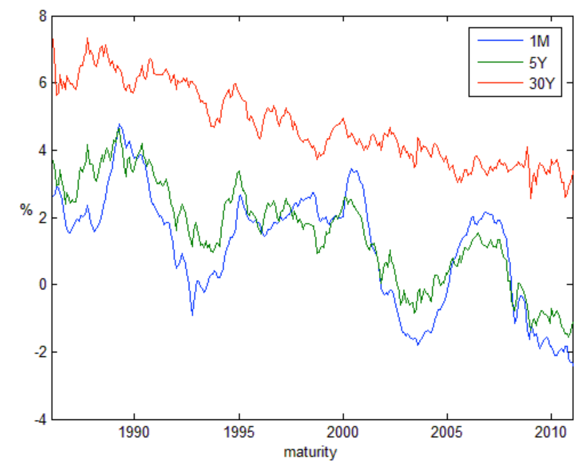


Fig. 8. We report the term structure of the sample standard deviation (blue line) and of the corresponding estimated model-implied standard deviation for the long memory model (green line) and short memory model (red line) for nominal yields (left panel), nominal forward rates (centre panel) and nominal returns (right panel). We used the parameters' values of Table 4 and Table 5 for the long memory and short memory case respectively.



(a) Real yields: term structure



(b) Real yields: time series

Fig. 9. We report the term structures (mean in green, standard deviation in blue) (left panel) and time series (right panel) of the model-implied real yields $r_{n,t}$:

$$r_{n,t} = \widetilde{A}_n + \widetilde{\mathbf{B}}'_{x,n}\mathbf{C}_{x,t} + \widetilde{\mathbf{B}}'_{z,n}\mathbf{C}_{z,t}.$$

The state variables $C_{x,t}, C_{z,t}$ are obtained by means of the Kalman filter and the MLE parameter estimates are plugged into the formulae for $\widetilde{A}_n, \widetilde{\mathbf{B}}_{x,n}, \widetilde{\mathbf{B}}_{z,n}$.

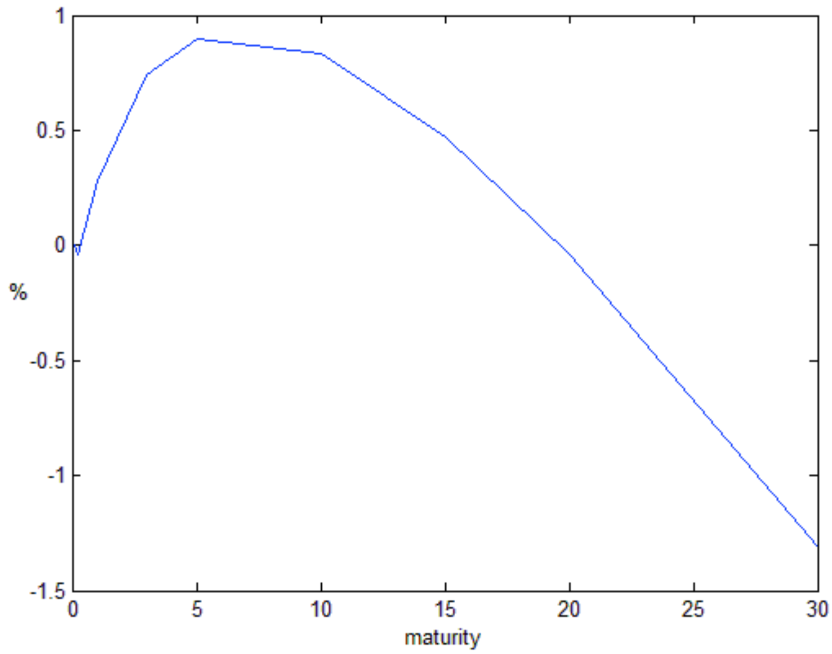


Fig. 10. We report the term structure of the average inflation risk premia $IP_{n,t}$:

$$IP_{n,t} = \frac{1}{n}\text{cov}_t[m_{t+1,t+n} - \pi_{t+1,t+n}, \pi_{t+1,t+n}],$$

obtained as the difference

$$IP_{n,t} = r^{\$}_{n,t} - c_n - r_{n,t} - \frac{1}{n}\text{EI}_{n,t}$$

and $EI_{n,t} = E_t(\log(\Pi_{t+n}/\Pi_t))$ defines the n-period expected inflation and c_n is the Jensen's inequality term.

Panel A: Yields and inflation

Maturity	1 M	3 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y	inflation
Mean	3.7005	3.9914	4.3444	4.8850	5.2703	5.9674	6.2928	6.4018	6.3122	3.6323
Std Dev	2.1791	2.3479	2.3912	2.2554	2.0673	1.7673	1.6356	1.5933	1.6233	4.2072
Skew	-0.1705	-0.1805	-0.2042	-0.2248	-0.1230	0.1071	0.1815	0.1687	0.1306	0.5403
Ex. Kurtosis	2.1902	2.1421	2.1106	2.1679	2.1431	2.0616	2.0349	1.9921	1.9392	4.0762
Min	0.0024	0.0024	0.1423	0.3401	0.8605	1.9843	2.6337	2.9225	2.5021	-21.7214
Max	8.6712	9.1104	9.6212	9.4564	9.2867	9.6419	9.8495	9.9562	10.1602	23.3429

Panel B: Forward rates

Mean	3.7005	4.1706	4.6714	5.5403	6.1731	6.9183	6.8874	6.5497	5.7320
Std Dev	2.1791	2.4416	2.4297	2.0420	1.7523	1.4532	1.4574	1.5948	1.9740
Skew	-0.1705	-0.1675	-0.2511	-0.1030	0.1623	0.3488	0.1547	-0.0119	0.2183
Ex. Kurtosis	2.1902	2.1757	2.1240	2.1204	2.0284	2.1558	2.0374	2.1781	2.3069
Min	0.0024	-0.1416	0.0874	0.9996	2.0859	3.7555	3.4967	2.1183	0.5501
Max	8.6712	10.1328	9.7205	9.5418	9.8925	10.4042	10.5341	10.6395	10.6858

Panel C: 1-month bond returns

Mean	3.7124	4.2295	4.9482	6.4328	7.6614	9.7314	10.9506	11.9248	13.9592
Std Dev	2.1724	2.4818	3.8183	11.2270	18.6264	35.1497	50.7653	64.8588	99.5842
Skew	-0.1732	-0.1230	0.4987	-0.0293	-0.1537	0.0347	0.1477	0.2601	0.4739
Ex. Kurtosis	2.2007	2.2141	3.0876	2.9654	3.0657	4.3073	5.1281	5.5565	5.5001
Min	0.0048	-0.0948	-3.1406	-26.7256	-47.6655	-116.7367	-174.2359	-218.2020	-299.6785
Max	8.6712	10.1616	17.4992	36.3941	60.5417	149.6504	232.1160	321.8883	495.4985

Table 1: Summary statistics for zero coupon monthly yields and inflation. The 1 and 3 month yields come from the Fama's Treasury Bills Term Structure Files, and the 1, 3, and 5 year yields come from the Fama-Bliss Discount Bond Files, 1952:06 - 2011:12, available from CRSP. The 10 to 30 year yields are obtained from Gurkaynak et al (2007), 1986:01-2011:12, available from the website of the Federal Reserve Board. All yields are continuously compounded. Inflation, 1947:01-2011:12, is calculated from the CPI All Urban Consumers.

	1 M	3 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y	inflation
constant	0.0448	0.0117	0.0094	0.0199	0.0271	0.0029	0.0042	0.0085	0.0243	0.5551
t−value	0.8047	0.4239	0.2870	0.4703	0.5473	0.0499	0.0669	0.1374	0.4154	3.1076
x_{t-1}	−0.0189	−0.0061	−0.0064	−0.0080	−0.0090	−0.0040	−0.0039	−0.0043	−0.0068	−0.1645
t−value	−1.4543	−1.0304	−0.9717	−1.0211	−1.0285	−0.4313	−0.4060	−0.4573	−0.7270	−4.1532

Table 2: Augmented Dickey-Fuller test of the 1, 3 month, 1, 3, 5, 10, 15, 20, and 30 year yields and inflation. We estimate the testing equation

$$\Delta x_t = \mu + \gamma x_{t-1} + \sum_{i=1}^{q} \delta_i \Delta x_{t-i} + \epsilon_t$$

where we minimize the Schwartz Information Criterion to determine the lags of Δx_t to be included in the testing regression. The null hypothesis is that there is a unit root: $\gamma = 0$. The 5% significance level for the Dickey-Fuller test with intercept is -2.87.

	1 M	3 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y	inflation
d	0.1680	0.2754	0.2231	0.1728	0.1545	0.1048	0.1094	0.1291	0.1633	0.3474
std.err.	0.1297	0.0613	0.0645	0.0682	0.0786	0.0924	0.1026	0.0928	0.0779	0.0357

Table 3: Estimates of the long memory parameter d from the $ARFIMA(1, d, 1)$ model:

$$(1 - \psi L)(1 - L)^d w_t = (1 + \theta L)\epsilon_t$$

where the observable w_t is equal to $r^{\$}_{n,t}$ for the above maturities and to inflation π_t. We use the Chan and Palma (1998) maximum likelihood estimator (see Appendix D). Standard errors are reported in small font.

ψ_x	θ_x	d_x	ψ_z	θ_z	d_z
0.8992	0.5877	0.2862	0.8698	0.9857	0.1878
0.0010	0.0136	0.0032	0.0014	0.0156	0.0021
μ_r			σ_x	σ_z	σ_π
0.0059			0.0055	0.0003	0.0313
0.0016			0.0001	0.0001	0.0013
λ_{x0}	λ_{z0}			β_{zx}	δ_z
$\times 1000$				$\times 100$	
-0.0295	-2.8880			0.0018	-0.0192
0.0008	0.0414			0.0001	0.0006
λ_{x1}	λ_{x2}	λ_{z1}	λ_{z2}		
	$\times 100000$				
0.0008	0.2673	-0.0291	-2.5723		
0.0001	0.0019	0.0001	0.0463		
σ_1	σ_3	σ_{12}	σ_{36}	σ_{60}	
0.0062	0.0038	0.0030	0.0007	0.0009	
0.0001	0.0002	0.0002	0.0002	0.0002	
σ_{120}	σ_{180}	σ_{240}	σ_{360}		
0.0007	0.0004	0.0011	0.0033		
0.0003	0.0001	0.0004	0.0001		

Table 4: We report the estimates of the long memory model with two factors, with measurement equations

$$\begin{pmatrix} r^{\$}_{n_1,t} \\ r^{\$}_{n_2,t} \\ \vdots \\ r^{\$}_{n_k,t} \\ \pi_t \end{pmatrix} = \begin{pmatrix} \widetilde{A}^{\$}_{n_1} \\ \widetilde{A}^{\$}_{n_2} \\ \vdots \\ \widetilde{A}^{\$}_{n_k} \\ \mu_\pi \end{pmatrix} + \begin{pmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{x,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{x,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{x,n_k} \\ 0 \end{pmatrix} \mathbf{C}_{xt-1} + \begin{pmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{z,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{z,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{z,n_k} \\ \mathbf{G}' \end{pmatrix} \mathbf{C}_{zt-1} + \begin{pmatrix} \epsilon_{n_1,t} \\ \epsilon_{n_2,t} \\ \vdots \\ \epsilon_{n_k,t} \\ \varepsilon_{\pi,t} \end{pmatrix}$$

where $\widetilde{A}^{\$}_n, \widetilde{\mathbf{B}}^{\$}_{x,n}, \widetilde{\mathbf{B}}^{\$}_{z,n}$ are defined in Theorem 4.2, and transition equations

$$\mathbf{C}_{x,t+1} = \mathbf{F}\mathbf{C}_{xt} + \mathbf{h}_x \varepsilon_{x,t+1}, \mathbf{C}_{z,t+1} = \mathbf{F}\mathbf{C}_{zt} + \mathbf{h}_z(\varepsilon_{z,t+1} + \beta_{zx}\varepsilon_{x,t+1}),$$

with $\epsilon_{x,t} \sim NID(0, \sigma_x^2), \epsilon_{z,t} \sim NID(0, \sigma_z^2), \epsilon_{\pi,t} \sim NID(0, \sigma_\pi^2)$ mutally independent and where $\mathbf{F}$ and $\mathbf{h}_x, \mathbf{h}_z$ are defined in (28) and (30) respectively. The factors $x_t = \mathbf{G}'\mathbf{C}_{xt}, z_t = \mathbf{G}'\mathbf{C}_{zt}$, with $\mathbf{G} = (1,\ 0,\ 0...)'$, are ARFIMA$(1, d, 1)$

$$(1 - \psi_z L)(1 - L)^{d_z} z_t = (1 + \theta_z L)\epsilon_{z,t}, \quad (1 - \psi_x L)(1 - L)^{d_x} x_t = (1 + \theta_x L)\epsilon_{x,t}.$$

The innovations to yields satisfy $\epsilon_{n,t} = (\widetilde{\mathbf{B}}^{\$\prime}_{x,n}\mathbf{h}_x + \beta_{zx}\widetilde{\mathbf{B}}^{\$\prime}_{z,n}\mathbf{h}_z)\varepsilon_{x,t} + \widetilde{\mathbf{B}}^{\$\prime}_{z,n}\mathbf{h}_z\varepsilon_{z,t} + \nu_{n,t}$, with $\nu_{n,t} \sim NID(0, \sigma_n^2)$. Robust standard errors are reported in small font. The model is estimated by the approximate maximum likelihood estimator proposed by Chan and Palma (1998) with the truncation lag set to 60. See Appendix D. The sample period is 1986:01 to 2011:12.

ψ_x	θ_x	d_x	ψ_z	θ_z	d_z
0.9759	0.3026	–	0.9637	0.8264	–
0.0037	0.0427		0.0021	0.0282	
μ_r			σ_x	σ_z	σ_π
0.0206			0.0018	0.0008	0.0324
0.0013			0.0002	0.0001	0.0017
λ_{x0}	λ_{z0}			β_{zx}	δ_z
$\times 1000$				$\times 100$	
−0.1830	−0.3329			−0.0034	0.0306
0.0010	0.0424			0.0001	0.0048
λ_{x1}	λ_{x2}	λ_{z1}	λ_{z2}		
	$\times 100000$				
−0.0399	−0.1136	−0.0121	−0.2555		
0.0001	0.0009	0.0011	0.0142		
σ_1	σ_3	σ_{12}	σ_{36}	σ_{60}	
0.0068	0.0042	0.0027	0.0007	0.0007	
0.0004	0.0003	0.0002	0.0004	0.0001	
σ_{120}	σ_{180}	σ_{240}	σ_{360}		
0.0010	0.0009	0.0015	0.0046		
0.0016	0.0014	0.0010	0.0011		

Table 5: We report the estimates of the short memory model with two factors, with measurement equations

$$\begin{pmatrix} r^{\$}_{n_1,t} \\ r^{\$}_{n_2,t} \\ \vdots \\ r^{\$}_{n_k,t} \\ \pi_t \end{pmatrix} = \begin{pmatrix} \widetilde{A}^{\$}_{n_1} \\ \widetilde{A}^{\$}_{n_2} \\ \vdots \\ \widetilde{A}^{\$}_{n_k} \\ \mu_\pi \end{pmatrix} + \begin{pmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{x,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{x,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{x,n_k} \\ 0 \end{pmatrix} \mathbf{C}_{xt-1} + \begin{pmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{z,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{z,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{z,n_k} \\ \mathbf{G}' \end{pmatrix} \mathbf{C}_{zt-1} + \begin{pmatrix} \epsilon_{n_1,t} \\ \epsilon_{n_2,t} \\ \vdots \\ \epsilon_{n_k,t} \\ \varepsilon_{\pi,t} \end{pmatrix}$$

where $\widetilde{A}^{\$}_n, \widetilde{\mathbf{B}}^{\$}_{x,n}, \widetilde{\mathbf{B}}^{\$}_{z,n}$ are defined in Theorem 4.2, and transition equations

$$\mathbf{C}_{x,t+1} = \mathbf{F}\mathbf{C}_{xt} + \mathbf{h}_x \varepsilon_{x,t+1}, \mathbf{C}_{z,t+1} = \mathbf{F}\mathbf{C}_{zt} + \mathbf{h}_z(\varepsilon_{z,t+1} + \beta_{zx}\varepsilon_{x,t+1}),$$

with $\epsilon_{x,t} \sim NID(0, \sigma_x^2), \epsilon_{z,t} \sim NID(0, \sigma_z^2), \epsilon_{\pi,t} \sim NID(0, \sigma_\pi^2)$ mutally independent and where $\mathbf{F}$ and $\mathbf{h}_x, \mathbf{h}_z$ are defined in (28) and (30) respectively. The factors $x_t = \mathbf{G}'\mathbf{C}_{xt}, z_t = \mathbf{G}'\mathbf{C}_{zt}$, with $\mathbf{G} = (1,\ 0,\ 0...)'$, are ARMA$(1,1)$

$$(1 - \psi_x L)x_t = (1 + \theta_x L)\epsilon_{x,t}, \ \ (1 - \psi_z L)z_t = (1 + \theta_z L)\epsilon_{z,t}.$$

The innovations to yields satisfy $\epsilon_{n,t} = (\widetilde{\mathbf{B}}^{\$\prime}_{x,n}\mathbf{h}_x + \beta_{zx}\widetilde{\mathbf{B}}^{\$\prime}_{z,n}\mathbf{h}_z)\varepsilon_{x,t} + \widetilde{\mathbf{B}}^{\$\prime}_{z,n}\mathbf{h}_z\varepsilon_{z,t} + \nu_{n,t}$, with $\nu_{n,t} \sim NID(0, \sigma_n^2)$. Robust standard errors are reported in small font. The model is estimated by the approximate maximum likelihood estimator proposed by Chan and Palma (1998) with the truncation lag set to 60. See Appendix D. The sample period is 1986:01 to 2011:12.

	dependent variable			
	PC_1	PC_2	PC_3	PC_4
regressors				
x	0.9617	0.0263	0.0024	0.0001
z	0.9468	0.0446	0.0001	0.0003
x and z	0.9920	0.9243	0.0112	0.0010

Table 6: We report the regression R^2 from projecting each of the first four principal components, extracted from the set of nominal yields in our sample, on the filtered factor x_t (first row), on the filtered factor z_t (second row) and on both x_t, z_t (third row). The sample period is 1986:01 to 2011:12.

Maturity	1 M	3 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y
Mean	1.1385	1.1288	1.1047	1.2610	1.5438	2.3205	2.9885	3.5793	4.6126
Std Dev	2.2422	2.2539	2.2090	1.9465	1.7139	1.4410	1.3396	1.3443	1.4181
skewness	−0.3969	−0.3935	−0.3959	−0.3458	−0.1970	0.0370	0.1151	0.1208	0.2541
excess Kurtosis	−0.7528	−0.7584	−0.7588	−0.8080	−0.9382	−0.9897	−0.9680	−1.0480	−1.1438
Min	−3.4420	−3.5061	−3.5324	−2.7312	−1.8234	−0.7674	0.0016	0.6025	1.7792
Max	5.7061	5.8617	5.8025	5.1860	4.9165	5.1039	5.7504	6.4154	7.7525
1st lag Autocorr.	0.9579	0.9539	0.9469	0.9458	0.9445	0.9384	0.9336	0.9340	0.9391
12th lag Autocorr.	0.7412	0.7418	0.7507	0.7827	0.8037	0.8186	0.8208	0.8277	0.8403
24th lag Autocorr.	0.3779	0.3860	0.4296	0.5299	0.6062	0.6857	0.7166	0.7370	0.7524

Table 7: Descriptive statistics of the estimated real yields implied by the long memory model:

$$r_{n,t} = A_n + \mathbf{B}'_{n,x}\mathbf{C}_{x,t} + \mathbf{B}'_{n,x}\mathbf{C}_{x,t}.$$

Maximum likelihood parameters' estimates are plugged into $A_n, \mathbf{B}_{n,x}, \mathbf{B}_{n,z}$ and the state variables are obtained from the Kalman filter. Data in percent per annum. The sample period is 1986:01 to 2011:12.

Panel A: value of the log-likelihood function:		
model	LM	SM
$log - likelihood$	13549	13323
Panel B: likelihood ratio test:		
LR		252
$p{-}value$		0

Table 8: Panel A reports the log-likelihood value corresponding to the estimated long memory (LM) and short memory (SM) models, with two latent factors. Panel B reports the likelihood ratio test of the null hypothesis: $H_0\, d_x = d_z = 0$, viz. that the SM is the correct model. The p-value is reported in small font. The test is distributed like a chi-square with two degrees of freedom under H_0.

yield	1 M	3 M	1 Y	3 Y	5 Y
forecast horizon			Panel A − LM specification		
1 M	**24.66**	**23.56**	37.79	57.92	**33.10**
3 M	**55.65**	**59.50**	76.93	91.60	**62.97**
6 M	**101.91**	**106.36**	**121.07**	129.84	**94.83**
1 Y	**180.29**	**183.80**	**188.78**	**176.82**	**127.06**
			Panel B − SM specification		
1 M	26.30	25.45	**37.12**	**48.88**	34.07
3 M	61.91	64.98	**75.41**	**86.01**	67.49
6 M	113.34	116.41	124.34	**129.53**	105.89
1 Y	195.70	197.37	195.82	187.22	151.71

Table 9: The table reports the root mean square error (RMSE) of the the in-sample forecasts of the long memory (LM) and short memory (SM) models, respectively in panel A and B. The RMSE statistics are reported for different forecasting horizons: 1, 3, 6 and 12 months. For calculation of the forecast errors we use the last 10 years of the sample. The forecasts are in-sample because are based on the parameter estimates obtained from the full sample. We start the forecasts in 2000:12. We obtain 120 forecasts for 1 month horizon, 118 forecasts for 3 month horizon, 115 forecasts for 6 month horizon, and 109 forecasts for 12 month horizon. The forecasts are obtained by with the Kalman filter. The smallest RMSE among the LM and SM models is highlighted in bold. The RMSE is reported in percent.

yield	1 M	3 M	1 Y	3 Y	5 Y
forecast horizon	Panel A: LM specification				
1 M	30.88	24.03	47.51	80.02	32.59
3 M	54.05	60.10	88.37	111.76	61.11
6 M	106.00	113.60	139.57	149.62	90.38
1 Y	197.15	202.99	215.11	196.43	118.26
	Panel B: SM specification				
1 M	28.57	24.57	34.95	51.75	36.46
3 M	62.41	67.29	82.13	97.06	77.17
6 M	124.09	129.55	142.73	152.02	126.61
1 Y	223.41	226.94	230.98	227.21	190.72
	Panel C: AP specification				
1 M	51.37	41.69	29.91	40.86	32.00
3 M	105.41	89.21	68.75	74.94	59.10
6 M	149.05	133.08	110.92	105.11	87.22
1 Y	201.37	192.50	167.61	143.34	109.90
	Panel D: ABW specification				
1 M	33.85	45.35	85.32	72.41	30.92
3 M	68.46	82.32	111.57	90.61	54.60
6 M	111.58	121.53	137.27	107.98	74.96
1 Y	162.07	165.73	161.18	117.75	88.08
	Panel E: DL specification				
1 M	35.43	28.19	24.88	35.68	32.76
3 M	63.18	59.98	58.78	70.22	62.24
6 M	100.79	99.73	101.80	108.33	94.04
1 Y	158.57	159.50	158.36	148.88	118.49

Table 10: The table reports the root mean square error (RMSE) of the out-of-sample forecasts of the long memory (LM) model and of the short memory (SM) models, respectively in panel A and B. Panels C,D,E reports the RMSE corresponding to the Ang and Piazzesi (2003) (AP), to the Ang et al (2008) (ABW) and to the Diebold and Li (2006) (DL) models. The out-of-sample forecast errors are calculated with a rolling estimation widow of 607 observations, based on the last 5 years of the sample. In the first step we estimate the model using the period 1952:06 to 2000:12, and using the last recursion of the Kalman filter to make predictions for 4 different horizons. In the second step we estimate the model using the period 1952:07 to 2001:01. We repeat the procedure 120 times. We obtain 120 forecasts for 1 month horizon, 118 forecasts for 3 month horizon, 115 forecasts for 6 month horizon, and 109 forecasts for 12 month horizon. The statistics are reported in percent for different forecasting horizons: 1, 3, 6 and 12 months.