Article:

Golinski, Adam orcid.org/0000-0001-8603-1171 and Zaffaroni, Paolo (2016) Long Memory Affine Term Structure Models. Journal of Econometrics. pp. 33-56. ISSN 0304-4076

https://doi.org/10.1016/j.jeconom.2015.09.006

Long Memory Affine Term Structure Models*

Adam Goliński **Paolo Zaffaroni**

This draft: 6th September 2015

Abstract

We develop a Gaussian discrete time essentially affine term structure model with long memory state variables. 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 in closed-form the dynamic and cross-sectional implications 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. An empirical application of our model is presented.

JEL classification: G12, C58 ,C32.

Key words: Gaussian essentially affine model, long memory, state space, P and Q measures.

*We are grateful to the Editor (Yacine Ait-Sahalia), the Associate Editor and two anonymous referees for very useful suggestions. We are solely responsible for any errors. Goliński: 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 and University of Rome La Sapienza.

1. Introduction

One of the main challenges in modelling the term structure of interest rates is the fact that nominal observed yields are extremely persistent. In fact, they are essentially non-distinguishable from a non-stationary series: any 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 economic 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, forwards 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.

Our long memory model belongs to the class of essentially affine (in the sense of Duffee (2002)) Gaussian term structure models with multiple, possibly latent, factors. We establish the closed-form solution of the general model and, relying on its state space representation, show how to carry out estimation by maximum likelihood and Kalman filtering when latent state variables are allowed for. These achievements are non trivial because an important feature of long memory models is to be non-Markov implying infinite-dimensional state variables.

Our approach shares the many virtues of the powerful class of affine models, pioneered by Vasicek (1977) and Cox et al (1985) highly influential models and formally defined by Duffie and Kan (1996). First, closed-form solution for bond prices and yields can be 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.

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 $DA_M^{\mathbb{Q}}(N)$ of discrete time affine models spelled out by Le et al (2010), where M of the N factors (here $0 \leq M \leq N$) drive the stochastic volatility.[1] Gaussian affine models, whereby the unconditional distribution of the state vectors is normal, feature $M = 0$ (no stochastic

[1] 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.

volatility) and makes the $DA_0^{\mathbb{Q}}(N)$ class. A crucial feature of the $DA_0^{\mathbb{Q}}(N)$ class is that, under the risk-neutral (hereafter $\mathbb{Q}$) measure, the N state variables form a Markov system, possibly of higher yet finite order 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.[2] 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 forwards diminishes fast toward zero as maturity increases. Instead, the theoretical 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 $DA_0^{\mathbb{Q}}(N)$ class, a unit or even an explosive root emerges, of which the consequences are also at odds with the empirical evidence: the theoretical (conditional) volatility of yields and forwards either flattens out (in the unit root case) or increase sharply (in the explosive root case) across maturity. For returns a sharp increase is always obtained.

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. At the same time, the model-driven term structures of volatility, for yields and forwards, can be slowly decaying for intermediate maturities yet flattening out or even (slowly) increasing for long maturities. Instead, the model-driven volatility term structure will diverge for returns. Unlike the Markov case, these implications are now compatible with mean-reversion. More importantly, these are the features observed in the data. As we shall see, long memory can be obtained by allowing the dimension of state vector, N, to become infinite, spanning the $DA_0^{\mathbb{Q}}(\infty)$ 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.

Long memory has been explored by Comte and Renault (1996), who develop a continuous time long memory model of the term structure, and by Duan and Jacobs (1996) where long memory enters through the volatility of the state variables. More recently, Abbritti et al (2015) and Osterrieder (2013) proposed Gaussian term structure models with observed-only state variables whose dynamics follow a suitably restricted vector autoregression with long memory. Latent factors are not permitted in either models which can be seen as different, particular cases, of our general approach.[3] In particular, Abbritti et al (2015) emulate long memory by a long finite-order vector autoregression of order k (with k up to 100) with a long memory parameterization of the coefficients. Theoretical properties of the model are standard since, by all means, it is an affine model with a finite set of observed

[2]Theorem 1 of Chan and Palma (1998) shows this result in the general set up of a linear state space with a finite dimensional state vector.

[3] Given observability of the state variables, both Abbritti et al (2015) and Osterrieder (2013) estimate the model in two separate steps, first the dynamic parameters of the state variables by maximum likelihood-type estimators, and then the market prices of risk parameters by minimizing the squared pricing errors. In contrast, we always use maximum likelihood for both pre-estimation of the dynamic parameters of the observed factors, as well as for estimation of the dynamic parameters of the latent factors, the latent factors themselves and the market of prices parameters.

state variables satisfying a k-th order VAR. The market price of risk dynamics is specified as an affine function of the state variables vector and, due to the long lags specification, depends on $O(k^2)$ parameters, most of which are zeroed for practical estimation. Osterrieder (2013) considers fractional cointegrating restrictions of the dynamics of the state variables, which must exhibit the same degree of long memory, using recent advances in the theory of fractional cointegration. This model provides a genuine long memory specification but the analysis is simplified by the assumption of constant market prices of risk, implying the equivalence of the $\mathbb{Q}$ and $\mathbb{P}$ measures in terms of second moments.

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 such as ours (see Cheridito (2003)). 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 macro variables 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, forwards and returns. This paves the ground for the model presented in Section 4: a multi-factor discrete time essentially affine non-Markov Gaussian term structure model with long memory. Section 4.2 provides in closed-form the analytical characterization of the time series and cross-sectional properties, in terms of volatility term structure, for model-implied yields, forwards and holding period returns, under various forms of the market price of risk. A review of different approaches to tackle the high persistence of observed nominal yields and their analogies with our long memory approach are discussed in Section 5. Section 6 discusses theoretical underpinnings of long memory in real and nominal yields, leaving some formal details to Appendix A. Section 7 presents estimation results for a version of our model that includes realised inflation and real activity as observed state variables. This makes such specifications of our model akin to term structure models that merge yields and macroeconomic data, such as the $DA_0^{\mathbb{Q}}(N)$-type models of Ang and Piazzesi (2003), Rudebush and Wu (2008) and Hordhalh et al (2008) among others.[4] 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

[4]Although not pursued in the current paper, including inflation is instrumental for recovering the canonical decomposition of nominal yields into the term structure of real yields, inflation expectation and inflation risk premia. 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 (2014) and Christensen and Gillan (2012) 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)).

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. We verify in Section 7 that the above described features of the empirical distribution of zero coupon bonds are extremely well matched by the estimated model. Final remarks make Section 8. Appendix A explains the mechanics of 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 technical description of the Kalman filter and an approximate maximum likelihood estimator for long memory processes is relegated to Appendix C. Appendix D contains some technical lemmas and the proofs of the main theorems.

2. Some stylised facts of nominal bonds

We now highlight the well established strong degree of dynamic persistence that characterises certain specific aspects of the empirical distribution of nominal bonds. We consider the term structure of nominal yields, forwards and holding period returns. 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 forwards 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, when looking more carefully, both the time series and the cross-sectional evidence appears at odds with the quantitative predictions of such term structure model built around both stationary and non-stationary Markov state variables.

We use a data set comprised of monthly observations of nominal yields $y^{\$}_{n,t}$ on zero coupon bonds with maturities n equal to 1 month and $1, 3, 5, 10, 15, 20$ and 30 years. The 1-month yield comes from the Fama's Treasury Bills Term Structure files while, for all other maturities, yields are extracted from the data of Gurkaynak et al (2007). We consider the period from January 1986 to December 2011.[5] Yields

$$y^{\$}_{n_i,t} = -\frac{1}{n_i} log P^{\$}_{n_i t}$$

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

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

and holding period returns

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

[5] Gurkaynak et al (2007) report the 30-year yield curve estimates since 25 November 1985, which is approximately when the US Treasury first started issuing 30-year bonds. This determined our choice of the beginning of our sample period.

referring to them as $f^{\$}_{n_i,t}$ and $r^{\$}_{n_i,t}$ in the standard, monthly, case when $n_{i+1} - n_i = 1$. The latter (monthly figure) is the case considered in our empirical analysis. 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, decreases and then slightly increases again. A similar pattern is obtained in terms of forwards, the main difference being that for forwards their volatility term structure raises even more for long maturities 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.[6] This is evident when performing unit root tests, illustrated in Table 2, where we present the results for the standard Augmented Dickey-Fuller (ADF) unit root test. The null hypothesis of a unit root is not rejected for nominal yields across all maturities.

[**Insert Table 2 near here**]

We propose to assess the persistence of nominal bonds 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 non-stationarity from the one of mean-reversion.

Figure D and Figure D plot the periodogram ordinates near the zero frequency, respectively, for yields (blue line) and forwards (green line) averaged across maturity,[7] and for returns at 1- (blue line), 10- (green line) and 30-year maturity (red line), where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is

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

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

[**Insert Figure 1 near here**]

The strength of using the periodogram comes essentially from the fact that it is a nonparametric measure, that uses the information of the entire string of sample autocorrelations of the data.[8] More generally, 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 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,

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

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

[8] In contrast, if one estimates an AR(1) model on the data, the autoregressive parameter is only made by the sample first-order autocorrelation coefficient.

as such, is a clearcut measure of long run persistence.[9] Instead, the local behaviour of the periodogram, as one moves away from the zero frequency, provides indication on the degree of mean-reversion. Finally, cycles induce local peaks at the corresponding frequencies within the interval $[-\pi, \pi]$.

To provide a benchmark, any stationary autoregressive moving average (ARMA) process implies a bounded spectral density, flattening out near the zero frequency. We plot the spectral density for AR(1) model with unit variance with autoregressive parameter equal to 0.90 (purple line) and 0.99 (light blue line), in both Figures D and D together with the periodogram of yields, forwards and returns.

The results are strikingly clear. Yields and forwards are very persistent, much more than an AR(1) model with coefficient equal to 0.99. A similar feature, although with a smaller discrepancy, applies to holding period returns with 1-year maturity. In contrast, as the maturity lengthens, returns appear much less persistent. Indeed the persistence diminishes (monotonically) as the maturity increases: the 10-year return appears approximately described by an AR model with a positive autoregressive coefficient whereas the 30-year return appears no persistent at all. This comparison is compelling: even a value of the AR coefficient as large as 0.99 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. On the other hand, an AR(1) model with a coefficient so close to unity, would be hard to be distinguished from a unit root using any conventional unit root test. The problem with the unit root paradigm is that it does 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, as discussed below.

We summarize this finding as follows.

Stylized Fact 1. *Nominal yields and forwards, at all maturities, and nominal holding period returns, for small maturities, are highly persistent and mean reverting across time: their periodogram displays a peak near the zero frequency and quickly diminishes when away from the zero frequency. The persistence of nominal holding period returns diminishes substantially with maturity.*

As an alternative, more precise, characterization of persistence found in the data, one can state that nominal yields, forwards and holding period returns (for short maturities) have an approximately linear negatively sloped log periodogram near the origin, slowly decaying as the frequency increases. Anticipating matters explained subsequently, if the data are characterized by a unit root, such negative slope would be approximately minus two. If instead the data were generated by a stationary ARMA this slope would be zero. In practice, careful examination of the data reveals a slope smaller than zero. We will explain these concepts below once we articulate the notion of long memory.

Figure D displays the term structure of the sample standard deviations of yields (blue line), forwards (green line) and returns (red line). As observed in Table 1, for yields the curve

[9] 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).

initially increases up to 1-year maturity, then it decays and, finally, it slightly increases for longer maturities. Forward rates have a similar pattern, although they show a more substantial increase toward the end of the term structure. Clearly for both yields and forwards the volatility is not vanishing at long maturity. Instead, the term structure of the sample standard deviation of holding period returns raises steeply with maturity. If non-stationarity is suspected, one can instead consider the term structure of the sample standard deviations of the *first difference* of yields, forwards and returns. The results are presented in Figure D, where for the first difference of yields and forwards their sample standard deviation is multiplied by 10, to better see the pattern. It turns out that the term structures of volatility of first-differenced yields, forwards and returns show essentially the same pattern (although with a different scale) as for the raw quantities, with now (first-differenced) forwards exhibiting a more clear raise at long maturities, as opposed to yields.

[**Insert Figure 2 near here**]

These observations lead to:

Stylized Fact 2. *The term structure of the sample standard deviation of nominal yields and forwards increases at short maturities with hump at around* 1*-year. They then decrease but eventually slowly increase for very long maturities. The term structure of the sample standard deviation of nominal holding period returns rises sharply 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 both features.

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. For the sake of expository purpose, consider the discrete time version of the Vasicek (1977) model, a one-factor Gaussian model, here applied to nominal yields. The price of a nominal zero coupon bond issued at time t which expires n periods ahead satisfies the no-arbitrage condition

$$P^{\$}_{n,t} = E_t\left(e^{m^{\$}_{t+1}} P^{\$}_{n-1,t+1}\right), \tag{1}$$

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 no-arbitrage implies existence of the (nominal) pricing kernel $e^{m^{\$}_{t+1}}$, the exponent of which, for this model, has the simple form

$$-m^{\$}_{t+1} = \delta_0 + \frac{1}{2}\lambda_t^2\sigma_x^2 + x_t + \lambda_t \varepsilon_{x,t+1} \tag{2}$$

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

$$x_t = \psi_x x_{t-1} + \varepsilon_{x,t}, \quad \text{where the } \varepsilon_{x,t} \text{ are } NID(0, \sigma_x^2), \tag{3}$$

and the market price of risk is affine in the factor:

$$\lambda_t = \lambda_0 + \lambda_1 x_t, \tag{4}$$

with $\lambda_0, \lambda_1, \delta_0, \psi_x, \sigma_x^2$ constant parameters. Stationarity of x_t requires $\mid \psi_x \mid < 1$.

By the standard recursive method one obtains that bond yields $y^{\$}_{n,t}$, forward rates $f^{\$}_{n,t}$ and holding one-period returns $r^{\$}_{n,t}$ satisfy, respectively,

$$y^{\$}_{n,t} = n^{-1}(A^{\$}_n + B^{\$}_n x_t), \tag{5}$$

$$f^{\$}_{n,t} = A^{\$}_{n+1} - A^{\$}_n + (B^{\$}_{n+1} - B^{\$}_n) x_t, \tag{6}$$

$$r^{\$}_{n,t} = A^{\$}_n - A^{\$}_{n-1} + B^{\$}_n x_{t-1} - B^{\$}_{n-1} x_t, \tag{7}$$

where, in turn, the affine function coefficients satisfy the well-established Riccati difference equations:

$$A^{\$}_n = A^{\$}_{n-1} + \delta_0 - \lambda_0 \sigma_x^2 B^{\$}_{n-1} - \frac{1}{2}(B^{\$}_{n-1})^2 \sigma_x^2 \quad \text{and} \quad B^{\$}_n = 1 + \psi_x^{\mathbb{Q}} B^{\$}_{n-1}, \tag{8}$$

with initial conditions $A^{\$}_0 = B^{\$}_0 = 0$, where we define the $\mathbb{Q}$-measure autoregressive coefficient $\psi_x^{\mathbb{Q}} = \psi_x - \lambda_1 \sigma_x^2$.

Consider first the stationary case $|\psi_x| < 1$.[10] Clearly yields, forwards and returns are affine transformations 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, forwards and returns are,[11] for $-\pi \le \lambda < \pi$,

$$s_{y_n}(\lambda) = (\frac{B^{\$}_n}{n})^2 \, s_x(\lambda),$$

$$s_{f_n}(\lambda) = (B^{\$}_n - B^{\$}_{n-1})^2 \, s_x(\lambda),$$

$$s_{r_n}(\lambda) = (1 - \lambda_1 \sigma_x^2)^2 s_x(\lambda) + (B^{\$}_{n-1})^2 \frac{\sigma_x^2}{2\pi} - 2(1 - \lambda_1 \sigma_x^2) B^{\$}_{n-1} \psi_x^{-1} \frac{\sigma_x^2}{2\pi} \Re \left(\frac{1}{1 - \psi_x e^{\imath\lambda}} - 1 \right),$$

where $\Re(.)$ denotes the real part of a complex number and $s_x(\lambda)$ indicates the spectral density of the AR(1) state variable (3), equal to $\sigma_x^2 / (2\pi |1 - \psi_x e^{\imath\lambda}|^2)$. By easy derivations, the slope

[10] Stationarity of yields, forwards and returns is driven by the physical measure autoregressive coefficients ψ_x. The $\mathbb{Q}$-measure autoregressive coefficient $\psi_x^{\mathbb{Q}}$ only enters into the construction of the loadings $A^{\$}_n, B^{\$}_n$ and, in particular, determines the behaviour of the variances for large n.

[11] For returns $r^{\$}_{n,t}$ the additional terms in the spectral density are due to the fact that $r^{\$}_{n,t}$ can be represented as $A^{\$}_n - A^{\$}_{n-1} + (1 - \lambda_1 \sigma_x^2) x_{t-1} - B^{\$}_{n-1} \varepsilon_{x,t}$. However the behaviour of the first and third term in $s_{r_n}(\lambda)$ are identical near the zero frequency.

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

$$\begin{aligned}
\frac{dlog\, s_{y_n}(\lambda)}{d\log\lambda} &\sim -\frac{2\psi_x}{(1-\psi_x)^2}\lambda^2,\\
\frac{dlog\, s_{f_n}(\lambda)}{d\log\lambda} &\sim -\frac{2\psi_x}{(1-\psi_x)^2}\lambda^2,\\
\frac{dlog\, s_{r_n}(\lambda)}{d\log\lambda} &\sim -2\left(\frac{(1-\lambda_1\sigma_x^2)^2}{(1-\psi_x)^2} - B_{n-1}^{\$}\right)^{-2}\left(\frac{(1-\lambda_1\sigma_x^2)^2\psi_x}{(1-\psi_x)^2} - \frac{B_{n-1}^{\$}(1-\lambda_1\sigma_x^2)}{(1-\psi_x)^2}\right)\lambda^2,
\end{aligned}$$

where $\sim$ indicates asymptotic equivalence.[12] 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.

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

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

where $\mathrm{var}_t(\cdot)$ is the physical measure variance operator conditional on the information available up to time t. Notice that the behaviour of the variances across maturity n is driven by the $\mathbb{Q}$-measure autoregressive coefficient $\psi_x^{\mathbb{Q}}$. Since $B_n^{\$} \sim 1/(1-\psi_x^{\mathbb{Q}})$ for large n when $|\psi_x^{\mathbb{Q}}| < 1$, it follows that, as $n \to \infty$,

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

An identical pattern is obtained for the unconditional variances. Finally, consider now the case $|\psi_x^{\mathbb{Q}}| \geq 1$. Obviously this does not imply that the model is truly non-stationary under the physical measure since stationarity depends on ψ_x. The term structure of conditional volatility for yields, forwards and returns will now be $\mathrm{var}_{t-1}(y_{t,n}^{\$}) = \sigma_x^2$, $\mathrm{var}_{t-1}(f_{t,n}^{\$}) = \sigma_x^2$, $\mathrm{var}_{t-1}(r_{t,n}^{\$}) = n^2\sigma_x^2$ when $\psi_x^{\mathbb{Q}} = 1$ whereas $\mathrm{var}_{t-1}(y_{t,n}^{\$}) \sim \left(\frac{\sigma_x^2}{(\psi_x^{\mathbb{Q}})^2-1}\right) n^{-1}(\psi_x^{\mathbb{Q}})^{2n}$, $\mathrm{var}_{t-1}(f_{t,n}^{\$}) \sim (\psi_x^{\mathbb{Q}})^{2n}\sigma_x^2$, $\mathrm{var}_{t-1}(r_{t,n}^{\$}) \sim \left(\frac{\sigma_x^2}{(\psi_x^{\mathbb{Q}})^2-1}\right)(\psi_x^{\mathbb{Q}})^{2n}$ when $\psi_x^{\mathbb{Q}} > 1$. Therefore, the volatility curves of yields and forwards will either decay very quickly (towards zero) or diverge very quickly for large maturities, depending on whether $\psi_x^{\mathbb{Q}}$ is smaller or bigger than unity. The model is purposely extremely stylized, but it shares the main implications in terms of persistence and 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 odds with the empirical evidence surveyed in Section 2. The stationary case generates a stronger than needed degree of mean-reversion

[12] We say that $a_n \sim b_n$, where $b_n \neq 0$, when $a_n/b_n \to 1$ as $n \to \infty$.

whereas non-stationarity, either a unit or explosive root, rules out mean-reversion altogether. Moreover, postulating a unit root invalidates 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 stationarity. 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 can long memory models 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.[13] 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 C. 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 forecasts. Moreover, the model can accomodate latent factors, which in turn can be optimally recovered by the Kalman filter. The possibility of allowing latency of the factors is particularly important for the purpose of term structure modelling, since it opens up the possibility of estimating the real term structure, the expected inflation term structure and the associated risk premia.

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 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. General multi-factor model

In this section we will refer to nominal yields for illustrative purposes but the model will apply to either nominal and real yields. The interpretation of the results would clearly differ. However, adopting some slight modifications, the model can also be used to decompose nominal yields in real yields, inflation expectations and inflation risk premia based on the same data set.

Assume that the (nominal) short rate is driven by K factors $\mathbf{x}_t = (x_{1,t}, ..., x_{K,t})'$:

$$y_{t,1}^{\$} = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t, \tag{10}$$

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

with coefficients $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_K)'$ and intercept δ_0. Some, none or all the factors could be latent, although this is not relevant in terms of representation. We then assume that the (nominal) stochastic discount factor $m_t^{\$}$ is a quadratic function of the K factors:

$$-m_{t+1}^{\$} = y_{t,1}^{\$} + \frac{1}{2}\boldsymbol{\lambda}_t' \boldsymbol{\Sigma} \boldsymbol{\lambda}_t + \boldsymbol{\lambda}_t' \boldsymbol{\varepsilon}_{t+1} \tag{11}$$

through the market prices of risk $\boldsymbol{\lambda}_t$, which are affine in the state variables

$$\boldsymbol{\lambda}_t = \begin{pmatrix} \lambda_{1,t} \\ \vdots \\ \lambda_{K,t} \end{pmatrix} = \boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_1 \mathbf{x}_t \tag{12}$$

for a $K \times 1$ vector $\boldsymbol{\lambda}_0$ and a $K \times K$ matrix $\boldsymbol{\lambda}_1 = (\lambda_{1.1} ... \lambda_{1.K})$ with jth column $\boldsymbol{\lambda}_{1.j}$. Formulation (12) qualifies the model as 'essentially' affine. The vector of innovations $\boldsymbol{\varepsilon}_t$ is assumed i.i.d. with

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{K,t} \end{pmatrix} \sim NID\left(\mathbf{0}, \boldsymbol{\Sigma}\right). \tag{13}$$

Expression (11) for the pricing kernel is implied by the existence of a conditionally log-normal stochastic process $\alpha_t = \alpha_{t-1} \exp(-0.5\boldsymbol{\lambda}_{t-1}' \boldsymbol{\Sigma} \boldsymbol{\lambda}_{t-1} - \boldsymbol{\lambda}_{t-1}' \boldsymbol{\varepsilon}_t)$ 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 $\mathbb{Q}$ measure (see Harrison and Kreps (1979)). Hereafter, we shall specify all model equations and parameters in terms of the physical measure, unless stated otherwise. Since the price of any (nominal) asset that does not pay dividends is a martingale under $\mathbb{Q}$ (once adjusted by $e^{-y_t^{\$}}$), for nominal zero coupon bond prices $P_{n,t}^{\$} = E_t^{\mathbb{Q}}[e^{-y_t^{\$}} P_{n-1,t+1}^{\$}] = E_t[e^{-y_t^{\$}} P_{n-1,t+1}^{\$} \alpha_{t+1}/\alpha_t] = E_t[P_{n-1,t+1}^{\$} e^{m_{t+1}^{\$}}]$.

To close the model one needs to specify the dynamics of the factors (under the physical measure). In order to introduce long memory, we need to make a distinction between factors and state variables. The dynamics of each factor $\mathbf{x}_{j,t}$, with $1 \le j \le K$, is more conveniently represented by the infinite-dimensional state vectors $\mathbf{C}_{j,t}$ which obey an infinite-dimensional VAR(1) model

$$\mathbf{C}_{j,t+1} = \mathbf{F}\mathbf{C}_{j,t} + \mathbf{h}_j \varepsilon_{j,t+1}, \qquad 1 \le j \le K, \tag{14}$$

for an infinite-dimensional vector $\mathbf{h}_j$ and a double-infinite dimensional matrix $\mathbf{F}$. Notice that the innovations in (14) are the same as in (13) and are, therefore, contemporaneously correlated through $\boldsymbol{\Sigma}$. This implies that each state vector $\mathbf{C}_{j,t}$ is influenced by the overall set of lagged state vectors $\mathbf{C}_{1,s}, \dots, \mathbf{C}_{K,s}$ for every $s \le t$. Equations (14) represent the transition equations of the state-space of the model used for the Kalman filter recursion. Obviously we could have written the K equations jointly as $\mathbf{C}_{t+1} = \mathbf{F}^*\mathbf{C}_t + \mathbf{h}^* \varepsilon_{t+1}$ for certain matrices $\mathbf{F}^*$ and $\mathbf{h}^*$ suitably restricted, but it is more convenient to rely on (14). The relationship between factors and state variables is simply

$$x_{j,t} = \mathbf{G}'\mathbf{C}_{j,t}, \;\; 1 \le j \le K, \tag{15}$$

for an infinite dimensional vector $\mathbf{G} = (1, 0, 0 \cdots)'$ with all zeros from the second row onwards. Recall that the elements of $\boldsymbol{\varepsilon}_t$ are in general cross-correlated (unless $\boldsymbol{\Sigma}$ is diagonal) and thus the $\mathbf{C}_{j,t}$ are not independent across j.

Despite the infinite dimension of the state variables, the model can be solved along 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 (12) and the state variable dynamics (14), the no-arbitrage zero coupon bond prices $P^{\$}_{n,t}$ satisfy, by Gaussianity (13),*

$$p^{\$}_{n,t} = -A^{\$}_n - \mathbf{B}^{\$\prime}_{1,n}\mathbf{C}_{1,t} - .... - \mathbf{B}^{\$\prime}_{K,n}\mathbf{C}_{K,t},$$

where $p^{\$}_{n,t} = \ln P^{\$}_{n,t}$ and the coefficients satisfy the Riccati recursions

$$A^{\$}_n = A^{\$}_{n-1} + \delta_0 - \frac{1}{2}\mathbf{B}^{\$\prime}_{n-1}\boldsymbol{\Sigma}\mathbf{B}^{\$}_{n-1} - \mathbf{B}^{\$\prime}_{n-1}\boldsymbol{\Sigma}\boldsymbol{\lambda}_0, \quad (16)$$

$$\mathbf{B}^{\$}_{j,n} = \left(\delta_j - \mathbf{B}^{\$\prime}_{n-1}\boldsymbol{\Sigma}\boldsymbol{\lambda}_{1.j}\right)\mathbf{G} + \mathbf{F}'\mathbf{B}^{\$}_{j,n-1}, \quad 1 \le j \le K. \quad (17)$$

setting the K dimensional vector

$$\mathbf{B}^{\$}_n = (\mathbf{B}^{\$\prime}_{1,n}\mathbf{h}_1, ..., \mathbf{B}^{\$\prime}_{K,n}\mathbf{h}_K)'.$$

Note that $A^{\$}_n$ is scalar, $\mathbf{B}^{\$}_n$ is K dimensional and the $\mathbf{B}^{\$}_{j,n}$ are infinite-dimensional vectors for every $1 \le j \le K$. These coefficients must be interpreted as being evaluated under the $\mathbb{Q}$-measure unless in (12) one sets $\boldsymbol{\lambda}_0 = \mathbf{0}$ for $A^{\$}_n$ or $\boldsymbol{\lambda}_1 = \mathbf{0}$ for the $\mathbf{B}^{\$}_{j,n}$. For these special 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. As explained below, the dynamic properties of the latent factors $x_{j,t}$ depend on the chosen parameterization for $\mathbf{h}_j$ which, in turn, determines the degree of persistence and mean-reversion of the model.

Nominal yields would then be obtained as

$$y^{\$}_{n,t} = -n^{-1}p^{\$}_{n,t} = n^{-1}A^{\$}_n + n^{-1}\mathbf{B}^{\$\prime}_{1,n}\mathbf{C}_{1,t} + ... + n^{-1}\mathbf{B}^{\$\prime}_{K,n}\mathbf{C}_{K,t}, \quad (18)$$

The short term interest rate (10) will then be equal to the one-period (nominal) yield $y^{\$}_{1,t}$, obtained when $A^{\$}_0 = 0$ and $\mathbf{B}^{\$}_{j,n} = \mathbf{0}$ for all $1 \le j \le K$. Nominal forward rates and holding period returns are given by

$$f^{\$}_{n,t} = A^{\$}_{n+1} - A^{\$}_n + (\mathbf{B}^{\$}_{1,n+1} - \mathbf{B}^{\$}_{1,n})'\mathbf{C}_{1,t} + ... + (\mathbf{B}^{\$}_{K,n+1} - \mathbf{B}^{\$}_{K,n})'\mathbf{C}_{K,t}, \quad (19)$$

$$r^{\$}_{n,t} = A^{\$}_n - A^{\$}_{n-1} + (\mathbf{B}^{\$\prime}_{1,n}\mathbf{C}_{1,t-1} - \mathbf{B}^{\$\prime}_{1,n-1}\mathbf{C}_{1,t}) + ... + (\mathbf{B}^{\$\prime}_{K,n}\mathbf{C}_{K,t-1} - \mathbf{B}^{\$\prime}_{K,n-1}\mathbf{C}_{K,t}). \quad (20)$$

4.2. Persistence characterization

The solution of the model is obtained without the need to specify whether the state variables are stationary or not, let alone when long memory is assumed or not. Indeed, only conditional Gaussianity 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 the vectors $\mathbf{h}_j$ which define both the degree of memory, and possibly of stationarity, of the model factors $x_{j,t}$ through (14). These choices define the time series and cross-sectional properties of yields, forwards 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 C)

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & & \cdots \\ 0 & 0 & 1 & 0 & \\ \vdots & \vdots & & \ddots & \ddots \end{bmatrix}, \tag{21}$$

By Gaussianity the factors can be expressed as linear processes in the i.i.d. innovations $\varepsilon_{j,t}$ of (13):

$$x_{j,t} = \sum_{i=0}^{\infty} \phi_{j,i}\varepsilon_{j,t-i}, \;\; 1 \leq j \leq K. \tag{22}$$

Recall that the $\varepsilon_{j,t-i}$ are contemporaneously cross-correlated though the covariance matrix $\mathbf{\Sigma}$. Stacking together the coefficients $\phi_{j,i}$ gives the infinite dimensional vector

$$\mathbf{h}_j = (1 \;\; \phi_{j,1} \;\; \phi_{j,2} \;\; \phi_{j,3} ...)', \qquad 1 \leq j \leq K. \tag{23}$$

Stationarity of factor $x_{j,t}$ follows if

$$\sum_{i=0}^{\infty} \phi_{j,i}^2 < \infty. \tag{24}$$

As explained below, the stationarity condition (24) includes a wide range of possibilities in terms of the degree of persistence, in turn expressed by the rate at which the coefficients $\phi_{j,i}$ go to zero. We briefly summarise such possibilities including the case when the stationarity condition (24) is violated. Given (22), factor $x_{j,t}$ will be defined short memory if

$$\sum_{i=0}^{\infty} \mid \phi_{j,i} \mid < \infty. \tag{25}$$

Alternatively, factor $x_{j,t}$ is said to be long memory if

$$\sum_{i=0}^{s} \mid \phi_{j,i} \mid \to \infty \text{ as } s \to \infty. \tag{26}$$

Note that short memory (25) implies stationarity (24) since summability is stronger than square summability. However, long memory (26) does not necessarily imply 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 (24) is violated and yet

$$\phi_{j,i} \to 0 \text{ as } i \to \infty, \tag{27}$$

and the case when even mean-reversion (27) does not occur. A simple example of this last, extreme, circumstance is given by the basic model of Section 3 when factor $x_t = x_{1,t}$ is a random walk, namely $\phi_{1,i} = 1$ for all i.

4.2.1. Short memory

We now check that the simple model of Section 3 is nested within the general solution of Section 4.1 . Set $K = 1$ and $x_t = x_{1,t}$, $\varepsilon_t = \varepsilon_{1,t}$ with $\delta_1 = 1$. Now the infinite dimensional vector (23) equals

$$\mathbf{h}_1 = (1 \;\; \psi_x \;\; \psi_x^2 \;\; \psi_x^3 ...)', \tag{28}$$

where ψ_x is the autoregressive parameter in (3). By standard arguments model (3) can be re-written as

$$x_{1,t} = \sum_{i=0}^{\infty} \psi_x^i \varepsilon_{1,t-i}, \tag{29}$$

implying that, obviously, the AR(1) satisfies the linearity assumption (22) with coefficients $\phi_{1,i} = \psi_x^i$. When $|\psi_x| < 1$ then the short memory condition (25) 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 (27) fails.

One just needs to find the scalar sequence A_n and the infinite dimensional sequences $\mathbf{B}_{1,n}^{\$}$, solution of the recurrence equations (16)-(17), and verify that indeed the basic affine model (5) is re-obtained. By (21), recursion (17) becomes

$$\mathbf{B}_{1,n}^{\$} = (1 - B_{n-1}^{\$} \kappa_1)\mathbf{G} + \mathbf{F}'\mathbf{B}_{1,n-1}^{\$}$$

setting $\kappa_1 = \Sigma\lambda_1$, $B_n^{\$} = \mathbf{B}_{1,n}^{\$\prime}\mathbf{h}_1$, with initial condition $\mathbf{B}_{1,0}^{\$} = \mathbf{0}$ yielding $\mathbf{B}_{1,n} = (b_n b_{n-1} \ldots b_1 0 \ldots)'$ where

$$\mathbf{B}_{1,1}^{\$} = (1, 0, \ldots)', \mathbf{B}_{1,2}^{\$} = (1 - \kappa_1, 1, 0, \ldots)', \mathbf{B}_{1,3}^{\$} = (1 - \kappa_1(\psi_x + 1 - \kappa_1), 1 - \kappa_1, 0, \ldots)', \ldots \tag{30}$$

A few algebraic steps give $\mathbf{B}_{1,n}^{\$\prime}\mathbf{h}_1 = 1 + \psi_x^{\mathbb{Q}} + ... + (\psi_x^{\mathbb{Q}})^{n-1} = (1 - (\psi_x^{\mathbb{Q}})^n)/(1 - \psi_x^{\mathbb{Q}})$ for every n $\geq$ 1 which in turn gives $A_n^{\$} = A_{n-1}^{\$} + \delta_0 - \lambda_0 \sigma_x^2 B_{n-1}^{\$} - \frac{1}{2}\sigma_x^2 (B_{n-1}^{\$})^2$, which coincides exactly with (8). Notice that $\mathbf{C}_{1,t}$ can be expressed as

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

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

$$\mathbf{B}_{1,n}^{\$\prime}\mathbf{C}_{1,t} = \sum_{i=0}^{n-1} b_{n-i} E_t(x_{1,t+i}) = \sum_{i=0}^{n-1} b_{n-i} \left(\sum_{j=i}^{\infty} \psi_x^j \varepsilon_{1,t+i-j} \right) = \sum_{i=0}^{n-1} b_{n-i} \psi_x^i \left(\sum_{j=i}^{\infty} \psi_x^{j-i} \varepsilon_{1,t+i-j} \right) = \mathbf{B}_{1,n}^{\$\prime}\mathbf{h}_1 x_{1,t}$$

which coincides with $B_n^{\$} x_{1,t}$ re-obtaining the solution of Section 3. This shows that the general solution (18) and the particular one (5) coincide when (28) holds.

4.2.2. Long memory

A particularly convenient long memory parameterization, that nests both stationary ARMA as well as the random walk is the ARFIMA model. In particular, the generic factor $x_{j,t}$ follows a stationary ARFIMA$(1, d, 1)$ model (see Brockwell and Davis (1991), Definition 12.4.2) when

$$(1 - \psi_j L)(1 - L)^{d_j} x_{j,t} = (1 + \theta_j L)\varepsilon_{j,t}, \tag{31}$$

where the autoregressive and moving average coefficients ψ_j, θ_j satisfy the usual stationarity and invertibility conditions

$$\mid \psi_j \mid < 1, \ \mid \theta_j \mid < 1, \text{ with } \psi_j \neq \theta_j, \tag{32}$$

and d_j is a real number such that

$$-1/2 < d_j < 1/2. \tag{33}$$

When (32) and (33) hold, it can be shown (see Brockwell and Davis (1991), Theorem 12.4.2) that $x_{j,t}$ admits the linear representation (22) with coefficients $\phi_{j,i} = \phi_{j,i}(\ \xi_j)$ satisfying

$$\sum_{i=0}^{\infty} \phi_{j,i} L^i = (1 + \theta_j L)(1 - \psi_j L)^{-1}(1 - L)^{-d_j} \tag{34}$$

function of the vector $\xi_j = (\psi_j, \theta_j, d_j)'$. To discuss the stationarity and memory properties of the factor $x_{j,t}$, we use the property

$$\phi_{j,i} \sim c\, i^{d_j - 1} \quad \text{as } i \to \infty, \tag{35}$$

which stems from (34) for any $d_j < 1$, for a constant c. Stationarity (24) then follows when (33) holds. Short memory (25) requires $d_j \leq 0$ and long memory $d_j > 0$.[14] As a particular case of short memory, stationary ARMA$(1,1)$ is obtained for $d_j = 0$. Although stationarity implies mean-reversion, the opposite is not necessarily true since mean-reversion (27) simply requires $d_j < 1$. Finally, when $d_j = 1$ one obtains the non stationary ARIMA$(1,1,1)$ process, a special case of which is the random walk (when $\psi_j = \theta_j = 0$). Specification (31) extends to ARFIMA(p, d, q) whenever $\psi_j(L)(1 - L)^{d_j} x_{j,t} = \theta_j(L)\varepsilon_{j,t}$ for polynomials $\psi_j(L), \theta_j(L)$ of order p, q, respectively, with roots bigger than one in absolute value. This will be the general specification adopted for the factors $x_{j,t}$ in the empirical illustration below.

Alternative definitions of long memory when $0 < d_j < 1/2$, equivalent to (35) for linear stationary processes, are in terms of autocovariance function and spectral density, respec-

[14] Case $d_j < 0$ is technically defined as anti-persistence, but it can be thought of as a special case of short memory.

tively expressed as

$$cov(x_{j,t}, x_{j,t+u}) \sim c\, u^{2d_j - 1}, \text{ as } u \to \infty, \text{ and } s_j(\lambda) \sim c\, \lambda^{-2d_j}, \text{ as } \lambda \to 0. \tag{36}$$

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

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

Our interest is in the characterization of the physical measure, namely the 'true' distribution, of observed bond prices and transformation of such as yields, forwards 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 distribution of bond prices 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 $\mathbf{B}_{j,n}^{\$}$, which are evaluated under the $\mathbb{Q}$ measure, and of the state variables $\mathbf{C}_{j,t}$, which are evaluated under the $\mathbb{P}$ measure.[15]

The results below indicate a clear dichotomy, namely that the $\mathbb{P}$ measure's parameters determine the 'long-run' time series properties of bond prices whereas the $\mathbb{Q}$ measure's parameters contribute to the 'long maturity' cross-sectional properties of bond prices. In other words, the dynamic persistence induced by the model on the physical measure 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. We will refer to these results, with a somewhat abuse of terminology since we are always referring to physical measure's characteristics, as holding 'under the $\mathbb{P}$' and 'under the $\mathbb{Q}$ measure' respectively.

To proceed, a key observation is that when the matrix $\mathbf{F}$ satisfies (21) (see Appendix C), which we assume for both short and long memory parameterizations, then the K recursions (17) in the infinite-dimensional loadings $\mathbf{B}_{j,n}^{\$}$, with $1 \leq j \leq K$, can in fact be reduced into a recursion of a scalar sequence. In particular, by direct evaluation the loadings to the jth factor, with $1 \leq j \leq K$, will satisfy the recursion:

$$\begin{aligned}
&\mathbf{B}_{j,n}^{\$} = (b_{j,n}\, b_{j,n-1}\, \ldots\, b_{j,1}\, 0\, \ldots)' \text{ with} \\
&b_{j,1} = \delta_j, \\
&b_{j,l} = \delta_j - \kappa_{j1}(\sum_{s=0}^{l-2} b_{1,l-s-1}\phi_{1,s}) - \kappa_{j2}(\sum_{s=0}^{l-2} b_{2,l-s-1}\phi_{2,s}) - \ldots - \kappa_{jK}(\sum_{s=0}^{l-2} b_{K,l-s-1}\phi_{2,s}), \;\; l \geq 2,
\end{aligned} \tag{37}$$

where we set the K dimensional vector

$$\kappa_j = (\kappa_{j1} \ldots \kappa_{jK})' = \mathbf{\Sigma}\lambda_{1.j}, \;\; 1 \leq j \leq K. \tag{38}$$

[15] Here we are not interested in deriving the loading coefficients $A_n^{\$}$ and the $\mathbf{B}_{j,n}^{\$}$ under the $\mathbb{P}$ measure nor the distribution of the state variables $\mathbf{C}_{j,t}$ under the $\mathbb{Q}$ measure.

Recursion (37) is highly nonlinear since the jth coefficient $\mathbf{B}^{\$}_{j,n}$ depends not only on the elements of $\mathbf{B}^{\$}_{j,n-1}, \mathbf{B}^{\$}_{j,n-2}, ...$ but also on the elements of all the others $\mathbf{B}^{\$}_{k,n-1}, \mathbf{B}^{\$}_{k,n-2}, ...$ for $k \neq j$, every $1 \leq k \leq K$. Useful insights can be obtained by looking at the one-factor case, $K = 1$. By recursive substitution one gets

$$
\begin{aligned}
b_{1,1} &= 1, \\
b_{1,2} &= 1 - \kappa_{1,1}\phi_{1,0}, \\
b_{1,3} &= 1 - \kappa_{1,1}(\phi_{1,0} + \phi_{1,1}) + \kappa_{1,1}^2\phi_{1,0}^2, \\
b_{1,4} &= 1 - \kappa_{1,1}(\phi_{1,0} + \phi_{1,1} + \phi_{1,2}) + \kappa_{1,1}^2(\phi_{1,0}^2 + 2\phi_{1,0}\phi_{1,1}) - \kappa_{1,1}^3\phi_{1,0}^3, \\
b_{1,5} &= 1 - \kappa_{1,1}(\phi_{1,0} + \phi_{1,1} + \phi_{1,2} + \phi_{1,3}) + \kappa_{1,1}^2(\phi_{1,0}^2 + \phi_{1,1}^2 + 2\phi_{1,0}\phi_{1,1} + 2\phi_{1,0}\phi_{1,2}) - \kappa_{1,1}^3(\phi_{1,0}^3 + 3\phi_{1,0}^2\phi_{1,1}) + \kappa_{1,1}^4\phi_{1,0}^4, \\
b_{1,6} &= ....
\end{aligned} \quad (39)
$$

We need to distinguish between evaluation of the $b_{1,l}$ under the $\mathbb{P}$ and the $\mathbb{Q}$ measures. The first case is obtained when $\kappa_{1,1} = 0$, which in turn follows when $\boldsymbol{\lambda}_1 = \mathbf{0}$ in (12), namely for a constant market price. This does not, of course, imply that bond prices are evaluated under the $\mathbb{P}$ measure.[16] In this case $b_{1,l} = 1$ for every $l = 1, 2, ...$ and one obtains a simple solution to bond prices, as formalized below. When instead $\kappa_{1,1} \neq 0$ then the $b_{1,l}$, now evaluated under the $\mathbb{Q}$ measure, have a more cumbersome expression. Important implications can nevertheless be derived in both cases: by looking at the recursion above, it is evident that the behaviour of the $b_{1,l}$ as l increases, depends on the interaction between powers of the slow (hyperbolic) increase of the partial sum terms $\sum_{l=0}^{k} \phi_{1,l}$ and the fast (exponential) decay of powers of the term $\kappa_{1,1}$ in particular when $|\kappa_{1,1}| < 1$. For instance, whereas the latter term can dominate for small and intermediate maturities, the former can dominate for long maturities since $\sum_{l=0}^{k} \phi_{1,l} \sim ck^{d_1}$ as k increases when (35) holds. See Lemma 2 in Appendix D. This gives rise to a remarkable degree of flexibility of our long memory affine model in fitting the volatility term structures of yields, forwards and returns.

With the $b_{j,l}$ at hand, for every $1 \leq l \leq K$, the general quasi-closed solution of the model under the $\mathbb{Q}$ measure follows. In fact, recalling

$$\mathbf{h}_j = (\phi_{j,0}\, \phi_{j,1}\, \phi_{j,2} ...)', \quad (40)$$

where $\phi_{j,i}$ are the linear representation coefficients of the factor $x_{j,t}$ in (22), one gets $\mathbf{B}^{\$\prime}_{j,n}\mathbf{h}_j = \sum_{i=0}^{n-1} b_{j,n-i}\phi_{j,i} = \Phi_{j,n,0}$, where we set

$$\Phi_{j,n,l} = \sum_{i=0}^{n-1} b_{j,n-i}\phi_{j,i+l}, \quad \text{for every } l \geq 0. \quad (41)$$

Plugging the $\Phi_{j,n,0}$ into (16) provides the $A^{\$}_n$, namely the first moment of the (log) bond

[16] By Gaussianity of the model, the distribution of bond prices only depends on the first two moments. The mean is evaluated under the $\mathbb{P}$ measure when $\boldsymbol{\lambda}_0 = \mathbf{0}$ whereas the variance requires $\boldsymbol{\lambda}_1 = \mathbf{0}$. 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.

prices. Next, since $E_t(x_{j,t+i}) = \sum_{l=0}^{\infty} \phi_{j,l+i}\varepsilon_{j,t-l}$ for all $i = 0, 1, ...$ then (see Appendix C)

$$\mathbf{B}_{j,n}^{\$\prime}\mathbf{C}_{j,t} = \sum_{i=0}^{n-1} b_{j,n-i}E_t(x_{j,t+i}) = \sum_{i=0}^{n-1} b_{j,n-i}\left(\sum_{l=0}^{\infty} \phi_{j,l+i}\varepsilon_{j,t-l}\right) = \sum_{l=0}^{\infty} \Phi_{j,n,l}\varepsilon_{j,t-l}, \qquad 1 \le j \le K,$$

the variance of which provide the second moment of (log) bond prices. Combining terms we get an alternative closed-form solution to (18)-(19)-(20) for the the term structure of yields, forward rates and return, summarized in the following theorem.

Theorem 4.2. *For the pricing kernel (11), the market price of risk (12) and the state variable dynamics (14), under Gaussianity (13), the term structure of yields, forward rates and returns are given by, respectively,*

$$y_{n,t}^{\$} = -n^{-1}p_{n,t}^{\$} = n^{-1}A_n^{\$} + \sum_{i=0}^{\infty} \Delta_{1,n,i}^{y}\varepsilon_{1,t-i} + ... + \sum_{i=0}^{\infty} \Delta_{K,n,i}^{y}\varepsilon_{K,t-i}, \tag{42}$$

$$f_{n,t}^{\$} = p_{n,t}^{\$} - p_{n+1,t}^{\$} = A_{n+1}^{\$} - A_n^{\$} + \sum_{i=0}^{\infty} \Delta_{1,n,i}^{f}\varepsilon_{1,t-i} + ... + \sum_{i=0}^{\infty} \Delta_{K,n,i}^{f}\varepsilon_{K,t-i}, \tag{43}$$

$$r_{n,t}^{\$} = p_{n-1,t}^{\$} - p_{n,t-1}^{\$} = A_n^{\$} - A_{n-1}^{\$} + \sum_{i=0}^{\infty}\Delta_{1,n,i}^{r}\varepsilon_{1,t-i} + ... + \sum_{i=0}^{\infty}\Delta_{K,n,i}^{r}\varepsilon_{K,t-i}, \tag{44}$$

where for each $1 \le j \le K$

$$\Delta_{j,n,l}^{y} = n^{-1}\Phi_{j,n,l}, \qquad l \ge 0, \tag{45}$$

$$\Delta_{j,n,l}^{f} = \Phi_{j,n+1,l} - \Phi_{j,n,l} = b_{j,1}\phi_{j,n+l} + \sum_{i=0}^{n-1}(b_{j,n+1-i} - b_{j,n-i})\phi_{j,i+l}, \; l \ge 0, \tag{46}$$

$$\Delta_{j,n,0}^{r} = -\Phi_{j,n-1,0}, \;\; \Delta_{j,n,l}^{r} = \Phi_{j,n,l-1} - \Phi_{j,n-1,l} = b_{j,n}\phi_{j,l-1}, \qquad l \ge 1. \tag{47}$$

and the $\Phi_{j,n,l}$ *are defined in (41).*

The above formulae apply for any specification of the market prices of risk, hence either under the $\mathbb{P}$ or $\mathbb{Q}$ measure. However, they greatly simplify under the $\mathbb{P}$ measure since, setting $\boldsymbol{\lambda}_1 = \mathbf{0}$, by solving the recursion (37) the coefficients $\mathbf{B}_{j,n}^{\$}$ turn out to be parameters-free, in particular equal to

$$\mathbf{B}_{j,n}^{\$} = (\underbrace{1....1}_{n\, terms}0...)', \;\; 1 \le j \le K \text{ and for every n } \ge 1.$$

Hence now (41) simplifies to $\Phi_{j,n,l} = \sum_{i=0}^{n-1} \phi_{j,i+l}$ and the $\Delta_{j,n,l}^{y}, \Delta_{j,n,l}^{f}, \Delta_{j,n,l}^{r}$ change accordingly. In particular

$$\Delta^{y}_{j,n,l} \quad = n^{-1}(\sum_{i=0}^{n-1}\phi_{j,i+l}), \qquad l \geq 0, \tag{48}$$

$$\Delta^{f}_{j,n,l} \quad = \phi_{j,n+l}, \qquad l \geq 0, \tag{49}$$

$$\Delta^{r}_{j,n,0} \quad = -(\sum_{i=0}^{n-1}\phi_{j,i}), \;\; \Delta^{r}_{j,n,l} = \phi_{j,l-1}, l \geq 1. \tag{50}$$

Unlike (18)-(19)-(20), the formulae of Theorem 4.2 will not be used to quantify model-implied yields $y^{\$}_{n,t}$, forwards $f^{\$}_{n,t}$ and returns $r^{\$}_{n,t}$ but rather to derive their conditional and unconditional second order properties as shown below. Note that the i.i.d. innovations $\varepsilon_{1,t}, ..., \varepsilon_{K,t}$ are in general cross-correlated.

Theorem 4.3. *Under the assumptions of Theorem 4.2:*
(i) yields $y^{\$}_{n,t}$ have conditional variance

$$\text{var}_{t-1}(y^{\$}_{t,n}) = (\Delta^{y}_{1,n,0}, ..., \Delta^{y}_{K,n,0})\mathbf{\Sigma}(\Delta^{y}_{1,n,0}, ..., \Delta^{y}_{K,n,0})'; \tag{51}$$

(ii) (stationary case) when the coefficients (41) satisfy $\sum_{i=0}^{\infty}(\Delta^{y}_{1,n,i}, ..., \Delta^{y}_{K,n,i})(\Delta^{y}_{1,n,i}, ..., \Delta^{y}_{K,n,i})', < \infty$, yields $y^{\$}_{n,t}$ have spectral density

$$s_{y_n}(\lambda) = \frac{1}{2\pi}\Big(\sum_{l=0}^{\infty}(\Delta^{y}_{1,n,l}, ..., \Delta^{y}_{K,n,l})e^{\imath\lambda l}\Big)\mathbf{\Sigma}\Big(\sum_{l=0}^{\infty}(\Delta^{y}_{1,n,l}, ..., \Delta^{y}_{K,n,l})e^{-\imath\lambda l}\Big)', \tag{52}$$

and unconditional variance

$$\text{var}(y^{\$}_{t,n}) = \sum_{i=0}^{\infty}(\Delta^{y}_{1,n,i}, ..., \Delta^{y}_{K,n,i})\mathbf{\Sigma}(\Delta^{y}_{1,n,i}, ..., \Delta^{y}_{K,n,i})'; \tag{53}$$

(iii) (non-stationary case) when the coefficients (41) do not satisfy the stationarity condition in (ii) but the first-differenced coefficients:

$$\bar{\Delta}^{y}_{j,n,0} = \Delta^{y}_{j,n,0}, \bar{\Delta}^{y}_{j,n,i} = \Delta^{y}_{j,n,i} - \Delta^{y}_{j,n,i-1}, \text{ for every } i \geq 1 \text{ and } 1 \leq j \leq K, \tag{54}$$

satisfy $\sum_{i=1}^{\infty}(\bar{\Delta}^{y}_{1,n,i}, ..., \bar{\Delta}^{y}_{K,n,i})(\bar{\Delta}^{y}_{1,n,i}, ..., \bar{\Delta}^{y}_{K,n,i})', < \infty$, the first differences of yields $\Delta y^{\$}_{n,t} = y^{\$}_{n,t} - y^{\$}_{n,t-1}$ have spectral density

$$s_{\Delta y_n}(\lambda) = \frac{1}{2\pi}\Big(\sum_{l=0}^{\infty}(\bar{\Delta}^{y}_{1,n,l}, ..., \bar{\Delta}^{y}_{K,n,l})e^{\imath\lambda l}\Big)\mathbf{\Sigma}\Big(\sum_{l=0}^{\infty}(\bar{\Delta}^{y}_{1,n,l}, ..., \bar{\Delta}^{y}_{K,n,l})e^{-\imath\lambda l}\Big)', \tag{55}$$

and unconditional variance

$$\text{var}(\Delta y^{\$}_{t,n}) = \sum_{i=0}^{\infty}(\bar{\Delta}^{y}_{1,n,i}, ..., \bar{\Delta}^{y}_{K,n,i})\mathbf{\Sigma}(\bar{\Delta}^{y}_{1,n,i}, ..., \bar{\Delta}^{y}_{K,n,i})'. \tag{56}$$

The same formulae apply to forwards and returns by substituting the $\Delta^y_{j,n,l}, \bar{\Delta}^y_{j,n,l}$ *with the* $\Delta^f_{j,n,l}, \bar{\Delta}^f_{j,n,l}$ *and* $\Delta^r_{j,n,l}, \bar{\Delta}^r_{j,n,l}$ *respectively.*

These formulae are extremely general since derived for generic specifications of the coefficients $\phi_{j,i}$ with arbitrary memory. We can now fully characterise the persistence of yields, forwards and returns when long memory is allowed for. Stationary ARFIMA $x_{j,t}$ are included as a special, parametric, case.

Theorem 4.4. *Assume that for every* $1 \leq j \leq K$

$$\phi_{j,s} \sim cs^{d_j-1} \text{ as } s \to \infty \text{ with } 0 < d_j < 1/2 \tag{57}$$

and

$$|\phi_{j,s+1} - \phi_{j,s}| \leq cs^{-1}\phi_{j,s} \text{ for any } s \geq S, \text{ some finite } S. \tag{58}$$

Under either the $\mathbb{P}$ *and* $\mathbb{Q}$ *measure, the spectral densities of yields* $y^{\$}_{n,t}$*, forwards* $f^{\$}_{n,t}$ *and returns* $r^{\$}_{n,t}$ *satisfy:*

$$s_{y_n}(\lambda) \sim c\lambda^{-2\underline{d}}, \quad s_{f_n}(\lambda) \sim c\lambda^{-2\underline{d}}, \quad s_{r_n}(\lambda) \sim c\lambda^{-2\underline{d}} \qquad \text{as } \lambda \to 0^+,$$

setting $\underline{d} = max(d_1, ...d_K)$*.*

The model spectral densities of yields, forwards and returns have a peak at zero frequency. Alternatively, taking logarithm, it follows that $log\, s_{y_n}(\lambda) \sim -2\underline{d} \log \lambda, log\, s_{f_n}(\lambda) \sim -2\underline{d} \log \lambda$ and $log\, s_{r_n}(\lambda) \sim -2\underline{d} \log \lambda$ for $\lambda \to 0^+$. This shows that the model log-spectral densities are all negatively sloped near the zero frequency, the more the larger the long memory parameters $\underline{d}$. Our model is potentially able to match *Stylized Fact* 1. The degree of memory will not depend on n although away from zero frequency the spectral densities of $y^{\$}_{n,t}$, $f^{\$}_{n,t}$ and $r^{\$}_{n,t}$ 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 (36)).

The degree of memory or, alternatively, of non-stationarity implied by the physical measure for yields, forwards and rates does not depend on the form of the $\mathbb{Q}$ measure since the parameters $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$, governing the market price of risk, although contributing in general to the physical measure, do not affect these particular aspects of the dynamic properties of the model. This result does not depend on the long memory assumption but holds true for any specification of the essentialy affine model. In contrast, the cross-sectional properties of yields differ markedly depending on whether the $\mathbb{P}$ or the $\mathbb{Q}$ measure holds. The next theorem illustrates the long maturity behaviour of both the conditional and unconditional variance for yields, forwards and returns under the $\mathbb{P}$ measure. The corresponding $\mathbb{Q}$-measure term structure properties are presented subsequently.

Theorem 4.5. *Assume (57) and (58). Under the* $\mathbb{P}$ *measure, setting* $\underline{d} = max(d_1, ...d_K)$*, as* $n \to \infty$*:*
(i) the conditional variances of yields $y^{\$}_{t,n}$*, forwards* $f^{\$}_{t,n}$ *and returns* $r^{\$}_{t,n}$ *satisfy*

$$\text{var}_{t-1}(y^{\$}_{t,n}) = O(n^{2\underline{d}-2}),\ \text{var}_{t-1}(f^{\$}_{t,n}) = O(n^{2\underline{d}-2}),\ \text{var}_{t-1}(r^{\$}_{t,n}) = O(n^{2\underline{d}});$$

(ii) (stationary case) the unconditional variances of yields $y^{\$}_{t,n}$, forwards $f^{\$}_{t,n}$ and returns $r^{\$}_{t,n}$ satisfy

$$\mathrm{var}(y^{\$}_{t,n}) = O(n^{2\underline{d}-1}),\ \mathrm{var}(f^{\$}_{t,n}) = O(n^{2\underline{d}-1}),\ \mathrm{var}(r^{\$}_{t,n}) = O(n^{2\underline{d}});$$

(iii) (non-stationary case) when (57) is replaced by:

$$\phi_{j,s} \sim cs^{d_j-1} \ \textit{as}\ s \to \infty\ \textit{with}\ 0 < d_j < 3/2, \tag{59}$$

for every $1 \leq j \leq K$, the unconditional variances of the first difference of yields $\Delta y^{\$}_{t,n} = y^{\$}_{t,n} - y^{\$}_{t-1,n}$, forwards $\Delta f^{\$}_{t,n} = f^{\$}_{t,n} - f^{\$}_{t-1,n}$ and returns $\Delta r^{\$}_{t,n} = r^{\$}_{t,n} - r^{\$}_{t-1,n}$ satisfy

$$\mathrm{var}(\Delta y^{\$}_{t,n}) = \mathrm{var}_{t-1}(y^{\$}_{t,n}) + O(n^{2\underline{d}-3}), \qquad \mathrm{var}(\Delta f^{\$}_{t,n}) = \mathrm{var}_{t-1}(f^{\$}_{t,n}) + O(n^{2\underline{d}-3}), \qquad \mathrm{var}(\Delta r^{\$}_{t,n}) = \mathrm{var}_{t-1}(r^{\$}_{t,n}) + O(1).$$

Under the $\mathbb{P}$ measure with long memory factors, the term structure of volatility for yields and forwards declines to zero at the same rate when mean-reversion holds, namely for $\underline{d} = max(d_1, ...d_K) < 1$.[17] Under the same conditions, these term structures diverge, with maturity, for returns as long as long memory is manifested ($\underline{d} > 0$). An important difference emerges for returns: the term structures of conditional and unconditional variances have the same limiting behaviour for long maturity n. Since the conditional variance is determined by the coefficients $\Delta^r_{j,n,0}$ to the innovations $\varepsilon_{j,t}$ for all $1 \leq j \leq K$, this means that the unconditional variance of returns is dominated by the variance of these i.i.d. components, namely a linear combination (of K terms) of the *square* of the $\Delta^r_{j,n,0}$ coefficients. This arises because it is evident from (50) that the coefficients $\Delta^r_{j,n,l}$ for every $l \geq 1$ have a different behaviour for large n from the case $l = 0$, in particular the former have a smaller order of magnitude for large n. On the other hand, the autocovariances of returns are function of the *level* of the $\Delta^r_{j,n,0}$ coefficients. This implies that for large maturity n it could be difficult to measure the persistence of returns accurately, as measured by, say, autocorrelations (ratio of autocovariances to variance) since it will be masked by the variance of this non persistent, in fact i.i.d., term associated with the contemporaneous innovations. More precisely, one could find evidence of little persistence in the returns data, the smaller the larger the maturity n is, simply because the persistent component is smaller (in terms of variance), in relative terms, when compared with the non-persistent component as n gets large.[18] We will verify this conjecture empirically with the data.

Comparing the results of Theorem 4.5 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, smoothly modulated by the magnitude of $\underline{d}$. For returns, short memory ruled out divergence altogether (under

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

[18] Inspecting the return formula with $K = 1$, namely $r^{\$}_{n,t} = A^{\$}_n - A^{\$}_{n-1} + B^{\$}_n x_{t-1} - B^{\$}_{n-1} x_t$, one might think that for large n it can be approximated by $-B^{\$}_n(x_t - x_{t-1})$, suggesting that stationarity will be achieved for large n. However, a closer analysis reveals that this is not the case within the class of Gaussian affine models. Consider the simple model of Section 3 when $\psi_x = 1$, a random walk state variable. Then is easily follows that $r^{\$}_{n,t} = A^{\$}_n - A^{\$}_{n-1} - n\epsilon_{x,t} + x_t$. Therefore for large n the persistence of returns does not change. However, the degree of persistence will be more difficult to measure, the larger is n, since its conditional variance will diverge.

stationarity), which instead occurs when $\underline{d} > 0$. When $\underline{d} = 1$, namely when at least one factor has a unit root, the non-stationary results of Section 3 are re-obtained since the unit root is a special case of long memory.

If non-stationarity is suspected, the conditional variances are well defined but not the unconditional variances. Part (iii) of Theorem 4.5 shows the behaviour of the variance of the first-differenced yields, forwards and returns. The stationarity condition is now relaxed to $d_j < 3/2$ for every $1 \leq j \leq K$ as indicated in (59). An interesting feature emerges: the behaviour for large n of the conditional variances of the raw data coincide (reported in part (i)), as order of magnitude, with the corresponding behaviour of the unconditional variances for the first-differenced data (reported in part (iii)). This implies that for large n it could be difficult to measure the degree of memory of the data, once one takes the first difference, because the persistence will be masked by the variance of non-persistent component associated with the contemporaneous innovations $\varepsilon_{j,t}, 1 \leq j \leq K$. Notice that this latter quantity represents the conditional variance of the first-differenced data, in turn exactly equal to the conditional variance of the data themselves, without differencing.[19] As explained above, this pitfall was observed for raw returns but now we find it arising for first-differenced yields and forwards as well. This phenomenon is due to the way bond data characteristics, namely yields, forwards and return, depend on the maturity n. We now present the $\mathbb{Q}$ measure results.

Theorem 4.6. *Assume (57) and (58) and set* $\underline{d} = max(d_1, ...d_K)$. *Under the* $\mathbb{Q}$ *measure, when for every* $1 \leq j \leq K$*:*

$$b_{j,s} \sim -\kappa_{j1}\delta_1(\sum_{i=0}^{s} \phi_{1,i}) - ... - \kappa_{jK}\delta_K(\sum_{i=0}^{s} \phi_{K,i}) \text{ as } s \to \infty, \tag{60}$$

(i) the conditional variance of yields $y^{\$}_{t,n}$*, forwards* $f^{\$}_{t,n}$ *and returns* $r^{\$}_{t,n}$ *as* $n \to \infty$ *satisfy:*

$$\text{var}_{t-1}(y^{\$}_{n,t}) = O(n^{4\underline{d}-2}),\ \text{var}_{t-1}(f^{\$}_{n,t}) = O(n^{4\underline{d}-2}),\ \text{var}_{t-1}(r^{\$}_{n,t}) = O(n^{4\underline{d}});$$

(ii) (stationary case) the unconditional variances of yields $y^{\$}_{n,t}$*, forwards* $f^{\$}_{n,t}$ *and returns* $r^{\$}_{n,t}$ *as* $n \to \infty$ *satisfy:*

$$\text{var}(y^{\$}_{t,n}) = O(n^{4\underline{d}-1}),\ \text{var}(f^{\$}_{t,n}) = O(n^{4\underline{d}-1}),\ \text{var}(r^{\$}_{t,n}) = O(n^{4\underline{d}}).$$

(iii) (non-stationary case) When (57) is replaced by (59), the unconditional variances of the first difference of yields $\Delta y^{\$}_{t,n} = y^{\$}_{t,n} - y^{\$}_{t-1,n}$*, forwards* $\Delta f^{\$}_{t,n} = f^{\$}_{t,n} - f^{\$}_{t-1,n}$ *and returns* $\Delta = r^{\$}_{t,n} = r^{\$}_{t,n} - r^{\$}_{t-1,n}$ *satisfy*

$$\text{var}(\Delta y^{\$}_{t,n}) = \text{var}_{t-1}(y^{\$}_{t,n}) + O(n^{4\underline{d}-3}), \quad \text{var}(\Delta f^{\$}_{t,n}) = \text{var}_{t-1}(f^{\$}_{t,n}) + O(n^{4\underline{d}-3}), \quad \text{var}(\Delta r^{\$}_{t,n}) = \text{var}_{t-1}(r^{\$}_{t,n}) + O(n^{2\underline{d}}).$$

Under the $\mathbb{Q}$ measure, long memory has an even stronger effect on the large maturity behaviour of the volatility term structures. As before, for returns the conditional and uncon-

[19] Note that $var_{t-1}(X_t) = var_{t-1}(X_t - X_{t-1})$ always holds for every random variable with finite conditional variance. Therefore there is no need to establish the conditional variance of the first-differenced data.

ditional variances diverge at the same rate, but more prominently than under the previous $\mathbb{P}$ measure case. In terms of conditional variances of yields and forward rates, their term structures tend to be negatively sloped when stationarity ($0 < \underline{d} < 1/2$) holds but diverging otherwise, including the mean-reversion case ($1/2 < \underline{d} < 1$). Instead, for their unconditional variance divergence can already occur even within the stationary case, as long as there is enough long memory ($\underline{d} > 1/4$). Notice that these are large n characterizations so that even if these variance term structures are now all turning positively sloped for large n, these could initially decline for short and intermediate maturities depending on the other parameters' value.

Therefore, under the $\mathbb{Q}$ measure the long memory model achieves a great deal of flexibility for the volatility term structure of yields, forwards and returns. Those closed-form results rely on condition (60) which can be easily verified numerically. In turn, the latter appears to require sufficiently small κ_j, $1 \leq j \leq K$, by (39).

The same pitfall observed under the $\mathbb{P}$ measure is manifested here. The degree of persistence of returns is masked, appearing weaker than it should, by the variance of the contemporaneous innovations $\varepsilon_{j,t}, 1 \leq j \leq K$, for large n. The same holds for yields and forwards when looking at their first differences.

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

5. Alternative approaches to model the persistence of nominal bonds

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

Besides long memory, 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.[20] Stationarity is typically imposed in the estimation. It is well known that the ordinary least squares (OLS) estimates of the maximal autoregressive 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 (2014), who impose the

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

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. An attractive, parametric, nonlinear alternative is obtained by means of allowing regime switching state variables which is very effective in capturing persistence. Indeed, 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.[21]

In this respect, regime switching and long memory models are both able to account for the persistence of observed yields, implying an asymptotic behaviour of the autocovariances such as (36).[22] The two models can instead differ with respect to the term structure of volatility, where is appears more cumbersome to find a suitable parameterization akin to *Stylized Fact 2* for regime switching models.[23]

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, 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. Moreover, note that although the long memory model is truncated for practical estimation, such truncation will

[21] Diebold and Inoue (2001) illustrate by means of Monte Carlo experiments that, when the $p_{1,1} = p_{2,2} = 0.95$ (they consider $K = 2$) and for samples between 200 and 400 observations, long memory is manifested with estimates of the memory parameter well in the stationary region (33). 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 (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 [Table 3](2013)).

[22] 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.

[23] A stylized regime switching affine term structure model is $y^{\$}_{n,t}(s_t) = c^{\$}_n(s_t) + B^{\$}_n x_t(s_t)$ where the latent variable s_t follows a K-state Markov chain, and the single factor $x_t(s_t)$ follows an ARMA with switching parameters driven by s_t. Ang et al (2008) clarify that the loading $B^{\$}_n$ must be regime-invariant in order to preserve a closed-form solution. Moreover, $B^{\$}_n$ declines rapidly to zero under stationarity for large n. Therefore, *Stylized Fact 2* can only be accounted for by a suitable parameterization of the intercept term $c^{\$}_n(s_t)$. This could prove very cumbersome and a closed-form expression of the volatility term structure is not warranted.

be asymptotically negligible as the sample size increases (see Appendix C).

6. 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 $y_{n,t}$, expected inflation and inflation risk premium:

$$y^{\$}_{n,t} = c_n + y_{n,t} + \frac{1}{n}\mathrm{E}_t \ln\left(\frac{\Pi_{t+n}}{\Pi_t}\right) + IP_{n,t}, \tag{61}$$

where Π_t defines the price index, $IP_{n,t}$ 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 $y_{n,t}$, respectively. Both channels are able to induce the form of long memory observed empirically in the data.

6.1. *Inflation channel*

Recent research shows 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 (see Altissimo et al (2009)), a similar result will apply to US inflation.[24] 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

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

where $\pi_{i,t}$ is the ith sector inflation rate, u_t is the i.i.d. common dynamic shock and $\varepsilon_{i,t}$ is the i.i.d. idiosyncratic component, assumed independent from u_t at any leads and lags.[25] The autoregressive coefficients $\psi_{\pi,i}$ are assumed i.i.d. with a common distribution over the stationary region ensuring that $-1 < \psi_{\pi,i} < 1$ for any sub-sector i. Although the $\varepsilon_{i,t}$ appear to dominate the variance of the individual $\pi_{i,t}$, the common factor appears to explain a large part of the aggregate CPI inflation dynamics. In fact $var(J^{-1}\sum_{i=1}^{J}\varepsilon_{i,t})$ is estimated to be much smaller than, about one fourth of the average variance of the idiosyncratic components $\varepsilon_{i,t}$.[26] At the same time, by well-known aggregation results (see the seminal

[24] For US data this has been documented by Hyung et al (2006), Cheung and Chung (2009) and Bos et al (2014) among others.

[25] In Altissimo et al (2009) $\varepsilon_{i,t}$ are modelled as ARMA, mutually independent from u_s for any t, s but the same aggregation result illustrated here carries through.

[26] 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

works of Robinson (1978) and Granger (1980) and the generalisation of Zaffaroni (2004)) under mild conditions

$$N^{-1}\sum_{i=1}^{N}\pi_{i,t} \to_2 \mu_\pi + \sum_{k=0}^{\infty}\phi_{\pi,k}u_{t-k}, \quad \text{as } N \to \infty, \tag{62}$$

where μ_π and $\phi_{\psi,k}$, $k = 0, 1, ...$ are the limit (cross-sectional) averages of the $\mu_{\pi,i}$ and $\psi^k_{\pi,i}$, $k = 0, 1, ...$ respectively, and $\to_2$ 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)), (62) occurs and the impulse response of the common shock u_t to CPI inflation satisfies $\phi_{\pi,k} \sim c\, k^{d_\pi - 1}$ as $k \to \infty$, which, recalling (35), is coherent with π_t exhibiting long memory with memory parameter $d_\pi > 0$:

$$cov(\pi_t, \pi_{t+k}) \sim c\, k^{2d_\pi - 1} \text{ as } k \to \infty. \tag{63}$$

Note that the expected inflation term in (61) 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)).

6.2. Real rate channel

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

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

where $u_t \sim NID(0, 1)$, $\varepsilon_{j,t} \sim NID(0, \sigma_j^2)$ mutually independent one of another and $-1 < \psi_{x,j} < 1$ for all $j = 1, ..., K$. Under suitable assumptions on the pricing kernel akin to (2), real bond yields satisfy the affine relationship

$$y_{t,n} = a_n + \sum_{j=1}^{K} n^{-1}B_{j,n}x_{j,t}, \tag{64}$$

where, in particular, the n-varying coefficients $B_{j,n}$, $j = 1, .., K$ satisfy $B_{j,n} = (1 - \psi^n_{x,j})/(1 - \psi_{x,j})$. Since Litterman and Sheinkman (1991), the large majority of estimated affine models considers up to three factors, that is $1 \leq K \leq 3$. This approach is essentially dictated by statistical consideration since the number of parameters to be estimated increases rapidly with K. On the other hand, a small K 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

$J = 404$ sectors in France, Germany, Italy only, with a correlation above 80%.

of Robinson (1978) and Granger (1980) as K increases. In particular, as illustrated in Appendix A, letting $K \to \infty$ leads to a semiparametric specification of an affine term structure model with long memory yields $y_{n,t}$:

$$\operatorname{cov}(y_{n,t}, y_{n,t+k}) \sim c\, k^{2d-1} \text{ as } k \to \infty. \tag{65}$$

with memory parameter d satisfying $0 < d < 1/2$. This semiparametric specification is characterised by an infinite number of coefficients, akin to the $\phi_{x,j}$ of (22), unrestricted except for the long memory property (35) (see Appendix A). A natural parameterization is then represented by the ARFIMA model with coefficients (34), so that estimation and inference on a finite, small, number of parameters can be carried out.

In conclusion, both (63) and (65) imply long memory in the nominal yields $y^{\$}_{n,t}$ through (61). 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.

7. Long memory affine term structure models: empirical example

To demonstrate the potential of the model in capturing the dynamic persistence of the data and the shape of their volatility term structures, we estimate a five factors model of the nominal term structure. Details on the estimation and filtering approach are described in the next section. We use the monthly data on nominal yields of Section 2. The data sample goes from January 1986 to December 2011. To motivate further the long memory parameterization of our model, Table 3 reports the long memory parameter estimates obtained by fitting the log-periodogram estimator of Robinson (1995) to nominal yields. It turns out that the memory parameter for yields is positive and significant, well within the mean-reverting yet non-stationary region, namely between 0.5 and 1.

[Insert Table 3 near here]

Let us consider the multi-factor model presented in Section 4.1 with $K = 5$ factors. We consider two observed factors, namely a real activity factor and an inflation factor, denoted by g_t and π_t respectively.[27] The remaining three factors, denoted by $x_{1,t}, x_{2,t}$ and $x_{3,t}$, are assumed latent. We estimate the model in two steps. First we estimate the parameters associated with the observed factors only. In particular, assuming that the observed and latent factors are mutually orthogonal, we estimate the intercept and the loadings $\delta_0, \delta_g, \delta_\pi$

[27]Following Ang and Piazzesi (2003), the real activity factor is obtained as the first principal component of the employment and unemployment rate, industrial production growth and HELP index whereas the inflation factor is the first principal component of the PPI and core CPI. Source: PPI and core CPI, unemployment and employment from U.S. Bureau of Labor Statistics, industrial production from Board of Governors of the Federal Reserve System, HELP from Barnichon (2010). All variables are standardized before calculating their principal component.

in (10) by OLS, projecting the one-month nominal yield $y^{\$}_{1,t}$ on the two observed factors.[28] The dynamics parameters, in particular the long memory parameters of the observed factor, and their associated filtered state variables (necessary for the second estimation step), can be obtained without any reference to the term structure model, but rather by straight application of the approximate maximum likelihood estimator.[29] See Appendix C for details.

To estimate the remaining parameters, associated with the latent factors and the market prices of risk, we plug the parameters estimated in the first stage and the filtered state variables associated to the observed factors, namely $\widehat{\mathbf{C}}_{\pi,t}$ for the inflation factor and $\widehat{\mathbf{C}}_{g,t}$ for real activity factor,[30] into the state space of the term structure model:

$$\begin{pmatrix} y^{\$}_{n_1,t} \\ \vdots \\ y^{\$}_{n_a,t} \end{pmatrix} = \begin{pmatrix} n_1^{-1} A^{\$}_{n_1} \\ \vdots \\ n_a^{-1} A^{\$}_{n_a} \end{pmatrix} + \begin{pmatrix} n_1^{-1} B^{\$\prime}_{\pi,n_1} \\ \vdots \\ n_a^{-1} \mathbf{B}^{\$\prime}_{\pi,n_a} \end{pmatrix} \widehat{\mathbf{C}}_{\pi,t} + \begin{pmatrix} n_1^{-1} B^{\$\prime}_{g,n_1} \\ \vdots \\ n_a^{-1} B^{\$\prime}_{g,n_a} \end{pmatrix} \widehat{\mathbf{C}}_{g,t} + \sum_{j=1}^{3} \begin{pmatrix} n_1^{-1} B^{\$\prime}_{j,n_1} \\ \vdots \\ n_a^{-1} B^{\$\prime}_{j,n_a} \end{pmatrix} \mathbf{C}_{j,t} + \begin{pmatrix} \nu_{n_1,t} \\ \vdots \\ \nu_{n_a,t} \end{pmatrix},$$

with transition equations (14), where we refer to Theorem 4.1 for the definition of $A^{\$}_n$ and the $\mathbf{B}^{\$}_{j,n}$. The $\nu_{n,t} \sim NID(0, \sigma^2_\nu)$ are measurement errors introduced to enhance the flexibility of the model, here assumed to have equal variance for the sake of simplicity. The model is then estimated by means of the approximate maximum likelihood estimator based on the Kalman recursions.[31] The latent state variables $\mathbf{C}_{j,t}$ are estimated with the Kalman filter, imposing that their corresponding innovation variances σ^2_j are equal to unity to achieve identification. As done for the observed factors, we fit ARFIMA(p, d, q) to each latent factors and select the orders $0 \le p \le 2$ and $0 \le q \le 2$ by using the BIC criteria. We also impose that the innovations' cross-correlations are zero, namely diagonal $\mathbf{\Sigma}$. Finally, we assume that $\boldsymbol{\lambda}_1$ is block-diagonal, implying that observed and latent factors determine their own market prices of risk.

[Insert Table 4 near here]

Table 4 presents the estimates of the model parameters when long memory is allowed for all five factors. Standard errors, obtained by numerical evaluation of the Hessian matrix, are reported in small font. Regarding the mean parameters, the unconditional mean of the one-period nominal rate δ_0 equals 3.7%. The three latent factors display statistically large and positive long memory parameters. The first latent factor, in particular, exhibits

[28]The orthogonality assumption is not required but it allows to simplify the estimation procedure. Alternatively, one can use the OLS estimates as initial values when estimating all the model parameters jointly. Such assumption is typically assumed in the term structure literate, see for instance Ang and Piazzese (2003).

[29]Each factor is assumed to be an ARFIMA(p, d, q) where the orders p, q, with $0 \le p \le 2$ and $0 \le q \le 2$, are selected by using the BIC criteria.

[30]In particular, from the first stage one gets estimates of $\delta_0, \delta_\pi, \delta_g$ as well as σ^2_π, σ^2_g and $\mathbf{h}_\pi, \mathbf{h}_g$. The hat $\widehat{}$ indicates the filtered, and thus observed, value of the corresponding state variable.

[31]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. In particular, their 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 (35) with the exponent $-(d_j + 1)$ replacing $d_j - 1$. This rules out using ordinary least squares.

non-stationarity together with mean reversion, with its parameters being between 1/2 and 1 whereas the other two latent factors appear stationary, with their long memory parameter well within the range 0 and 1/2. The first two latent factors also manifest a large AR parameter, close but significantly below unity. Concerning the two observed factor, the inflation factor appears stationary with a long memory parameter of about 1/4, in agreement with previous findings in the literature. The real activity factor appears short memory yet with a relatively large autoregressive parameter. In terms of the estimates of the price of risk parameters, both the intercepts vector $\boldsymbol{\lambda}_0$ and the slopes matrix $\boldsymbol{\lambda}_1$ are significant. This suggests that the data strongly reject the statement by which the $\mathbb{P}$ and the $\mathbb{Q}$ measures coincide. Indeed, as illustrated below, the combination of non zero $\boldsymbol{\lambda}_1$ parameters together with the long memory feature of the model drastically enhance the goodness of fit of the model in terms of volatility term structures. Moreover, we notice that the the third latent factor dominates the other two latent factors in determining the volatility of bond characteristics, although all three factors have a significant impact on both the mean and the variance of yields, forwards and returns.

[**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 the long memory parameter d equal to zero for all the five factors. Estimation is still carried out by maximum likelihood with the the Kalman recursions. [32] The reason for it is twofold: on one hand the factors are assumed to obey an ARMA structure, as opposed to an AR structure, on the other, three of the factors are latent. Therefore, we cannot afford the computational ease of the estimation methods of model with non-latent factors that obey an AR structure, described in Section 5. The short memory results are reported in Table 5. Not surprisingly, the AR coefficients are all statistically large and positive, in fact the first two latent factors have estimated AR coefficients at about 0.99. This would imply a unit root in bond characteristics, ruling out any form of mean reversion. This contrasts with our long memory specification which allows to disentagle non-stationarity from mean reversion. Noticeably, the fit of the short memory 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 but some indication is already obtained by looking at the estimated variance of the measurement errors, assumed equal for all maturities. This equals 9 and 12 basis points for the long memory and short memory models respectively.

Turning again to the long memory specification, the latent factors appear to be a rotation of the 'level', 'slope' and 'curvature' factors as expressed by the three dominant static principal components, extracted from the nominal yields. This is evident from Table 6 which reports the regression adjusted-R^2 associated with the projection of each of the first five principal components on the observed factors, g_t and π_t, as well as on the filtered values $\hat{x}_{j,t}$ of the latent factors, both individually and jointly. The goodness of fit is virtually zero when projecting the fifth principal component on either of the regressors. Instead, the second, the first and the third filtered latent factors explains most of the 'level', 'slope' and 'curvature'

[32] All the other aspects of the estimation are left unchanged, so that in particular ARMA(p, q) models are fitted with $0 \leq p, q \leq 2$, adopting the BIC for model selection.

factors, respectively. The two observed factors appear more strongly related to the 'level' factor than to the other two factors, although this level of correlation is somewhat modest when compared with the effect of the second filtered latent factor $\hat{x}_{1,t}$. The real activity factor appears also related to the fourth principal component.

[Insert Table 6 near here]

7.1. 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 3 plots the logarithm of the periodogram for (standardised) nominal yields and forwards, averaged across maturities, and for nominal 1-year returns together with the (logarithm of the) theoretical spectral density:

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

with $\underline{d} = 0.884$ as from Table 4, where LM indicates the long memory model. Analogously, we will refer to SM as the short memory model. 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 $\underline{d}$, will be important to achieve a good fit of the model across all frequencies. Both Table 4 and Figure 3 confirm that long memory is an important feature of the bond data remarkably accounted for by our model, inducing a degree of persistence that well agrees with *Stylized Fact 1.*

[Insert Figure 3 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 4 reports the term structures of the sample standard deviation (blue diamonds) together with both the long memory (black line) and short memory (red line) model-implied standard deviation for the *first difference* in nominal yields (left panel), nominal forwards (centre panel) and nominal returns (right panel). We need to consider first difference due to the apparent non stationarity of the data that is arising from the estimates, with the maximal long memory parameter estimated above 3/4. 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 4 near here]

The long maturity shape of the volatility term structures depends on the magnitude of the coefficients in $\boldsymbol{\lambda}_1$ which ensure that the $b_{j,l}$ are well-behaved for large j and, moreover, satisfy condition (60). 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 raising again for long maturities for yields and especially for forwards. Instead the short memory model implies declining curves for long maturities, in agreement with estimated autoregressive coefficients close to, yet smaller than, unity. For nominal returns, the long memory model is able to produce a monotonically increasing term structure without violating mean-reversion. 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, unlike when estimating the market prices of risk parameters by minimizing the squared pricing errors. This is obvious when one looks at the very short maturities.[33]

7.2. Term premia

An important use of term structure models consists in quantifying term premia and their time variation. In turn, accurate estimation of term premia are instrumental when testing that expected (excess) bond returns are not forecastable. Alternatively, time-variation in term premia rules out the construct of the expectation hypothesis. There are two, related, notion of term premia, namely the yield term premium $ytp_{n,t} = y^{\$}_{n,t} - \frac{1}{n}\sum_{i=0}^{n-1} E_t(y^{\$}_{1,t+i})$ and the forward term premium $ftp_{n,t} = f^{\$}_{n,t} - E_t(y^{\$}_{1,t+n})$. Thus, to determine both term premia one needs to use the model-implied forecast of the short term interest rate, given by

$$E_t(y^{\$}_{1,t+i}) = A^{\$}_1 + \mathbf{B}^{\$\prime}_{1,1} E_t(\mathbf{C}_{1,t+i}) + \ldots + \mathbf{B}^{\$\prime}_{K,n} E_t(\mathbf{C}_{K,t+i}),$$

where $A^{\$}_1 = \delta_0, \mathbf{B}^{\$}_{j,1} = \delta_j \mathbf{G}$ and $E_t(\mathbf{C}_{j,t+i}) = \mathbf{F}^i \mathbf{C}_{j,t}$ for every $1 \leq j \leq K$.

We consider now the 5 year-for-5 year forward term premia (investing for five years time for a period of five years) evaluated using both the LM and SM models. These are reported in Figure 5. We also plot the forward term premia of Wright (2011) as well as of Bauer et al (2012), here indicated as BRW. One can observe that our SM term premia is rather close to the figures of Wright (2011). For the sake of comparison with other methods, we report quarterly figures for the period from March 1990 to March 2009. The LM term premia co-moves in a similar fashion but, as expected, it it less smooth since it takes into better account the long term volatility of (short term) yields. Finally, the BRW appears much smoother than the other figures, possibly due to the bias correction. The same conclusions are obtained by looking at the cross-correlations between these four measures of forward term premia, with the SM and Wright (2011) being the closest, and with the BRW being (mildly) negatively correlated with the other measures, reported in Table 7. We also report the sample standard deviation of all these forward term premia estimates: as expected, the LM estimate has the largest variance, hence it is the less smooth of all, whereas the BRW has the smallest variance. The SM and Wright estimates are somewhere in between.

[Insert Figure 5 near here]

[Insert Table 7 near here]

[33] By allowing a richer parameterization, one can always fit specific part of the volatility curve such as the short end. This is not the aim of the current exercise.

7.3. Statistical performance

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, for the null hypotheses that the LM and SM are equally close to to the data generating process, using Vuong (1989) test since the obtained specifications for the SM and LM models are non-nested. The short memory model is rejected at 0.1% confidence level in favour of the long memory model. This agrees with the estimates of the measurement errors, much larger for the short memory model.

[**Insert Table 8 near here**]

We have also evaluated the in-sample goodness of fit of the long and short memory models. We first estimate the model using the overall period, from January 1986 to December 2011. Using the estimated parameters and filtered state variables, we evaluate the fitted yield curve for the various maturities. Table 9 shows that, in terms of the root mean square error (RMSE) statistic, the long memory model always provides a superior goodness of fit. The RMSE for the long memory model is at least 30% smaller than for the short memory case, often even smaller.

[**Insert Table 9 near here**]

Out-of-sample

We compare the out of sample forecasting performance of the LM and SM models. The RMSFE (root mean square forecasting error) of the forecasts are reported in Table 10. The evaluation period goes from January 2001 to December 2011, namely the last ten years of our sample. Each month we re-estimate the model, using a rolling window of 192 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 tends to dominate the short memory one in all cases except for the 30-year yield. Their performance is very similar for the 1-year yield.

[**Insert Table 10 near here**]

8. Final remarks

In this paper we introduce a long memory essentially affine model of the term structure, a discrete time essentially affine Gaussian factor term structure model with long memory factors, designed to account for the strong persistence in observed yields and inflation. We provide the closed-form solution of the model and its second-order properties. A detailed characterisation of the long memory implications in terms of the $\mathbb{P}$ and $\mathbb{Q}$ measures' parameters is presented, for both the dynamic and cross-sectional characteristics of the model. Despite the infinite dimensional state variables, we show how estimation of the model can still be 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 one empirical application 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 this respect, allowing for time-varying market prices of risk greatly enhances the relevance of long memory. Given the capability of the long memory model to induce non-negligible volatility of long term yields and forwards, its theoretical and empirical implications in terms of term premia dynamics appear important. Several generalizations are of interest. 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 K factor affine term structure model

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

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

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

where $-1 < \psi_{x,j} < 1$, $u_t \sim NID(0,1)$, $\varepsilon^*_{j,t} \sim NID(0,1)$ mutually independent from 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 $j = 1, ..., K$. No-arbitrage implies the K cross-equation restrictions

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

We wish to evaluate the limiting behaviour of $y_{t,n}$ as $K \to \infty$ 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 $j = 1, ..., K$, mutually independent from one another. Note that by letting $K \to \infty$ the parameters γ_j and σ^2_j must both be $O_p(K^{-1})$, a simple form of which consists of:

$$\gamma_j = \frac{\gamma^*_j}{K}, \;\; \sigma_j = \frac{\sigma^*_j}{K^{\frac{1}{2}}}, \tag{67}$$

where γ^*_j, σ^*_j are i.i.d. random parameters such that $0 <\mid E\gamma^*_j \mid< \infty$ and $0 < E\sigma^{*2}_j < \infty$. To see why (67) is required, note that the variance of $y_{n,t}$ conditional on parameters $\theta =$

$(\theta_1, ..., \theta_K)$ satisfies

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

and (67) ensures that $var(y_{n,t})$ would not increase just because a larger number K of factors $x_{j,t}$ is considered. In other words, the larger is K, 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 $j = 1, ..., K$ 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 $\varepsilon_{j,t}$:

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

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: [34]

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

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

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

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

where

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

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

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

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

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

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 K, real yields $y_{n,t}$ can be expressed, net of constant terms, as the sum of $U_{n,t}$ and $E_{n,t}$, with coefficients satisfying (35) 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 (65), namely

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

holds for some $0 < d < 1/2$ under the above conditions. Note that the limit of $y_{n,t}$ can be viewed as a semiparametric affine model since $U_{n,t}$ and $E_{n,t}$ are function of the infinite sequences of coefficients $\phi_{n,k}, \upsilon_{n,k}$, $k = 0, \ldots$ 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 $H > 1/2$ since

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

Expression (71) is analogue to (36) by setting $H = d + 1/2$. [35] 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 mild forms of arbitrage such as 'free lunch with vanishing risk'.[36] The previous results assumed a frictionless 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 returns 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. 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

$$x_t = \frac{\Theta(L)}{\Psi(L)}(1-L)^{-d}\varepsilon_t = \sum_{i=0}^{\infty} \varphi_i \varepsilon_{t-i},$$

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

$$\begin{aligned} \mathbf{X}_{t+1} &= \mathbf{F}\mathbf{X}_t + \mathbf{h}\varepsilon_t, \\ x_t &= \mathbf{G}'\mathbf{X}_t + \varepsilon_t, \end{aligned} \tag{72}$$

[35] It can be shown that the discrete-time process r_τ, $\tau = 0, \pm 1, ...$ admits a representation (22) with coefficients satisfying (26) and (35).

[36] 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).

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

$$\mathbf{X}_t = \begin{bmatrix} E[x_t \mid x_{t-1}, x_{t-2}, ...] \\ E[x_{t+1} \mid x_{t-1}, x_{t-2}, ...] \\ E[x_{t+2} \mid x_{t-1}, x_{t-2}, ...] \\ \vdots \end{bmatrix},$$

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 $(x_1, ...x_T)$ the exact Gaussian likelihood function can be obtained through the usual Kalman recursion, based on the first T components of the Kalman equations (72). 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 differences satisfy

$$x_t - x_{t-1} = \sum_{i=0}^{\infty} \psi_i^* \varepsilon_t, \;\; \psi_i^* = \psi_i - \psi_{i-1}, \;\; \psi_0^* = 1.$$

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

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

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

$$\mathbf{Y}_{t+1} = \begin{bmatrix} 0 & & \mathbf{I}_m \\ 0 & \cdots & 0 \end{bmatrix} \mathbf{Y}_t + \begin{bmatrix} \psi_1^* \\ \vdots \\ \psi_m^* \end{bmatrix} \varepsilon_t, \tag{73}$$

$$z_t = [\,1\,0 \cdots 0\,]\,\mathbf{Y}_t + \varepsilon_t. \tag{74}$$

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 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

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

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):

$$\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{h}\varepsilon_t$$

where now

$$\mathbf{X}_t = \begin{bmatrix} E[x_t \mid x_t, x_{t-1}, ...] \\ E[x_{t+1} \mid x_t, x_{t-1}, ...] \\ E[x_{t+2} \mid x_t, x_{t-1}, ...] \\ \vdots \end{bmatrix},$$

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

Appendix D. Proof of Theorems 4.3, 4.4, 4.5 and 4.6

We first establish two preliminary lemmas.

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

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

Setting

$$\Phi_{n,0} = \sum_{j=0}^{n-1} \phi_j \tag{75}$$

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

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

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 D.2. *For a finite d assume*

$$\phi_i \sim ci^{d-1} \textit{ as } i \to \infty.$$

Setting

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

then

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

and

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

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 $i \le n$ then $\sum_{j=0}^{n-1}(i+j)^{d-1} \sim c((n+i)^d - i^d) \sim cn^d$ since $n^d \le ((n+i)^d - i^d) \le n^d(2^d - 1)$. 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 $0 < \tilde{n} < n$,

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

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

Proof of Theorem 4.3. Consider the solution for yields (42), which we rewrite as

$$y_{n,t}^{\$} = -n^{-1}p_{n,t}^{\$} = n^{-1}A_n^{\$} + \sum_{i=0}^{\infty}(\Delta_{1,n,i}^{y}, ..., \Delta_{K,n,i}^{y})\varepsilon_{t-i}.$$

Hence, $y_{n,t}^{\$}$ has a linear representation in the i.i.d. innovation ε_t. Then (52) and (53) follow from standard arguments of multivariate time series. For example, for the spectral density result (53) apply Brockwell and Davis (1991), Theorem 4.10.1. The proof applies to forwards and returns. For first-differenced quantities, one simply needs to rewrite yields as:

$$\begin{aligned} y_{n,t}^{\$} - y_{n,t-1}^{\$} &= \Delta_{1,n,0}^{y}\varepsilon_{1,t} + \sum_{i=1}^{\infty}\Delta_{1,n,i}^{y}\varepsilon_{1,t-i} + \ldots + \Delta_{K,n,0}^{y}\varepsilon_{K,t} + \sum_{i=1}^{\infty}\Delta_{K,n,i}^{y}\varepsilon_{K,t-i} \\ &\quad - \sum_{i=0}^{\infty}\Delta_{1,n,i}^{y}\varepsilon_{1,t-i-1} + \ldots + \sum_{i=0}^{\infty}\Delta_{K,n,i}^{y}\varepsilon_{K,t-i-1} \\ &= \Delta_{1,n,0}^{y}\varepsilon_{1,t} + \sum_{i=1}^{\infty}(\Delta_{1,n,i}^{y} - \Delta_{1,n,i-1}^{y})\varepsilon_{1,t-i} + \ldots + \Delta_{K,n,0}^{y}\varepsilon_{K,t} + \sum_{i=1}^{\infty}(\Delta_{K,n,i}^{y} - \Delta_{K,n,i-1}^{y})\varepsilon_{K,t-i} \\ &= \bar{\Delta}_{1,n,0}^{y}\varepsilon_{1,t} + \sum_{i=1}^{\infty}\bar{\Delta}_{1,n,i}^{y}\varepsilon_{1,t-i} + \ldots + \bar{\Delta}_{K,n,0}^{y}\varepsilon_{K,t} + \sum_{i=1}^{\infty}\bar{\Delta}_{K,n,i}^{y}\varepsilon_{K,t-i}. \end{aligned}$$

Then apply the same arguments used for raw data. 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 D.2 can be strengthen to

$$\Phi_{j,n,l} \sim cl^{d_j-1} \text{ as } l \to \infty \text{ for all } 1 \le j \le K.$$

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

The autocovariance of $y_{n,t}^{\$}$ satisfies

$$\text{cov}(y_{n,t}^{\$}, y_{n,t+u}^{\$}) = \sum_{i=0}^{\infty}\mathbf{\Delta}_i^{y\prime}\mathbf{\Sigma}\mathbf{\Delta}_{i+u}^{y} \sim cu^{2\underline{d}-1} \text{ as } u \to \infty,$$

setting $\mathbf{\Delta}_i^{y} = (\Delta_{1,n,i}^{y}...\Delta_{K,n,i}^{y})'$. To show this, we use a truncation argument as follows. First recall that $\text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+u})$ can be written as the sum of K^2 terms, namely $\sum_{a,b=1}^{K} \sigma_{a,b}\Big(\sum_{i=0}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i+u}\Big)$ setting $\sigma_{a,b}$ equal to the (a,b)th term of $\mathbf{\Sigma}$. For the (a,b)th term, ignoring term $\sigma_{a,b}$, from $\sum_{i=0}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i+u} = \sum_{i=0}^{u} \Delta^y_{a,n,i}\Delta^y_{b,n,i+u} + \sum_{i=u+1}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i+u}$ one gets, as $u \to \infty$,

$$\sum_{i=0}^{u} \Delta^y_{a,n,i}\Delta^y_{b,n,i+u} \sim c\Delta^y_{b,n,u} \sum_{i=0}^{u} \Delta^y_{a,n,i} \sim cu^{d_a+d_b-1}$$

and, likewise,

$$\sum_{i=u+1}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i+u} \sim c \sum_{i=u+1}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i} \sim cu^{d_a+d_b-1}.$$

Moreover, by (58), $\text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+u})$ satisfies the quasi-monotonic convergence condition $\left|\text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+u}) - \text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+u+1})\right| = O(u^{-1}|\text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+u})|)$ and the bounded variation condition $\sum_{k=u}^{\infty} \left|\text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+k}) - \text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+k+1})\right| = O(|\ \text{cov}(y^{\$}_{n,t}, y^{\$}_{n,t+u}\ |)$ as $u \to \infty$. In fact, since $|\Delta^y_{a,n,i+1} - \Delta^y_{a,n,i}| \leq ci^{-1}|\Delta^y_{a,n,i}|$ by elementary calculations,

$$\sum_{i=0}^{\infty} |\Delta^y_{a,n,i}||\Delta^y_{b,n,i+u} - \Delta^y_{b,n,i+u+1}| = \sum_{i=0}^{u} |\Delta^y_{a,n,i}||\Delta^y_{b,n,i+u} - \Delta^y_{b,n,i+u+1}| + \sum_{i=u+1}^{\infty} |\Delta^y_{a,n,i}||\Delta^y_{b,n,i+u} - \Delta^y_{b,n,i+u+1}$$

$$\leq c|\Delta^y_{b,n,u} - \Delta^y_{b,n,u+1}| \sum_{i=0}^{u} |\Delta^y_{a,n,i}| + c \sum_{i=u+1}^{\infty} |\Delta^y_{a,n,i}||\Delta^y_{b,n,i} - \Delta^y_{b,n,i+1}|$$

$$\leq cu^{-1}|\Delta^y_{b,n,u}| \sum_{i=0}^{u} |\Delta^y_{a,n,i}| + c \sum_{i=u+1}^{\infty} i^{-1}|\Delta^y_{a,n,i}||\Delta^y_{b,n,i}| \leq cu^{-1}|\Delta^y_{a,n,u}||u^{d_b}| + c \sum_{i=u+1}^{\infty} i^{d_a+d_b-3} \leq cu^{-1}|u^{d_a+d_b-}$$

$$\leq cu^{-1} \sum_{i=0}^{\infty} |\Delta^y_{a,n,i}||\Delta^y_{b,n,i+u}|.$$

Therefore, the conditions of Yong (1974), Lemma III-12, hold concluding the proof. The same proof apply to the spectral density of $f^{\$}_{n,t}$ and $r^{\$}_{n,t}$. QED

Proof of Theorem 4.5. Part (i): the results easily follow by applying Lemma D.1 to the conditional variances formulae (51).

Part (ii): for the unconditional variances, use Lemma D.2 together with a truncation argument as follows. Since $\text{var}(y^{\$}_{n,t}) = \sum_{a,b=1}^{K} \sigma_{a,b} \sum_{i=0}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i}$, considering the (a,b)th element

$$\sum_{i=0}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i} = \sum_{i=0}^{n} \Delta^y_{a,n,i}\Delta^y_{b,n,i} + \sum_{i=n+1}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i},$$

then

$$\sum_{i=0}^{n} \Delta^y_{a,n,i}\Delta^y_{b,n,i} = O(n^{-2} \sum_{i=0}^{n} (n^{d_a} + i^{d_a})(n^{d_b} + i^{d_b})) = O(n^{d_a+d_b-1}),$$

and

$$\sum_{i=n+1}^{\infty} \Delta^y_{a,n,i}\Delta^y_{b,n,i} = O(\sum_{i=n+1}^{\infty} i^{d_a+d_b-2}) = O(n^{d_a+d_b-1}).$$

A similar reasoning applies to $\mathrm{var}(f^{\$}_{t,n})$. For $\mathrm{var}(r^{\$}_{t,n})$ notice that since $d_j < 1/2$ then $\sum_{l=0}^{\infty} \phi_{j,l}^2$ are bounded for every $1 \leq j \leq K$.

Part (iii): since $y^{\$}_{n,t} - y^{\$}_{n,t-1} = \Big(\Delta^y_{1,n,0}\varepsilon_{1,t} + \sum_{i=1}^{\infty}(\Delta^y_{1,n,i} - \Delta^y_{1,n,i-1})\varepsilon_{1,t-i}\Big) + \ldots + \Big(\Delta^y_{K,n,0}\varepsilon_{K,t} + \sum_{i=1}^{\infty}(\Delta^y_{K,n,i} - \Delta^y_{K,n,i-1})\varepsilon_{1,t-i}\Big) = \Big(\Delta^y_{1,n,0}\varepsilon_{1,t} + \sum_{i=1}^{\infty} \tilde{\Delta}^y_{1,n,i}\varepsilon_{1,t-i}\Big) + \ldots + \Big(\Delta^y_{K,n,0}\varepsilon_{K,t} + \sum_{i=1}^{\infty} \tilde{\Delta}^y_{K,n,i}\varepsilon_{1,t-i}\Big)$, one needs to find the asymptotic behaviour of the $\tilde{\Delta}^y_{a,n,i} = \Delta^y_{a,n,i} - \Delta^y_{a,n,i-1}$ as i and n diverge, for every $1 \leq a \leq K$. Using the results of part (ii)

$$\tilde{\Phi}_{a,n,i} = \Phi_{a,n,i} - \Phi_{a,n,i-1} \sim c \begin{cases} n i^{d_a - 2} & \text{as } n/i \to 0, \\ i^{d_a - 1} & \text{as } i/n \to 0, \end{cases}$$

used below into

$$\tilde{\Delta}^y_{a,n,i} = n^{-1}\tilde{\Phi}^y_{a,n,i}.$$

It easily follows that the stationarity condition for $y^{\$}_{n,t} - y^{\$}_{n,t-1}$ for every given n, namely square-summability of the $\tilde{\Delta}^y_{a,n,i}$, is now $\underline{d} < 3/2$. The same condition emerges for stationarity of the first difference of forwards $f^{\$}_{n,t} - f^{\$}_{n,t-1}$ and rates $r^{\$}_{n,t} - r^{\$}_{n,t-1}$.

The result for yields follows since the (a,b)th term of $\mathrm{var}(y^{\$}_{n,t} - y^{\$}_{n,t-1})$ is $\sigma_{ab}\Big(\Delta^y_{a,n,0}\Delta^y_{b,n,0} + \sum_{i=1}^{\infty} \tilde{\Delta}^y_{a,n,i}\tilde{\Delta}^y_{b,n,i}\Big)$. For forward rates we use

$$\tilde{\Delta}^f_{a,n,i} = \tilde{\Phi}_{a,n+1,i} - \tilde{\Phi}_{a,n,i} \sim c \begin{cases} i^{d_a - 2} & \text{as } n/i \to 0, \\ 0 & \text{as } i/n \to 0, \end{cases}$$

and for returns

$$\tilde{\Delta}^r_{a,n,i} = \tilde{\Phi}_{a,n,i-1} - \tilde{\Phi}_{a,n-1,i} \sim c \begin{cases} i^{d_a - 2} + (n-1) i^{d_a - 3} & \text{as } n/i \to 0, \\ i^{d_a - 2} & \text{as } i/n \to 0, \end{cases}$$

However, as for part (ii), the dominating term of $\mathrm{var}_{t-1}(r^{\$}_{n,t} - r^{\$}_{n,t-1})$, as n diverges, is $\Delta^r_{a,n,0}\Delta^r_{b,n,0}$ as opposed to $\sum_{i=1}^{\infty} \tilde{\Delta}^r_{a,n,i}\tilde{\Delta}^r_{b,n,i}$. QED

Proof of Theorem 4.6. Part (i): consider the $\Phi_{a,n,l}$ for any $1 \leq a \leq K$. Since $b_{a,l} \sim c l^{\underline{d}}$ as $l \to \infty$, re-writing

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

one gets $I = \sum_{i=0}^{[n/2]} b_{a,n-i}\phi_{a,i+l} \sim c n^{\underline{d}} \sum_{i=0}^{[n/2]} (i+l)^{d_a - 1} \sim c n^{\underline{d}}(([n/2]+l)^{d_a} - l^{d_a})$. Hence

$$I \sim \begin{cases} c n^{\underline{d} + d_a} & \text{as } l/n \to 0, \\ c n^{\underline{d}+1} l^{d_a - 1} & \text{as } n/l \to 0, \end{cases} \quad \text{and } II \sim c(l+n)^{\underline{d}-1} n^{d_a + 1}.$$

For the conditional variance of yields $y^{\$}_{n,t}$ the result follows simply by using the above formulae with $l = 0$ into $\Delta_{a,n,l}$ for all $1 \leq a \leq K$.

Part (ii): a truncation argument leads to the unconditional variance result applying the above formulae for the two cases $0 \leq l \leq n$ and $l > n$ to $\sum_{l=0}^{\infty} \Delta_{a,n,l}^{y} \Delta_{b,n,l}^{y} = \sum_{l=0}^{n} \Delta_{a,n,l}^{y} \Delta_{b,n,l}^{y} + \sum_{l=n+1}^{\infty} \Delta_{a,n,l}^{y} \Delta_{b,n,l}^{y}$. In fact as $n \to \infty$

$$\sum_{l=0}^{n} \Delta_{a,n,l}^{y} \Delta_{b,n,l}^{y} \sim n^{-2} \sum_{l=0}^{n} n^{d_a+d_b} (n^{\underline{d}} + n(l+n)^{\underline{d}-1})^2 \sim c n^{d_a+d_b} n^{2\underline{d}-1}$$

and

$$\sum_{l=n+1}^{\infty} \Delta_{a,n,l}^{y} \Delta_{b,n,l}^{y} \sim n^{-2} \sum_{l=n+1}^{\infty} (n^{\underline{d}+1} l^{d_a-1} + (l+n)^{\underline{d}-1} n^{d_a+1})(n^{\underline{d}+1} l^{d_b-1} + (l+n)^{\underline{d}-1} n^{d_b+1}) \sim c n^{d_a+d_b} n^{2\underline{d}-1}.$$

With respect to forwards $f_{n,t}^{\$}$ the conditional variance result follows from

$$\Phi_{a,n+1,0} - \Phi_{a,n,0} \sim c((n+1)^{\underline{d}+d_a} - n^{\underline{d}+d_a}) \sim c n^{\underline{d}+d_a-1},$$

whereas for their unconditional variance we use

$$\Phi_{a,n+1,l} - \Phi_{a,n,l} \sim c n^{d_a} (n^{\underline{d}-1} + n(l+n)^{\underline{d}-2} + (l+n)^{\underline{d}-1}) \sim c n^{\underline{d}+d_a-1} \text{ when } 0 \leq l \leq n,$$

and

$$\Phi_{a,n+1,l} - \Phi_{a,n,l} \sim c(l^{d_a-1} n^{\underline{d}} + n^{d_a+1}(l+n)^{\underline{d}-2} + n^{d_a}(l+n)^{\underline{d}-1}) \sim c(l^{d_a-1} n^{\underline{d}} + n^{d_a+1} l^{\underline{d}-2} + n^{d_a} l^{\underline{d}-1}) \text{ when } l > n,$$

with a truncation argument into $\sum_{l=0}^{\infty} \Delta_{a,n,l}^{f} \Delta_{b,n,l}^{f}$. Finally, for returns $r_{n,t}^{\$}$ the result follows straightforwardly substituting I and II with $l = 0$ into their conditional variance expression, whereas in terms of their unconditional variance one uses

$$\Phi_{a,n,l-1} - \Phi_{a,n-1,l} \sim c(n^{d_a+\underline{d}} - (n-1)^{d_a+\underline{d}}) + (l+n)^{\underline{d}-1}(n^{d_a+1} - (n-1)^{d_a+1}) \sim n^{\underline{d}+d_a-1} \text{ when } 0 \leq l \leq n,$$

and

$$\begin{aligned} \Phi_{a,n,l-1} - \Phi_{a,n-1,l} \quad & \sim c((l-1)^{d_a-1} n^{\underline{d}+1} - l^{d_a-1}(n-1)^{\underline{d}+1}) + (l+n)^{\underline{d}-1}(n^{d_a+1} - (n-1)^{d_a+1}) \\ & \sim c(l^{d_a-1} n^{\underline{d}} + l^{d_a-2} n^{\underline{d}+1} + n^{d_a} l^{\underline{d}-1}) \text{ when } l > n, \end{aligned}$$

since $(l+1)^d - l^d = d(l+\varepsilon)^{d-1} \sim c l^{d-1}$ for some $0 < \varepsilon < 1$ as $l \to \infty$.

Part (iii): the proof follows precisely the proof of part (iii) to Theorem 4.5. We simply need to use

$$\tilde{\Phi}_{a,n,i} = \Phi_{a,n,i} - \Phi_{a,n,i-1} \sim c \begin{cases} n^{\underline{d}+1}(n+i)^{d_a-2} + n^{\underline{d}+1} i^{d_a-2} & \text{as } n/i \to 0, \\ n^{\underline{d}+1}(n+i)^{d_a-2} & \text{as } i/n \to 0, \end{cases}$$

For yields, we then use the previous result into

$$\tilde{\Delta}_{a,n,i}^{y} = n^{-1} \tilde{\Phi}_{a,n,i}^{y}.$$

For forward rates we now obtain

$$\tilde{\Delta}^f_{a,n,i} = \tilde{\Phi}_{a,n+1,i} - \tilde{\Phi}_{a,n,i} \sim c \begin{cases} n^{\underline{d}}(n+i)^{d_a-2} + n^{\underline{d}+1}(n+i)^{d_a-3} + n^{\underline{d}} i^{d_a-2} & \text{as } n/i \to 0, \\ n^{\underline{d}}(n+i)^{d_a-2} + n^{\underline{d}+1}(n+i)^{d_a-3} & \text{as } i/n \to 0, \end{cases}$$

and for returns

$$\tilde{\Delta}^r_{a,n,i} = \tilde{\Phi}_{a,n,i-1} - \tilde{\Phi}_{a,n-1,i} \sim c \begin{cases} n^{\underline{d}}(n+i-1)^{d_a-2} + n^{\underline{d}} i^{d_a-2} + n^{\underline{d}+1} i^{d_a-3} & \text{as } n/i \to 0, \\ n^{\underline{d}}(n+i-1)^{d_a-2} & \text{as } i/n \to 0, \end{cases}$$

In all cases the stationarity condition is now $d_j < 3/2$ for every $1 \leq j \leq K$. QED

References

[1] Abbritti M. and Gil-Alana L. and Lovcha Y. and Moreno A., 2015, "Term structure persistence", *Journal of Financial Econometrics* , forthcoming.

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

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

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

[5] 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.

[6] 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.

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

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

[9] Barnichon, R., 2010, "Building a composite Help-Wanted Index", Economics Letters, vol 109, 175-178.

[10] Barr, D.F., and Campbell J.Y., 1997, "Inflation, real interest rates, and the bond market: a study of UK nominal and index-linked government bond prices", Journal of Monetary Economics, vol 39, 361-383.

[11] 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.

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

[13] Bos, C.S., Koopman, S.J. and Ooms, M., 2014, "Long memory with stochastic variance model: A recursive analysis for US inflation", Computational Statistics & Data Analysis, vol 76, 144-157.

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

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

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

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

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

[19] Chernov, M. and Mueller, P., 2012, "The term structure of inflation expectations", Journal of Financial Economics, vol 106, 367-394.

[20] Cheung, Y.W. and Chung, S.K., 2009, "A Long Memory Model with Mixed Normal GARCH for US Inflation Data", working paper.

[21] Christensen J.H.E. and Gillan J.M., 2012, "Could the U.S. Treasury benefit from issuing more TIPS?" FRBSF working paper.

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

[23] 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.

[24] 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.

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

[26] 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.

[27] D'Amico, S., Kim, D.H. and Min W., 2014, "Tips from TIPS: the informational content of Treasury inflation-protected security prices", FEDS Working Paper No. 2014-24.

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

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

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

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

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

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

[34] Evans, M.D., 1998, "Real rates, expected inflation, and inflation risk premia", Journal of Finance, vol 53, 187-218.

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

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

[37] 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.

[38] 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.

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

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

[41] Haubrich, J., Pennacchi, G.G. and Ritchken, P., 2012, "Inflation expectations, real rates, and risk premia: evidence from inflation swaps", Review of Financial Studies, vol 25, 1588-1629.

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

[43] Hyung, N., Franses, P.H. and Penm, J., 2006, “Structural breaks and long memory in US inflation rates: do they matter for forecasting”, Research in International Business and Finance, vol 20 , 95-110.

[44] 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.

[45] Joslin S., Priebsch M. and Singleton K. J., 2014, “Risk premiums in dynamic term structure models with unspanned macro risks”, Journal of Finance, vol 69, 1197-1233.

[46] 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.

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

[48] 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.

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

[50] 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.

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

[52] Osterrieder D., 2013, “Interest rates with long memory: a generalized affine term-structure model”, CREATES Research Paper No. 2013-17.

[53] Pennacchi, G.G., 1991, “Identifying the dynamics of real interest rates and inflation: evidence using survey data”, Review of Financial Studies, vol 4, 53-86.

[54] Robinson, P.M., 1978, “Statistical Inference for a random coefficient autoregressive model”, Scandinavian Journal of Statistics, vol 5, 163168.

[55] Robinson, P.M., 1995, “Log-Periodogram Regression of Time Series with Long Range Dependence”, Annals of Statistics, vol 23, 10481072.

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

[57] 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.

[58] 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.

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

[60] Vuong Q.H., 1989, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", Econometrica, vol 57, 307-333

[61] Wright J.H., 2011, "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset", American Economic Review, vol 101, 1514-34.

[62] Yong, C.H., 1974, "Asymptotic behaviour of trigonometric series", Chinese University of Hong Kong, monograph.

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

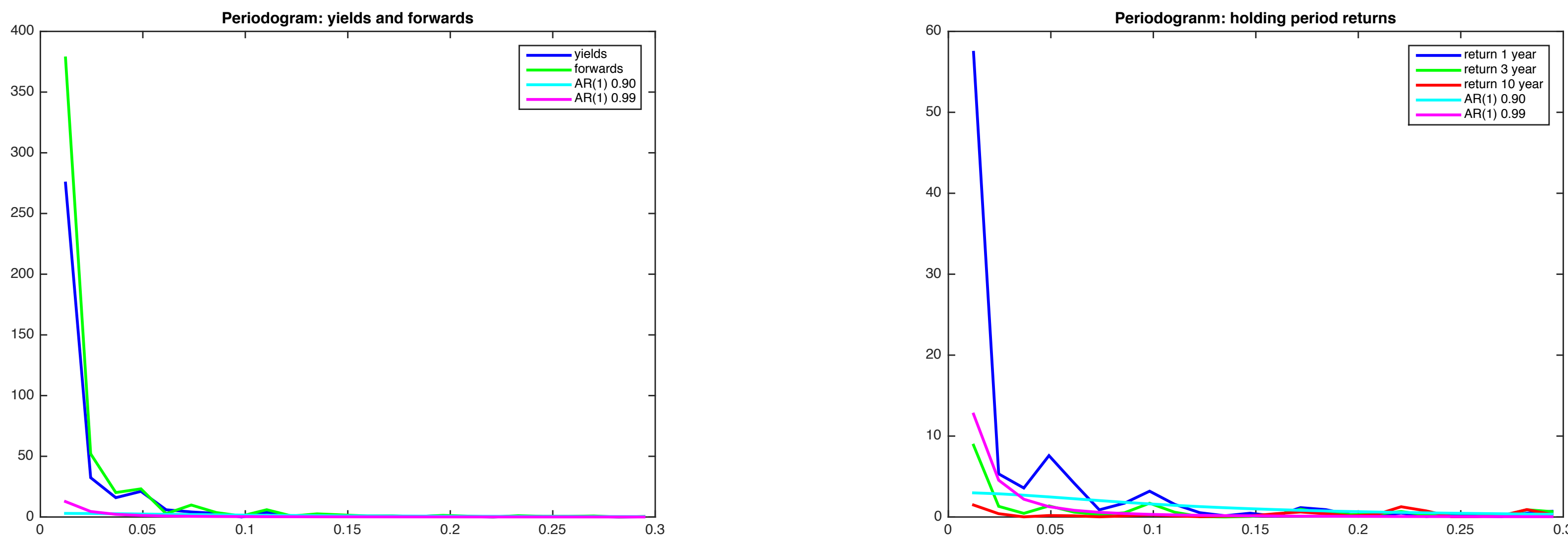

Fig. 1. Panel (a) shows the plot of the periodogram ordinates near the zero frequency for nominal yields (blue), forwards (green) averaged across maturities. Panel (b) shows the plot of the periodogram ordinates near the zero frequency for returns with maturities 1-year (blue), 3-year (green), 10-year (red), 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 on both panels 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 0.90 (light blue line), 0.99 (purple line). On the horizontal axis the numbers $1 \leq j \leq 25$ refer to the first 25 ordinates of the frequencies $\lambda_j = 2\pi j/512$.

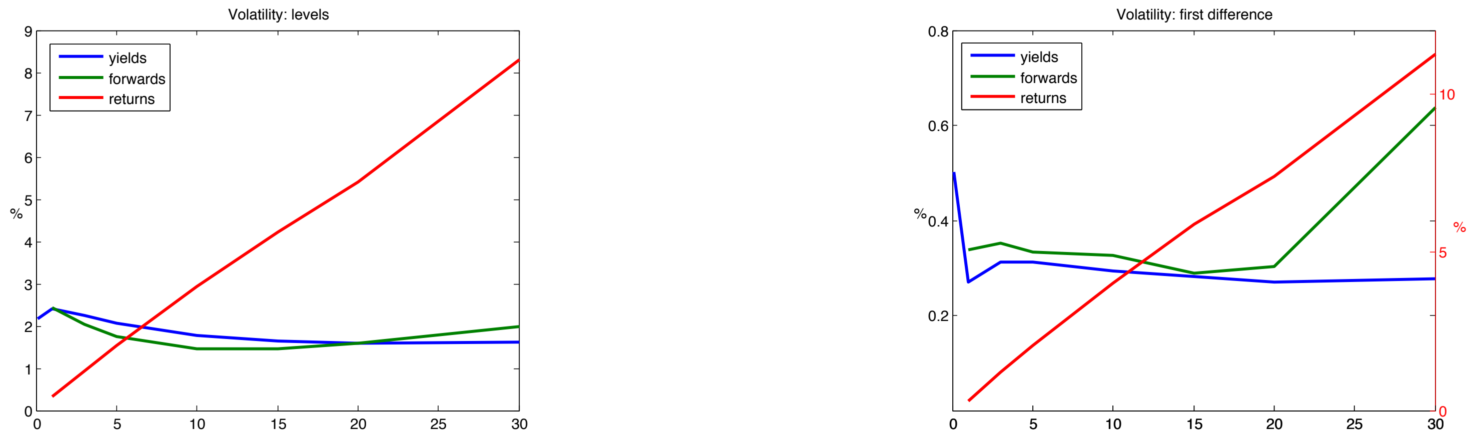

Fig. 2. We plot the sample standard deviation across maturity for nominal yields, forwards and returns in levels (left panel) and in first differences (right panel), where for a sample of generic observables $(w_1, ...w_T)$ the sample standard deviation is defined as

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

For the right hand side panel, the scale for the first-differenced returns is reported on the right hand side vertical axis and the scale for first-differenced yields and forwards is reported on the left hand side vertical axis.

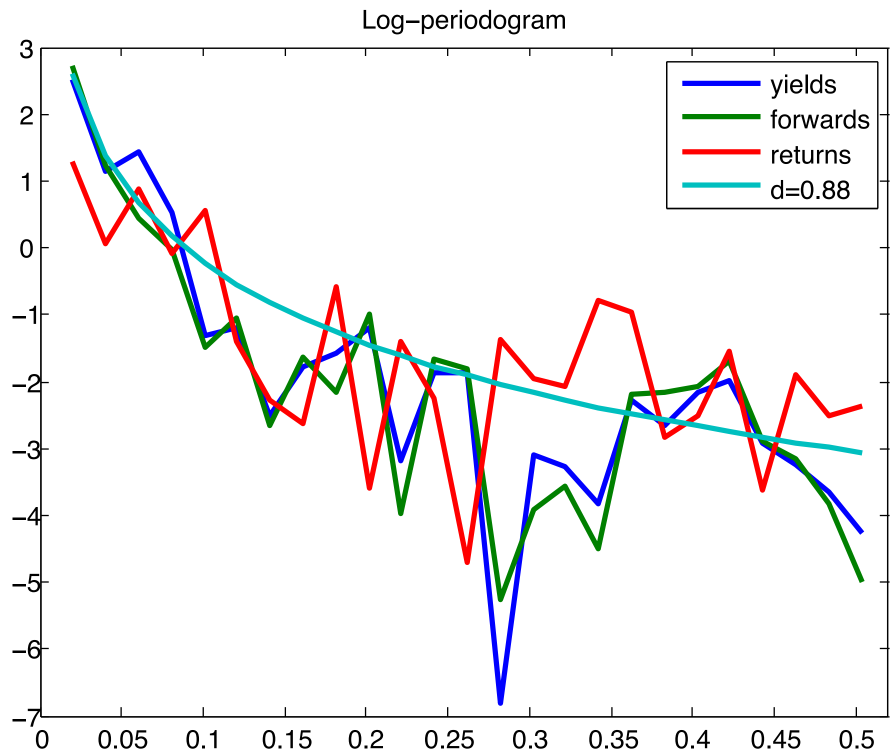

Fig. 3. We plot the logarithm of the periodogram ordinates near the zero frequency for nominal yields (blue line) and forwards (green line), averaged across maturity, and 1 year-returns (red line), where for a sample of generic observables $(w_1, ...w_T)$ the periodogram is $I_w(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} w_t e^{\imath\lambda t}\right|^2$, $-\pi < \lambda \leq \pi$ together with the spectral density (light blue line)

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

with long memory parameter $d = 0.88$. On the horizontal axis the numbers $1 \leq j \leq 25$ refer to the first 25 frequencies $\lambda_j = 2\pi j/256$.

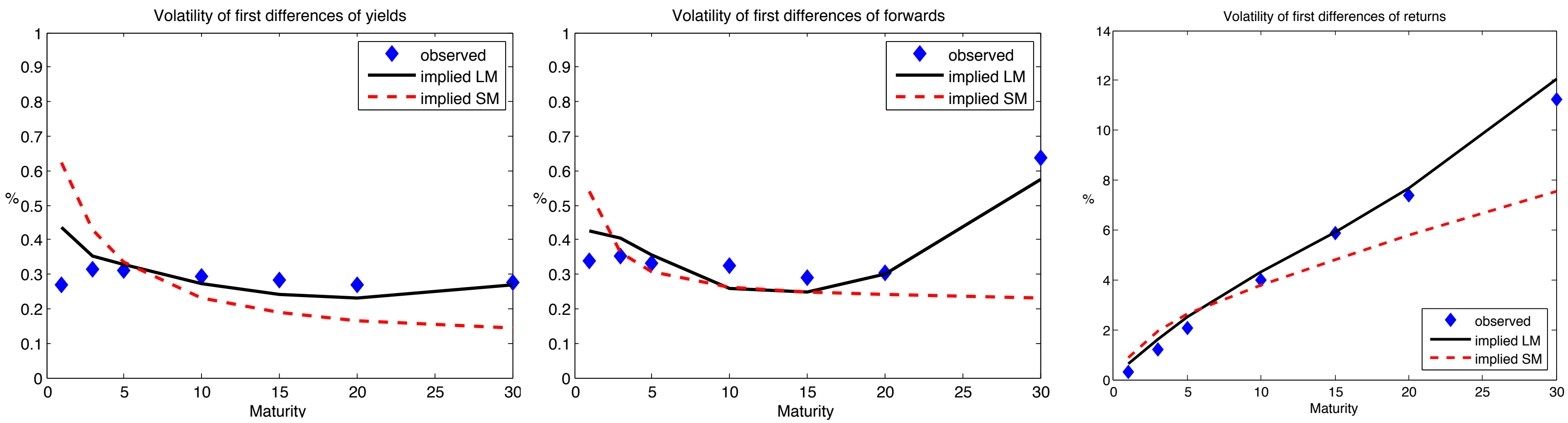

Fig. 4. We report the term structure of the sample standard deviation (blue line) and of the corresponding estimated model-implied standard deviation for the long memory model (green line) and short memory model (red line) for the *first difference* of nominal yields (left panel), nominal forwards (centre 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.

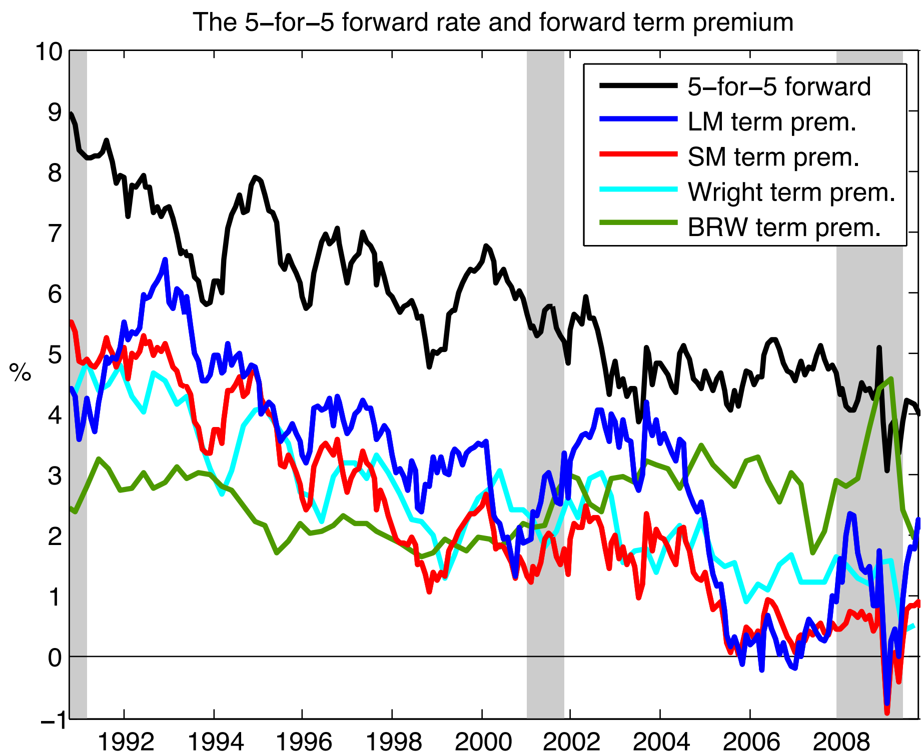

Fig. 5. We plot the forward term premia 5year-for-5year $ftp_{n,t} = f^{\$}_{n,t} - E_t(y^{\$}_{1,t+n})$ implied by our model (LM blue line, SM red line) where the forecast of the 1-month yield is obtained as

$$E_t(y^{\$}_{1,t+i}) = A^{\$}_1 + \mathbf{B}^{\$\prime}_{1,1} E_t(\mathbf{C}_{1,t+i}) + \ldots + \mathbf{B}^{\$\prime}_{K,n} E_t(\mathbf{C}_{K,t+i}),$$

with $A^{\$}_1 = \delta_0, \mathbf{B}^{\$}_{j,1} = \delta_j \mathbf{G}$ and $E_t(\mathbf{C}_{j,t+i}) = \mathbf{F}^i \mathbf{C}_{j,t}$ for every $1 \leq j \leq K$. We used the parameters' values of Table 4 and Table 5 for the LM and SM case respectively. We also plot the Wright (2011) estimate (light blue line) and the Bauer et al (2012) estimate (green line) of the forward term premium. The shaded areas represent recessions determined by the NBER's Business Cycle Committee. We report figures in quarterly frequency over the period March 1990 to March 2009.

Maturity	1 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y
					Panel A: Yields			
Mean	3.70	4.39	4.90	5.29	5.97	6.30	6.40	6.31
Std Dev	2.18	2.41	2.26	2.07	1.77	1.64	1.59	1.62
Skew	-0.17	-0.22	-0.22	-0.12	0.11	0.18	0.17	0.13
Ex. Kurtosis	-0.81	-0.90	-0.84	-0.86	-0.94	-0.97	-1.01	-1.06
Min	0.00	0.13	0.39	0.88	1.98	2.63	2.92	2.50
Max	8.67	9.66	9.46	9.32	9.64	9.85	9.96	10.16
					Panel B: Forward rates			
Mean		4.67	5.54	6.17	6.92	6.89	6.55	5.73
Std Dev		2.43	2.04	1.75	1.45	1.46	1.59	1.97
Skew		-0.25	-0.10	0.16	0.35	0.15	-0.01	0.22
Ex. Kurtosis		-0.88	-0.88	-0.97	-0.84	-0.96	-0.82	-0.69
Min		0.09	1.00	2.09	3.76	3.50	2.12	0.55
Max		9.72	9.54	9.89	10.40	10.53	10.64	10.69
					Panel C: Returns			
Mean		0.41	0.54	0.64	0.81	0.91	0.99	1.16
Std Dev		0.32	0.94	1.55	2.93	4.23	5.40	8.30
Skew		0.50	-0.03	-0.15	0.03	0.15	0.26	0.47
Ex. Kurtosis		0.09	-0.03	0.07	1.31	2.13	2.56	2.50
Min		-0.26	-2.23	-3.97	-9.73	-14.52	-18.18	-24.97
Max		1.46	3.03	5.05	12.47	19.34	26.82	41.29

Table 1: Summary statistics for zero coupon monthly yields, forward rates and one-month holding period returns. The 1 to 30 year yields are obtained from Gurkaynak et al (2007) available from the website of the Federal Reserve Board. The one-month holding period returns require yields with maturity 1 year and 11 month, 2 year and 11 month up to 29 year and 11 month, which are also available from the same source. The 1 month yield comes from the Fama's Treasury Bills Term Structure Files available from CRSP. All yields are continuously compounded. The data sample is 1986:01-2011:12.

	1 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y
μ	0.045	0.009	0.020	0.027	0.0030	0.004	0.009	0.024
$t-value$	0.804	0.287	0.470	0.547	0.050	0.0670	0.137	0.415
γ	−0.019	−0.006	−0.008	−0.009	−0.004	−0.004	−0.004	−0.007
$t-value$	−1.454	−0.972	−1.021	−1.029	−0.431	−0.406	−0.457	−0.727

Table 2: Augmented Dickey-Fuller test of the 1 month, 1, 3, 5, 10, 15, 20, and 30 year yields. 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. We refer to Table 1 for a description of the dataset.

	1 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y
d	0.93	0.80	0.71	0.66	0.64	0.62	0.60	0.69

Table 3: Estimates of the long memory parameter, and its standard error, based on the log-periodogram regression of Robinson (1995). The bandwidth is equal to 17, namely the square root of the sample size. The standard error is 0.16. The sample period is 1986:01 to 2011:12.

	ψ	d	θ_1	θ_2	σ
inflation		0.2667 0.0401	$-$	$-$	0.0091 0.0004
real activity	0.9371 0.0222	$-$	-0.6896 0.0620	0.1962 0.0629	0.0088 0.0004
latent 1	0.8752 0.0007	0.8836 0.0043	$-$		1
latent 2	0.9915 0.0002	0.1912 0.0089	-0.5094 0.0135		1
latent 3	$-$	0.4451 0.0021	$-$		1
δ_0	0.0370 0.0011				
$\delta_1' \times 100$	70.8251 10.5560	51.0703 7.4003	0.0037 0.0000	0.4951 0.0008	0.6697 0.0030
σ_v	0.0009 0.0001				
λ_0'	300.3477 23.6467	73.4471 20.1468	1.0638 0.0446	-1.5065 0.1385	-1.2429 0.0892
λ_1	-4050.3335 31.0808	-1064.5465 102.1187	0	0	0
	1155.7683 18.9950	-463.5397 60.4596	0	0	0
	0	0	0.0003 0.0000	0.0082 0.0003	0.0622 0.0001
	0	0	0.0065 0.0000	0.0238 0.0009	0.1894 0.0003
	0	0	0.0038 0.0000	-0.0082 0.0014	-0.0364 0.0005

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

$$\begin{pmatrix} y^{\$}_{n_1,t} \\ y^{\$}_{n_2,t} \\ \vdots \\ y^{\$}_{n_9,t} \end{pmatrix} = \begin{pmatrix} n_1^{-1}A^{\$}_{n_1} \\ n_2^{-1}A^{\$}_{n_2} \\ \vdots \\ n_9^{-1}A^{\$}_{n_9} \end{pmatrix} + \begin{pmatrix} n_1^{-1}\mathbf{B}^{\$\prime}_{\pi,n_1} \\ n_2^{-2}\mathbf{B}^{\$\prime}_{\pi,n_2} \\ \vdots \\ n_9^{-1}\mathbf{B}^{\$\prime}_{\pi,n_9} \end{pmatrix} \hat{\mathbf{C}}_{\pi,t} + \begin{pmatrix} n_1^{-1}\mathbf{B}^{\$\prime}_{g,n_1} \\ n_2^{-1}\mathbf{B}^{\$\prime}_{g,n_2} \\ \vdots \\ n_9^{-1}\mathbf{B}^{\$\prime}_{g,n_9} \end{pmatrix} \hat{\mathbf{C}}_{g,t} + \sum_{j=1}^{3} \begin{pmatrix} n_1^{-1}\mathbf{B}^{\$\prime}_{j,n_1} \\ n_2^{-1}\mathbf{B}^{\$\prime}_{j,n_2} \\ \vdots \\ n_9^{-1}\mathbf{B}^{\$\prime}_{j,n_9} \end{pmatrix} \mathbf{C}_{j,t} + \begin{pmatrix} \nu_{n_1,t} \\ \nu_{n_2,t} \\ \vdots \\ \nu_{n_9,t} \end{pmatrix}$$

where the coefficients $A^{\$}_n, \mathbf{B}^{\$}_{l_j,n}$, for $1 \leq j \leq 5$ are given in Theorem 4.1. The $\mathbf{C}_{l_j,t}$ are the state variables, for $1 \leq j \leq 5$, satisfying the transition equations

$$\mathbf{C}_{l_j,t+1} = \mathbf{F}\mathbf{C}_{l_j,t} + \mathbf{h}_{l_j}\varepsilon_{l_j,t+1},$$

for $l_j \in \{\pi,\, g,\, 1,\, 2,\, 3,\}$ with $\varepsilon_{l_j,t} \sim NID(0, \sigma^2_{l_j})$ mutally independent and where $\mathbf{F}$ and $\mathbf{h}_{l_j}$ are defined in (21) and (23) respectively. The factors $x_{l_j,t} = \mathbf{G}'\mathbf{C}_{l_j,t}$, with $\mathbf{G} = (1,\, 0,\, 0...)'$, are ARFIMA$(p,d,q)$, with $0 \leq p, q \leq 2$ (selected with the BIC criteria):

$$\psi_{l_j}(L)(1-L)^{d_{l_j}} x_{l_j,t} = \theta_{l_j}(L)\varepsilon_{l_j,t}.$$

The measurement errors satisfy $\nu_{n,t} \sim NID(0, \sigma^2_\nu)$. All innovations are mutually independent. Robust standard errors are reported in small font. The model is estimated by using the approximate maximum likelihood estimator proposed by Chan and Palma (1998) with the truncation lag set to 60. See Appendix C. The sample period is 1986:01 to 2011:12. The filtered values $\hat{\mathbf{C}}_{\pi,t}, \hat{\mathbf{C}}_{g,t}$, together with the estimates of δ_0 and $\delta_{l_j}, \psi_{l_j}, \theta_{l_j}, d_{l_j}, \sigma_{l_j}$, with $l_j \in \{\pi, g\}$, are obtained by means of preliminary univariate estimation with the approximate maximum likelihood estimation.

	ψ	d	θ_1	θ_2	σ
inflation	0.9493	–	−0.9759	–	0.0091
	0.0335		0.0761		0.0004
real activity	0.9371	–	−0.6896	0.1962	0.0088
	0.0222		0.0620	0.0629	0.0004
latent 1	0.9961	–	−0.1543	–	1
	0.0001		0.0122		
latent 2	0.9967	–	−0.5834	–	1
	0.0001		0.0018		
latent 3	0.8478	–	−0.3717	–	1
	0.0066		0.0243		
δ_0	0.0370				
	0.0011				
$\delta_1' \times 100$	70.8251	51.0703	0.0001	0.8565	0.4565
	10.5560	7.4003	0.0008	0.0010	0.0029
σ_v	0.0012				
	0.0001				
λ_0'	287.1659	114.0139	0.0812	−2.8640	−0.9052
	3.5804	9.3998	0.0077	0.0194	0.0598
Λ_1	−3737.8014	−1529.1777	0	0	0
	1.5629	6.3800			
	1032.0323	25.5759	0	0	0
	3.1958	9.3681			
	0	0	0.0030	0.0016	0.0614
			0.0000	0.0000	0.0001
	0	0	0.0260	0.0245	0.1810
			0.0001	0.0001	0.0001
	0	0	0.0042	−0.0222	0.0109
			0.0006	0.0002	0.0077

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

$$\begin{pmatrix} y^{\$}_{n_1,t} \\ y^{\$}_{n_2,t} \\ \vdots \\ y^{\$}_{n_k,t} \\ \pi_t \\ g_t \end{pmatrix} = \begin{pmatrix} \widetilde{A}^{\$}_{n_1} \\ \widetilde{A}^{\$}_{n_2} \\ \vdots \\ \widetilde{A}^{\$}_{n_k} \\ \mu_\pi \\ \mu_g \end{pmatrix} + \begin{pmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{\pi,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{\pi,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{\pi,n_k} \\ \mathbf{G}' \\ 0 \end{pmatrix} \mathbf{C}_{\pi,t} + \begin{pmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{g,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{g,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{g,n_k} \\ 0 \\ \mathbf{G}' \end{pmatrix} \mathbf{C}_{g,t} + \sum_i \begin{pmatrix} \widetilde{\mathbf{B}}^{\$\prime}_{l_i,n_1} \\ \widetilde{\mathbf{B}}^{\$\prime}_{l_i,n_2} \\ \vdots \\ \widetilde{\mathbf{B}}^{\$\prime}_{l_i,n_k} \\ 0 \\ 0 \end{pmatrix} \mathbf{C}_{l_i,t} + \begin{pmatrix} \nu_{n_1,t} \\ \nu_{n_2,t} \\ \vdots \\ \nu_{n_k,t} \\ 0 \\ 0 \end{pmatrix}$$

where $\mathbf{C}_{l_i,t}$ is the state vector of the $i-$th latent factor. $\widetilde{A}^{\$}_n, \widetilde{\mathbf{B}}^{\$}_{i,n}$ are defined in Theorem 4.1, and transition equations

$$\mathbf{C}_{i,t+1} = \mathbf{F}\mathbf{C}_{i,t} + \mathbf{h}_i \varepsilon_{i,t+1},$$

for $i \in \{\pi,\, g,\, l_1,\, l_2,\, l_3,\}$ with $\epsilon_{i,t} \sim NID(0, \sigma_i^2)$ mutally independent and where $\mathbf{F}$ and $\mathbf{h}_i$ are defined in (21) and (23) respectively. The factors $x_{l_j,t} = \mathbf{G}'\mathbf{C}_{l_j,t}$, with $\mathbf{G} = (1,\, 0,\, 0...)'$, are ARMA$(p, q)$, with $0 \leq p, q \leq 2$ (selected with the BIC criteria):

$$\psi_{l_j}(L) x_{l_j,t} = \theta_{l_j}(L) \epsilon_{l_j,t}.$$

The measurement errors satisfy $\nu_{n,t} \sim NID(0, \sigma_n^2)$. All innovations are mutually independent. 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 C. The sample period is 1986:01 to 2011:12.

Regressor:	Regressand: PC_1	PC_2	PC_3	PC_4	PC_5
inflation	0.1886	-0.0020	0.0081	-0.0006	0.0049
real activity	0.1697	0.0200	0.0226	0.2011	-0.0018
latent 1	0.3403	0.4148	0.0023	0.0613	-0.0023
latent 2	0.7043	0.1556	-0.0001	0.0524	-0.0031
latent 3	0.1649	0.0251	0.3356	0.0149	-0.0030
macro (infl.+real act.)	0.3006	0.0211	0.0392	0.1998	0.0049
latent all	0.9407	0.6756	0.3812	0.1316	-0.0075
macro + latent 1	0.4900	0.5402	0.0475	0.2156	0.0025
macro + latent 2	0.8305	0.1849	0.0408	0.2900	0.0027
macro + latent 3	0.3080	0.0602	0.6743	0.3095	0.0030

Table 6: We report the regression adjusted R^2 from projecting each of the first five principal components, extracted from the set of nominal yields in our sample, on the macro variables and filtered factors. The sample period is 1986:01 to 2011:12.

	LM	SM	Wright	BRW
Std.dev.	1.64	1.56	1.14	0.60
	Correlations:			
LM	1			
SM	0.8939	1		
Wright	0.7988	0.9244	1	
BRW	-0.0994	-0.0866	-0.0987	1

Table 7: Standard deviations and correlation matrix of the forward term premium implied by the LM and SM models and the term premium calculated by Wright (2011) and Bauer, Rudebusch and Wu (2014) (here denoted by BRW). The term premium calculated by Wright was replicated by Bauer, Rudebusch and Wu and both series are available on their paper's AEA website. We align the monthly estimates of the LM and SM to the quarterly Wright's and BRW's term premia over the period 1990:3 to 2009:3.

	LM	SM
k	37	38
ℓ_{infl}	1,023.1623	1,022.8581
$\ell_{real.act.}$	1,032.4668	1,032.4668
ℓ_{yields}	12,063.7244	11,733.8974
ℓ_{Total}	14,119.3535	13,789.2223
BIC	-27,941.0202	-27,272.7122
$LR\ test$	7.6819	
$p{-}value$	0.0000	

Table 8: Estimation results of the long memory (LM) and short memory (SM) models. In particular rows we report the total number of parameters (k), log-likelihood of inflation, real activity, yields and the total log-likelihood of the model. In the second last row we report the Bayesian Information Criterion (BIC). The last row reports the likelihood ratio test of Vuong (1989) for the null hypotheses that the non-nested LM and SM are equally close to to the data generating process, The sample period is 1986:01 to 2011:12.

	average	1 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30 Y
LM	6.88	2.53	8.32	6.13	6.19	6.19	7.54	5.71	12.42
SM	8.62	2.61	11.07	7.04	7.67	9.25	8.77	6.10	16.45

Table 9: The table reports the RMSE of the *in-sample* fitted values corresponding to the long memory (LM) and short memory (SM) models, where for each maturity n and model M

$$RMSE_n \equiv \left(\frac{1}{T}\sum_{t=1}^{T}\left(y^{\$}_{n,t} - \hat{y}^{\$,M}_{n,t}\right)^2\right)^{\frac{1}{2}},$$

where $y^{\$}_{n,t}$ is the observed (nominal) yield for maturity n and period t and $\hat{r}^{\$,M}_{n,t}$ is the fitted (nominal) yield for maturity n and model $M \in \{LM, SM\}$. These in-sample fitted values are constructed as

$$\hat{y}^{\$,M}_{n,t} = \hat{A}^{\$}_n + \hat{\mathbf{B}}^{\$\prime}_{g,n}\hat{\mathbf{C}}_{g,t} + \hat{\mathbf{B}}^{\$\prime}_{\pi,n}\hat{\mathbf{C}}_{\pi,t} + \hat{\mathbf{B}}^{\$\prime}_{1}\hat{\mathbf{C}}_{1,t} + \hat{\mathbf{B}}^{\$\prime}_{2}\hat{\mathbf{C}}_{2,t} + \hat{\mathbf{B}}^{\$\prime}_{3}\hat{\mathbf{C}}_{3,t},$$

where $\hat{A}^{\$}_n$ and the $\hat{\mathbf{B}}^{\$}_{j,n}$ are obtained by plugging the ML estimates into the formulas of Theorem 4.1 and $\hat{\mathbf{C}}_{j,t}$ denote the Kalman filtered values. The statistics, reported in basis points, are evaluated on the period 1986:01 to 2011:12.

yield	1 M	1 Y	3 Y	5 Y	10 Y	15 Y	20 Y	30Y
forecast horizon	Panel A − LM specification							
1 M	46.22	28.42	34.74	33.60	29.58	28.61	28.41	30.73
3 M	63.97	56.13	60.70	56.94	47.33	44.24	44.16	48.00
6 M	94.47	91.68	88.79	81.44	66.62	60.59	59.22	64.74
1 Y	154.55	146.38	123.19	100.43	71.76	63.46	59.78	73.67
	Panel B − SM specification							
1 M	50.32	30.42	36.78	34.54	30.70	30.06	29.70	31.46
3 M	64.91	56.63	64.05	58.67	48.29	46.26	47.48	44.62
6 M	95.23	91.30	93.33	83.91	68.01	63.84	67.86	56.24
1 Y	155.23	146.47	131.64	108.86	77.62	71.24	86.07	54.41

Table 10: The table reports the RMSFE of the *out-of-sample* fitted values corresponding to the long memory (LM) and short memory (SM) models, where for each maturity n and model M

$$RMSFE_{n,h} \equiv \left(\frac{1}{120-h+1} \sum_{t=1}^{120-h+1} \left(y^{\$}_{n,t+h} - \hat{y}^{\$,M}_{n,t+h|t} \right)^2 \right)^{\frac{1}{2}},$$

where $y^{\$}_{n,t+h}$ and $\hat{y}^{\$,M}_{n,t+h|t}$ are the observed and the *out-of-sample* predicted (nominal) yield for maturity n and period $t+h$, respectively using model $M \in \{LM, SM\}$. These out-of-sample forecasts are constructed as

$$\hat{y}^{\$,M}_{n,t+h|t} = \hat{A}^{\$}_n + \hat{\mathbf{B}}^{\$\prime}_{g,n}\hat{\mathbf{C}}_{g,t+h|t} + \hat{\mathbf{B}}^{\$\prime}_{\pi,n}\hat{\mathbf{C}}_{\pi,t+h|t} + \hat{\mathbf{B}}^{\$\prime}_1\hat{\mathbf{C}}_{1,t+h|t} + \hat{\mathbf{B}}^{\$\prime}_2\hat{\mathbf{C}}_{2,t+h|t} + \hat{\mathbf{B}}^{\$\prime}_3\hat{\mathbf{C}}_{3,t+h|t},$$

where $\hat{A}^{\$}_n$ and the $\hat{\mathbf{B}}^{\$}_{j,n}$ are obtained by plugging rolling ML estimates into the formulas of Theorem 4.1 and $\hat{\mathbf{C}}_{j,t+h|t}$ denote the h-steps ahead Kalman filter *out-of-sample* forecasts based on all available information up to period t. In the first step we estimate the model using the period 1986:01 to 2001:12, equivalent to a rolling window of 192 observations, and using the Kalman filter to make *out-of-sample* predictions across the four horizons. We repeat the procedure 120 times, obtaining 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.