



ELSEVIER

Journal of Econometrics 74 (1996) 149–176

JOURNAL OF  
Econometrics

## Nonlinear interest rate dynamics and implications for the term structure

Gerard A. Pfann<sup>a</sup>, Peter C. Schotman<sup>\*, b</sup>, Rolf Tschernig<sup>c</sup>

<sup>a</sup>*Department of Econometrics, University of Limburg, 6200 MD Maastricht, The Netherlands*

<sup>b</sup>*LIFE, University of Limburg, 6200 MD Maastricht, The Netherlands*

<sup>c</sup>*Institut für Statistik und Ökonometrie, Humboldt Universität, 10178 Berlin, Germany*

---

### Abstract

This paper explores nonlinear dynamics in the time series of the short-term interest rate in the United States. The proposed model is an autoregressive threshold model augmented by conditional heteroskedasticity. The performance of the model is evaluated by considering its implications for the term structure of interest rates. The nonlinear dynamics imply a form of nonlinearity in the levels relation between the long and the short rate.

*Key words:* Nonlinear dynamics; Term structure of interest rates; SETAR models

*JEL classification:* C22; E43

---

### 1. Introduction and background

The dynamic process of the short-term interest rate receives considerable attention in the finance literature, where it is the main input in models of the term structure of interest rates. In most of this literature it is assumed that the short rate follows some autoregressive process with possibly conditional heteroskedastic errors. Campbell and Shiller (1984) analyze linear discrete time ARIMA processes, and their implications for the term structure. The Cox, Ingersoll, and Ross (1985),

---

\* Corresponding author.

Gerard Pfann gratefully acknowledges the hospitality of the Graduate School of Business of the University of Chicago, where part of this research was done, and the financial support of the William J. Fulbright Commission and of the Royal Netherlands Academy of Arts and Sciences. Rolf Tschernig gratefully acknowledges the hospitality and support of the University of Limburg. Part of his research was also carried out within the Sonderforschungsbereich 373 at the Humboldt University, Berlin, which is funded by the Deutsche Forschungsgemeinschaft. We thank Larry Christiano, Wouter den Haan, John Faust, Daniel Peña, Simon Potter, Neil Shephard, and the referees for helpful comments, but all errors are our own.

Merton (1973), and Vasicek (1977) models are examples from the continuous time finance literature. Most of these popular interest models in finance are based on univariate linear AR(1) processes, with or without a unit root, and with varying forms of heteroskedasticity. Chan, Karolyi, Longstaff, and Sanders (1992, CKLS) have empirically investigated the class of univariate autoregressive linear models.

There are, however, several indications for nonlinear dynamics in interest rates, both in the mean as well as in the variance. For example, Hamilton (1988) applies a Markov switching regime model to monthly U.S. short-term interest rate data, and finds that this model fits the data better than a linear autoregressive model. Granger (1993) reports regressions on monthly data showing that the U.S. short-term interest rate depends in a nonlinear way on the spread between long and short rates. Anderson (1994) provides additional evidence for this type of nonlinear effects. Kozicki (1994) finds different responses to positive and negative shocks. Naik and Lee (1993) and Das (1993) link the nonlinearities to changes in economic regime and stochastic jumps, respectively. This evidence raises questions about the appropriateness of a linear process to fit the short-term U.S. interest rate.

Additional evidence for nonlinear dynamics is obtained from the term structure of interest rates. Linear time series models for the short-term interest rate also imply linearity of the term structure relation between the yield on a long-term bond and the short rate. *Nonlinear* dynamics will imply a *nonlinear* equilibrium relation between the level of the short-term interest rate and the long-term interest rates. For instance, linear dynamics cannot explain the empirical fact that the long-term interest rate is less responsive to shocks in the short rate, when the short rate is high compared to when it is at low levels. We propose a nonlinear model for the short rate that is consistent with the empirical fact that the ratio of the volatilities of long and short rates decrease as the short-term interest rate rises.

A visual inspection of the U.S. data, shown in Fig. 1, suggests that interest rates behave differently in the episode around between October 1979 and mid 1982, when interest rates were high and also extremely volatile. The episode around 1974 has similar features, although both the level as well as the variance were less dramatic. The important modelling decision is how to treat the few episodes when the interest rate is clearly behaving differently. Do we deem these observations as outliers and throw them out, or should we specify a separate regime for periods with high interest rates? What is the probability that such an exceptional episode recurs? We follow the latter approach, and investigate the hypothesis that the interest rate dynamics deviate from a random walk at high levels of the interest rate. Such kind of regime-specific interest rate behavior can result from monetary policy of the Federal Reserve Bank, and could be motivated by fixed costs, in terms of reputation or credibility, that are associated with changes in policy. Linear models are not able to capture this kind of dynamics and nonlinear models are required. In particular, an autoregressive model with different regimes seems a promising alternative.

The main purpose of this paper is to explore the impact of possible nonlinear dynamics in the univariate time series of the short-term interest rate for term structure models. We expect term structure models to fit better if nonlinearities are allowed for in the stochastic model of the short-term interest rate.

We will take the results of CKLS (1992) as our point of departure. In this paper we will use a dataset similar to CKLS (1992). This also means that we confine ourselves to univariate models. For the univariate models we can still solve the term structure model for the yield on long-term bonds using numerical integration techniques. In a multivariate model with the yield on long-term bonds as one of the forecasting variables for the short-term interest rate, the formal solution of the equilibrium term structure becomes technically more involved, and is beyond the scope of this paper.

An important issue in interest rate models are the long-memory properties. Since in the term structure models the yield on a ten-year bond depends on expected short-term interest rates over a ten-year horizon, long-term expectations and hence the long-memory properties become crucially important. Under a linear stationary autoregressive process the long-term expectations quickly converge to the unconditional mean, while for the random walk [or an  $I(1)$  process in general] the long-term expectations will depend on the current interest rate. Parsimonious models under both hypotheses appear inconsistent with the data. Both fractional integration as well as nonlinear dynamics are alternatives that can provide the required long-memory properties in a parsimonious model. The advantage of the nonlinear threshold model that we use in this paper is that it can handle time-varying persistence of shocks. The main characteristic of the model we develop is that the random walk is correct most of the time, except in extraordinary circumstances.

The plan of the paper is as follows. In Section 2 we explain the theoretical implications of models with linear interest rate dynamics. Section 3 presents the empirical puzzles that linear dynamics have difficulties to account for. The empirical evidence is based on monthly observations of the U.S. three-month T-bill rate and the ten-year government bond rate for the period January 1962 through June 1990. Section 4 discusses threshold autoregressive models; empirical results with these models are presented in Section 5. This includes the development of a Gibbs sampling algorithm that allows formal (Bayesian) statistical inference on the threshold parameters. Section 6 investigates the implied term structures. Section 7 concludes.

## **2. Term structure models**

Our motivation for considering nonlinear interest rate dynamics stems from some empirical puzzles that cannot be solved by the widely-used term structure models like Vasicek (1977) or Cox, Ingersoll, and Ross (1985, CIR). These term

structure models belong to the class of so-called one-factor models. They are built on time series processes of the form

$$dr = \mu(r)dt + \sigma(r)dZ, \quad (1)$$

where  $r$  is the instantaneous spot rate and  $Z$  is standard Brownian motion. The drift function  $\mu(r)$  and the volatility function  $\sigma(r)$  both depend solely on the single state variable  $r(t)$ .<sup>1</sup> For the CIR and Vasicek models the drift function is linear in  $r$ :

$$\mu(r) = \kappa(\theta - r), \quad (2)$$

with  $\kappa$  the mean reversion parameter and  $\theta$  the unconditional mean of the (nominal) interest rate. The volatility is of the form

$$\sigma(r) = \sigma r^\gamma, \quad (3)$$

with  $\gamma = 0$  for the Vasicek model and  $\gamma = \frac{1}{2}$  for the CIR model. The models imply that the yield on a discount bond with time to maturity  $\tau$  is given by

$$R(t, \tau) = a(\tau) + b(\tau)r(t), \quad (4)$$

with the functions  $a(\tau)$  and  $b(\tau)$  depending on the parameters of the time series process defined in (1), (2), and (3), and the price of risk. A linear relationship between a long-term yield and the spot rate exist only if the drift function is linear as in (2).<sup>2</sup>

In discrete time an equation like (4) is obtained for an AR(1) process under the simplifying assumption of the expectations hypothesis. In discrete time  $r_t$  denotes a one-period interest rate, and  $R_t^{(n)}$  the yield on a discount bond with  $n$  periods to maturity. According to the expectations model,

$$R_t^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} E_t[r_{t+i}] + \phi^{(n)}, \quad (5)$$

which states that the yield on an  $n$ -period bond is given by the average expected future one-period short rates  $r_{t+i}$  plus a possible term premium  $\phi^{(n)}$ .<sup>3</sup> A discrete time approximation to the diffusion process [(1) to (3)] is given by a linear AR(1) process (possibly with heteroskedasticity):

$$r_t - r_{t-1} = (1 - \rho)(\theta - r_{t-1}) + \sigma_t \varepsilon_t, \quad (6)$$

<sup>1</sup> As a matter of notation, we will denote time for continuous time variables in parentheses as in  $r(t)$  and for discrete time variables by a subscript as in  $r_t$ .

<sup>2</sup> See Duffie (1992, Ch. 7). There are also conditions on the volatility function that we will not explore here.

<sup>3</sup> This model is used in most of the monetary economics literature of the term structure. See Mankiw (1986) and Campbell and Shiller (1987) for extensive empirical work with this model.

with  $\rho$  the first-order autocorrelation coefficient and  $\varepsilon_t$  white noise. By direct calculation of the expected future short rates we find the long-term rate as

$$R_t^{(n)} = \psi^{(n)} + w^{(n)}r_t, \quad (7)$$

where

$$w^{(n)} = \frac{1 - \rho^n}{n(1 - \rho)},$$

and  $\psi^{(n)}$  is a constant related to the term premium and the parameters of the short rate process. For coupon bonds, which we will use in the empirical work, most of the empirical literature has followed Shiller (1979)'s linearised version of the expectations hypothesis that replaces (5) by

$$Y_t^{(n)} = \frac{1 - \delta}{1 - \delta^n} \sum_{i=0}^{n-1} \delta^i E_t[r_{t+i}], \quad (8)$$

where  $Y_t^{(n)}$  is the yield to maturity on the bond and where  $\delta = (1 + \bar{R})^{-1}$  is a constant discount factor.<sup>4</sup> For the yield  $Y_t$  the AR(1) process implies that the coefficient  $w$  equals

$$w = \frac{1 - \delta}{1 - \delta\rho} \frac{1 - (\delta\rho)^n}{1 - \delta^n}. \quad (9)$$

Eqs. (4), (7), and (8) are deterministic and will not hold exactly. Deviations can be due to model errors or to omitted factors. However, the residuals of (7) or (8) should be unrelated to the state variable  $r_t$ . In particular the coefficient  $w$  should be a constant, and not depend on  $r$ .

In the next section we will empirically test this implication, and show that linearity is not a valid assumption for single-factor models. The type of systematic variation in  $w$  will indicate in what direction to modify the linear AR(1) model (6).

### 3. Data and stylized facts

The data series in this paper consists of monthly observations of the U.S. three-month T-bill rate and the ten-year government bond rate for the period January 1962 through June 1990; the series are shown in Fig. 1.<sup>5</sup> Table 1 provides summary statistics of the level and first difference of the data series, with and without the influential period October 1979 through October 1982. The lower

<sup>4</sup> When there can be no confusion we will omit the superscripts <sup>(n)</sup> from now on.

<sup>5</sup> All data were kindly provided by the Federal Reserve Bank of Minneapolis. Observations are sampled on the last trading day of the month.

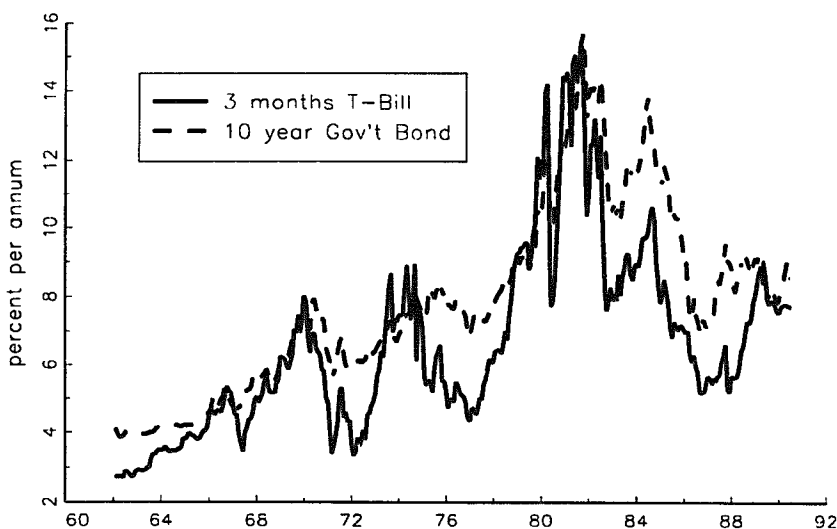


Fig. 1. Interest rates.

U.S. three-month T-bill rate (solid line) and ten-year government bond yield (dashed line); monthly data, 1962–1990.

two panels of the table split the sample according to whether the lagged short-term interest rate is above or below 8.5%. When considering first moments of the levels of the short- and long-term interest rate, a martingale appears to be a good model. A unit root can never be rejected, while first differences appear only slightly autocorrelated.<sup>6</sup>

In contrast, the volatility of interest rates appears to depend on the level of interest rates. As is evident from panels C and D in the table, interest rates are more volatile in periods when the level of the short rate is also high, consistent with the CIR model, and also emphasized in Chan, Karolyi, Longstaff, and Sanders (1992, CKLS). In fact, the well-known difference in interest rate volatility between low and high levels motivates the sample split. The positive skewness of the levels is consistent with the unconditional distribution implied by the Cox, Ingersoll, and Ross (1985) model. Another standard feature of the data is the extremely high kurtosis of the first differences, partly due to heteroskedasticity.

<sup>6</sup> Formal unit root tests (not reported) have been performed using various treatments of transient dynamics and correcting for heteroskedasticity. To perform a unit root test for the subsamples created according to the value of the lagged short-term interest, one cannot use standard critical values. Presumably the Perron (1989) critical values for models with structural breaks are more appropriate. Formal tests, considering heteroskedasticity, for the significance of the autocorrelation in the first differences cannot reject the null hypothesis of no autocorrelation.

Table 1  
Summary statistics

	Mean	Std. dev	$\hat{\rho}_1$	Min.	Max.	Skewness	Kurtosis
(A) Full sample							
$r_t$	6.64	2.73	0.971	2.70	15.52	1.04	3.95
$Y_t$	7.87	2.76	0.988	3.85	15.84	0.60	2.80
$\Delta r_t$	0.01	0.62	0.116	-3.85	2.40	-1.35	11.76
$\Delta Y_t$	0.01	0.38	0.108	-1.88	1.59	-0.44	6.53
(B) 60:1 – 79:9 and 82:10 – 90:6 ( $T = 315$ )							
$r_t$	5.98	1.92	0.961	2.70	10.63	0.25	2.17
$Y_t$	7.28	2.23	0.982	3.85	13.91	0.43	2.88
$\Delta r_t$	0.02	0.42	-0.044	-2.81	1.68	-1.40	11.70
$\Delta Y_t$	0.01	0.30	0.009	-1.13	1.09	-0.34	4.62
(C) Sample conditional on $r_{t-1} \leq 8.5\%$ ( $T = 273$ )							
$r_t$	5.61	1.62	0.875	2.70	8.93	0.01	2.03
$Y_t$	6.91	1.93	0.917	3.85	11.91	0.19	2.49
$\Delta r_t$	0.05	0.36	0.143	-1.38	1.68	-0.18	6.14
$\Delta Y_t$	0.01	0.30	0.003	-1.13	1.09	-0.34	4.62
(D) Sample conditional on $r_{t-1} > 8.5\%$ ( $T = 69$ )							
$r_t$	10.74	2.30	0.681	6.12	15.52	0.48	2.20
$Y_t$	11.71	2.10	0.683	6.90	15.84	-0.31	2.23
$\Delta r_t$	-0.12	1.16	0.195	-3.85	2.40	-0.70	4.30
$\Delta Y_t$	0.03	0.61	0.132	-1.13	0.74	-0.69	4.99

Short-term interest rate  $r_t$  is the three-month T-bill rate; the long rate  $Y_t$  is the ten-year government bond rate. Interest rates are measured as percent per annum.  $\hat{\rho}_1$  is the first-order autocorrelation. Skewness is the sample third moment scaled by  $\hat{\sigma}^3$ ; kurtosis is the sample fourth moment scaled by  $\hat{\sigma}^4$  and minus 3.

A drawback of such linear processes is, however, that they cannot explain a puzzle in the term structure literature, related to the volatility of long- versus short-term interest rates.<sup>7</sup> To see why, let us first return to the term structure model (7). This linear model implies that the conditional volatility of innovations in the long rate is proportional to the conditional volatility of the short rate with proportionality factor  $w^{(n)}$ , which in turn crucially depends on the autocorrelation coefficient  $\rho$ . If the short rate is a martingale ( $\rho = 1$ ), the coefficient  $w^{(n)}$  is equal to one for all  $n$  and the changes in the long rates should be as volatile as changes in the short rate. Yet Table 1 shows that the sample ratio of the two standard deviations is considerably smaller than one:  $0.38/0.62 = 0.61$ .

<sup>7</sup> For an overview of this literature see Shiller (1979), Campbell and Shiller (1987), and recently den Haan (1995).

On the other hand, for the empirical first-order autocorrelation coefficient in monthly data,  $\hat{\rho} = 0.97$ , the coefficient  $w^{(n)}$  decreases rapidly as a function of  $n$ . For example, for a ten-year coupon bond, setting  $\delta = (1 + 0.07)^{-1/12} = 0.994$ , we obtain  $w^{(120)} = 0.33$ . In this case, the level of the long rate should be much less volatile – since  $w = 0.33$  – than the level of the short rate. But the actual sample standard deviations (2.73 and 2.76, respectively) are about equal, therefore grossly violating this implication.

The differences in implications between the unit root and a stationary AR(1) have sparked some of the immense attention to tests of the unit root hypothesis. Shea (1989) and Backus and Zin (1993) consider a fractionally integrated model as an intermediate process. A fractionally integrated process can reconcile the unconditional relative volatilities of the interest rates with long memory properties in both time series.

However, fractional integration cannot explain why the volatility ratio falls as the interest rate level rises. Panels C and D of Table 1 show that, if the short rate is below 8.5%, the implied value of  $w$  is  $0.30/0.36 = 0.83$ , whereas for high values of  $r_{t-1}$  we find  $w = 0.61/1.16 = 0.53$ . As a crude test for the significance of the change we used Seemingly Unrelated Regression to relate the squared changes in the short and long rates to a dummy variable  $D_t$  being equal to one if  $r_{t-1} \leq 8.5\%$  and zero otherwise ( $t$ -values in parentheses):

$$(\Delta r_t)^2 = \underset{(2.0)}{0.135} D_t + \underset{(9.7)}{1.323} (1 - D_t),$$

$$(\Delta Y_t)^2 = \underset{(4.6)}{0.091} D_t + \underset{(9.4)}{0.365} (1 - D_t). \quad (10)$$

A Wald test of the restriction that the ratio of the parameters on  $D_t$  equals the ratio of the parameters on  $(1 - D_t)$  gives  $W(1) = 8.94$ , rejecting the null hypothesis.

For a slightly more sophisticated test we explicitly model the conditional heteroskedasticity as in CKLS (1992). CKLS consider the model

$$r_t - r_{t-1} = (1 - \rho)(r_{t-1} - \mu) + \sigma r_{t-1}^{\gamma} \varepsilon_t, \quad (11)$$

where the conditional volatility depends on the level of the short rate. We estimated (11) jointly for the long and short rate by quasi maximum likelihood, assuming a constant correlation between the innovations to the short and long rates ( $t$ -values in parentheses):

$$\begin{aligned} \Delta r_t &= \underset{(2.1)}{0.115} - \underset{(1.4)}{0.016} r_{t-1} + \underset{(7.8)}{0.033} r_{t-1}^{\hat{\gamma}_1} \varepsilon_t, & \hat{\gamma}_1 &= \underset{(20.8)}{1.42}, \\ \Delta Y_t &= \underset{(1.8)}{0.09} - \underset{(1.7)}{0.011} Y_{t-1} + \underset{(7.9)}{0.038} r_{t-1}^{\hat{\gamma}_2} \eta_t, & \hat{\gamma}_2 &= \underset{(14.4)}{1.15}, \end{aligned} \quad (12)$$

with the main result that  $\hat{\gamma}$  is smaller for the long rate than it is for the short rate.<sup>8</sup> A test for the null hypothesis that  $\gamma_1 = \gamma_2$  yields  $W(1) = 6.93$ , rejecting the null. The ratio of the two volatilities is

$$h_t = \frac{0.038}{0.033} r_{t-1}^{-0.27}, \quad (13)$$

and is a decreasing function of  $r_{t-1}$ , ranging from 0.95 at  $r = 2\%$  to 0.55 at  $r = 15\%$ . At low interest levels the relative volatility would imply near random walk behavior for the short rate, while it would be more consistent with stationary dynamics at high values.

This evidence of a falling volatility ratio is at odds with the assumption of a linear mean function as this would require the volatility ratio to remain constant. This is why it also cannot be explained by a fractionally integrated process. Our interpretation is that a falling volatility ratio indicates that the amount of mean reversion in interest rates depends on the level of the interest rate. Nominal interest rates are close to a random walk until they reach high values, when the comovements of long and short rates seem to indicate that short rates become mean-reverting.

The same information is conveyed by the low-frequency components of the data. Plotting the short rate against the long rate we would expect to see observations scattered around a 45° line in case of a random walk, and scattered around a much flatter line with slope equal to  $w$  in case of a stationary AR(1). Under the alternative of nonlinear dynamics, however, and consistent with the volatility evidence above, we would expect the slope to be close to one for small values of  $r_t$ , whereas the slope flattens out as  $r_t$  increases.

In order to investigate the change of slope in the levels relation we estimated a piecewise linear regression between the yield on ten-year government bonds  $Y_t$  and the three-month T-bill rate  $r_t$ :

$$Y_t = \frac{1.19}{(0.58)} + \frac{1.02}{(0.11)} r_t - \frac{0.58}{(0.27)} \max(r_t - c, 0) + u_t, \quad (14)$$

$$\hat{\sigma} = 1.18, \quad R^2 = 0.82, \quad DW = 0.13,$$

where  $c = 10.8\%$  is the estimated breakpoint and  $u_t$  an error term.<sup>9</sup> The slope is almost equal to one at low interest rates, but (significantly) lower for high values

<sup>8</sup> The dependence of the volatility on the level of the interest rate implies that we cannot have an exact unit root. See Broze, Scaillet, and Zakoian (1993) for exact ergodicity conditions for this model. A problem with the estimated model is that  $\gamma > 1$  violates the conditions. Pagan, Hall, and Martin (1994), Koedijk, Nissen, Schotman, and Wolff (1994), and Aït-Sahalia (1995) discuss issues in estimating  $\gamma$ . It appears that the estimate of  $\gamma$  becomes smaller when GARCH effects are taken into account, and also that  $\gamma$  decreases to below one for high interest rates.

<sup>9</sup> Parameter estimates have been obtained by nonlinear least squares. Newey–West standard errors, conditional on  $c$  and using twelve lags, are in parentheses.

of the short-term interest rate. In this sense the data support the hypothesis of a regime shift at high interest rate levels.<sup>10</sup>

Summarising, the term structure data suggest nonlinear interest rate dynamics, with dynamic properties depending on the level of the short-term interest rate. Next section purports to specify a suitable nonlinear model for the short-term interest rate.

#### 4. Threshold autoregressive models

##### 4.1. Specification

A simple way to approximate a nonlinear function is the piecewise linear approximation. In this section we consider several variants of the self-exciting threshold autoregressive (SETAR) model to fit the dynamics of the three-month T-bill interest rate. We also introduce a modification of the SETAR model by adding proportional heteroskedasticity of the CKLS type [see Eq. (11)]. Statistical inference is Bayesian and proceeds through the Gibbs sampler.

Let a set of threshold parameters  $c_j$  ( $j = 0, 1, \dots, J$ ) partition the real line into  $J$  adjacent regimes or regions  $[c_{j-1}, c_j)$  with  $c_j < c_{j+1}$ ,  $c_0 = -\infty$ , and  $c_J = +\infty$ . At time  $t$  the  $j$ th regime is active if the realisation  $d$  periods ago lies within  $[c_j, c_{j+1})$ ;  $d$  is called the delay parameter. Let  $y_t$  denote the dependent variable and  $x_{jt}$  the vector of explanatory variables in regime  $j$ . A self-exciting threshold model SETAR( $K_1, \dots, K_J$ ) with  $J$  regimes can be written as<sup>11</sup>

$$y_t = \left\{ \begin{array}{ll} \beta'_1 x_{1t} + \sigma_1 \varepsilon_t & \text{if } c_0 \leq y_{t-d} < c_1 \\ \beta'_2 x_{2t} + \sigma_2 \varepsilon_t & \text{if } c_1 \leq y_{t-d} < c_2 \\ \dots & \dots \\ \beta'_J x_{Jt} + \sigma_J \varepsilon_t & \text{if } c_{J-1} \leq y_{t-d} < c_J \end{array} \right\}, \quad \text{with } \varepsilon_t \sim N(0, 1). \quad (15)$$

We will mostly concentrate on models with only first-order dynamics. In that case the conditional mean for time  $t$  in regime  $j$  is determined by a linear AR(1) specification with parameter vector  $\beta_j = (\beta_{j0} \ \beta_{j1})'$  and with  $x'_{jt} = (1, y_{t-1})'$ . Furthermore the delay parameter  $d$  is equal to one.

In the case of two-regime models the abrupt shifts of regimes can be refined by using the Smooth Transition Autoregressive Model (STAR) discussed in Granger

<sup>10</sup> Of course the single-factor models do not explain all variation in long-term yields. However, we do wish to capture all those fluctuations that are related to the level of the short-term interest rate. Although the errors are likely to be highly autocorrelated, the unit root hypothesis for yield spreads has always been rejected in the literature; see, for example, Campbell and Shiller (1987) and Hall, Anderson, and Granger (1992). Eq. (14) is therefore not likely to be spurious.

<sup>11</sup> Brockwell and Davis (1991) provide a good introduction to threshold models. An extensive analysis is given by Tong (1990).

and Teräsvirta (1993). These models include some gradual shift from one regime to another, and thus additionally require some transition function.

SETAR processes also introduce a specific kind of conditional heteroskedasticity by allowing the variance parameters  $\sigma_j$  ( $j = 1, \dots, J$ ) to vary across regimes. In this basic SETAR model conditional heteroskedasticity can only be driven by changes in the regime, i.e., by changes in  $y_{t-1}$ .

In order to allow for conditional heteroskedasticity of the CKLS type [see Eq. (11)], which has found widespread application in models of interest rate dynamics in the finance literature, we introduce the SETAR model with proportional heteroscedasticity (SETAR-PH model). With  $r_t$  denoting the short-term interest rate this model is written as

$$r_t = \beta_{j0} + \beta_{j1}r_{t-1} + \sigma_j \varepsilon_t \quad \text{if } r_{t-1} \in [c_{j-1}, c_j), \quad j = 1, \dots, J. \quad (16)$$

In the SETAR-PH model the shift in regimes is automatically determined by the nonlinearity in the mean only. Volatility is related to the level of the lagged variable and is not subject to regime shifts.

In general, determining the stationarity conditions for SETAR models is difficult. For first-order models with  $J$  regimes [SETAR(1, ..., 1)] and with  $d = 1$ , Chan, Petruccielli, Tong, and Woolford (1985) derive sufficient conditions for ergodicity, which only depend on the parameters of the two outermost regimes:

$$\begin{aligned} \beta_{11} < 1, & \quad \beta_{J1} < 1, & \quad \beta_{11}\beta_{J1} < 1; \\ \beta_{11} = 1, & \quad \beta_{J1} < 1, & \quad \beta_{10} > 0; \\ \beta_{11} < 1, & \quad \beta_{J1} = 1, & \quad \beta_{J0} < 0; \\ \beta_{11} = 1, & \quad \beta_{J1} = 1, & \quad \beta_{J0} < 0 < \beta_{10}; \\ \beta_{11}\beta_{J1} = 1, & \quad \beta_{11} < 1, & \quad \beta_{J0} + \beta_{J1}\beta_{10} > 0. \end{aligned} \quad (17)$$

For the SETAR-PH model we have the additional complication that the process is nonergodic if the volatility elasticity parameter  $\gamma$  is greater than one (see Broze, Scaillet, and Zakoian, 1993).

#### 4.2. Estimation

The statistical inference for this kind of threshold models poses some difficulties, because the likelihood function is discontinuous with respect to the threshold parameters. The concentrated likelihood function is flat for values of a threshold parameter between successive observations of the re-ordered data series  $\tilde{y}_i$  ( $\tilde{y}_1 \leq \tilde{y}_2 \leq \dots \leq \tilde{y}_T$  for the  $T$  re-ordered observations  $\tilde{y}_i = y_{t_i}$ ). While least squares or ML estimates can be obtained by a (possibly multidimensional) grid search, correct standard errors are not available.

This and other difficulties with classical sampling theory are discussed in detail in Pole and Smith (1985) and Geweke and Terui (1993) for the standard SETAR model. Both papers advocate a Bayesian approach, which we will pursue here.

However, the exact methods of these papers are not applicable if parameters are restricted across regimes. We therefore use a simulation method to obtain the marginal posterior densities of all the parameters in the model. Since the SETAR model is piecewise linear conditional on the threshold, it is well-suited for a Gibbs sampler. The Gibbs sampler can also deal with the type of heteroskedasticity in the SETAR-PH model.

The Gibbs sampler is a simulation method that produces a sample of dependent draws from the posterior distribution.<sup>12</sup> The algorithm cycles through a series of conditional posteriors. The method is effective in models where alternative factorizations of the joint posterior in conditional densities and a marginal density produce a set of conditional densities from which it is computationally easy to generate random drawings. As the Gibbs sampler is a Bayesian procedure we need to specify a prior for our parameters. All results have been obtained with the standard flat prior on all regression parameters. For the scale parameters  $\sigma_i$  we use the uninformative inverted Gamma prior  $p(\sigma) \propto \sigma^{-1}$ . For the thresholds we use a uniform prior. Details of the simulation method are given in Appendix A.

## 5. Empirical results

The specification of SETAR models requires the number of regimes a priori. We start with the simplest possible model with two regimes:

$$\Delta r_t = \begin{cases} \alpha_{10} + \alpha_{11}r_{t-1} + \sigma_1\varepsilon_t, & r_{t-1} < c, \\ \alpha_{20} + \alpha_{21}r_{t-1} + \sigma_2\varepsilon_t, & r_{t-1} \geq c. \end{cases} \quad (18)$$

Using the Gibbs sampler we obtain the results in the first row of Table 2. The first regime is a random walk and the upper regime implies mean reversion. This is exactly the type of dynamics that we anticipated in Section 2. At low levels the short rate behaves like a random walk, while it becomes mean-reverting at high levels. The process as a whole is stationary according to the criteria of Eq. (17).

The threshold is estimated very precisely. However, the regime shift is largely the result of the variance shift at high interest rate levels. It is the big difference between the  $\sigma$ 's that identifies the regimes here. The possible nonlinearity in the dynamics is dominated by the change in the second moment.

In the remainder of this section we examine some variations on the basic model to investigate the interaction between the heteroskedasticity and the non-linear dynamics. One way to separate the change in the dynamic structure and the variance shift is to estimate a three-regime SETAR model with parameter restrictions such that the breaks in  $\beta_j$  and  $\sigma_j$  are only determined by the first and

<sup>12</sup> The Gibbs sampler is described in Casella and George (1991) and Tierney (1991). Examples of applications are Geweke (1994) and McCulloch and Tsay (1994).

Table 2  
SETAR models

$$\Delta y_t = \alpha_{j0} + \alpha_{j1} y_{t-1} + \sigma_j y_{t-1}^2 \varepsilon_t, \quad y_{t-1} \in I_j(c_1, c_2)$$

Model	Threshold	$\alpha_0$	$\alpha_1$	$\sigma$	$\gamma$	$c$
SETAR	$r_{t-1} < c$	0.04 (0.08)	−0.00 (0.01)	0.37 (0.02)	0	8.44 (0.10)
	$r_{t-1} \geq c$	0.64 (0.71)	−0.07 (0.15)	1.15 (0.10)		
SETAR-TWO	$\alpha: r_{t-1} <, \geq c_1$	0.06 (0.10)	−0.00 (0.02)	0.37 (0.02)	0	10.32 ( $c_1$ ) (1.10)
	$\sigma: r_{t-1} <, \geq c_2$	4.27 (2.75)	−0.34 (0.19)	1.12 (0.10)		8.44 ( $c_2$ ) (0.11)
SETAR-PH	$r_{t-1} < c$	0.05 (0.02)	−0.00 (0.01)	0.027 (0.005)	1.54 (0.10)	10.80 (0.86)
	$r_{t-1} \geq c$	4.68 (3.18)	−0.37 (0.24)	0.027 (0.005)	1.54 (0.10)	
SETAR-SUB	$r_{t-1} < c$	0.03 (0.02)	−0.00 (0.01)	0.024 (0.02)	1.62 (0.57)	9.24 (0.93)
	$r_{t-1} \geq c$	4.33 (9.70)	−0.44 (0.95)	0.024 (0.02)	1.62 (0.57)	
SETAR-LOG	$\ln r_{t-1} < \ln c$	0.0067 (0.0054)		0.074 (0.003)	0	9.34 (2.47)
	$\ln r_{t-1} \geq \ln c$	−0.0047 (0.0257)		0.074 (0.003)	0	

The table reports posterior means and standard deviations (in parentheses) obtained by the Gibbs sampler for various specifications of the self-exciting threshold autoregression model. SETAR refers to the basic single-threshold model (18); SETAR-TWO refers to the model with separate thresholds for the mean and variance (19); SETAR-PH is the proportional heteroskedasticity model (16) with a single threshold; SETAR-SUB is the proportional heteroskedasticity model (16) with a single threshold estimated over the subsample starting in November 1982; SETAR-LOG is a model for  $\ln r_t$ .

second moments, respectively:

$$\Delta r_t = \left\{ \begin{array}{ll} \alpha_{10} + \alpha_{11} r_{t-1} & \text{if } r_{t-1} < c_1 \\ \alpha_{20} + \alpha_{21} r_{t-1} & \text{if } r_{t-1} \geq c_1 \end{array} \right\} + \left\{ \begin{array}{ll} \sigma_1 \varepsilon_t & \text{if } r_{t-1} < c_2 \\ \sigma_2 \varepsilon_t & \text{if } r_{t-1} \geq c_2 \end{array} \right\}. \quad (19)$$

The results are in the row labelled SETAR-TWO of Table 2. The average number of observations in the three regimes are 269 (= 79%), 35 (= 10%), and 37 (= 11%), respectively. The estimates of  $\sigma$  are the same as in the simpler model, but the autocorrelation parameter in the highest regime is much different. The variance threshold is exactly the same as the simple threshold in (18); the change in the dynamics takes place at a higher level of  $r_t$ . This model illustrates the consequences of restricting the variance to be the same across regimes, so that regime shifts can only be identified by a shift in the conditional mean. Because of

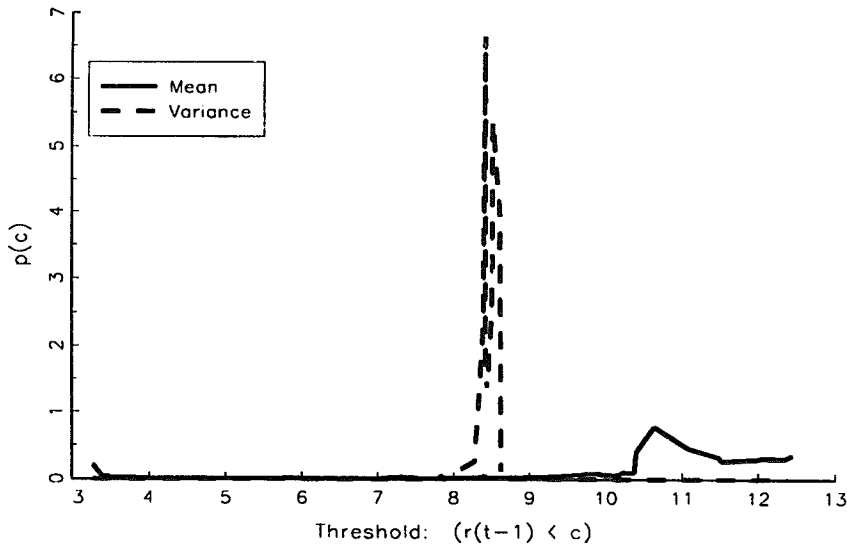


Fig. 2. Posterior density of thresholds.

The solid line shows the posterior density of the threshold for the mean; the dashed line is the posterior density for the variance threshold in the SETAR-TWO model (19).

the higher threshold for the break in the dynamic structure, the AR parameter is also much lower than in model (18). In the region between 8.44% and 10.32% the short rate still behaves like a random walk, but is already more volatile than at lower levels. As before, the estimated process is stationary. The high posterior standard deviation of the AR parameter in the high interest regime can be attributed to the small number of observations.

In comparison to the precision of the variance threshold  $c_2$  the precision of the mean threshold  $c_1$  appears to be much weaker. Fig. 2 shows the posterior densities of the two thresholds, and conveys similar information of a very precisely determined variance threshold and of much weaker evidence for the mean threshold.

Another way to separate mean and volatility effects in identifying the threshold is to combine a SETAR model for the mean dynamics and the proportional heteroskedasticity specification as in the SETAR-PH model (16) introduced above. This takes us back to a two-regime model

$$\Delta r_t = \alpha_{j0} + \alpha_{j1}r_{t-1} + \sigma r_{t-1}^j \varepsilon_t, \quad j = 1, 2, \quad (20)$$

with volatility being a continuous function of  $r_{t-1}$ .

The results are reported in the row labelled SETAR-PH of Table 2. Changing the volatility specification leaves the nonlinear mean dynamics almost unchanged, except for a somewhat higher threshold value and an improved precision of the

threshold estimate. The main effect of separating the mean and variance switches in the specifications SETAR-TWO and SETAR-PH compared to the basic SETAR model is that the stationary regime starts at a higher threshold level, but at the same time shows much stronger mean reversion. Allowing for a nonlinear mean does not alter inference on  $\gamma$ . It remains around 1.5 as in CKLS and in model (12) above.

In the sample the second regime primarily manifests itself in the period 1980–82, the same sample period that causes trouble in every empirical study of U.S. interest rates. Since only 10% of the observations are in the high regime, and since these observations are concentrated in a short period, the nonlinearity may be spurious, and only picking up a few outliers.<sup>13</sup> As a check on the robustness of the results we re-estimated the model over the last part of the sample, starting in November 1982 after the interest rate had come down from double digits to the 8% region. The results in row 4 of Table 2 are surprisingly similar to the full-sample model, except for the much larger standard deviations on the parameter estimates of the AR process in the high-interest-rate regime which, again, can be attributed to the very few number of only 17 observations in the higher regime.

The value of  $\gamma$  indicates that the variance of the interest rate innovations is strongly related to the level of the interest rate. A standard econometric procedure to deal with this form of heteroskedasticity is to take logarithms of the data. Taking the log of  $r_t$  also automatically guarantees positive interest rates by introducing a specific interaction between the conditional mean and conditional variance. It removes proportional heteroskedasticity completely only for  $\gamma = 1$ . In the finance literature time series models for the log of the short rate have been advocated in Black, Derman, and Toy (1990). When we estimated a SETAR(1,1) model for  $\ln r_t$ , the autoregressive parameter in both regimes turned out to be very close to unity, which led us to simplify the model to a combination of two geometric random walks, that we will denote as SETAR-LOG:

$$\Delta \ln r_t = \begin{cases} \alpha_{10} & \text{if } r_{t-1} < c \\ \alpha_{20} & \text{if } r_{t-1} \geq c \end{cases} + \sigma \varepsilon_t. \quad (21)$$

The results for this model are reported in row 5 of Table 2. Although both regimes are nonstationary, the upward drift in the first regime, and the downward drift in the second regime ensure that the joint process for  $\ln r_t$  is stationary, see the conditions (17). Although this is the most parsimonious threshold model

<sup>13</sup> Detection of outliers would already be a good motivation for estimating nonlinear models, though. Also note that we think the nonlinearity is nonspurious because of the term structure evidence presented in Section 3. Deleting observations to obtain a random walk will result in completely different term structure implications than the nonlinear specifications. This argument is similar to the treatment of the stock market crash in determining the risk of investment in the stock market.

that describes the main features of the data, the threshold is not very precisely determined here.<sup>14</sup>

Conditional on the posterior mean of the parameters, the residuals of the various SETAR models do not show any linear autocorrelation. The residuals do exhibit ARCH-type heteroskedasticity, though, indicating that neither the CKLS-type conditional variance model nor the regime-switching SETAR-TWO model is fully adequate, and that a separate volatility factor as in Longstaff and Schwartz (1992) might be important as a second factor.

As a further test of the time series specification of the SETAR models we considered higher-order dynamics. For the basic SETAR model as well as for the SETAR-PH model we used maximum likelihood to estimate all models with two regimes and with AR orders of 1 through 5 in each regime. We also considered all values of the delay parameter from  $d = 1$  to  $d = 3$ . These  $5 \times 5 \times 3 = 75$  models were ranked according to the Schwarz criterion, and for both the basic SETAR as well as the SETAR-PH model the first-order specification came out as best.

We also considered the logistic first-order STAR model proposed in Granger and Teräsvirta (1993). For the SETAR-PH model the least squares estimates (not reported here) implied a near infinite speed of transition, so that the model is almost indistinguishable from a SETAR model.<sup>15</sup>

Summarising the results, heteroskedasticity and nonlinearity interfere in the identification of a threshold in the basic SETAR models. The SETAR-PH model with proportional heteroskedasticity and the three-regime SETAR-TWO model appear promising as these do not show signs of serious misspecifications. Finally, the log model with two geometric random walks is suggested as a very parsimonious alternative. All models have two regimes with distinct dynamics in common. The next section explores the implications of these models for the term structure.

## 6. Implied term structures

In this section we investigate the term structure implications of the various threshold models estimated in the previous section above. We will use the expectations model (8) to generate theoretical values for the long-term interest rate.

<sup>14</sup> The posterior standard deviation for the threshold is so large because the posterior density of the threshold is bimodal, with the secondary mode around  $r = 3\%$ . Yet a model with two different thresholds appears overparameterized, and leads to extremely flat posteriors on all parameters except  $\sigma$ .

<sup>15</sup> Kozicki (1994) found evidence for smooth transition in a different specification. In her model regime switches depend on the change of the interest rate instead of the level, implying a two-factor model.

That means that we compute expected three-months interest rates over long forecast horizons, and construct the theoretical ten-year Government bond rate  $\hat{Y}_t$  conditional on different initial conditions  $r_t$  for the short rate:

$$\hat{Y}_t = \frac{1 - \delta}{1 - \delta^{40}} \sum_{k=0}^{39} \delta^k E_t[r_{t+3k}], \quad (22)$$

where  $\delta$  is the quarterly equivalent  $(0.994)^3$  of the monthly discount factor that we used before (see Section 3).

Since there are no closed form solutions for long horizon forecasts, we have to use numerical methods. For a model with only first-order dynamics the forecasts can easily be generated by approximating the SETAR model by a finite-state Markov chain. For each drawing  $\theta^{(j)}$  from the Gibbs sampler of the parameter vector  $\theta$  the Markov chain is used to construct the long-term yield for each initial condition  $r_t = x_i$  ( $i = 1, \dots, M$ ) as a function  $f_i$  of the parameters and the initial condition:

$$\hat{Y}_i(\theta^{(j)}) = f_i(\theta^{(j)}). \quad (23)$$

Averaging over all realized drawings  $\theta^{(j)}$  ( $j = 1, \dots, N$ ) of the Gibbs sampler we obtain the posterior mean and standard deviation of the implied long-term yield. The details of the algorithm are described in Appendix B. This section discusses the results.

Fig. 3 contains the main results for the levels relation. The figure shows the implied long-term yield as a function of the current short-term rate and a one-standard-error band (standard-errors are due to the parameter uncertainty). For comparison the same figure also shows the piecewise linear regression (14) of Section 3 and the long-term yield implied by a linear autoregression with monthly autocorrelation  $\rho = 0.97$ . The double threshold SETAR-TWO model (with separate thresholds for the mean and the variance) and the logarithmic SETAR-LOG model exhibit the same qualitative implications. For both models the implied levels relation becomes flatter as the short-term interest rises, exactly as we find in the data.<sup>16</sup> For low values of  $r$  the slope is close to one, the value implied by cointegration. For the double-threshold model the slope of the long rate converges quickly to a constant in the high-interest region, due to the strong estimated mean reversion in the high regime. For the SETAR-LOG model the convergence

<sup>16</sup> One possibility for obtaining nonlinearities in the levels relation even if the underlying dynamics are linear is by using the exact local expectations model instead of the linear version. The exact local expectations hypothesis states that  $(1 + R^{(n)})^{-n} = E_t \prod_k (1 + r_{t+k})^{-1}$  (see Ingersoll, 1987). If volatility is low, there is not much difference with the linearized version (22). Using the Markov chain method we checked the differences with the local expectations hypothesis, and found almost the same results.

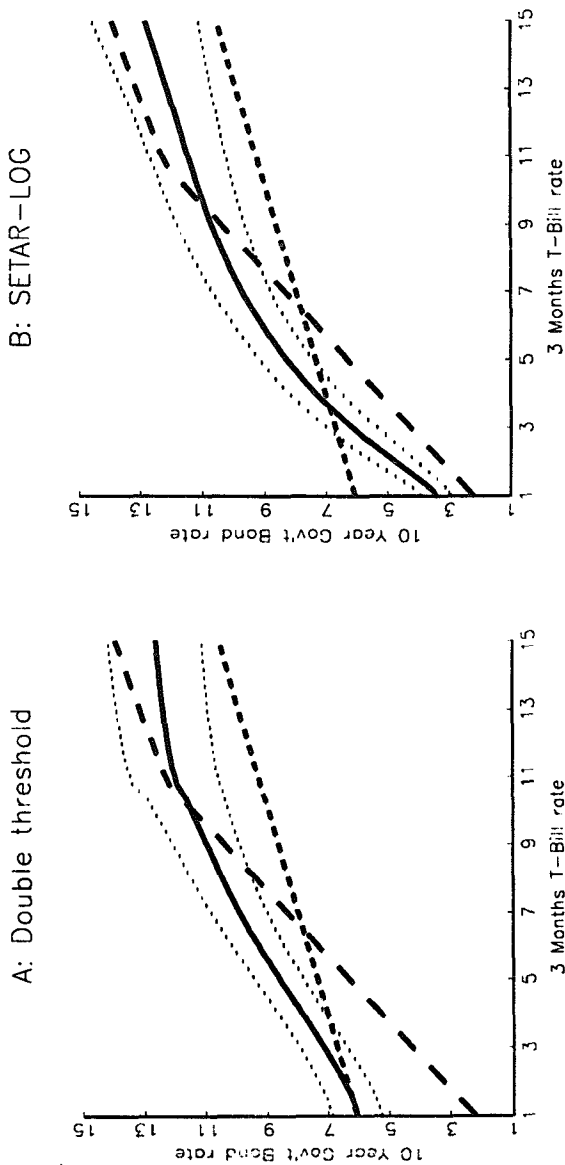


Fig. 3. Implied long-term yield.

The left panel of the figure shows the long-term yield implied by the expectations model (8) for the SETAR-TWO model with separate threshold for mean and variance. The solid line is the posterior mean of the implied long-term yield; the dotted lines one posterior standard deviation above and below; the piecewise dashed line is the estimated levels relation (14); the short dashes represent the long-term yield for a linear AR(1) with  $\rho = 0.97$ . The right panel shows the analogous relation for the SETAR-LOG model.

is much slower, implying that shocks are much more persistent for this model.<sup>17</sup> The linear AR(1) model does much worse, since a first-order autocorrelation of 0.97 implies a much flatter slope than either of the nonlinear SETAR models. The SETAR models, and especially the SETAR-LOG model, exhibit considerably more persistence than the linear AR(1) based on the monthly first-order autocorrelation. Standard errors for the implied yield are large, however.

The implied volatility ratio can be computed analogously. Conditional on the short rate being at  $r_t = x_t$ , the one-period conditional variance is directly available from the SETAR model. To obtain the conditional variance of the long rate we again use the Markov chain approximation (see Appendix B for details). For each parameter drawing  $\theta^{(j)}$  produced by the Gibbs sampler we then compute the ratio of the volatilities, denoted  $R_t(\theta^{(j)})$ , and average to obtain the posterior mean and standard deviation of the ratio.

Fig. 4 shows the results. For comparison the figure also shows the volatility ratio estimated from the actual data on the long-term rate [see Eq. (13) in Section 3]. Both nonlinear models imply a decreasing volatility ratio, as we find in the actual data. In the SETAR-TWO model the volatility of the long-term rate drops suddenly as soon as the short rate passes the threshold of 8.44%. At high short-term interest rates the model implies that a ten-year government bond rate is almost constant relative to the high volatility of the short rate. The smoothness of the long rate at high levels is due to the strong mean reversion in the high regime.

Fig. 4 also indicates that the SETAR-LOG model implies more persistence than the SETAR-TWO model. The posterior mean of the volatility ratio for the SETAR-LOG model is close to the estimates from the actual long-term yield data. However, the standard deviations around this mean are extremely large. A sensitivity analysis with selective resampling from the Gibbs results revealed that the large standard errors for the SETAR-LOG model are due to the uncertainty in the threshold parameter  $c$  (see Table 2). Computing the posterior mean of the volatility ratio using only those  $\theta^{(j)}$  for which  $c^{(j)}$  is above 7% produces a smoother picture with much lower standard errors.<sup>18</sup>

The conditional volatility results are in line with the evidence for the levels. The two types of results are theoretically related, since the volatility ratio is in effect nothing but an estimate of the first-order derivative  $\partial f / \partial r$ .

<sup>17</sup> We did not construct the implied long rate for the proportional heteroskedasticity model, since that model is nonstationary for  $\gamma > 1$ . However, if we truncate the range for the interest rate to (0,30%) in the Markov approximation, the implied long rate is very similar to the double-threshold model, since both models have the same strong mean reversion in the high-interest-rate regime.

<sup>18</sup> For all models we also redid the Gibbs run under the restricted prior that all parameter draws, conditional on the threshold parameter(s), produce a stationary dynamic process as defined by the conditions (17). These conditions are easy to impose. For the SETAR-LOG model this implies that the drift should always be positive in the lower regime and negative in the higher regime. The results are hardly different from those reported.

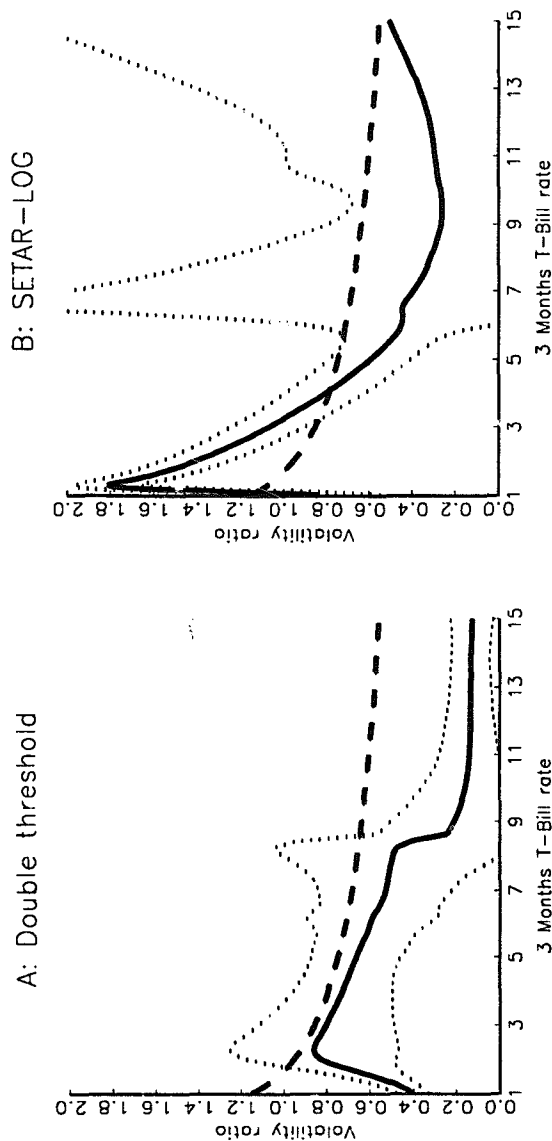


Fig. 4. Implied volatility ratio.

The figure shows the conditional volatility of the long-term yield relative to the conditional volatility for the SETAR-TWO model (left panel) and the SETAR-LOG model (right panel). The solid line is the posterior mean of this ratio; the dotted lines are one posterior standard deviation above and below. The dashed line represents the estimated volatility ratio in Eq. (13).

Table 3  
Implied moments

	Mean	s.e.	Std. dev.	s.e.
(A) SETAR-TWO model				
$r_t$	9.35	(1.89)	2.98	(0.79)
$Y_t$	9.69	(1.96)	1.37	(0.78)
$\Delta r_t$	0.00	(—)	0.84	(0.17)
$\Delta Y_t$	0.00	(—)	0.25	(0.07)
(B) SETAR-LOG model				
$r_t$	8.19	(2.50)	3.40	(1.30)
$Y_t$	8.50	(2.60)	2.06	(1.14)
$\Delta r_t$	0.00	(—)	0.67	(0.19)
$\Delta Y_t$	0.00	(—)	0.29	(0.12)

The entries are the posterior means of the moments of the long- and short-term interest rates implied by the Markov chain approximation of the three-month T-bill rate process. The columns labelled 's.e.' report the posterior standard deviations.

Finally we consider the implied first and second moments of both interest rates. Table 3 shows the implied means and variances of the long and short rate, and first differences of these rates. These moments are the model counterparts of the data moments in the first panel of the summary statistics in Table 1. Both SETAR models imply a higher unconditional mean and variance of the short rate than we actually see in the sample. The large posterior standard deviation of the unconditional mean indicates that it is not estimated very precisely, though. The large standard deviations are due to the near nonstationarity of the process, which prohibits precise inference on moments related to the level.

The standard deviation of the first differences is estimated more precisely. Since there is no drift the mean change is zero by default. The standard deviations show that the SETAR models are not sufficiently close to nonstationarity to generate enough volatility in the long rate. The ratio of the standard deviations ( $0.25/0.84 = 0.30$  for the SETAR-TWO model and  $0.29/0.67 = 0.43$  for SETAR-LOG) is still below the sample value.

The implied moments in Table 3 are directly related to the graphical evidence in Figs. 3 and 4. The slope of the function  $Y = f(r)$  in Fig. 3 is still flatter on average than what is found in the data, and the volatility ratio in Fig. 4 is below the ratio found in the data for most values of the  $r$ .

The logarithmic model comes closest to the type of nonlinearity and persistence that is required to explain the behavior of long rates. The slope of the function  $Y = f(r)$  in Fig. 3 is close to one initially (up to  $r = 6\%$ ), but it flattens out too early. Increasing the threshold parameter for this model would improve its term structure fit. But given the small number of observations on high interest rates, such a high threshold value does not follow from a univariate time series model for the short rate; but neither can we rule it out.

## 7. Conclusions

In this paper we have explored the scope of nonlinear dynamics in short-term interest rates and its implications for the term structure. Our choice of SETAR models was guided by the empirical analysis of three-months U.S. T-bill rates and its relation to the ten-year government bond rate. We found evidence for the presence of two regimes with distinct dynamics in the mean. Until interest rates reach double digits they behave like a random walk. At higher levels, however, they show a mean-reverting tendency. In specifying the nonlinear models we also accounted for the strong heteroskedasticity.

Nonlinear interest rate dynamics have asset pricing implications for the term structure of interest rates. We investigated the term structures arising from the threshold autoregressive models. One of the implications is that the levels relation between the short rate and a long rate is no longer linear (as in a cointegration model). The SETAR model predicts that the slope coefficient of the short rate is close to one when the short rate is low, but gradually decreases at higher levels due to the mean reversion. This implication is present in the data. The mean reversion at high levels can also explain why long rates are less volatile relative to the short rate at high levels of the short rate.

The term structure implications have not been formally tested in this paper. For a test of the implications we would have to estimate a simultaneous model for the long- and the short-term interest rate, consisting of the time series process of the short rate, and the implied relation between long and short rates to provide overidentifying restrictions. Given the highly nonlinear way by which the parameters of the short rate process enter the implied process for the long rate, the Gibbs sampling methods in this paper are not suited for this problem, and we leave it for future research.

One limitation of the present model is that there is only a single state variable, and consequently that there is a single threshold variable that triggers regime switches. This restriction may be crucial as a graph of the data suggests that there are several episodes where the interest rate shows mean reversion tendencies. It would therefore be interesting to introduce a time-varying threshold, where the threshold is related to the rate of inflation and a business cycle variable like the unemployment rate. Such a model, however, requires at least two factors and could be motivated by direct observations on policy target rates as in Balduzzi, Bertola, and Foresi (1993), who have data on the Fed targets for the period 1985 to 1991. Another route to improve the model is the generalization to a multivariate time series model containing both the short rate as well as the long rate. This would lead to a nonlinear counterpart of the linear cointegration models of Campbell and Shiller (1987) and Hall, Andersen, and Granger (1992).

## Appendix A: The Gibbs sampler

A general formulation of the threshold models we consider is

$$y_t = \beta_j' x_t + \sigma_j w_t^\gamma \varepsilon_t \quad \text{if } I_t = j, \quad (24)$$

where  $\beta_j$  is a  $K$ -vector of parameters,  $x_t$  a  $K$ -vector of explanatory variables,  $w_t$  a scalar variable, and  $I_t$  is an indicator variable specifying which regime will be active at time  $t$ , depending on the vector of threshold parameters  $c$ . A flat prior is assumed for  $\beta_j$ ,  $\gamma$ , and  $\ln \sigma_j$ . The prior on  $\gamma$  is restricted to  $\gamma \geq 0$ . The prior for the threshold(s) depends on the particular specification of the threshold function and will be discussed for the individual models. Some of the parameters in  $\beta_j$ ,  $\sigma_j$  can be equal across regimes, or even be known a priori; the modifications to the algorithm will be mentioned briefly. Given the data set  $Y$ , the Gibbs sampler for the general SETAR model with heteroskedasticity consists of four steps:

*1. Regression Step:* Conditional on  $\gamma$ ,  $\sigma$ , and  $c$  the distribution of  $\beta_j$  is normal with mean  $\hat{\beta}_j$  and covariance matrix  $V(\hat{\beta}_j)$  given by

$$\hat{\beta}_j = (X_j' W_j X_j)^{-1} (X_j' W_j y_j),$$

$$V(\hat{\beta}_j) = \sigma_j^2 (X_j' W_j X_j)^{-1},$$

where  $X_j$  is a  $T_j \times K$  matrix of the observations on  $x_t$  in regime  $j$ ,  $W_j$  a  $T_j \times T_j$  diagonal matrix with  $w_t^{-2\gamma}$  on its main diagonal, and  $y_j$  a  $T_j$  vector with observations on the dependent variable  $y_t$  in regime  $j$ .

*2. Inverted Gamma Step:* Conditional on  $c$ ,  $\beta$ , and  $\gamma$ , the distribution of  $\sigma_j$  is inverted Gamma with parameters  $T_j - K$  and  $s_j^2 = e_j' e_j$ , where  $e_j = W_j^{1/2}(y_j - X_j \beta_j)$ . Under the assumption  $\sigma_1 = \sigma_2$  the sum over all  $e_i^2$  replaces the regime dependent sums.

*3. Proportional Heteroskedasticity Step:* This step is only executed for the proportional heteroskedasticity model. Conditional on  $c$ ,  $\beta$ , and  $\sigma$ , the density of the heteroskedasticity parameter  $\gamma$  follows from the conditional likelihood function

$$\ln L(\gamma | c, \beta, \sigma, Y) = -\gamma \sum_{t=1}^T \ln w_t - \frac{1}{2} \sum_{j=1}^2 \sum_{t \in J_j} \left( \frac{u_t}{\sigma_j w_t^\gamma} \right)^2, \quad (25)$$

where  $u_t$  is the unscaled residual  $y_t - \beta_j x_t$  and  $J_j = \{t: I_t = j\}$ . Since there is no direct way to sample from (25), an accept/reject algorithm is used. At each iteration the conditional posterior is approximated by a Student- $t$  distribution with mean equal to the conditional mode, precision equal to the second-order derivative at the mode, and degrees of freedom equal to 4. A new value  $\gamma$  is generated from

the Student- $t$   $p(\gamma)$  and is accepted with probability

$$\Pr(\text{accept}) = k \frac{L(\gamma)}{p(\gamma)},$$

where  $k$  is a normalizing constant such that  $\Pr(\text{accept})$  is between zero and one over the whole range of  $\gamma$ .

**4a. Threshold Step:** The algorithm for drawing a new threshold value depends on the specific model. We will discuss the models that we used in Section 4, starting with the simplest case of a single threshold.

In the single threshold model the sets of the indicator variable are defined as  $J_1 = \{t: r_{t-1} < c\}$  and  $J_2 = \{t: r_{t-1} \geq c\}$ . The prior on  $c$  is assumed to be uniform on a bounded interval  $(c_l, c_u)$  with boundaries that at a minimum leave enough observations in the upper and lower regimes to do a least squares regression. Conditional on  $\beta$ ,  $\gamma$ , and  $\sigma$ , the density of the threshold  $c$  is a step function which is discontinuous at the sample points of the interest rate  $r_{t-1}$ .

$$\ln L(c|\beta, \gamma, \sigma, Y) = -\frac{1}{2} \sum_{j=1}^2 \sum_{t \in J_j} \left( \frac{u_t}{\sigma_j^2 w_t^j} \right)^2. \quad (26)$$

The function (26) is evaluated in all admissible points  $c \in (c_l, c_u)$ . Numerical integration gives the cumulative density  $F(c)$ . A new threshold value is obtained by drawing a uniform random number  $U$  and setting  $c^{(i)} = F^{-1}(U)$ .

**4b. Separate Thresholds for Mean and Variance:** For the model with separate thresholds for the variance and the dynamics [see Eq. (19)] step 4a is replaced by two other steps. Under flat priors for both thresholds and assuming  $\gamma = 0$ , the joint density of  $c_1$  and  $c_2$  is written as

$$\ln L(c_1, c_2|\beta, \sigma, Y) = -\sum_{i=1}^2 \sum_{j=1}^2 \sum_{t \in J_{ij}} \left( \frac{e_t^{(i)}}{2\sigma_j} \right)^2, \quad (27)$$

where  $J_{11} = \{t: r_{t-1} < c_1 \text{ and } r_{t-1} < c_2\}$ ,  $J_{12} = \{t: r_{t-1} \geq c_1 \text{ and } r_{t-1} < c_2\}$ ,  $J_{21} = \{t: r_{t-1} < c_1 \text{ and } r_{t-1} \geq c_2\}$ ,  $J_{22} = \{t: r_{t-1} \geq c_1 \text{ and } r_{t-1} \geq c_2\}$ , and where  $e_t^{(i)} = y_t - \beta_i' x_t$ . One of the sets  $J_{ij}$  will be empty. New values of  $c_1$  and  $c_2$  are drawn from the conditional densities of one threshold given the other. Both conditional densities are step functions with steps at each sample point  $r_{t-1} \in (c_l, c_u)$ . The procedure to draw from these densities is analogous to step 4a above.

The Gibbs sampler produces a series of  $i = 1, \dots, N$  dependent drawings by cycling through the conditional posteriors. Quantities like the expectation of certain

functions of the parameters are computed as the average

$$E[g(\theta)] = \frac{1}{N} \sum_{i=1}^N g(\theta^{(i)}). \quad (28)$$

Some quantities are easier and more accurately computed by using the analytical expression for the conditional expectation, and averaging over the conditional expectation (see Geweke, 1994). For example, the expected value of  $\beta_j$  is computed as the average conditional mean

$$E[\beta_j] = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_j^{(i)},$$

and its variance as the average conditional variance plus the variance of the conditional mean,

$$V[\beta_j] = \frac{1}{N} \sum_{i=1}^N V[\hat{\beta}_j^{(i)}] + \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_j^{(i)} - E[\beta_j])(\hat{\beta}_j^{(i)} - E[\beta_j])'.$$

This procedure is also used to estimate the marginal posterior densities. For the marginal posterior of the threshold  $c$  we average over the conditional densities,

$$p(c|Y) \propto \sum_{i=1}^N L(c|\beta^{(i)}, \gamma^{(i)}, \sigma^{(i)}, Y).$$

For the threshold models in Section 4 the number of iterations was set, after some experimentation, at  $N = 10000$ , after which the numerical standard errors are negligible.

## Appendix B: Construction of the yield curve

The expectations model for the term structure introduced in Sections 2 and 6 is represented here for the relation between the three-month rate  $r_t$  and a long-term rate on a coupon bond with maturity  $3n$  months:

$$Y_t = \frac{1 - \delta}{1 - \delta^n} \sum_{i=0}^{n-1} \delta^i E_t[r_{t+3i}], \quad (29)$$

where  $\delta$  is the discount factor in the Shiller (1979) linearization. Eq. (29) is analogous to (8), but  $\delta$  is now a discount factor for a three-month period, reflecting the three months time to maturity of the short-term interest rate.

The model with first-order dynamics for the short-term interest rate  $r_t$  can be written

$$r_{t+3} = h(r_t) + \eta_{t+3}, \quad (30)$$

with  $h(r_t)$  the conditional expectation of  $r_{t+3}$  at time  $t$  and  $\eta_{t+3}$  the prediction error.

For the actual computations the process (30) is approximated by a finite-state Markov chain. Let  $x = (x_1, \dots, x_M)'$  be the  $(M \times 1)$  vector of states  $r_t$  at which the long rate will be constructed. The first stage of the algorithm consists of the estimation of the matrix of the one-month transition probabilities

$$\tilde{a}_{ij} = \Pr(r_{t+1} = x_j | r_t = x_i). \quad (31)$$

The SETAR models estimated in Section 4 provide an analytical expression for the one-month conditional densities, as they are all conditionally normal. Since the conditional densities of the short rate have smaller variance at low levels than at high levels the number of points at low levels needs to be larger. We therefore choose a logarithmic equidistant grid, i.e.,  $x_i/x_{i-1} = d$ . The transition probabilities are found as

$$\tilde{a}_{ij} = N\left(\frac{d^{1/2}x_j - \mu_i}{\omega_i}\right) - N\left(\frac{d^{-1/2}x_j - \mu_i}{\omega_i}\right), \quad (32)$$

where  $\mu_i$  and  $\omega_i^2$  are the conditional mean and variance at  $r_t = x_i$ . Let  $\vec{A} = (\tilde{a}_{ij})(i, j = 1, \dots, M)$  be the matrix of transition probabilities. Then the transition probabilities for a three-month period are available as  $A = \vec{A}^3$ .

The vector of conditional expectations  $E[r_{t+3k} | r_t = x_i]$  ( $i = 1, \dots, M$ ) is computed as  $A^k x$ , so that the yield on a coupon bond follows from (29) as

$$\begin{aligned} y = f(x) &= \frac{1 - \delta}{1 - \delta^n} \sum_{k=0}^{n-1} (\delta A)^k x \\ &= \frac{1 - \delta}{1 - \delta^n} (I - \delta A)^{-1} (I - (\delta A)^n) x, \end{aligned} \quad (33)$$

where  $y$  is an  $(M \times 1)$  vector of yields. The conditional variance  $V_t^2[Y_{t+1}]$  of the long-term yield follows as

$$V_t^2[Y_{t+1}] = \vec{A} \left( (y - \vec{A}y) \cdot (y - \vec{A}y) \right), \quad (34)$$

where  $x \cdot x$  denotes elementwise multiplication of two vectors. The conditional variance of the short rate is directly available from the SETAR parameters.

To compute the unconditional moments of long- and short-term interest rates, we need the stationary distribution of the short-term interest rate conditional on the parameters of the SETAR model. The interest rate distribution is available as the vector  $\pi$  solving the eigenvalue problem  $\pi' = \pi' A$ . Using  $\pi$  all moments can be calculated straightforwardly, for example the mean of  $y$  is given by  $\pi' y$ , with analogous expressions for other moments.

The finite-state Markov state method works very fast. In the computations we used a grid of 200 points with interest rates in the range between 1% and 30%. For the term structure implications we averaged the implied long-term yield over all realizations of the parameter vector  $\theta^{(k)}$  produced by the Gibbs sampler.

We also averaged over all implied volatility ratios. A particular draw  $\theta^{(k)}$  for which the term structure is constructed consists of a draw of the threshold(s), and conditional on the thresholds a draw of the  $\beta_j$  parameters and finally a draw of the volatility parameters.

## References

- Aït-Sahalia, Y., 1995, Testing continuous-time models of the spot interest rate, *Review of Financial Studies*, forthcoming.
- Anderson, H.M., 1994, Transaction costs and nonlinear adjustment towards equilibrium in the US Treasury bill market, Technical report (University of Texas, Austin, TX).
- Backus, D.K. and S.E. Zin, 1993, Long-memory inflation uncertainty: Evidence from the term structure of interest rates, *Journal of Money Credit and Banking* 25, 681–700.
- Balduzzi, P., G. Bertola, and S. Foresi, 1993, A model of target changes and the term structure of interest rates, NBER working paper 4347.
- Black, F., E. Derman, and W. Toy, 1990, A one factor model of interest rates and its application to Treasury bond options, *Financial Analysts Journal* 46, 33–39.
- Brockwell, P.J. and R.A. Davis, 1991, *Time series: Theory and methods* (Springer Verlag, Berlin).
- Broze, L., O. Scaillet, and J.-M. Zakoian, 1993, Testing for continuous time models of the short term interest rate, Discussion paper 9331 (CORE, Louvain-la-Neuve).
- Campbell, J.Y. and R.J. Shiller, 1984, A simple account of the behavior of long-term interest rates, *American Economic Review* 74, 44–54.
- Campbell, J.Y. and R.J. Shiller, 1987, Cointegration and tests of present value relations, *Journal of Political Economy* 95, 1062–1088.
- Casella, G. and E.I. George, 1991, Explaining the Gibbs sampler, *American Statistician* 46, 167–174.
- Chan, K.S., J.D. Petrucelli, H. Tong, and S.W. Woolford, 1985, A multiple-threshold AR(1) model, *Journal of Applied Probability* 22, 267–279.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1227.
- Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.
- Das, S.R., 1993, Mean rate shifts and alternative models of the interest rate: Theory and evidence, Working paper (Stern School of Business, New York, NY).
- Duffie, D., 1992, *Dynamic asset pricing theory* (Princeton University Press, Princeton, NJ).
- Geweke, J., 1994, Priors for macroeconomic time series and their application, *Econometric Theory* 10, 609–632.
- Geweke, J. and N. Terui, 1993, Bayesian threshold autoregressive models for nonlinear time series, *Journal of Time Series Analysis* 14, 441–454.
- Granger, C.W.J., 1993, *Modelling non-linear relationships between long-memory variables* (University of California, San Diego, CA).
- Granger, C.W.J. and T. Teräsvirta, 1993, *Modelling nonlinear economic relationships* (Oxford University Press, Oxford).
- Haan, W.J. den, 1995, The term structure of interest rates in real and monetary economies, *Journal of Economic Dynamics and Control* 19, forthcoming.
- Hall, A.D., H.M. Anderson, and C.W.J. Granger, 1992, A cointegration analysis of treasury bill yields, *Review of Economics and Statistics* 74, 116–126.
- Hamilton, J.D., 1988, Rational expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates, *Journal of Economic Dynamics and Control* 12, 385–423.

- Ingersoll, J.E., 1987, *Theory of financial decision making* (Rowman and Littlefield, Savage, MD).
- Koedijk, C.G., F.G.J.A. Nissen, P.C. Schotman, and C.C.P. Wolff, 1994, The dynamics of short-term interest rate volatility reconsidered, LIFE working paper 94-13 (University of Limburg, Maastricht).
- Kozicki, S., 1994, A nonlinear model of the term structure, Working paper (Federal Reserve Board, Washington, DC).
- Longstaff, F.A. and E.S. Schwartz, 1992, Interest rate volatility and the term structure: A two factor general equilibrium model, *Journal of Finance* 47, 1259–1282.
- Mankiw, N.G., 1986, The term structure of interest rates revisited, *Brookings Papers on Economic Activity* 1, 61–109.
- McCulloch, R.E. and R.S. Tsay, 1994, Bayesian analysis of autoregressive time series via the Gibbs sampler, *Journal of Time Series Analysis* 15, 235–250.
- Merton, R.C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141–183.
- Naik, V. and M.H. Lee, 1993, The yield curve and bond option prices with discrete shifts in economic regimes, Working paper (University of British Columbia, Vancouver).
- Pagan, A.R., A.D. Hall, and V. Martin, 1994, Exploring the relations between the finance and econometrics literatures on the term structure, Working paper (Australian National University, Canberra).
- Perron, P., 1989, The great crash, the oil price shock and the unit root hypothesis, *Econometrica* 57, 1361–1401.
- Pole, A.M. and A.F.M. Smith, 1985, A Bayesian analysis of some threshold switching models, *Journal of Econometrics* 29, 97–119.
- Shea, G.S., 1989, Uncertainty and implied variance bounds in long-memory models of the interest rate term structure, *Empirical Economics* 16, 387–412.
- Shiller, R.J., 1979, The volatility of long term interest rates and expectations models of the term structure, *Journal of Political Economy* 87, 1190–1219.
- Tierney, L., 1991, Exploring posterior distributions using Markov chains, Working paper (University of Minnesota, Minneapolis, MN).
- Tong, H., 1990, *Non-linear time series: A dynamical systems approach* (Clarendon Press, Oxford).
- Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.