Available online at www.sciencedirect.com
ScienceDirect
Journal of Empirical Finance 15 (2008) 567 – 581
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

Is long memory necessary? An empirical investigation of nonnegative interest rate processes ☆

Jin-Chuan Duan[a,1], Kris Jacobs[b,*,2]

[a] Risk Management Institute and Department of Finance and Accounting, National University of Singapore, and Rotman School of Management, University of Toronto, Canada
[b] Desautels Faculty of Management, McGill University, CIRANO and CIREQ, Canada

Accepted 26 June 2007
Available online 17 October 2007

Abstract

This paper analyzes a class of nonnegative processes for the short-term interest rate. The dynamics of interest rates and yields are driven by the dynamics of the conditional volatility of the pricing kernel. We study Markovian interest rate processes as well as more general non-Markovian processes that display "short" and "long" memory. These processes also display heteroskedasticity patterns that are more general than those of existing models. We find that deviations from the Markovian structure significantly improve the empirical performance of the model. Certain aspects of the long memory effect can be captured with a (less parsimonious) short memory parameterization, but a simulation experiment suggests that the implied term structures corresponding to the estimated long- and short-memory specifications are very different. We also find that the choice of proxy for the short rate affects the estimates of heteroskedasticity patterns.

JEL classification: G12
Keywords: Interest rate; GARCH; Heteroskedasticity; Long memory; Nonnegativity; Term structure

1. Introduction

Models of the term structure of interest rates are the focus of a substantial research effort. Much of the empirical work in this area has been performed using the affine class of term structure models (Duffie and Kan, 1996), which nests some early term structure models such as Vasicek (1977) and Cox, Ingersoll and Ross (1985). Affine models have some important advantages, most notably the existence of analytical solutions for bond prices and yields. However, it

☆ We would like to thank Peter Christoffersen, Nour Meddahi and Angelo Melino for helpful comments. The paper has also benefited from presentations at CIRANO, the 1999 NFA Meetings, the 1999 North American Summer Meetings of the Econometric Society and the 2000 North American Winter Meetings of the Econometric Society. Xiaofei Li provided excellent research assistance.
* Corresponding author.
E-mail address: kris.jacobs@mcgill.ca (K. Jacobs).
[1] Duan acknowledges support from a Direct Allocation Grant from the Hong Kong University of Science and Technology and a grant from the Research Grants Council of Hong Kong (HKUST6017/99H).
[2] Jacobs acknowledges FQRSC of Quebec and SSHRC of Canada for financial support.

0927-5398/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2007.06.003

has become clear that their empirical performance is unsatisfactory. Most importantly, no affine parameterization seems to be able to simultaneously match the short interest rate (the time series implications of the model) as well as the implied shape of the term structure (the cross sectional implications of the model).

There have been several new approaches in response to these empirical failures. One approach focuses exclusively on the model's performance in the cross-sectional dimension by modeling time-varying drifts, which allow exact matching of the term structure at one time point but requires repeated recalibrations over time to maintain a reasonable fit.[3] Another approach has extended the one-factor exponential affine framework to include multiple factors.[4] Although these multifactor extensions have enjoyed some empirical success, much remains to be done.[5] It may therefore prove interesting to explore other modeling approaches, such as deviations from the Markovian structure. Backus and Zin (1993) explore long memory using a discrete-time framework with Gaussian innovations. Their work indicates that departures from the short memory structure can substantially improve model performance, but a disadvantage of such an approach is that interest rates can become negative.

This paper uses a discrete-time setup to analyze a large class of term structure models that can be nested in a single framework. In general the framework is non-Markovian and allows for long memory, but testable restrictions can be imposed to obtain a short memory or Markovian model. The derivation of the interest rate dynamic is non-standard: the main difference with existing models is that the interest rate dynamic is driven by the time-varying volatility of the pricing kernel, as opposed to the time-varying mean of the pricing kernel. In principle, this modeling strategy can be implemented by modeling the volatility of the pricing kernel using a class of nonnegative processes. In this paper we use a chi-square dynamic for the volatility of the pricing kernel, which arises naturally from the extensive work done in the area of stock market volatility using the GARCH framework (see Engle, 1982; Bollerslev, 1986). The resulting dynamics for the short-term interest rate are tractable for empirical purposes. These dynamics are governed by innovations of chi-square distributions, in contrast with most existing interest rate models which are governed by normally distributed innovations. The advantage of the chi-square innovations is that interest rates can stay positive under suitable parameter restrictions, whereas it is not possible to ensure positive interest rates with normally distributed innovations.

The resulting class of interest rate dynamics is very general and includes a range of dynamics resembling those of models based on the ARMA specification. Building on the evidence of long memory in interest rates uncovered by Backus and Zin (1993), we also investigate a class of long-memory specifications. Further, to demonstrate the flexibility of our approach, we specify and test some interest rate dynamics with fairly general heteroskedasticity patterns. Specifically, we incorporate the CEV-type heteroskedasticity advocated by Chan et al. (1992) in our model by directly modeling the time variation in the conditional heteroskedasticity of the pricing kernel. Affine models do not explore this dimension, and therefore cannot model such heteroskedasticity patterns.

In this paper, we estimate and test a large number of specifications of the short-term interest rate based on this modeling approach. The empirical analysis highlights several interesting stylized facts and suggests that the class of models proposed in this paper may be able to match some empirical regularities of the term structure. We analyze two time series containing daily observations on the short-term interest rate, overnight Eurodollar rates and one-week Eurodollar rates. When we restrict our analysis to the Markovian subclass of models, these two time series seem to exhibit similar characteristics. However, when we use the same two data series to estimate and test non-Markovian models, the empirical results are quite different. Although formal statistical tests support deviations from the Markovian structure in both cases, the overnight rate series requires a much richer parameterization compared to the one-week rate series. The heteroskedasticity patterns displayed by the two series are also different: whereas the weekly interest rate series supports heteroskedasticity models that are inherently different from the typical term structure models, the results for overnight rates suggest that existing models are adequate. Finally, both interest rate series exhibit

[3] See Ho and Lee (1986), Hull and White (1990, 1993), Black, Derman and Toy (1990), Black and Karasinski (1991) and Heath, Jarrow and Morton (1992).

[4] See Litterman and Scheinkman (1991), Chen and Scott (1993), Pearson and Sun (1994) and Dai and Singleton (2000). Observable factors used in this context include the long yield (Brennan and Schwartz, 1979), the volatility of the short rate (Longstaff and Schwartz, 1992; Balduzzi et al., 1996), and the long-run mean of the short rate (Jegadeesh and Pennacchi, 1996; Balduzzi et al., 1996).

[5] Yet another literature investigates continuous-time and discrete-time models of the short rate outside the affine class of models. These specifications do not lead to closed form solutions for bond prices and yields (e.g. see Chan et al., 1992; Brenner et al., 1996; Andersen and Lund, 1997, 1999; Koedijk et al., 1997; Pfann et al., 1996) on the incorporation of more general heteroskedasticity patterns. See Hamilton (1988) and Gray (1996) on the importance of regime switching for interest rates).

long-memory characteristics. While a likelihood-based comparison indicates that the long memory effect can also be captured by a (substantially less parsimonious) short-memory parameterization, a small-scale simulation experiment suggests that the implied term structures corresponding to the estimated long- and short-memory specifications are very different.

The paper proceeds as follows. Section 2 presents the models for the short rate and discusses how the model dynamics are related to the conditional volatility of the pricing kernel. Section 3 extends this approach to obtain interest rate dynamics with more general heteroskedasticity patterns. Section 4 discusses the data and Section 5 presents the empirical results. Section 6 presents a simulation experiment that illustrates the quantitative impact of long memory for the term structure, and Section 7 concludes.

2. A class of nonnegative interest rate processes

We characterize the short-term interest rate dynamic starting from the specification of the pricing kernel. Let M_t denote the pricing kernel and let Ω_t denote the information set up to time t. For any asset with a price X_t and a dividend payment D_t at time t, equilibrium requires the following Euler equation to hold:

$$X_t = E[M_{t+1}(X_{t+1} + D_{t+1})|\Omega_t]. \tag{1}$$

Let r_t' denote the one-period risk-free interest rate (continuously compounded) at time t. Since the time-t value of a one-dollar time-$(t+1)$ payoff is $e^{-r_t'}$, we get the well-known result:

$$e^{-r_t'} = E[M_{t+1}|\Omega_t]. \tag{2}$$

Given this setup, the dynamic for the interest rate process is determined by the dynamic for the pricing kernel. We model the negative of the logarithm of the pricing kernel using a GARCH-in-mean process; that is, denoting $m_{t+1} = -\ln M_{t+1}$[6]

$$\begin{aligned} m_{t+1} &= \mu + \kappa q_{t+1} + \sqrt{q_{t+1}}\varepsilon_{t+1} \\ \varepsilon_{t+1}|\Omega_t &\sim N(0,1). \end{aligned} \tag{3}$$

We investigate the interest rate dynamic using two different classes of models for the conditional volatility of the pricing kernel. The GARCH class of processes has proven useful for describing stock price volatility, and therefore these processes may also be useful for describing the time variation of the volatility of the state variable that drives financial markets. The first class of models we investigate therefore assumes the GARCH(p,q) specification of Bollerslev (1986).[7]

$$q_{t+1} = \beta_0' + \sum_{j=1}^{p} \beta_j L^{j-1} q_t + \sum_{i=1}^{q} \alpha_i L^{i-1} q_t \varepsilon_t^2. \tag{4}$$

The volatility dynamic in (4) can accommodate fairly rich autocorrelation patterns.[8] However, it has recently been shown that volatility of market portfolio returns displays long memory, and that such a phenomenon cannot be adequately captured by (4).[9] Backus and Zin (1993), Shea (1997) and Connolly and Guner (1999) show that long memory processes can also be helpful for the empirical modeling of interest rate processes. Therefore, we study a

[6] The mechanism underlying these models is most clearly illustrated by formulating the dynamics in terms of the negative of the logarithm of the pricing kernel, see e.g. Campbell, Lo and MacKinlay (1997, chapter 11).

[7] Note that the GARCH(p,q) specification is often represented as $q_{t+1} = \beta_0' + \Sigma_{j=1}^{p} \beta_j L^{j-1} q_t + \Sigma_{i=1}^{q} \alpha_i L^{i-1} u_t^2$ with $u_t = \sqrt{q_t}\varepsilon_t$ and $\varepsilon_t|\Omega_{t-1} \sim N(0, 1)$. The representation in (4) is therefore completely standard.

[8] Although we limit ourselves to the linear GARCH(p,q) process here, our derivation can be easily generalized to GARCH(p,q) models with leverage effects such as the non-linear asymmetric GARCH(p,q) of Engle and Ng (1993), EGARCH(p,q) of Nelson (1991), or to more general models like the augmented GARCH(p,q) process proposed by Duan (1997).

[9] See the overview paper by Baillie (1996) for examples of long memory in the natural and social sciences and Baillie and Bollerslev (2000) for the importance of long memory for the foreign exchange market.

second class of models that assumes the FIGARCH(p,d,q) model of Baillie, Bollerslev and Mikkelsen (1996) for conditional volatility.[10]

$$q_{t+1} = \beta_0' + \sum_{j=1}^{p} \beta_j L^{j-1} q_t + \left[1 - \sum_{j=1}^{p} \beta_j L^j - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] q_{t+1}\varepsilon_{t+1}^2 \tag{5}$$

where the parameters in (4) and (5) are subject to the usual restrictions. It should be noted that the FIGARCH(p,d,q) dynamic in (5) nests the GARCH(p,q) dynamic in (4) by setting $d=0$. We present them as two separate specifications for the sake of clarity.[11] From the Euler equation for the interest rate in (2) and the specification of the pricing kernel in (3) we obtain

$$r_t' = \mu + (\kappa - 1/2) q_{t+1}. \tag{6}$$

This relationship between the interest rate and the conditional volatility is the key to the interest rate models derived below.[12] To simplify notation we define an adjusted interest rate, $r_t \equiv r_t' - \mu$ and $\beta_0 \equiv \beta_0'(\kappa - 1/2)$.[13] This yields the following models

Model 1. Short-memory interest rate dynamic:

$$r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \sum_{i=1}^{q} \alpha_i L^{i-1} r_{t-1}\varepsilon_t^2, \tag{7}$$

if the GARCH(p,q) dynamic in (4) is employed; and

Model 2. Long-memory interest rate dynamic:

$$r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \left[1 - \sum_{j=1}^{p} \beta_j L^j - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] r_t\varepsilon_{t+1}^2, \tag{8}$$

if the FIGARCH(p,d,q) dynamic in (5) is used. This particular interest rate dynamic has been previously derived in Duan and Jacobs (1996) using an equivalent assumption on the utility function of the representative agent.

In both models, the interest rate dynamic inherits the properties of the conditional volatility process for the pricing kernel. The long-memory interest rate dynamic of Model 2 differs substantially from the long-memory interest rate model investigated by Backus and Zin (1993). Because the innovation in this model is of the chi-square type, the interest rate in Model 2 can be made positive with proper parameter restrictions, whereas there is always a positive probability for the interest rate in the long-memory interest rate model of Backus and Zin (1993) to take on negative values.[14]

[10] Other forms of fractionally integrated GARCH process such as Bollerslev and Mikkelsen (1996) and McCurdy and Michaud (1996) can also be used to derive alternative long memory interest rate dynamics.

[11] It has been noted in the literature that the long memory specification in (5) implies $\beta_0'=0$ (e.g. see Chung (1999)) if a finite unconditional variance is desired. We have not as yet found a satisfactory long memory specification that resolves this issue. Whereas this is a problem for these models, it affects the estimate for β_0' but not the estimates for other parameters of interest in a significant way.

[12] Note that a similar relationship between the interest rate and the conditional volatility always holds. We can therefore generate arbitrary processes for the interest rate by an appropriate choice of volatility dynamic. However, the choice of volatility dynamic is restricted by its implications for the dynamics under the risk- neutral probability measure, such as the term structure. It will be shown in Section 6 that the choice of the GARCH(p,q) and FIGARCH(p,d,q) dynamics is critical in that respect because the innovations are simple functions of the normal distribution, which allows us to characterize the interest rate dynamic explicitly under the risk-neutral measure.

[13] Notice that in the relationship $r_t'=\mu+(\mathrm{k}-1/2)\mathrm{q}_{t-1}$ between the short interest rate and volatility, it is the presence of the GARCH-in-mean parameter κ that makes it possible for positivity of interest rates to be imposed, specifically for $\kappa>1/2$, under the reasonable assumption that $\mu>0$. The interpretation of this is that (up to a constant μ) we only model the conditional mean through the GARCH-in-mean effect. In a more realistic model one could also specify ARMA-type effects in the conditional mean. These effects would interact with the conditional volatility dynamic to produce equilibrium interest rates. However, in such a more realistic model one cannot ascertain that interest rates will stay positive.

[14] Interest rates can become negative in the Backus-Zin setup because of the assumption of Gaussian shocks within a discrete time framework.

For estimation purposes it is more convenient to express these interest rate dynamics with an innovation that has conditional mean 0 and variance 1. Let $\upsilon_{t+1} \equiv (1/\sqrt{2})(\varepsilon^2_{t+1} - 1)$. It is clear that υ_{t+1} has mean 0 and variance 1, conditional on Ω_t. Model 1 can thus be rewritten as

$$r_t = \beta_0 + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) L^{j-1} r_{t-1} + \sqrt{2} \sum_{i=1}^{\max(p,q)} \alpha_i L^{i-1} r_{t-1} \upsilon_t. \tag{9}$$

Note that if $p>q$, then $\beta_j=0$ for $j=q+1,\ldots, p$. Similarly, $\alpha_i=0$ for $i=p+1,\ldots, q$ if $q>p$. Model 2, on the other hand, can be written as

$$\begin{aligned} r_t = \beta_0 &+ \left[1 - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^{i-1}\right)(1-L)^d\right] r_t \\ &+ \sqrt{2}\left[1 - \sum_{j=1}^{p} \beta_j L^j - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] r_t \upsilon_{t+1}. \end{aligned} \tag{10}$$

To illustrate the implementation of the long memory interest rate model, it is also useful to express the fractional difference operation as an infinite sum. Using the FIGARCH(1,d,1) process as an example, Eq. (10) gives:

$$r_t = \beta_0 + \sum_{k=0}^{\infty} [\theta\pi_k(d) - \pi_{k+1}(d)] L^k r_{t-1} + \sqrt{2}\left(\sum_{k=0}^{\infty} [\theta\pi_k(d) - \pi_{k+1}(d)] L^\kappa - \beta_1\right) r_{t-1} \upsilon_t. \tag{11}$$

We have used in the above expression the fact that $(1-L)^d = \sum_{k=0}^{\infty} \pi_k(d) L^k$ with $\pi_0(d)=1$ and $\pi_k(d) = (-1)^k \prod_{i=1}^{k} \frac{d-i+1}{i}$ for $k \geq 1$.

3. Heteroskedasticity in interest rates processes

Inspection of expressions (7) and (8) illustrates that the interest rate processes derived in Section 2 allow for heteroskedasticity because the innovations depend on the level of the lagged interest rate. In this section we show that this approach can be easily extended to accommodate more general heteroskedasticity patterns, which are supported by existing research. It must be noted that in contrast to the approach in this paper, these existing studies start by assuming directly an *ad hoc* short interest rate process. Chan et al. (1992) study the following specification for the short interest rate r_t:

$$r_t = \gamma + \beta r_{t-1} + r_{t-1}^{\psi} v_t \tag{12}$$

where v_t is a disturbance term with conditional mean zero. Their empirical results show that the data indicate point estimates of ψ around 1.5, and that these estimates are statistically different from 1. Therefore, the data seem to indicate support for heteroskedasticity patterns that are different from the ones modeled in Section 2.[15]

It is straightforward to incorporate this type of heteroskedasticity into the model derived in Section 2. If we change the pricing kernel volatility dynamic to

$$q_{t+1} = \beta_0' + \sum_{j=1}^{p} \beta_j L^{j-1} q_t + \sum_{i=1}^{q} \alpha_i L^{i-1} q_t^{\psi} \varepsilon_t^2, \tag{13}$$

the corresponding interest rate dynamic becomes

Model 3. Short-memory interest rate dynamic with general heteroskedasticity:

$$r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \sum_{i=1}^{q} \alpha_i L^{i-1} r_{t-1}^{\psi} \varepsilon_t^2. \tag{14}$$

[15] Subsequent studies, significantly differing in empirical methodology and model specification, have confirmed these findings (see for example Ait-Sahalia, 1996; Campbell et al., 1997; Conley et al., 1997; Ahn and Gao, 1999). Andersen and Lund (1997), on the other hand, do not find evidence for these types of heteroskedasticity effects.

To obtain the long-memory version of the model, we simply adopt the following pricing kernel volatility dynamic:

$$q_{t+1} = \beta_0' + \sum_{j=1}^{p} \beta_j L^{j-1} q_t + \left[1 - \sum_{j=1}^{p} \beta_j L^j - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] q_{t+1}^{\psi} \varepsilon_{t+1}^2. \tag{15}$$

The resulting interest rate model is

Model 4. Long-memory interest rate dynamic with general heteroskedasticity:

$$r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \left[1 - \sum_{j=1}^{p} \beta_j L^j - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] r_t^{\psi} \varepsilon_{t+1}^2. \tag{16}$$

Again, it is more convenient for estimation purposes to express these interest rate dynamics with an innovation that has conditional mean 0 and variance 1. Letting $\upsilon_{t+1} \equiv (1/\sqrt{2})(\varepsilon_{t+1}^2 - 1)$ where v_{t+1} has mean 0 and variance 1, conditional on Ω_t, Model 3 can be rewritten as

$$r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \sum_{i=1}^{q} \alpha_i L^{i-1} r_{t-1}^{\psi} + \sqrt{2} \sum_{i=1}^{q} \alpha_i L^{i-1} r_{t-1}^{\psi} \upsilon_t \tag{17}$$

and Model 4 as

$$\begin{aligned} r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \left[1 - \sum_{j=1}^{p} \beta_j L^j - \left(1 - \sum_{i-1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] r_t^{\psi} \\ + \sqrt{2}\left[1 - \sum_{j-1}^{p} \beta_j L^j - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] r_t^{\psi} v_{t+1}. \end{aligned} \tag{18}$$

The incorporation of this type of heteroskedasticity in the context of our model is therefore technically trivial. Inspecting the issue from a different perspective, however, it shows the flexibility of an approach based on the specification of the conditional volatility dynamic of the pricing kernel. The essence of standard term structure models is that the interest rate dynamic is driven by the specification of the conditional mean of the pricing kernel. This paper argues that the most straightforward way of getting an internally consistent interest rate model is via modeling the conditional volatility of the pricing kernel. The current term structure literature does not follow this approach, however. Consequently, existing term structure models face difficulties in incorporating this type of heteroskedasticity into the specification of the interest rate process.

To emphasize this important point, we demonstrate in our discrete time environment the problems of incorporating heteroskedasticity when using a model that displays similarities with the well-known class of exponential affine term structure models. In the simplest of these exponential affine term structure models, the dynamic of the pricing kernel is given by

$$m_{t+1} = \theta + \phi m_t + \eta_{t+1} \tag{19}$$

$$\eta_{t+1} | \Omega_t \sim N\left(0, \sigma^2\right), \tag{20}$$

which yields the following AR(1) dynamic for the short-term interest rate r_t:

$$r_t = \theta(1-\phi) - \left(\sigma^2/2\right)(1-\phi) + \phi r_{t-1} + \phi \eta_t. \tag{21}$$

This interest rate model is similar to the continuous-time Vasicek model. It seems straightforward to insert heteroskedasticity into the interest rate dynamic by changing the dynamic of the pricing kernel to

$$m_{t+1} = \theta + \phi m_t + m_t^{\psi} \eta_{t+1}. \tag{22}$$

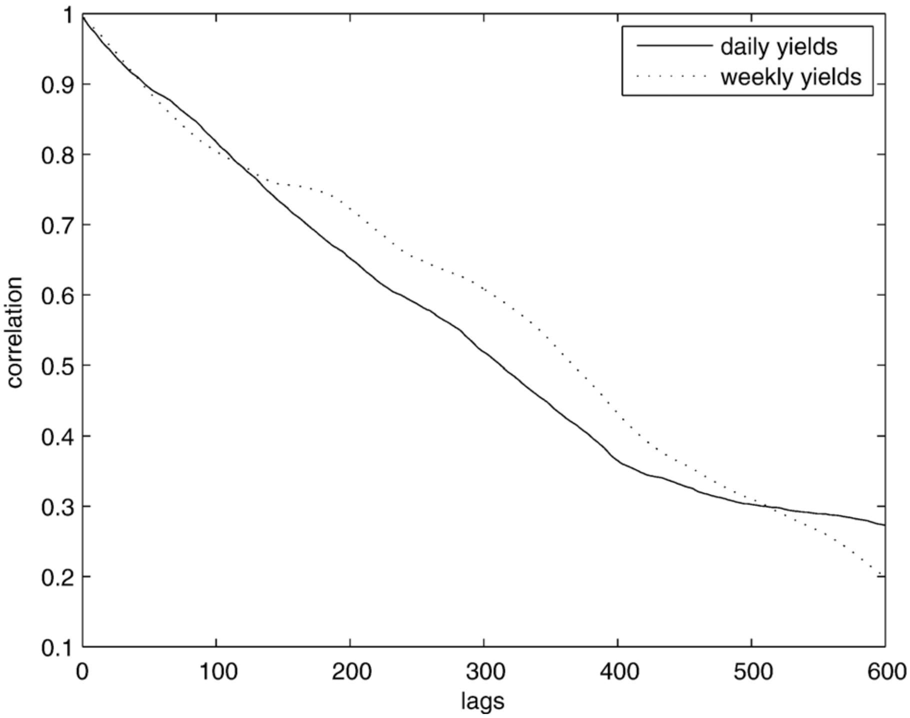

Fig. 1. Autocorrelation functions for annualized daily and weekly yields.

However, it can be easily verified that this dynamic yields an analytical solution for the interest rate only when $\psi=0.5$. The nature of heteroskedasticity in the resulting model is similar to the continuous-time Cox–Ingersoll–Ross (1985) model.[16] This example demonstrates that the potential for building heteroskedasticity into term structure models of this class is very limited. In contrast, the approach of this paper allows for a wider range of heteroskedasticity patterns.

4. Data

We present estimation and test results using two different data sets. The first data set is the one analyzed in Ait-Sahalia (1996). This data set contains daily observations on one-week Eurodollar rates from June 1, 1973 to February 25, 1995, totaling 5505 observations. The second data set contains daily observations on overnight Eurodollar rates from January 1, 1981 to May 30, 1997, for a total of 4280 observations.

In much of the term structure literature, there is a trade-off between using overnight rates versus one-week rates for the short-term interest rates. When estimating and testing continuous-time models, for instance, conceptually there is a clear advantage to using the overnight rates, because these data are a closer approximation to the theoretical construct of the model (see Chapman et al., 1999) for an extensive analysis of this issue). In this paper, the theory is formulated in discrete time. If the data has daily frequency, the overnight rate is, by the definition of the model, the unambiguous choice for the short-term interest rate. Overnight rates are, however, more prone to microstructure effects, which can considerably complicate the interpretation of the empirical findings. We therefore also estimate the models using the one-week rates. To illustrate the potential importance of long memory specifications for the data, Fig. 1 plots the sample autocorrelations up to lag 600 for the weekly and overnight data. For the weekly yields, the autocorrelation at lag 600 is 0.21. For the daily data, the autocorrelation at lag 600 is even higher at 0.28. It is clear that long memory specifications are very useful to capture this kind of empirical phenomena, because it may be hard to fit these characteristics of the data with standard short memory models.

[16] A comparison between discrete and continuous-time models is often tenuous. However, the characteristic of the model under study here has similar effects in both modeling approaches.

5. Estimation and test results

Estimation of the parameters in Models 1–4 is not possible by exact maximum likelihood estimation, because the innovation in these systems is governed by a χ^2 distribution with one degree of freedom, which has an unbounded likelihood at the origin. To the best of our knowledge it is not possible to restrict the parameter space to deal with this problem in an exact maximum likelihood setup.

We therefore estimate the parameters by maximizing the quasi-maximum likelihood. Because the model is nonstandard, we briefly outline the structure of the quasi log-likelihood. Note that the four models are defined in terms of r_t, which is equal to $r_t'-\mu$, with r_t' the short-term rate which is assumed to be observable. While we could estimate μ as part of the quasi log-likelihood setup, we proceed by simply assuming $r_t \equiv r_t'$, thus effectively setting μ equal to zero. The structure of the quasi log-likelihood is similar for the four models. For example, for Model 1, note from (9) that conditional on Ω_{t-1}, $u_t = \sqrt{2}\alpha_1 r_{t-1} \upsilon_t$ has mean 0 and variance 1. We can then use the initial conditions $v_{-q+1}=v_{-q+2}=\ldots=v_0=0$ and iterate on (9) to set up the conditional quasi log-likelihood

$$-\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \ln\left(2\alpha_1^2 r_{t-1}^2\right) - \frac{1}{2}\sum_{t=1}^{T} \frac{u_t^2}{2\alpha_1^2 r_{t-1}^2}. \tag{23}$$

The quasi log-likelihood for Models 2, 3 and 4 can be obtained in a similar fashion. White (1982) and Bollerslev and Wooldridge (1992) present sufficient conditions for consistency and asymptotic normality of the quasi maximum likelihood estimator, but these conditions are difficult to verify in practice.[17] Bollerslev and Wooldridge (1992) present some small sample evidence suggesting that the quasi-maximum likelihood estimator works well in a dynamic model governed by a χ^2 distribution with one degree of freedom, but this dynamic is different from the ones considered in this paper.

Estimation of Models 2 and 4 involves one additional complication. Inspection of expression (11) shows that one actually needs an infinite number of data points for the initial condition. In practice, existing studies truncate the infinite sum in (11) at a sufficiently high number. In our empirical analysis, the data required for the initial condition are truncated at 1000 lags. Implicit in this practice is our assumption that the first 1000 data points are enough to bring the system to its steady state so that the resulting likelihood function for either Model 2 or 4 is a good proxy for the true likelihood function. For Models 1 and 3, the initial condition does not require as many data points. In order to ensure meaningful comparisons between the results of Model 1 (3) and Model 2 (4), we therefore discard the first 1000 data points, giving 4505 and 3280 observations respectively in the two datasets.

Tables 1A–4B present the estimation results. Table 1A presents results for Model 1 (the short memory interest rate dynamics) obtained using the data on the one-week Eurodollar rates. Table 1B also uses the one-week Eurodollar rates but presents results for Model 2 (the long memory interest rate dynamic). Table 2 presents results for Models 1 and 2 using data on the overnight Eurodollar rates. Tables 3 and 4 present results for the one-week rates and overnight rates respectively using Models 3 and 4. In each table, robust asymptotic standard errors are indicated below the parameters. Besides point estimates and standard errors for the parameters, each table also contains a variety of robust Wald tests that test the joint significance of extra coefficients for higher order specifications. The notational convention for these tests is as follows: for a given row, α_k denotes the highest order coefficient for that specification and β_k denotes the highest order β coefficient. For example, in the row corresponding to the GARCH(3,3) dynamic, α_k stands for the α_3 coefficient and β_{k-1} stands for the β_2 coefficient. This notation explains the interpretation of the robust Wald statistics that are listed: for example, for the row corresponding to the GARCH(3,3) specification the column with $\beta_{k-1}=\beta_k=0$ presents the robust Wald statistic that $\beta_3=\beta_2=0$.[18] The last column in each table presents the value of the objective function at the optimum. It must be noted that because we estimate using quasi-maximum likelihood, the standard likelihood ratio test statistics are no longer valid and we resort to the robust Wald statistics.

Table 1A presents the results for Model 1 using the one-week Eurodollar rates. The first row presents results corresponding to the GARCH(1,1) dynamic. This model is of particular interest because the resulting interest rate

[17] See Lumsdaine (1996) for verification of a particular case.

[18] For each specification, we also constructed Wald statistics to test the hypothesis that $\beta_k=0$ for all k and either $\alpha_k=0$ or $\phi_k=0$ for all k. These hypotheses are always strongly rejected and therefore not reported.

Table 1
(A) Model 1 (short memory) parameter estimates for the equilibrium short rate dynamics using one-week Eurodollar rates (4505 observations)

Pricing kernel	β_1	β_2	β_3	β_0	α_1	α_2	α_3	$\alpha_k=\alpha_{k-1}=0$	$\beta_k=\beta_{k-1}=0$	$\alpha_k=\beta_k=0$	QLL
GARCH(1,1)	0.975	–	–	0.0214	0.0238	–	–	–	–	–	3790.9
	0.001	–	–	0.0084	0.0011	–	–	–	–	–	
GARCH(2,2)	1.173	−0.185	–	0.0116	0.0223	−0.0111	–	–	–	79.85	4089.4
	0.041	0.040	–	0.0057	0.0007	0.0012	–	–	–	0.000	
GARCH(3,3)	1.229	−0.311	0.069	0.0121	0.0223	−0.0122	0.0018	49.37	9.49	0.57	4091.4
	0.096	0.177	0.093	0.0069	0.0007	0.0022	0.0028	0.000	0.009	0.751	

(B) Model 2 (Long Memory) Parameter estimates for the equilibrium short rate dynamics using one–week Eurodollar (4505 observations)

Pricing kernel	β_1	β_2	β_3	β_0	ϕ_1	ϕ_2	ϕ_3	d	$\phi_k=\phi_{k-1}=0$	$\beta_k=\beta_{k-1}=0$	$\phi_k=\beta_k=0$	QLL
FIGARCH(1,d,1)	0.989	–	–	0.0106	0.999	–	–	0.0122	–	–	–	4058.0
	0.001	–	–	0.0071	0.001	–	–	0.0014	–	–	–	
FIGARCH(2,d,2)	1.140	−0.152	–	0.0119	1.159	−0.161	–	0.0029	–	–	7.620	4090.1
	0.056	0.056	–	0.0064	0.059	0.058	–	0.0039	–	–	0.022	
FIGARCH(3,d,3)	1.427	−0.912	0.472	0.0115	1.437	−0.917	0.479	0.0114	74.070	74.59	75.21	4136.7
	0.069	0.112	0.055	0.0077	0.070	0.114	0.056	0.0016	0.000	0.000	0.000	

Notes: Each table contains estimated coefficients, robust Wald statistics and the quasi log-likelihood for three different specifications of the pricing kernel. Standard errors are indicated under estimated coefficients. Significance levels are indicated under the Wald statistics. Both tables use 4505 observations on one-week Eurodollar rates in estimation. In the description of the Wald statistics, α_k denotes the highest order α coefficient, β_k denotes the highest order β coefficient and ϕ_k denotes the highest order ϕ coefficient in each specification.

model displays similarities with a number of models in the literature. First, note that the interest rate process inherits the characteristics of the conditional volatility process. Since the GARCH process has its conditional volatility behaving like an ARMA process, we can think of the resulting interest rate dynamic as an ARMA process with heteroskedastic features. In the case of the GARCH(1,1) specification, the resulting interest rate process is Markovian and bears resemblance to the models belonging to the exponential affine class of interest rate models.[19]

Estimation results for the interest rate process corresponding to the GARCH(1,1) dynamic indicate that all coefficients are very precisely estimated. The estimate for β_1 is large and for α_1 is small. The sum of α_1 and β_1 is very close to but smaller than one, which indicates a slow mean reversion (large persistence) and satisfies the stationarity requirement.[20]

These results are consistent with the standard empirical findings in the literature when Markovian processes are used in estimation. The interesting question is therefore whether extensions to non-Markovian processes are required by the data. In rows 2 and 3, we present the results for the GARCH(2,2) and GARCH(3,3) dynamics, respectively. The robust standard errors and Wald statistics (in most cases) support extensions to non-Markovian models. These empirical findings are therefore supportive of the results established by Backus and Zin (1994), who reject the Markovian assumption in exponential affine models.

Other specifications such as GARCH(2,1), GARCH(1,2), as well as higher order GARCH dynamics were investigated, but are not reported to keep the tables manageable. The main result is that for high orders of p and q (higher than 8), the robust Wald statistics indicate that the inclusion of extra coefficients does not improve the statistical fit of the model. As a rule, t-statistics and robust Wald statistics become smaller as we increase p and q.

Table 1B also uses the one-week data, but presents results for Model 2, which is based on the FIGARCH pricing kernel. We start by inspecting the first row, which shows that the FIGARCH(1,d,1) dynamic yields an estimate of the long memory parameter d of 0.0122. Even though the estimated value for d is rather small, it is statistically different

[19] Our reference to the ARMA process is tenuous, because the positivity requirement on the conditional volatility of the pricing kernel imposes restrictions on the parameters of the interest rate process. These restrictions do not apply if one directly writes down an ARMA process for the short-term interest rate without deriving it using the framework used in this paper.

[20] Bollerslev (1986) presents sufficient conditions on β_1 and α_1 of Eq. (5) for positivity of the conditional volatility. His conditions require β_1 and α_1 to be non-negative. These conditions can actually be relaxed when one goes beyond the GARCH(1,1) model. Nelson and Cao (1992) provided a weaker set of requirements that can ensure positive conditional volatility under GARCH(p,q). Our estimation results require invoking Nelson and Cao's (1992) conditions because some parameter estimates turn out to be negative. We have verified that the parameter values presented in Table 1 do satisfy the conditions for positive conditional variances.

Table 2
(A) Model 1 (short memory) parameter estimates for the equilibrium short rate dynamics using overnight Eurodollar rates (3280 observations)

Pricing kernel	β_1	β_2	β_3	β_0	α_1	α_2	α_3	α_6	$\beta_k=\beta_{k-1}=0$	$\alpha_k=\beta_k=0$	QLL
GARCH(1,1)	0.971	–	–	0.0538	0.0245	–	–	–	–	–	3591.4
	0.002	–	–	0.0223	0.0012	–	–	–	–	–	
GARCH(2,2)	0.942	0.034	–	0.0389	0.0242	−0.0033	–	–	–	21.14	3630.1
	0.061	0.058	–	0.0199	0.0012	0.0024	–	–	–	0.000	
GARCH(3,3)	0.819	0.158	0.003	0.0267	0.0239	−0.0015	−0.0046	–	6.72	9.33	3676.8
	0.077	0.078	0.008	0.0159	0.0012	0.0021	0.0016	–	0.035	0.009	

(B) Model 2 (long memory) parameter estimates for the equilibrium short rate dynamics using overnight Eurodollar rates (3280 observations)

Pricing kernel	β_1	β_2	β_3	β_0	ϕ_1	ϕ_2	ϕ_3	d	$\phi_k=\phi_{k-1}=0$	$\beta_k=\beta_{k-1}=0$	$\phi_k=\beta_k=0$	QLL
FIGARCH(1,d,1)	0.987	–	–	0.0132	0.998	–	–	0.0129	–	–	–	3679.2
	0.002	–	–	0.0088	0.001	–	–	0.0022	–	–	–	
FIGARCH(2,d,2)	0.663	0.325	–	0.0111	0.669	0.330	–	0.0182	–	–	97.980	3721.6
	0.079	0.079	–	0.0088	0.079	0.079	–	0.0019	–	–	0.000	
FIGARCH(3,d,3)	0.664	−0.186	0.508	0.0151	0.670	−0.182	0.510	0.0174	20.060	20.01	17.73	3755.1
	0.230	0.241	0.121	0.0099	0.211	0.242	0.121	0.0013	0.000	0.000	0.000	

Notes: Each table contains estimated coefficients, robust Wald statistics and the quasi log-likelihood for three different specifications of the pricing kernel. Standard errors are indicated under estimated coefficients. Significance levels are indicated under the Wald statistics. Both tables use 3280 observations on overnight Eurodollar rates in estimation. In the description of the Wald statistics, α_k denotes the highest order α coefficient, β_k denotes the highest order β coefficient and ϕ_k denotes the highest order ϕ coefficient in each specification.

from zero as indicated by its robust standard error.[21] The data seem therefore supportive of long memory in the interest rate dynamic. This finding confirms that the Markovian assumption is quite a poor assumption for the one-week interest rate data. The remaining rows in the table indicate what happens if we increase p and q in the specification of the FIGARCH(p,d,q) dynamic. Importantly, they confirm the findings of Table 1A: as we increase p and q, the standard errors increase rapidly. What happens to the estimate of the long-memory parameter d as p and q increase? This question is important because of the finding in row 1 that the long-memory parameter is significantly estimated. Whereas long-memory processes are of interest because they allow us to capture important empirical phenomena with a rather parsimonious parameterization, their implementation in term structure models is far from obvious. It would therefore be interesting if we could capture nonzero autocorrelation at long horizons by increasing p and q in the specification of the FIGARCH(p,d,q) dynamic. Inspection of Table 1B indicates that the estimates of d differ substantially between the FIGARCH(1,d,1) and FIGARCH(3,d,3) cases on the one hand, and the FIGARCH(2,d,2) case on the other hand. When investigating p and q higher than 5, we found negative point estimates of d that were insignificantly estimated (not reported). By increasing p and q in the FIGARCH(p,d,q) specification, it therefore seems possible to capture the empirical phenomena that give rise to a long memory parameter that is significantly different from zero in the FIGARCH(1,d,1) dynamic. The extra AR and MA coefficients capture a large part of the correlation that was initially captured by the long memory parameter, and the remaining correlation captured by the long-memory parameter is not statistically significant. This does not necessarily indicate that the estimate for d in the FIGARCH(1,d,1) dynamic is spurious. In terms of predictive accuracy, it is likely that one will have to pay a price for estimating large numbers of MA parameters. Ray (1991), for example, provides an analysis of the predictive performance of long-memory and competing short memory models, but analyzes AR models only.

Table 2 presents the results for Models 1 and 2 using the overnight Eurodollar rates. To some extent, the estimates support the conclusions drawn from those presented in Table 1. Row 1 in Table 2A indicates that the data support deviations from Markovian models. Row 1 in Table 2B indicates that the data support the presence of long memory. Again, the estimate for the coefficient that captures long memory is small but significantly different from zero. Finally, investigation of orders of p and q higher than five confirms the presence of a negative long-memory parameter, which in many cases is insignificantly estimated (not reported).

[21] The estimate of the long-memory parameter is much smaller than that obtained in other studies of interest rates (Backus and Zin, 1993). This is not necessarily surprising because our model is specified differently and the parameter d does not have exactly the same interpretation.

Table 3
(A) Model 3 (short memory) parameter estimates for the equilibrium short rate dynamics using one-week Eurodollar rates (4505 observations)

Pricing kernel	β_1	β_2	β_3	β_0	α_1	α_2	α_3	ψ	$\alpha_k=\alpha_{k-1}=0$	$\beta_k=\beta_{k-1}=0$	$\alpha_k=\beta_k=0$	QLL
GARCH(1,1)	0.970	–	–	0.0978	0.0115	–	–	1.3400	–	–	–	3933.4
	0.002	–	–	0.0161	0.0019	–	–	0.0066	–	–	–	
GARCH(2,2)	1.092	−0.108	–	0.0583	0.0092	−0.0040	–	1.4080	–	–	26.77	4303.8
	0.046	0.045	–	0.0095	0.0013	0.0008	–	0.0065	–	–	0.000	
GARCH(3,3)	1.161	−0.346	0.165	0.0708	0.0090	−0.0043	0.0015	1.4160	20.180	2.46	1.33	4313.0
	0.117	0.265	0.158	0.0182	0.0013	0.0011	0.0017	0.0065	0.000	0.293	0.513	

(B) Model 4 (long memory) parameter estimates for the equilibrium short rate dynamics using one-week Eurodollar rates (4505 observations)

Pricing Kernel	β_1	β_2	β_3	β_0	ϕ_1	ϕ_2	ϕ_3	d	ψ	$\phi_k=\phi_{k-1}=0$	$\beta_k=\beta_{k-1}=0$	$\phi_k=\beta_k=0$	QLL
FIGARCH(1,d,1)	0.986	–	–	0.0479	0.990	–	–	0.0054	1.397	–	–	–	4259.0
	0.002	–	–	0.0084	0.001	–	–	0.0009	0.065	–	–	–	
FIGARCH(2,d,2)	1.072	−0.088	–	0.0565	1.080	−0.092	–	0.0008	1.408	–	–	4.340	4304.5
	0.057	0.056	–	0.0112	0.059	0.058	–	0.0017	0.058	–	–	0.114	
FIGARCH(3,d,3)	1.220	−0.638	0.399	0.0633	1.225	−0.640	0.401	0.0040	1.408	178.330	179.78	85.76	4351.5
	0.026	0.050	0.045	0.0110	0.026	0.051	0.046	0.0007	0.040	0.000	0.000	0.000	

Notes: Each table contains estimated coefficients, robust Wald statistics and the quasi log-likelihood for three different specifications of the pricing kernel. Standard errors are indicated under estimated coefficients. Significance levels are indicated under the Wald statistics. Both tables use 4505 observations on one-week Eurodollar rates in estimation. In the description of the Wald statistics, α_k denotes the highest order α coefficient, β_k denotes the highest order β coefficient and ϕ_k denotes the highest order ϕ coefficient in each specification.

A comparison of Tables 1A and 2A indicates that there is an interesting difference between the empirical regularities for the weekly rates as opposed to the overnight rates. Whereas in Table 1A the standard errors increase significantly as we increase p and q in the GARCH(p,q) specification, this is not the case in Table 2A. The same can be said about the comparison between Tables 1B and 2B. Consequently, the robust Wald statistics in Table 2 indicate that the inclusion of extra coefficients is supported by the data. Investigation of higher orders of p and q indicated that very high orders are needed to provide an adequate statistical fit for the overnight rates (not reported). The highest order that we investigated is $p=q=15$, and the test statistics indicated that including α_{15} and β_{15} improves the statistical fit of the model. Perhaps this finding is not surprising: as mentioned above, the time series of overnight rates contains important microstructure effects, and is consequently harder to fit with a parsimonious parameterization. These findings are of interest for the study of term structure models. They seem to indicate that the performance for a given model may be critically affected by the choice of the short-term interest rate.[22]

Table 3 presents the estimation results for Models 3 and 4, using data on one-week Eurodollar rates. The result corresponding to the modified GARCH(1,1) dynamic is presented in row 1 of Table 3A. The point estimate for ψ is 1.340. This heteroskedasticity parameter is very precisely estimated. The result based on the robust standard error shows that it is significantly different from 1. This result is largely consistent with the finding by Chan et al. (1992), although our estimate for ψ is slightly lower than theirs. The other results in Table 3A can be summarized very briefly, because they largely confirm the findings of Table 1. The data favor larger values for p and q, and therefore support departures from the Markovian model. However, for higher-order modified GARCH dynamics, the standard errors are fairly large and the robust Wald statistics in some cases favor a more parsimonious parameterization. The estimated values for ψ in rows 2 and 3 are very similar to the estimate in row 1. The estimation using the long-memory model in Table 3B also yields a value of ψ similar to those in Table 3A. The estimates for the α_i parameters in Table 3A are very different from those in Table 1. The estimates for the β_i parameters are relatively similar, however. Point estimates of the long-memory parameter d in Table 3B are lower than those in Table 1B. For FIGARCH specifications with high orders of p and q, the estimated d parameter is negative and insignificantly estimated (not reported).

Table 4 presents empirical results for Models 3 and 4 using the overnight Eurodollar rates. Interestingly, the point estimates for the parameter ψ are very different from those in Table 3. Moreover, the robust standard errors suggest

[22] See Chapman, Long and Pearson (1999) on the importance of the choice of short rate.

Table 4
(A) Model 3 (short memory) parameter estimates for the equilibrium short rate dynamics using overnight Eurodollar rates (3280 observations)

Pricing kernel	β_1	β_2	β_3	β_0	α_1	α_2	α_3	ψ	$\alpha_k=\alpha_{k-1}=0$	$\beta_k=\beta_{k-1}=0$	$\alpha_k=\beta_k=0$	QLL
GARCH(1,1)	0.972	–	–	0.0420	0.0268	–	–	0.9500	–	–	–	3592.7
	0.005	–	–	0.0412	0.0051	–	–	0.1270	–	–	–	
GARCH(2,2)	0.976	0.002	–	0.0226	0.0275	−0.0048	–	0.9280	–	–	9.26	3632.6
	0.013	0.013	–	0.0299	0.0052	0.0020	–	0.1280	–	–	0.010	
GARCH(3,3)	0.841	0.165	−0.024	0.0085	0.0282	−0.0027	−0.0064	0.9050	1.810	3.73	8.20	3680.9
	0.144	0.454	0.309	0.1145	0.0053	0.0030	0.0110	0.1010	0.405	0.154	0.017	

(B) Model 4 (long memory) parameter estimates for the equilibrium short rate dynamics using overnight Eurodollar rates (3280 observations)

Pricing kernel	β_1	β_2	β_3	β_0	ϕ_1	ϕ_2	ϕ_3	d	ψ	$\phi_k=\phi_{k-1}=0$	$\beta_k=\beta_{k-1}=0$	$\phi_k=\beta_k=0$	QLL
FIGARCH(1,d,1)	0.989	–	–	0.0002	1.002	–	–	0.0165	0.882	–	–	–	3686.1
	0.003	–	–	0.0131	0.004	–	–	0.0043	0.131	–	–	–	
FIGARCH(2,d,2)	0.671	0.318	–	0.0022	0.678	0.324	–	0.0217	0.906	–	–	63.820	3726.0
	0.088	0.087	–	0.0109	0.088	0.087	–	0.0044	0.128	–	–	0.000	
FIGARCH(3,d,3)	0.678	−0.223	0.533	0.0036	0.685	−0.218	0.535	0.0206	0.901	19.130	19.15	18.64	3760.1
	0.352	0.406	0.136	0.0382	0.357	0.407	0.138	0.0052	0.200	0.000	0.000	0.000	

Notes: Each table contains estimated coefficients, robust Wald statistics and the quasi log-likelihood for three different specifications of the pricing kernel. Standard errors are indicated under estimated coefficients. Significance levels are indicated under the Wald statistics. Both tables use 3280 observations on overnight Eurodollar rates in estimation. In the description of the Wald statistics, α_k denotes the highest order α coefficient, β_k denotes the highest order β coefficient and ϕ_k denotes the highest order ϕ coefficient in each specification.

that the more general heteroskedasticity pattern is not supported by the data. This result indicates that the heteroskedasticity properties of the overnight Eurodollar rates are different from the one-week Eurodollar rates. Given that the data do not support a more general heteroskedasticity pattern, it is not surprising that the other parameter estimates for Models 3 and 4 presented in Table 4 do not differ greatly from those for Models 1 and 2 presented earlier in Table 2.

6. The economic significance of long memory

Whereas the estimates of the long memory parameter d are statistically significant for a parsimonious parameterization of the short memory component, they are fairly small in magnitude. It is therefore of critical importance to investigate whether these small estimates of the long-memory parameter are economically significant. To investigate this issue, we present the results of a small-scale simulation experiment that investigates the term structure implications of the long-memory estimates obtained in the previous section.

The implications for the term structure can be investigated by exploiting the results in Duan (1996), who characterizes the dynamics under the risk-neutral price measure Q for this type of interest processes. From Theorem 3 in Duan (1996), we have that under the risk-neutral measure Q, $\varepsilon_{t+1}|\Omega_t \sim N(-\sqrt{q_{t+1}}, 1)$, or equivalently, using $r_t \equiv r_t' - \mu$ and $r_t' = \mu + (\kappa - 0.5)q_{t+1}$, $\varepsilon_{t+1}|\Omega_t \sim N\left(-\sqrt{\frac{1}{(\kappa-0.5)}r_t}, 1\right)$. Then defining $\varepsilon^*_{t+1} = \varepsilon_{t+1} + \sqrt{\frac{1}{(\kappa-0.5)}r_t}$ we have $\varepsilon^*_{t+1}|\Omega_t \sim N(0,1)$ under Q. We can therefore easily obtain pricing results for our models. For example the specification under the risk-neutral pricing measure Q for the short memory dynamic (7) in Model 1 is

$$r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \sum_{i=1}^{q} \alpha_i L^{i-1} r_{t-1} \left(\varepsilon^*_t - \sqrt{\frac{1}{(\kappa-0.5)} r_{t-1}}\right)^2 \tag{24}$$

and for the long memory dynamic (8) in Model 2 we get

$$r_t = \beta_0 + \sum_{j=1}^{p} \beta_j L^{j-1} r_{t-1} + \left[1 - \sum_{j=1}^{p} \beta_j L^j - \left(1 - \sum_{i=1}^{\max(p,q)} \phi_i L^i\right)(1-L)^d\right] r_t \left(\varepsilon^*_{t+1} - \sqrt{\frac{1}{(\kappa-0.5)} r_t}\right)^2 \tag{25}$$

The risk-neutral dynamics for Models 3 and 4 can be obtained in the same fashion. Theorem 2 and Corollary 1 in Duan (1996) demonstrate that this measure Q is the equivalent martingale measure, which means that the yield to maturity at time t for a default-free zero-coupon bond of maturity τ is given by

$$R_t(\tau) = -\frac{1}{\tau}\ln\left(E^Q\left(e^{-\sum_{i=t}^{t+\tau-1} r_i}|\Omega_t\right)\right) \tag{26}$$

using the interest rate dynamics under the risk-neutral pricing measure Q given by (24) and (25). Note that we do not have an analytical expression for (26). However, the expressions for the short-term interest rate under risk neutrality in (24) and (25) make it straightforward to compute (26) using Monte Carlo simulation.

In order to investigate if the estimates of the long memory parameter d are statistically significant, we conduct the following experiment. We use the estimates for the overnight Eurodollar rates for Models 1 and 2 contained in Table 2. We analyze the implied term structure corresponding to the last day of the sample, May 30, 1997, by computing the yields for maturities ranging from one day ($\tau=1$) to 10 years ($\tau=3650$) for two interest rate dynamics: the GARCH(1,1) and FIGARCH(1,d,1). In addition, we report results for the FIGARCH(1,d,1) where we set $d=0$ while not altering the other parameters, to illustrate that the differences between the GARCH(1,1) and the FIGARCH(1,d,1) term structure are not exclusively due to the d parameter. We corroborated our results by repeating this exercise for other days in the sample, but we only report the results for the last day of the sample because it is the cleanest experiment, i.e., no information from the period over which the yields are computed is used in the estimation of the parameters. We set the number of simulation paths in the yield computation to 5000. Increasing the number of simulations above 5000 does not materially change the results. The simulation of the long-memory dynamic is implemented in the same way as in the estimation stage, namely by truncating the infinite sum in (11) at 1000. Finally, we set the price of risk, $\frac{1}{(\kappa-0.5)}$, equal to 150 in the simulation. It is important to note that the price of risk cannot be determined simply from the time series of interest rates. Our estimates in Tables 1 through 4 identify the regression intercept β_0, but this intercept is equal to $\beta_0'(\kappa-0.5)$ and therefore does not identify the price of risk.

Fig. 2 presents the implied term structures. These results clearly demonstrate that a small estimate of the long memory parameter d can have a significant economic impact. The term structure generated by the FIGARCH(1,d,1)

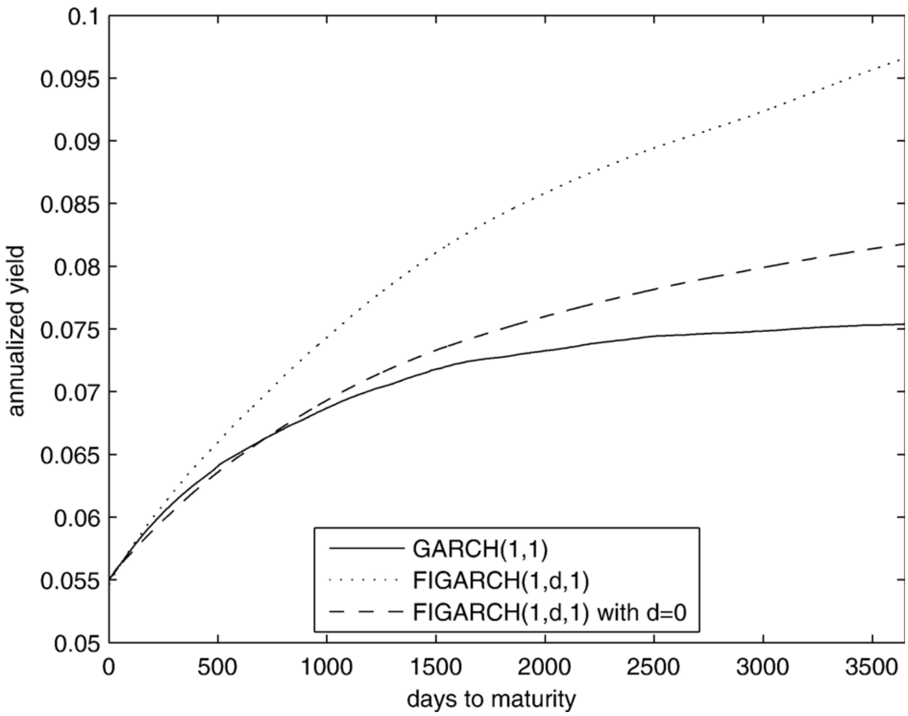

Fig. 2. Implied term structures on May 30, 1997 under different model specifications.

dynamic is significantly steeper than the one generated by the GARCH(1,1) model. When setting $d=0$ in the FIGARCH(1,d,1) specification while not altering the other parameter values, the yield curve falls in between the GARCH(1,1) and FIGARCH(1,d,1) curves, because the β_0, β_1 and ϕ_1 coefficients also partially determine the steepness of the FIGARCH(1,d,1) yield curve. Obviously, these results are only meant to be suggestive of the distinctive characteristics of the long-memory dynamic. A more extensive analysis of the implications of this model for the term structure and the ability of different parameterizations to achieve a satisfactory cross-sectional and time-series fit is left for future work.

7. Conclusion

This paper analyzes the empirical performance of different specifications for the short-term interest rate. The primary focus of the paper is a statistical in-sample analysis of these specifications, but we also analyze the implications of some of our estimates for the term structure. We use a non-standard approach that specifies the conditional volatility of the pricing kernel as a nonnegative process, which yields nonnegative interest rates under certain parameter restrictions. Specifically, we model the conditional volatility of the pricing kernel as a GARCH-type process, implying a chi-square innovation for interest rates. The resulting class of interest rate models includes Markovian and non-Markovian interest rate dynamics, which can display short or long memory. It is also straightforward in this framework to incorporate more general heteroskedasticity patterns that are known to be consistent with interest rate data.

The empirical analysis, using overnight and one-week Eurodollar rates, indicates that the data support deviations from the Markovian assumption. The long memory parameter is small but very precisely estimated. While our results indicate that we can improve the fit of the short-memory models and reduce the statistical significance of the long-memory parameter by increasing the number of short-memory parameters, a simulation experiment shows that the term structure implications of long-memory and short-memory interest rate dynamics are very different. Another empirical finding is that for one-week Eurodollar rates, the data support heteroskedasticity patterns different from those within the exponential affine class. For the overnight Eurodollar rates, this is not the case. Interestingly, the overnight and one-week rates also differ in other dimensions, but the two interest rate series produce similar outcomes when the interest rate model is restricted to be Markovian.

Our empirical results seem to indicate that the class of interest rate models proposed in this paper can parsimoniously capture some important aspects of the data. The ability of these models to describe the cross-sectional variations of the term structure and to price term structure derivatives is left for future research. Another finding that deserves attention is the apparent ability of richly parameterized short memory models to mimic certain aspects of the performance of long memory models, such as the time series in-sample fit of the short rate. It is well-known that parsimonious parameterizations can be helpful out-of-sample, and therefore it will be instructive to compare the predictive (out-of-sample) performance of short and long memory models as opposed to their descriptive (in-sample) performance.

References

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

Ahn, D.-H., Gao, B., 1999. "A parametric nonlinear model of term structure dynamics". Review of Financial Studies 12, 721–762.

Andersen, T., Lund, J., 1997. Estimating continuous-time stochastic volatility models of the short-term interest rate. Journal of Econometrics 72, 343–377.

Andersen, T., Lund, J., 1999. Stochastic volatility and mean drift in the short-term interest rate diffusion: sources of steepness, level and curvature in the yield curve. Working paper. Northwestern University.

Backus, D., Zin, S., 1993. Long-memory inflation uncertainty: evidence from the term structure of interest rates. Journal of Money, Credit and Banking 25, 681–700.

Backus, D., Zin, S., 1994. Reverse engineering the yield curve. Working paper. New York University.

Baillie, R., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 31, 5–59.

Baillie, R., Bollerslev, T., 2000. The forward premium anomaly is not as bad as you think. Journal of International Money and Finance 19, 471–488.

Baillie, R., Bollerslev, T., Mikkelsen, H., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3–30.

Balduzzi, C., Das, S., Foresi, S., Sundaram, R., 1996. A simple approach to three factor affine term structure models. Journal of Fixed Income 6, 43–53.

Black, F., Karasinski, P., 1991. Bond and option pricing when short rates are lognormal. Financial Analysts Journal, 52–59.

Black, F., Derman, E., Toy, W., 1990. A one-factor model of interest rates and its application to treasury bond options. Financial Analysts Journal 33–39.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Mikkelsen, H., 1996. Modelling and pricing long-memory in stock market volatility. Journal of Econometrics 73, 151–184.
Bollerslev, T., Wooldridge, J., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. Econometric Reviews 11, 143–172.
Brennan, M., Schwartz, E., 1979. A continuous-time approach to the pricing of bonds. Journal of Banking and Finance 3, 133–155.
Brenner, R., Harjes, R., Kroner, K., 1996. Another look at models of the short term interest rate. Journal of Financial and Quantitative Analysis 31, 85–107.
Campbell, J., Lo, A., MacKinlay, A., 1997. The Econometrics of Financial Markets. Princeton University Press.
Chan, K., Karolyi, G., Longstaff, F., Sanders, A., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209–1227.
Chapman, D., Long, J., Pearson, N., 1999. Using proxies for the short rate: when are three months like an instant? Review of Financial Studies 12, 763–806.
Chen, R.-R., Scott, L., 1993. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. Journal of Fixed Income 3, 14–31.
Chung, C.-F., 1999. Estimating the fractionally integrated GARCH model. Working Paper. National Taiwan University.
Conley, T., Hansen, L., Luttmer, E., Scheinkman, J., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525–577.
Connolly, R., Guner, N., 1999. Long memory characteristics of the distributions of treasury security yields, returns and volatility. Working paper. University of North Carolina, Chapel Hill.
Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.
Dai, Q., Singleton, K., 2000. Specification Analysis of Affine Term Structure Models. Journal of Finance 55, 1943–1978.
Duan, J., 1996. Term Structure and Bond Option Pricing under GARCH. Working paper, Hong Kong University of Science and Technology.
Duan, J., 1997. Augmented GARCH(p,q) process and its diffusion limit. Journal of Econometrics 79, 97–127.
Duan, J., Jacobs, K., 1996. A simple long-memory equilibrium interest rate model. Economics Letters 53, 317–321.
Duffie, D., Kan, R., 1996. A yield–factor model of interest rates. Mathematical Finance 6, 379–406.
Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.
Engle, R., Ng, V., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.
Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.
Hamilton, J., 1988. Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. Journal of Economic Dynamics & Control 12, 385–423.
Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: a new methodology. Econometrica 60, 77–105.
Ho, T., Lee, S.-B., 1986. Term structure movements and pricing interest rate contingent claims. Journal of Finance 41, 1011–1020.
Hull, J., White, A., 1990. Pricing interest-rate derivative securities. Review of Financial Studies 3, 573–592.
Hull, J., White, A., 1993. One-factor interest rate models and the valuation of interest-rate derivative securities. Journal of Financial and Quantitative Analysis 28, 235–254.
Jegadeesh, N., Pennacchi, G., 1996. The behavior of interest rates implied by the term structure of Eurodollar futures. Journal of Money, Credit and Banking 28, 426–446.
Koedijk, K., Nissen, F., Schotman, C., Wolff, C., 1997. The dynamics of short term interest rate volatility reconsidered. European Finance Review 1, 105–130.
Litterman, R., Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income 1, 54–61.
Longstaff, F., Schwartz, E., 1992. Interest rate volatility and the term structure: a two-factor general equilibrium model. Journal of Finance 47, 1259–1282.
Lumsdaine, R., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64, 575–596.
McCurdy, T., Michaud, P., 1996. Capturing long memory in the volatility of equity returns: a fractionally integrated asymmetric power ARCH model. Working paper. Queen's University.
Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Nelson, D., Cao, C., 1992. Inequality constraints in the univariate GARCH model. Journal of Business and Economic Statistics 10, 229–235.
Pearson, N., Sun, T.-S., 1994. Exploiting the conditional density in estimating the term structure: an application to the Cox, Ingersoll, and Ross Model. Journal of Finance 49, 1270–1304.
Pfann, G., Schotman, C., Tschernig, R., 1996. Nonlinear interest rate dynamics and implications for the term structure. Journal of Econometrics 74, 149–176.
Ray, B., 1991. Modeling long-memory processes for optimal long-range prediction. Journal of Time Series Analysis 14, 511–525.
Shea, G., 1997. Uncertainty and implied variance bounds in long-memory models of the interest rate term structure. Empirical Economics 16, 287–312.
Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.
White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrics 50, 1–25.