Journal of Economic Dynamics and Control 12 (1988) 385–423. North-Holland

RATIONAL-EXPECTATIONS ECONOMETRIC ANALYSIS OF CHANGES IN REGIME
An Investigation of the Term Structure of Interest Rates

James D. HAMILTON*

University of Virginia, Charlottesville, VA 22901, USA

Received October 1987, final version received December 1987

This paper is of interest both for its methodological contribution of new tools for analyzing rational-expectations models and for its substantive conclusions concerning the term structure of interest rates during the monetary experiment of October 1979.

The paper studies systems subject to changes in regime, interpreted here as occasional, discrete shifts in the parameters governing the time series behavior of exogenous economic variables. The specification is shown to be quite tractable both theoretically and empirically. The technique is used to analyze yields on three-month Treasury bills and ten-year Treasury bonds during 1962 to 1987. A constant-parameter linear model for short-term rates is shown to be inconsistent both with the univariate time series properties of short rates and with the observed bivariate relation between long and short rates under the expectations hypothesis of the term structure. An alternative nonlinear model that admits the possibility of changes in regime affords a much better description of the univariate process for short rates. Moreover, the cross-equation restrictions implied by the expectations hypothesis of the term structure are consistent with the nonlinear specification. Indeed, the residuals of the restricted relation have a standard error of only 0.8 basis points. This is a third less than that of a completely unrestricted linear regression of long rates on short rates, and compares with an unconditional standard deviation of long rates of 142 basis points.

I conclude that once the recognition by bond traders of changes in regime is taken into account, the expectations hypothesis of the term structure of interest rates holds up fairly well for these data.

1. Introduction and summary

This paper develops new technical tools for theoretical analysis and empirical estimation of rational-expectations models. The task is to model changes in regime in a way that satisfies the internal consistency requirement of the rational-expectations hypothesis. If the regime can change, private actors in the economy must assign some probability to the possibility of such

*I have benefited from helpful comments on an earlier draft of this paper by Marjorie Flavin, Ron Michener, Carl Walsh, anonymous referees, and participants at workshops at the University of Virginia, Duke University, Ohio State University, UCLA, and the University of Washington. I am further grateful to Bonnie Garrett and David Wilcox for assistance in obtaining the data for this project and to the NSF for research support under Grant SES-8720731.

0165-1889/88/$3.50

change, and form forecasts as best they can under the circumstances. The paper develops a fully integrated system for describing these forecasts and their influence on the behavior of endogenous variables. Despite its nonlinear nature, the system is shown to be quite tractable and amenable to empirical estimation subject to the cross-equation restrictions implied by the rational-expectations hypothesis. The technique is used to analyze the term structure of interest rates and its response to the change in Fed operating procedures initiated in October 1979.

That changes in regime may be very important for the evolution of interest rates has been emphasized in a number of recent studies. Mankiw, Miron and Weil (1987) examined the effects of the establishment of the Federal Reserve in 1914. Peek and Wilcox (1987) considered the possibility of a change in policy with each new Federal Reserve chairman. The particular focus of the present paper is on the change in Fed operating procedure in October 1979, an episode studied by Antoncic (1986), Huizinga and Mishkin (1986), Hardouvelis and Barnhart (1987), and Walsh (1987), among others. While all of these researchers have underscored the importance of changes in monetary policy, none developed an explicit model of how bond traders form their perceptions of changes in regime and what probabilities traders assign to the prospect of future changes. This paper, by contrast, develops a fully specified model of rational-expectations learning by bond traders in which parameters characterizing the process for interest rates and changes in regime can be estimated by full information maximum likelihood subject to the cross-equation restrictions imposed by the rational-expectations hypothesis. The specification thus permits extension of the approach of Hansen and Sargent (1980) to environments subject to changes in regime.

The technique is applied to quarterly yields on three-month Treasury bills and ten-year Treasury bonds over the period 1962:1 through 1987:3. The principal conclusions are as follows. (1) An AR(4) process with constant coefficients is inconsistent with the univariate process for short rates over this period. (2) The rational-expectations forecast of future short rates based on a constant-parameter AR(4) process for short rates is likewise inconsistent with the historical correlation between long and short rates under the expectations hypothesis of the term structure of interest rates. (3) An alternative nonlinear process that allows the possibility of changes in regime offers a vastly superior fit to the univariate data on short rates. The maximum likelihood estimates associate the alternative regime with the period 1979:4 through 1982:3, and characterize it as a period with average short rates twice as high and with a standard deviation four times as great as in the period preceding. (4) The rational-expectations forecasts of future short rates under the two regimes can be constructed from the maximum likelihood estimates. When the response of long rates to short rates is restricted to be this rational-expectations forecast, the residuals have a standard deviation of only 0.8 basis points. This is a third less than that associated with a completely unconstrained linear OLS regression of long rates on short rates, and compares with an unconditional standard deviation of long rates over this period of 142 basis points. Relaxing the cross-equation restrictions on the nonlinear representation does not lead to a statistically significant improvement in the likelihood function.

I conclude that once the recognition by bond traders of changes in regime is taken into account, the expectations hypothesis of the term structure of interest rates holds up fairly well for these data.

The plan of the paper is as follows. Section 2 presents the basic model of changes in regime, develops the rational-expectations nonlinear filter that a bond trader could use to learn about such changes in regime, and presents an algorithm for maximum likelihood estimation of parameters. Section 3 reviews the expectations hypothesis of the term structure of interest rates and discusses maximum likelihood estimation subject to the cross-equation restrictions imposed by rational expectations. Empirical results and substantive conclusions are presented in section 4, with concluding discussion offered in section 5.

2. A model of changes in regime

2.1. Stochastic specification

Following my 1987 paper, consider a variable y_t whose stochastic process is given by

$$y_t - \mu(S_t) = \phi_1[y_{t-1} - \mu(S_{t-1})] + \phi_2[y_{t-2} - \mu(S_{t-2})] + \cdots + \phi_m[y_{t-m} - \mu(S_{t-m})] + \sigma(S_t)v_t, \tag{2.1}$$

with $v_t \sim$ i.i.d. N(0,1) and $[1 - \phi_1 z^1 - \phi_2 z^2 - \cdots - \phi_m z^m] \neq 0$ for any z: $\|z\| \leq 1$. This differs from the usual linear AR(m) specification in that both the constant term around which the process is defined $[\mu(S_t)]$ and the standard deviation of its innovation $[\sigma(S_t)]$ are functions of the regime operative at date t. The regime is indexed by the discrete-valued variable S_t; for example, $S_t = 1$ means that the process was in regime 1 at date t.

The assumption in this paper is that bond traders recognize the possibility of changes in regime and incorporate this into their forecasts for the future. To arrive at a rational-expectations solution of a model incorporating (2.1), we must (a) specify the subjective probabilities that bond traders associate with future regime changes, and (b) verify that these assumed probabilities are consistent with all information that bond traders in fact have available at date t. A very tractable structure for accomplishing this task is to model S_t as the outcome of an unobserved discrete-time, discrete-state Markov process. This paper employs the simplest example of such a specification – a two-state ($S_t = 0$ or 1), first-order Markov process:

$$\begin{aligned} &\mathrm{P}[S_t = 1 \mid S_{t-1} = 1] = p, \quad \mathrm{P}[S_t = 0 \mid S_{t-1} = 1] = 1 - p, \\ &\mathrm{P}[S_t = 1 \mid S_{t-1} = 0] = 1 - q, \quad \mathrm{P}[S_t = 0 \mid S_{t-1} = 0] = q, \end{aligned} \tag{2.2}$$

for which we can (without loss of generality) parameterize

$$\mu(S_t) = \alpha_0 + \alpha_1 S_t, \quad \sigma(S_t) = \omega_0 + \omega_1 S_t. \tag{2.3}$$

I further specify that v_t is independent of S_{t-j} for all j, and normalize by defining state 1 to be the state with the higher variance (achieved by setting $\omega_1 > 0$).

Three aspects of the Markov process (2.2) will be important below. First, I assume that bond traders do not observe the current regime S_t directly, but instead must form inference about it based on observation of $\{y_t\}$. If, hypothetically, one could observe the regime directly, the Markov structure means that S_{t-1} would summarize all information available at date $t-1$ that would be useful for forecasting S_t:

$$\begin{aligned} &\mathrm{P}[S_t = s_t \mid S_{t-1} = s_{t-1}] \\ &= \mathrm{P}[S_t = s_t \mid S_{t-1} = s_{t-1}, S_{t-2} = s_{t-2}, \ldots, v_{t-1}, v_{t-2}, \ldots]. \end{aligned} \tag{2.4}$$

Second, note that (2.2) can be written

$$s_t = (1-q) + \lambda s_{t-1} + w_t,$$

with

$$\lambda = -1 + p + q, \tag{2.5}$$

where conditional on $s_{t-1} = 1$,

$$\begin{aligned} w_t &= 1 - p && \text{with probability } p, \\ w_t &= -p && \text{with probability } 1 - p, \end{aligned}$$

whereas conditional on $s_{t-1} = 0$,

$$\begin{aligned} w_t &= -(1 - q) && \text{with probability } q, \\ w_t &= q && \text{with probability } 1 - q. \end{aligned}$$

From these one deduces[1]

$$E[S_{t+j}|S_t = s_t] = \rho + \lambda^j(s_t - \rho), \tag{2.6}$$

where

$$\rho = \frac{1-q}{(1-p)+(1-q)}. \tag{2.7}$$

Finally, notice that the unconditional distribution of S_t is given by $P[S_t = 1] = \rho$ (and of course $P[S_t = 0] = 1 - \rho$).

The process defined by (2.1) and (2.2) turns out to be stationary provided that p and q are both strictly less than unity.[2] By Wold's decomposition, there exists a constant-parameter linear representation for (2.1)–(2.2) of the form

$$y_t = \kappa + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots, \tag{2.8}$$

where the u_t's are uncorrelated but not independent. What this means in practice is that while one could forecast the series y_t on the basis of a linear representation such as (2.8), these forecasts are suboptimal; forecasts that make use of a nonlinear function of $y_{t-1}, y_{t-2}, \ldots$ are superior.

The optimal nonlinear forecast can be thought of as arising from a two-step procedure. First we ask, what is the optimal inference about the current state based on what we have actually observed; in other words, what is $P[S_t = 1|y_t, y_{t-1}, \ldots]$? The answer to this question is provided by the nonlinear filter presented shortly. One can then use the output of the filter to generate optimal future forecasts of y_t, as is done in the present value calculations of section 3.

The principal tool used in this investigation is a straightforward adaptation of the nonlinear filter introduced in my 1987 paper. That paper also describes the filter's antecedents in work by Goldfeld and Quandt (1973), Liptser and Shiryayev (1977), Neftci (1982,1984), and Cosslett and Lee (1985). The filter allows us both to estimate parameters of the univariate process for y_t and to characterize optimal forecasts of future values of y_t. The latter will then be used in section 3 below to generate the cross-equation restrictions implied by the expectations hypothesis of the term structure of interest rates.

[1] These are standard results. See for example Chiang (1980, p. 160).

[2] See the discussion in my 1987 paper. There I introduced a very similar process in which y_t represented the first difference of the raw data $\tilde{y}_t$, and the specification was suggested as an alternative approach to describing the nonstationarity of the raw data $\tilde{y}_t$.

2.2. Filtering

Suppose for the moment that we knew with certainty the particular numerical values for the parameters $(\phi_1, \phi_2, \ldots, \phi_m, p, q, \alpha_0, \alpha_1, \omega_0, \omega_1)$ in eqs. (2.1)–(2.3). Even if we knew these parameters, we would still be unsure as to whether the process was in state 0 or 1 at date t, because the state variable s_t is presumed unobserved by the econometrician. This subsection summarizes my (1987) nonlinear filter for solving the second problem of inference. Throughout this subsection on filtering, the parameters $(\phi_1, \phi_2, \ldots, \phi_m, p, q, \alpha_0, \alpha_1, \omega_0, \omega_1)$ are treated as known constants, and the task is solely to draw inference about the historical sequence of states $\{s_t\}$ based on the observed sequence of data $\{y_t\}$. In the following subsection, I then show how the parameters $(\phi_1, \phi_2, \ldots, \phi_m, p, q, \alpha_0, \alpha_1, \omega_0, \omega_1)$ can be estimated by maximum likelihood.

My procedure for drawing inference about $\{s_t\}$ is an iterative one. Given an initial inference about s_{t-1} based on data observed through date $t-1$ (and given knowledge of the parameters $[\phi_1, \phi_2, \ldots, \phi_m, p, q, \alpha_0, \alpha_1, \omega_0, \omega_1]$), iteration t produces an inference about s_t based on data observed through date t (this latter inference being based on the same values of $[\phi_1, \phi_2, \ldots, \phi_m, p, q, \alpha_0, \alpha_1, \omega_0, \omega_1]$ as were used in iteration $t-1$ but using one more observation on y).

Specifically, the filter accepts as input the joint conditional probability

$$\mathrm{p}(s_{t-1}, s_{t-2}, \ldots, s_{t-m} | y_{t-1}, y_{t-2}, \ldots, y_0),$$

and has as output

$$\mathrm{p}(s_t, s_{t-1}, \ldots, s_{t-m+1} | y_t, y_{t-1}, \ldots, y_0),$$

along with, as a byproduct, the conditional likelihood of y_t:

$$\mathrm{p}(y_t | y_{t-1}, y_{t-2}, \ldots, y_0).$$

Note well the notation: $[s_t, s_{t-1}, \ldots, s_{t-m+1}]$ refers to the m most recent observations on S, whereas $[y_t, y_{t-1}, \ldots, y_0]$ denotes the complete history of y observed through date t. I further let $p(x)$ denote $\mathrm{Prob}[X = x]$ for x a discrete-valued variable or the density function $f(x)$ for x continuous.

Step 1. Recalling the Markov property (2.4), calculate

$$\begin{aligned}&\mathrm{p}(s_t, s_{t-1}, \ldots, s_{t-m} | y_{t-1}, y_{t-2}, \ldots, y_0)\\ &= \mathrm{p}(s_t | s_{t-1}) \times \mathrm{p}(s_{t-1}, s_{t-2}, \ldots, s_{t-m} | y_{t-1}, y_{t-2}, \ldots, y_0),\end{aligned}$$

where $p(s_t|s_{t-1})$ is given by (2.2). Here (as in all the steps to follow) the second term on the right-hand-side is known from the preceding step of the filter [in this case, $p(s_{t-1}, s_{t-2}, \ldots, s_{t-m}|y_{t-1}, y_{t-2}, \ldots, y_0)$ is known from the input to the filter, which in turn represents the output of the iteration at date $t-1$].

Step 2. Calculate the joint conditional density-distribution of y_t and $(S_t, S_{t-1}, \ldots, S_{t-m})$:

$$\begin{aligned}
&p(y_t, s_t, s_{t-1}, \ldots, s_{t-m}|y_{t-1}, y_{t-2}, \ldots, y_0)\\
&= p(y_t|s_t, s_{t-1}, \ldots, s_{t-m}, y_{t-1}, y_{t-2}, \ldots, y_0)\\
&\quad \times p(s_t, s_{t-1}, \ldots, s_{t-m}|y_{t-1}, y_{t-2}, \ldots, y_0),
\end{aligned}$$

where we know

$$\begin{aligned}
&p(y_t|s_t, s_{t-1}, \ldots, s_{t-m}, y_{t-1}, y_{t-2}, \ldots, y_0)\\
&= \frac{1}{\sqrt{2\pi}\,(\omega_0+\omega_1 s_t)} \exp\Bigg[-\frac{1}{2[\omega_0+\omega_1 s_t]^2}\big((y_t-\alpha_1 s_t-\alpha_0)\\
&\quad -\phi_1(y_{t-1}-\alpha_1 s_{t-1}-\alpha_0)-\cdots-\phi_m(y_{t-m}-\alpha_1 s_{t-m}-\alpha_0)\big)^2\Bigg].
\end{aligned}$$

Step 3. We then have

$$\begin{aligned}
&p(y_t|y_{t-1}, y_{t-2}, \ldots, y_0)\\
&= \sum_{s_t=0}^{1}\sum_{s_{t-1}=0}^{1}\cdots\sum_{s_{t-m}=0}^{1} p(y_t, s_t, s_{t-1}, \ldots, s_{t-m}|y_{t-1}, y_{t-2}, \ldots, y_0).
\end{aligned}$$

Step 4. Combining the results from steps 2 and 3, we can then calculate

$$\begin{aligned}
&p(s_t, s_{t-1}, \ldots, s_{t-m}|y_t, y_{t-1}, \ldots, y_0)\\
&= \frac{p(y_t, s_t, s_{t-1}, \ldots, s_{t-m}|y_{t-1}, y_{t-2}, \ldots, y_0)}{p(y_t|y_{t-1}, y_{t-2}, \ldots, y_0)}.
\end{aligned}$$

Step 5. The desired output is then obtained from

$$\begin{aligned}
&p(s_t, s_{t-1}, \ldots, s_{t-m+1}|y_t, y_{t-1}, \ldots, y_0)\\
&= \sum_{s_{t-m}=0}^{1} p(s_t, s_{t-1}, \ldots, s_{t-m}|y_t, y_{t-1}, \ldots, y_0).
\end{aligned}$$

2.3. Maximum likelihood estimation of parameters

Throughout the preceding iteration for all $t = m, m+1, \ldots, T$, the parameters $(\alpha_1, \alpha_0, p, q, \omega_0, \omega_1, \phi_1, \phi_2, \ldots, \phi_m)$ were treated as known, fixed constants. However, we can now see that one byproduct of the filter (generated at step 3) is evaluation of the conditional likelihood function associated with that choice of constants. The sample conditional log likelihood is

$$\ln \mathrm{p}(y_T, y_{T-1}, \ldots, y_m | y_{m-1}, y_{m-2}, \ldots, y_0)$$

$$= \sum_{t=m}^{T} \ln \mathrm{p}(y_t | y_{t-1}, y_{t-2}, \ldots, y_0),$$

which can be maximized numerically with respect to the unknown parameters $(\alpha_1, \alpha_0, p, q, \omega_0, \omega_1, \phi_1, \phi_2, \ldots, \phi_m)$.

2.4. Extensions and additional details

The simplest approach is to start the first iteration on the filter $(t = m)$ with the unconditional probability

$$\mathrm{p}(s_{m-1}, s_{m-2}, \ldots, s_0)$$

$$= \mathrm{p}(s_{m-1}|s_{m-2}) \times \mathrm{p}(s_{m-2}|s_{m-3}) \times \cdots \times \mathrm{p}(s_1|s_0) \times \mathrm{p}(s_0),$$

where $\mathrm{p}(s_j|s_{j-1})$ is given by (2.2) and the final term $[\mathrm{p}(s_0)]$ is obtained from (2.7):

$$\mathrm{P}[S_0 = 1] = \rho, \qquad \mathrm{P}[S_0 = 0] = 1 - \rho.$$

For some applications, one might want to allow the possibility of a permanent change in regime (e.g., allow $q = 1$). In that case, one should not start the filter with the unconditional probability $\rho = (1-q)/[(1-p) + (1-q)]$, for if $q = 1$, then $\rho = 0$, and the filter would end up setting this and all subsequent probabilities $\mathrm{P}[S_t = 1 | y_t, y_{t-1}, \ldots]$ equal to zero as well. Instead one could treat $\rho_0 \equiv \mathrm{P}[S_0 = 1]$ as an unconstrained separate parameter to be estimated by maximum likelihood along with $(\alpha_1, \alpha_0, p, q, \omega_0, \omega_1, \phi_1, \phi_2, \ldots, \phi_m)$. In this paper I have imposed the restriction $\rho_0 = (1-q)/[(1-p) + (1-q)]$.

Technically, one would only have calculated the conditional likelihood function at step 3 if the filter were started at $t = m$ with $\mathrm{p}(s_{m-1}, s_{m-2}, \ldots, s_0 | y_{m-1}, y_{m-2}, \ldots, y_0)$ as input rather than $\mathrm{p}(s_{m-1}, s_{m-2}, \ldots, s_0)$ as suggested here. Calculating the former is considerably more effort and seems unlikely to make a material difference, though once one has made these calculations, exact maximum likelihood estimates rather than estimates that maximize an approximation to the conditional likelihood are easily derived.

Given the specification chosen for the start-up values [in which the conditional probability $\mathrm{p}(s_{m-1}, s_{m-2}, \ldots, s_0 | y_{m-1}, y_{m-2}, \ldots, y_0)$ is approximated by the unconditional probability $\mathrm{p}(s_{m-1}, s_{m-2}, \ldots, s_0)$], it might be best to seek to maximize $\ln\ \mathrm{p}(y_T, y_{T-1}, \ldots, y_{2m} | y_{2m-1}, y_{2m-2}, \ldots, y_0)$, say, rather than $\ln\ \mathrm{p}(y_T, y_{T-1}, \ldots, y_m | y_{m-1}, y_{m-2}, \ldots, y_0)$ in arriving at parameter estimates. I nevertheless use $\mathrm{p}(y_T, y_{T-1}, \ldots, y_m | y_{m-1}, \ldots, y_0)$ in the empirical application in section 4.

Clearly one can permit more complicated dynamics for the regime shift by replacing $\mathrm{p}(s_t | s_{t-1})$ in step 1 with some parameterized function $\mathrm{p}(s_t | s_{t-1}, s_{t-2}, \ldots, s_{t-m}, y_{t-1}, y_{t-2}, \ldots, y_0)$ and maximizing the likelihood function with respect to these parameters along with the others. The potential obstacle here is a numerical one. Identification of the parameters characterizing the dynamics of S_t separately from those of the Gaussian component depends on nonlinearities in the data (e.g., the difference between $\mathrm{E}[y_t | y_{t-1}]$ and $\mathrm{E}[y_t | y_{t-1}, y_{t-1}^2]$). There is a practical limitation on how complicated we can permit the dynamics for both the regime shift and the Gaussian component to become and still have hope of obtaining useful results.

It is also straightforward to replace ϕ_j in step 2 with $\phi_j(s_t)$, and so allow the autoregressive coefficients to shift along with the mean and variance. Similarly, more than two states can be accommodated by replacing $\sum_{s_{t-j}=0}^{1}$ in steps 3 and 5 with $\sum_{s_{t-j}=0}^{n_s}$ for n_s the number of states. Again the difficulties of such generalizations are chiefly numerical. At iteration t, the filter is updating n_s^{m+1} different probabilities; with 4 lags, the two-state filter is thus keeping track of $2^5 = 32$ different numbers; a three-state filter would calculate $3^5 = 243$. Rational-expectations applications are also considerably more complicated once one departs from the simple parameterization of regime shifts embodied in (2.1)–(2.3).

The filter was motivated as drawing an inference about the state s_t based on currently available information,

$$\mathrm{p}(s_t | y_t, y_{t-1}, \ldots, y_0)$$

$$= \sum_{s_{t-1}=0}^{1} \sum_{s_{t-2}=0}^{1} \cdots \sum_{s_{t-m+1}=0}^{1} \mathrm{p}(s_t, s_{t-1}, \ldots, s_{t-m+1} | y_t, y_{t-1}, \ldots, y_0).$$

Alternatively, one can obtain a more reliable inference about the lagged value of the state using currently available information. For example, using the output from step 4 of the basic filter, one can calculate an m-lag smoother:

$$\mathrm{p}(s_{t-m} | y_t, y_{t-1}, \ldots, y_0)$$

$$= \sum_{s_t=0}^{1} \sum_{s_{t-1}=0}^{1} \cdots \sum_{s_{t-m+1}=0}^{1} \mathrm{p}(s_t, s_{t-1}, \ldots, s_{t-m} | y_t, y_{t-1}, \ldots, y_0),$$

which summarizes the optimal inference based on information available at date t about the state the economy was historically in at date $t-m$.

A full-sample smoother is described in my 1987 paper.

3. The expectations hypothesis of the term structure of interest rates

3.1. Theoretical formulation

There are three potential sources of nonlinearity in specifying a theory of the term structure of interest rates. The first is nonlinearities in the utility function – investors may care about higher moments in addition to the mean return associated with a portfolio, as in Roll (1971). The second is nonlinearities arising from Jensen's inequality. For example, equating the price of a bond to the expected present value of its coupon payments (discounted using the future path of short rates) is not quite the same thing as equating the expected one-period yield from holding short-term and long-term bonds [see Cox, Ingersoll and Ross (1981)]. The third source of nonlinearity is that the optimal least-squares forecasts of future short-term rates may be a nonlinear rather than a linear function of past short rates.

In this paper I am concerned only with the third of these issues, primarily in the interests of keeping the scope of the project manageable. In ignoring nonlinearities in the utility function, the objective is to see how far one can go with a model in which investors are concerned only about mean returns and not the variance. Nonlinearities associated with Jensen's inequality may well be minor quantitatively [see Shiller (1979) and Campbell (1986)], and are suppressed with a straightforward application of Shiller's (1979) linearization of the expectations hypothesis of the term structure of interest rates.

Let t index the current quarter and r_t the nominal quarterly yield on a three-month T-bill purchased at date t. In the formulas used in this section, r_t is measured as a fraction of 1. By contrast, in the empirical estimates presented in section 4, r_t is reported in units of 100 basis points.

Consider an $(n/4)$-year bond to be redeemed at date $t+n$ for \$1 and paying a semiannual coupon of C beginning at date $t+2$. The perfect-foresight present value of such a bond is defined as

$$V_t^{(n)} \equiv \frac{C}{\prod_{j=0}^{1}(1+r_{t+j})} + \frac{C}{\prod_{j=0}^{3}(1+r_{t+j})} + \frac{C}{\prod_{j=0}^{5}(1+r_{t+j})} + \cdots + \frac{C}{\prod_{j=0}^{n-1}(1+r_{t+j})} + \frac{1}{\prod_{j=0}^{n-1}(1+r_{t+j})}. \tag{3.1}$$

Note that for constant interest rates ($r_{t+j}=\bar{r}$ for all j with certainty), the present value would equal the par value ($V_t^{(n)} = 1$) when its semiannual coupon (C) equals the return from rolling over two three-month T-bills ($\bar{r}^2 + 2\bar{r}$). Setting $C = (\bar{r}^2 + 2\bar{r})$ and taking a first-order Taylor series expansion of (3.1) around $r_t = r_{t+1} = \cdots = r_{t+n-1} = \bar{r}$ yields[3]

$$\begin{aligned} V_t^{(n)} \cong 1 &- \frac{1}{1+\bar{r}}\left[[r_t - \bar{r}] + [r_{t+1} - \bar{r}]\right] \\ &- \frac{1}{(1+\bar{r})^3}\left[[r_{t+2} - \bar{r}] + [r_{t+3} - \bar{r}]\right] \\ &- \cdots - \frac{1}{(1+\bar{r})^{n-1}}\left[[r_{t+n-2} - \bar{r}] + [r_{t+n-1} - \bar{r}]\right]. \end{aligned} \tag{3.2}$$

Shiller's linearization of the expectations hypothesis of the term structure equates the current price of the bond $P_t^{(n)}$ with the expected present value:

$$P_t^{(n)} = \mathrm{E}_t V_t^{(n)} + \phi^{(n)} \tag{3.3}$$

where $\phi^{(n)}$ denotes a potential term premium on long-term bonds.

I now use the definition of the 'yield to maturity' to rewrite the price of the bond in terms of its yield to maturity. The yield to maturity ($R_t^{(n)}$) for this ($n/4$)-year bond, measured at a semiannual rate, is defined by

$$\begin{aligned} P_t^{(n)} &\equiv \frac{C}{1+R_t^{(n)}} + \frac{C}{\left[1+R_t^{(n)}\right]^2} + \cdots + \frac{C}{\left[1+R_t^{(n)}\right]^{n/2}} + \frac{1}{\left[1+R_t^{(n)}\right]^{n/2}} \\ &= \frac{C}{R_t^{(n)}} + \frac{1}{\left[1+R_t^{(n)}\right]^{n/2}}\left[1 - \frac{C}{R^{(n)}}\right]. \end{aligned} \tag{3.4}$$

Again taking a first-order Taylor series expansion around $R_t^{(n)} = C = (\bar{r}^2 + 2\bar{r})$, we obtain

$$P_t^{(n)} \cong 1 + \left[\frac{-1}{\bar{r}^2 + 2\bar{r}} + \frac{1}{(\bar{r}^2 + 2\bar{r})(1+\bar{r})^n}\right] \times \left[R_t^{(n)} - \bar{r}^2 - 2\bar{r}\right]. \tag{3.5}$$

[3] This differs slightly from Shiller's (1979, p. 1198) expression

$$V_t^{(n)} \cong 1 - \sum_{j=1}^{n} \frac{1}{(1+\bar{r})^j}\left[r_{t+j-1} - \bar{r}\right].$$

For Shiller's data sets, coupons are paid every period, whereas for mine, coupons are paid every other period. For a pure discount bond, a linearization of the term structure makes the present value a simple unweighted average of future short rates, whereas when coupons are paid every period one obtains Shiller's geometrically declining weights [see Shiller, Campbell and Schoenholtz (1983, p. 177)]. The case I analyze, with coupons paid every other period, is a hybrid of the two.

Note that 'yield to maturity' $R_t^{(n)}$ is *not* a holding-period yield; rather, it is simply an accounting construct calculated from the current price $P_t^{(n)}$ and coupon (C) of the bond. The yield to maturity is thus of course known with certainty at date t.

Equating (3.5) with the expectation of (3.2) plus $\phi^{(n)}$ yields[4]

$$\begin{aligned} R_t^{(n)} = & \left[\frac{\bar{r}^2 + 2\bar{r}}{1+\bar{r}}\right]\left[1 - \frac{1}{(1+\bar{r})^n}\right]^{-1} \\ & \times \left(\mathrm{E}_t[r_t + r_{t+1}] + \frac{1}{(1+\bar{r})^2}\mathrm{E}_t[r_{t+2} + r_{t+3}] \right. \\ & \left. + \cdots + \frac{1}{(1+\bar{r})^{n-2}}\mathrm{E}_t[r_{t+n-2} + r_{t+n-1}] \right) \\ & - \bar{r}^2 - (\bar{r}^2 + 2\bar{r})\left[1 - \frac{1}{(1+\bar{r})^n}\right]^{-1}\phi^{(n)}. \end{aligned} \tag{3.6}$$

It would be nice to have completely specified all of the factors that influence the long rate $R_t^{(n)}$, and pretend that eq. (3.6) holds with an R^2 of 1 and no error term. Unfortunately, such a pretension would be quite hopeless, given errors arising from linearization, measurement, nonconstant risk premia, and information available to bond traders but not the econometrician. In the empirical analysis to follow, I will assume that the actual yield to maturity on long-term bonds differs from the value predicted in (3.6) by an error term u_t. I specify that u_t follows a Gaussian white noise process that is independent of the process for short rates.

One could attempt a more realistic and refined analysis of this error term, for example, through explicit modeling of the response predicted by CAPM to the changes in the variance of returns implied by a shift from state 0 to state 1. I do not attempt such an exercise here, in order to keep the scope of the project manageable. Instead, my primary objective is to see how far one can go with a linearized model in which only expected returns matter to investors and in which the econometrician has available all the information used by bond traders in forming their forecasts of future short rates.[5]

[4]Alternatively, eq. (3.6) can be derived directly from Shiller's (1979) eq. (1), as I show in the appendix.

[5]For an analysis of conditional heteroskedasticity and risk aversion, see Engle, Lilien and Robins (1987). Explicit treatments of information available to bond traders but not the econometrician were provided by Flavin (1984) and Hamilton (1985).

The model to be estimated is thus

$$R_t^{(n)} = \kappa + \delta\big(\mathrm{E}_t[r_t + r_{t+1}] + \beta^2\mathrm{E}_t[r_{t+2} + r_{t+3}] + \beta^4\mathrm{E}_t[r_{t+4} + r_{t+5}]$$
$$+ \cdots + \beta^{n-2}\mathrm{E}_t[r_{t+n-2} + r_{t+n-1}]\big) + u_t, \tag{3.7}$$

where

$$\delta = \left[\frac{\bar{r}^2 + 2\bar{r}}{1+\bar{r}}\right]\left[1 - \frac{1}{(1+\bar{r})^n}\right]^{-1}, \tag{3.8}$$

$$\beta = \frac{1}{1+\bar{r}}, \tag{3.9}$$

and κ reflects any constant term premium built into the term structure.

3.2. *Empirical implementation – The case of a linear process assumed for short-term interest rates*

Before examining the estimation of the term structure relation under the nonlinear model described in section 2, I first review the standard linear case as developed by Sargent (1979) and Hansen and Sargent (1980). Suppose that short-term rates are modeled as a stable, linear AR(4) process:

$$r_t = k_r + a_1 r_{t-1} + a_2 r_{t-2} + a_3 r_{t-3} + a_4 r_{t-4} + e_{r,t}. \tag{3.10}$$

This can be written in vector form

$$\begin{bmatrix} r_t \\ r_{t-1} \\ r_{t-2} \\ r_{t-3} \end{bmatrix} = \begin{bmatrix} k_r \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} r_{t-1} \\ r_{t-2} \\ r_{t-3} \\ r_{t-4} \end{bmatrix} + \begin{bmatrix} e_{r,t} \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3.11}$$

Thus

$$\mathrm{E}[r_{t+j} | r_t, r_{t-1}, r_{t-2}, r_{t-3}]$$
$$= c_j + [1 \quad 0 \quad 0 \quad 0]A^j\,[r_t \quad r_{t-1} \quad r_{t-2} \quad r_{t-3}]', \tag{3.12}$$

where c_j is a constant and A^j denotes the (4×4) matrix of coefficients in (3.11) raised to the jth power. Substituting (3.12) into the term structure eq.

(3.7) yields[6]

$$R_t^{(n)} = k_R + [\delta \quad 0 \quad 0 \quad 0]$$
$$\times\big[[I+A] + \beta^2[A^2+A^3] + \cdots + \beta^{n-2}[A^{n-2}+A^{n-1}]\big]$$
$$\times[r_t \quad r_{t-1} \quad r_{t-2} \quad r_{t-3}]' + e_{R,t}. \tag{3.13}$$

Expressions (3.10) and (3.13) characterize the cross-equation restrictions relating the parameters of the univariate process for r_t to the predicted relation between the long-rate ($R_t^{(n)}$) and the short rate under the hypothesis of no changes in regime.

3.3. Empirical implementation – The case of a nonlinear process assumed for the short-term interest rate

I now consider the case where short-term rates r_t follow a nonlinear process of the type described in section 2. Letting $r_t = y_t$, that system can be conveniently written

$$r_t = \alpha_0 + \alpha_1 S_t + z_t, \tag{3.14}$$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + \phi_4 z_{t-4} + [\omega_0 + \omega_1 S_t]v_t, \tag{3.15}$$

$$v_t \sim \mathrm{N}(0,1),$$

with transition probabilities for S_t given by (2.2). From (3.14), optimal forecasts of r_{t+j} given an arbitrary information set Ω_t are given by

$$\mathrm{E}[r_{t+j}|\Omega_t] = \alpha_0 + \alpha_1 \mathrm{E}[S_{t+j}|\Omega_t] + \mathrm{E}[z_{t+j}|\Omega_t]. \tag{3.16}$$

Our ultimate task is to evaluate (3.16) when bond traders' information is confined to current and past observations on short rates,

$$\Omega_t = \{r_t, r_{t-1}, \dots\}.$$

This is most easily evaluated when we first consider (3.16) under a hypothetical alternative scenario in which the information set included observation of the regime directly,

$$\Omega_t^* = \{r_t, r_{t-1}, \dots, s_t, s_{t-1}, \dots\},$$

[6]One can of course 'simplify' the sum in (3.13) as $[I - \beta^n A^n][I+A][I-\beta^2A^2]^{-1}$, though this is no faster computationally and produces algorithmic difficulties when A has a root near β^{-1}. The algorithm was coded so as to exploit the identity matrix sub-block of A when multiplying A^{j-1} by A.

which amounts to assuming that the hypothetical forecaster also observes $z_t, z_{t-1}, \ldots$ directly.

Under this hypothetical scenario, we see from (2.6) that

$$\mathrm{E}_t\left[S_{t+j}|\Omega_t^*\right] = \rho + \lambda^j(s_t - \rho). \tag{3.17}$$

Likewise, from eq. (3.15) and reinterpretation of (3.12),

$$\mathrm{E}\left[z_{t+j}|\Omega_t^*\right] = [1 \quad 0 \quad 0 \quad 0]\Delta^j[z_t \quad z_{t-1} \quad z_{t-2} \quad z_{t-3}]', \tag{3.18}$$

where

$$\Delta = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \phi_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Substituting (3.17) and (3.18) into (3.16) and using (3.14) gives

$$\mathrm{E}\left[r_{t+j}|\Omega_t^*\right] = \alpha_0 + \alpha_1\{\rho + \lambda^j(s_t - \rho)\}$$
$$+ [1 \quad 0 \quad 0 \quad 0]\Delta^j \begin{bmatrix} r_t - \alpha_0 - \alpha_1 s_t \\ r_{t-1} - \alpha_0 - \alpha_1 s_{t-1} \\ r_{t-2} - \alpha_0 - \alpha_1 s_{t-2} \\ r_{t-3} - \alpha_0 - \alpha_1 s_{t-3} \end{bmatrix}. \tag{3.19}$$

Now, the critical feature of (3.19) that makes rational-expectations applications of the process so straightforward is that (3.19) is linear in $s_t, s_{t-1}, s_{t-2}, s_{t-3}$. This means that when one wishes to calculate the forecast of r_{t+j} given the information set $\Omega_t = \{r_t, r_{t-1}, \ldots\}$, one simply replaces $s_t, s_{t-1}, s_{t-2}, s_{t-3}$ in expression (3.19) with their conditional expectation given Ω_t,[7] or in other words, replace s_{t-j} with the inference drawn from iteration t

[7] In general

$$\mathrm{E}[Y|X] = \int \frac{f(z,x)}{f(x)}\left[\int \frac{y \cdot f(y,z,x)}{f(z,x)}\,\mathrm{d}y\right]\mathrm{d}z.$$

The term in brackets is

$$\left[\int \frac{y \cdot f(y,z,x)}{f(z,x)}\,\mathrm{d}y\right] = \mathrm{E}[Y|Z,X].$$

In the case where X is a subset of the vector Z, this in turn equals $\mathrm{E}[Y|Z]$. Where this expectation is a linear function as well,

$$\mathrm{E}[Y|Z] = \beta_0 + \beta_1 Z,$$

we obtain

$$\mathrm{E}[Y|X] = \int \frac{f(z,x)}{f(x)}[\beta_0 + \beta_1 z]\,\mathrm{d}z = \beta_0 + \beta_1 \int \frac{z \cdot f(z,x)}{f(x)}\,\mathrm{d}z = \beta_0 + \beta_1 \mathrm{E}[Z|X].$$

of the basic filter of section 2:

$$\mathrm{E}\left[r_{t+j}|\Omega_t\right] = \alpha_0 + \alpha_1\rho(1-\lambda^j) + \alpha_1\lambda^j \cdot \mathrm{P}[S_t = 1|r_t, r_{t-1}, \dots]$$
$$+ [1 \quad 0 \quad 0 \quad 0]\Delta^j$$
$$\times \begin{bmatrix} r_t - \alpha_0 - \alpha_1 \cdot \mathrm{P}[S_t = 1|r_t, r_{t-1}, \dots] \\ r_{t-1} - \alpha_0 - \alpha_1 \cdot \mathrm{P}[S_{t-1} = 1|r_t, r_{t-1}, \dots] \\ r_{t-2} - \alpha_0 - \alpha_1 \cdot \mathrm{P}[S_{t-2} = 1|r_t, r_{t-1}, \dots] \\ r_{t-3} - \alpha_0 - \alpha_1 \cdot \mathrm{P}[S_{t-3} = 1|r_t, r_{t-1}, \dots] \end{bmatrix}. \tag{3.20}$$

Substituting (3.20) into (3.7) gives

$$R_t^{(n)} = \kappa_R + \frac{\delta\alpha_1(1+\lambda)\left[1-(\beta\lambda)^{40}\right]}{1-(\beta\lambda)^2} \cdot \mathrm{P}[S_t = 1|r_t, r_{t-1}, \dots]$$
$$+ [\delta \quad 0 \quad 0 \quad 0] \times \left[[I + \Delta] + \beta^2[\Delta^2 + \Delta^3]\right.$$
$$\left. + \cdots + \beta^{n-2}[\Delta^{n-2} + \Delta^{n-1}]\right]$$
$$\times \begin{bmatrix} r_t - \alpha_0 - \alpha_1 \cdot P[S_t = 1|r_t, r_{t-1}, \dots] \\ r_{t-1} - \alpha_0 - \alpha_1 \cdot \mathrm{P}[S_{t-1} = 1|r_t, r_{t-1}, \dots] \\ r_{t-2} - \alpha_0 - \alpha_1 \cdot \mathrm{P}[S_{t-2} = 1|r_t, r_{t-1}, \dots] \\ r_{t-3} - \alpha_0 - \alpha_1 \cdot \mathrm{P}[S_{t-3} = 1|r_t, r_{t-1}, \dots] \end{bmatrix} + \varepsilon_{R,t}, \tag{3.21}$$

$$\mathrm{E}\left(\varepsilon_{R,t}^2\right) = \sigma_{\varepsilon_R}^2.$$

Maximum likelihood estimation subject to the cross-equation restrictions is then achieved as follows. From the j-lag smoother derived from step 4 at iteration t of the basic filter, we can calculate the values $\mathrm{P}[S_{t-j} = 1|r_t, r_{t-1}, \dots]$ appearing in (3.21). For given values of α_0, α_1, p, q, ϕ_1, ϕ_2, ϕ_3, ϕ_4, ω_0, ω_1, σ_{ε_R} and κ_r (and imposing δ and β a priori), $\varepsilon_{R,t}$ can thus be calculated as another byproduct of the basic filter. One then adds $-\frac{1}{2}\ln(2\pi) - \ln(\sigma_{\varepsilon_R}) - \varepsilon_{R,t}^2/(2\sigma_{\varepsilon_R}^2)$ to $\ln[p(r_t|r_{t-1}, r_{t-2}, \dots, r_0)]$ from step 3 of the basic filter to arrive at the conditional log likelihood of the bivariate process for $\{r_t, R_t\}$ as restricted by the hypothesis of rational expectations. The resulting expression can then be maximized numerically with respect to the parameters α_0, α_1, p, q, ϕ_1, ϕ_2, ϕ_3, ϕ_4, ω_0, ω_1, σ_{ε_R} and κ_R.

4. Empirical results

The analysis was conducted on quarterly U.S. data for 1962:1–1987:3. The series for the short-term rate (r_t) is the three-month Treasury bill rate as of the first business day of the quarter, reported on an effective yield basis at a quarterly rate.[8] The long rate (here denoted R_t) is the yield to maturity as of the first business day of the quarter on a ten-year Treasury bond at a semiannual rate.[9] These series are presented in table 1. Note that as a consequence of the units (r_t is the percentage earned over three months and R_t is the percentage earned over six months), R_t tends on average to be twice as large as r_t.

In all the specifications investigated, the short rate (r_t) is modeled as a function of past short rates, and the long rate (R_t) is modeled as a function of present and past short rates. The alternative specifications arise from choosing linear or nonlinear representations for one or both of these functions, and from whether or not the cross-equation restrictions implied by rational expectations are imposed. The linear representations for r_t and R_t take the form

$$r_t = k_r + a_1 r_{t-1} + a_2 r_{t-2} + a_3 r_{t-3} + a_4 r_{t-4} + e_{r,t}, \quad e_{r,t} \sim \mathrm{N}\left(0, \sigma_{e_r}^2\right), \tag{4.1}$$

$$R_t = k_R + b_0 r_t + b_1 r_{t-1} + b_2 r_{t-2} + b_3 r_{t-3} + e_{R,t}, \quad e_{R,t} \sim \mathrm{N}\left(0, \sigma_{e_R}^2\right), \tag{4.2}$$

and the associated cross-equation restriction is

$$\begin{aligned}\begin{bmatrix} b_0 & b_1 & b_2 & b_3 \end{bmatrix} = [\delta \quad 0 \quad 0 \quad 0] \\ \times\left[[\boldsymbol{I}+\boldsymbol{A}] + \beta^2[\boldsymbol{A}^2+\boldsymbol{A}^3]\right. \\ \left. + \cdots + \beta^{38}[\boldsymbol{A}^{38}+\boldsymbol{A}^{39}]\right],\end{aligned} \tag{4.3a}$$

[8] The raw data (rr_t) were Treasury bill discount rates reported at an annual rate and measured in units of 100 basis points. The conversion used was $r_t = 9125 \times rr_t/(36{,}000 - 91 \times rr_t)$. Data for rr_t are from the series RMGBS3D, carried on a daily basis in the data banks of the Board of Governors of the Federal Reserve System going back to 1962. I am grateful to David Wilcox and Bonnie Garrett for assistance in obtaining these series. Note that the formulas of section 3 measured r_t as a fraction of 1, whereas in this section on empirical results it is reported in units of 100 basis points. All numerical calculations have been adjusted appropriately.

[9] This is the Treasury's 'constant maturity' series, inferred by term structure interpolation from bonds whose maturity is closest to ten years from a given sample date. The raw data were reported at an annual rate and were converted to a semiannual rate by dividing by 2. The source for the data is the series RMGNB10D (see preceding footnote).

Table 1

Data used in analysis: yield to maturity on three-month T-bills (quarterly rate) and yield to maturity on ten-year Treasury bonds (semiannual rate).

Date (t)	Three-month bills (r_t)	Ten-year bonds (R_t)
1962:1	0.6891	1.3500
1962:2	0.6994	1.3700
1962:3	0.7456	1.4600
1962:4	0.6968	1.3650
1963:1	0.7354	1.4400
1963:2	0.7405	1.4500
1963:3	0.7662	1.5000
1963:4	0.8615	1.6850
1964:1	0.9028	1.7650
1964:2	0.9028	1.7650
1964:3	0.8899	1.7400
1964:4	0.9080	1.7750
1965:1	0.9751	1.9050
1965:2	1.0061	1.9650
1965:3	0.9777	1.9100
1965:4	1.0242	2.0000
1966:1	1.1563	2.2550
1966:2	1.1537	2.2500
1966:3	1.1901	2.3200
1966:4	1.3695	2.6650
1967:1	1.2420	2.4200
1967:2	1.0165	1.9850
1967:3	1.0501	2.0500
1967:4	1.1278	2.2000
1968:1	1.2940	2.5200
1968:2	1.3174	2.5650
1968:3	1.3616	2.6500
1968:4	1.3252	2.5800
1969:1	1.5809	3.0700
1969:2	1.5416	2.9950
1969:3	1.7197	3.3350
1969:4	1.8509	3.5850
1970:1	2.0485	3.9600
1970:2	1.6332	3.1700
1970:3	1.6699	3.2400
1970:4	1.5077	2.9300
1971:1	1.2550	2.4450
1971:2	0.9338	1.8250
1971:3	1.3382	2.6050
1971:4	1.1667	2.2750

Table 1 (continued)

Date (t)	Three-month bills (r_t)	Ten-year bonds (R_t)
1972:1	0.9441	1.8450
1972:2	0.9700	1.8950
1972:3	1.0605	2.0700
1972:4	1.1563	2.2550
1973:1	1.3408	2.6100
1973:2	1.6646	3.2300
1973:3	2.0432	3.9500
1973:4	1.8220	3.5300
1974:1	1.9430	3.7600
1974:2	2.1515	4.1550
1974:3	1.8799	3.6400
1974:4	1.7249	3.3450
1975:1	1.7800	3.4500
1975:2	1.4503	2.8200
1975:3	1.5495	3.0100
1975:4	1.6830	3.2650
1976:1	1.3356	2.6000
1976:2	1.2862	2.5050
1976:3	1.3773	2.6800
1976:4	1.2940	2.5200
1977:1	1.1252	2.1950
1977:2	1.1641	2.2700
1977:3	1.2888	2.5100
1977:4	1.5312	2.9750
1978:1	1.5913	3.0900
1978:2	1.6253	3.1550
1978:3	1.7931	3.4750
1978:4	2.0643	3.9900
1979:1	2.4433	4.7050
1979:2	2.4779	4.7700
1979:3	2.3158	4.4650
1979:4	2.6405	5.0750
1980:1	3.1827	6.0850
1980:2	3.9601	7.5150
1980:3	2.0802	4.0200
1980:4	2.9888	5.7250
1981:1	3.8562	7.3250
1981:2	3.2367	6.1850
1981:3	3.8562	7.3250
1981:4	3.8343	7.2850
1982:1	2.9726	5.6950
1982:2	3.4722	6.6200
1982:3	3.2853	6.2750
1982:4	1.8983	3.6750

Table 1 (continued)

Date (t)	Three-month bills (r_t)	Ten-year bonds (R_t)
1983:1	2.0485	3.9600
1983:2	2.2389	4.3200
1983:3	2.2627	4.3650
1983:4	2.2654	4.3700
1984:1	2.3344	4.5000
1984:2	2.5152	4.8400
1984:3	2.5311	4.8700
1984:4	2.6619	5.1150
1985:1	2.0248	3.9150
1985:2	2.1093	4.0750
1985:3	1.7721	3.4350
1985:4	1.8141	3.5150
1986:1	1.8325	3.5500
1986:2	1.6306	3.1650
1986:3	1.5416	2.9950
1986:4	1.3330	2.5950
1987:1	1.4268	2.7750
1987:2	1.4242	2.7700
1987:3	1.4503	2.8200

where

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{4.3b}$$

Throughout the analysis, $\delta = [(\bar{r}^2 + 2\bar{r})/(1+\bar{r})][1 - 1/(1+\bar{r})^{40}]^{-1}$ and $\beta = 1/(1+\bar{r})$ were imposed a priori on the basis of the sample mean of the short rate ($\bar{r} = 0.016964$) over 1962:1–1987:3; thus $\delta = 0.0686974$ and $\beta = 0.9833189$.

The nonlinear representation for r_t takes the form

$$r_t = \alpha_0 + \alpha_1 S_t + z_t, \tag{4.4a}$$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + \phi_4 z_{t-4} + [\omega_0 + \omega_1 S_t] v_t, \quad v_t \sim \mathrm{N}(0,1), \tag{4.4b}$$

$$\mathrm{P}[S_t = 1 | S_{t-1} = 1] = p, \tag{4.4c}$$

$$\mathrm{P}[S_t = 0 | S_{t-1} = 0] = q. \tag{4.4d}$$

The corresponding nonlinear representation for R_t is

$$R_t = \kappa_R + \beta_0 r_t + \beta_1 r_{t-1} + \beta_2 r_{t-2} + \beta_3 r_{t-3} + \gamma_0 \cdot \mathrm{P}[S_t = 1 | r_t, r_{t-1}, \ldots] + \gamma_1 \cdot \mathrm{P}[S_{t-1} = 1 | r_t, r_{t-1}, \ldots] + \gamma_2 \cdot \mathrm{P}[S_{t-2} = 1 | r_t, r_{t-1}, \ldots] + \gamma_3 \cdot \mathrm{P}[S_{t-3} = 1 | r_t, r_{t-1}, \ldots] + \varepsilon_{R,t}, \quad \varepsilon_{R,t} \sim \mathrm{N}(0, \sigma^2_{\varepsilon_R}), \tag{4.5}$$

where $\mathrm{P}[S_{t-j} = 1 | r_t, r_{t-1}, \ldots]$ is calculated from step 4 of the basic filter associated with the nonlinear univariate process for r_t [system (4.4)]. The associated cross-equation restrictions are

$$[\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3] = [\delta \quad 0 \quad 0 \quad 0] \times [[I + \Delta] + \beta^2 [\Delta^2 + \Delta^3] + \cdots + \beta^{38} [\Delta^{38} + \Delta^{39}]], \tag{4.6a}$$

$$\Delta = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \phi_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{4.6b}$$

$$\gamma_0 = \frac{\delta \alpha_1 (1 + \lambda) [1 - (\beta\lambda)^{40}]}{1 - (\beta\lambda)^2} - \alpha_1 \beta_0, \tag{4.6c}$$

$$\gamma_j = -\alpha_1 \beta_j, \qquad j = 2, 3, 4, \tag{4.6d}$$

$$\lambda = -1 + p + q. \tag{4.6e}$$

The basic results for the various specifications are summarized in table 2 and discussed in detail below.

Table 2

Summary of fit achieved by alternative specifications.

Model	Short rate (r_t)	Long rate (R_t)	Cross-equation restrictions?	Number of parameters	Log likelihood[a]
1	Linear [eq. (4.7)]	Linear [eq. (4.8)]	No	12	450.27
2	Linear [eq. (4.9)]	Linear [eq. (4.10)]	Yes	8	436.86
3	Nonlinear [table 3]	Linear [eq. (4.8)]	No	16	482.12
4	Nonlinear [table 6]	Nonlinear [eq. (4.11)]	Yes	12	508.26
5	Nonlinear [table 6]	Nonlinear [eq. (4.12)]	No	20	514.18

[a] The constant term $-N \ln(2\pi)$ has been omitted from all entries.

Model 1: Unrestricted linear representations for both r_t and R_t

The benchmark case is the unrestricted linear representation (4.1) and (4.2). OLS estimation for $t = 1963{:}1-1987{:}3$ yields (standard errors in parentheses):

$$\begin{aligned} r_t = {} & \underset{(0.0827)}{0.1627} + \underset{(0.0992)}{0.7138} r_{t-1} - \underset{(0.114)}{0.004} r_{t-2} + \underset{(0.114)}{0.469} r_{t-3} \\ & - \underset{(0.0984)}{0.2676} r_{t-4} + e_{r,t}, \qquad \hat{\sigma}_{e_r} = 0.311, \end{aligned} \tag{4.7}$$

$$\begin{aligned} R_t = {} & \underset{(0.00336)}{0.07298} + \underset{(0.00399)}{1.88043} r_t + \underset{(0.00457)}{0.00465} r_{t-1} + \underset{(0.00457)}{0.00719} r_{t-2} \\ & - \underset{(0.00396)}{0.00121} r_{t-3} + e_{R,t}, \qquad \hat{\sigma}_{e_R} = 0.0125. \end{aligned} \tag{4.8}$$

Model 2: Restricted linear representations for r_t and R_t

Eq. (4.8) indicates that 99.99% of the variance of the long rate can be accounted for by a linear regression on the short rate. Is this close dependence what one would expect if investors were rationally forecasting short rates on the basis of a linear representation such as (4.7)?

To investigate this possibility, I estimated the system (4.1)–(4.2) subject to the constraints (4.3):[10]

$$\begin{aligned} r_t = {} & \underset{(0.0353)}{0.01487} + \underset{(0.00241)}{0.992990} r_{t-1} - \underset{(0.00414)}{0.00141} r_{t-2} + \underset{(0.00397)}{0.00501} r_{t-3} \\ & - \underset{(0.00211)}{0.000888} r_{t-4} + e_{r,t}, \qquad \hat{\sigma}_{e_r} = \underset{(0.0262)}{0.35657}, \end{aligned} \tag{4.9}$$

$$\begin{aligned} R_t = {} & \underset{(0.00328)}{0.07300} + 1.8805 r_t + 0.00474 r_{t-1} + 0.00755 r_{t-2} \\ & - 0.00161 r_{t-3} + e_{R,t}, \qquad \hat{\sigma}_{e_R} = \underset{(0.000888)}{0.012503}\ . \end{aligned} \tag{4.10}$$

The parameters reported without standard errors in eq. (4.10) were generated from the parameters reported in eq. (4.9) by use of (4.3).

[10] Maximization was achieved by a Davidon–Fletcher–Powell routine. Standard errors were derived from the numerically evaluated second derivatives of the log likelihood function. I would like to thank Kent Wall for use of his DFP routine and Steve Stern for use of his second derivative program.

Note from table 2 that the standard likelihood ratio statistic for comparing the restricted and unrestricted linear specifications is $2(450.27 - 436.86) = 26.82$. This is distributed as $\chi^2(4)$ under the null hypothesis that the restrictions correctly characterize the data. The null hypothesis is thus rejected with a p-value[11] of 2.16×10^{-5}. The data are grossly inconsistent with the cross-equation restrictions implied by the expectations hypothesis of the term structure.

A comparison between the restricted estimated (4.9)–(4.10) and the unrestricted estimates (4.7)–(4.8) reveals why. If one were willing to believe that the short rate basically follows a random walk [as the restricted estimates in eq. (4.9) try to construe it to be], one would predict [as in eq. (4.10)] precisely the response of long rates to short rates as is found in the unrestricted regression [eq. (4.8)]. However, the actual process for short rates [eq. (4.7)] is too far from being a random walk to make this scenario plausible; the leading coefficient is too small and coefficients at lags 3 and 4 too large to be consistent with a simple random walk. Thus the behavior of long rates suggests that investors were not basing their forecasts on the assumption of a linear process for short rates.

Model 3: Unrestricted nonlinear representation for r_t, unrestricted linear representation for R_t

Not only is the response of long rates to short rates difficult to reconcile with the linear specification (4.1); I now show that the univariate behavior of short rates alone also offers overwhelming evidence against (4.1) in favor of a nonlinear model incorporating the possibility of changes in regime.

The nonlinear Markov model of changes in regime was fit to univariate data on r_t by means of the basic filter evaluation of the conditional log likelihood function as described in section 2. Maximum likelihood estimates are reported in table 3.[12] Note that data on long rates (R_t) were not used at all in constructing the parameter estimates in table 3.

These estimates speak to a very dramatic shift in the time series properties of short-term interest rates between states 0 and 1. State 1 is characterized by an average level of interest rates nearly twice as high as that in state 0 ($\alpha_0 + \alpha_1 = 2.8$ versus $\alpha_0 = 1.6$). Even more dramatic is the four-fold increase in the standard deviation of the Gaussian component of the process ($\omega_0 + \omega_1 = 0.73$ versus $\omega_0 = 0.18$).

[11] The p-value (also referred to as the marginal significance level) represents the probability that as large a difference would have been found given that the null hypothesis is true; $p < 0.05$ is the standard criterion for rejecting the null hypothesis.

[12] The probabilities p and q were parameterized as $\exp[-\theta_p^2]$ and $\exp[-\theta_q^2]$. This is desirable not to force the estimates to lie within this interval (since this leaves the maximization routine free to push p or q arbitrarily close to 0 or 1), but rather to ensure that the search procedure always evaluates a well-defined likelihood function.

Table 3

Maximum likelihood estimates of nonlinear Markov model as estimated from the behavior of three-month Treasury bill yields (r_t).

Parameter[a]	Maximum likelihood estimate	Standard error
α_0	1.633	0.208
α_1	1.188	0.230
p	0.9087	0.0818
q	0.9899	0.0102
ω_0	0.1752	0.0135
ω_1	0.5574	0.166
ϕ_1	0.8778	0.103
ϕ_2	0.0675	0.150
ϕ_3	0.1692	0.152
ϕ_4	-0.2108	0.0952

[a] Parameters are defined by eq. (4.4) and were estimated solely on the basis of observations on short rates $\{r_t\}$ for $t = 1962{:}1$–$1987{:}3$.

From the output of step 4 at iteration t of the basic filter, we can assign probabilities to whether the process was in regime 0 or regime 1 at date t based on information available at the time $P[S_t = 1 | r_t, r_{t-1}, \ldots, r_0]$. This series is reported in the column labelled '$j = 0$' in table 4. We can further obtain from step 4 a refined assessment of the probable state of the process in the preceding four quarters $P[S_{t-j} = 1 | r_t, r_{t-1}, \ldots, r_0]$, also reported in table 4. The maximum likelihood estimates associate the shift in regime between the two states very dramatically with the new monetary policy adopted in October of 1979. If one were only looking at the behavior of short rates, the change was not apparent at the time ($P[S_{1979:4} = 1 | r_{1979:4}, \ldots] = 0.0088$), but by January 1, 1980, evidence of a regime shift was fairly convincing ($P[S_{1980:1} = 1 | r_{1980:1}, \ldots] = 0.7553$). This recognition in 1980:1 would incidentally cause one to significantly revise upward the likelihood that in fact the regime change had begun the preceding quarter ($P[S_{1979:4} = 1 | r_{1980:1}, \ldots] = 0.3371$). By 1980:3, one would conclude that the shift in regime was indeed more likely than not to have begun in 1979:4 ($P[S_{1979:4} = 1 | r_{1980:3}, \ldots] = 0.5628$). The shift back to the original regime was immediately apparent to the filter in 1982:4.

The period 1979:4–1982:3 is thus identified as a time of dramatically higher and more volatile short-term interest rates than that seen before or since. Fig. 1 makes this point visually, depicting the short-term rate against the imputed regime of higher, more volatile interest rates. This dating of an apparent shift in the process for short-term interest rates corresponds precisely with a profound change in Federal Reserve operating procedures. Beginning in October 1979 and ending in October 1982, the Federal Reserve adopted a policy of targeting nonborrowed reserves, allowing interest rates to fluctuate across a broad range. Outside of this period, the policy has basically been one

Table 4

Inferred probability, using parameters of unrestricted univariate process for short-term rates (table 3) and based on information available at date t, that the economy was in the high interest rate, high volatility state at date t (column $j=0$), along with inference using table 3 parameters and based on information available at t about the historical state the economy had been in at dates $t-1$, $t-2$, $t-3$, and $t-4$ (columns $j=1,2,3,4$).

Date	$P[S_{t-j}=1 \mid r_t, r_{t-1}, \ldots]$				
(t)	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
1963:1	0.0253	0.0247	0.0324	0.0379	0.0372
1963:2	0.0069	0.0063	0.0061	0.0142	0.0143
1963:3	0.0022	0.0016	0.0014	0.0014	0.0033
1963:4	0.0011	0.0005	0.0003	0.0003	0.0003
1964:1	0.0009	0.0002	0.0001	0.0001	0.0001
1964:2	0.0008	0.0002	0.0001	0.0000	0.0000
1964:3	0.0008	0.0002	0.0001	0.0000	0.0000
1964:4	0.0008	0.0002	0.0000	0.0000	0.0000
1965:1	0.0008	0.0002	0.0000	0.0000	0.0000
1965:2	0.0008	0.0002	0.0000	0.0000	0.0000
1965:3	0.0008	0.0002	0.0000	0.0000	0.0000
1965:4	0.0008	0.0002	0.0000	0.0000	0.0000
1966:1	0.0010	0.0002	0.0000	0.0000	0.0000
1966:2	0.0008	0.0002	0.0000	0.0000	0.0000
1966:3	0.0008	0.0002	0.0000	0.0000	0.0000
1966:4	0.0012	0.0002	0.0000	0.0000	0.0000
1967:1	0.0010	0.0004	0.0001	0.0000	0.0000
1967:2	0.0019	0.0007	0.0003	0.0000	0.0000
1967:3	0.0010	0.0004	0.0002	0.0001	0.0000
1967:4	0.0010	0.0002	0.0001	0.0000	0.0000
1968:1	0.0018	0.0004	0.0001	0.0000	0.0000
1968:2	0.0010	0.0004	0.0001	0.0000	0.0000
1968:3	0.0009	0.0002	0.0001	0.0000	0.0000
1968:4	0.0008	0.0002	0.0001	0.0000	0.0000
1969:1	0.0026	0.0004	0.0001	0.0000	0.0000
1969:2	0.0012	0.0006	0.0001	0.0000	0.0000
1969:3	0.0018	0.0004	0.0002	0.0000	0.0000
1969:4	0.0014	0.0005	0.0001	0.0000	0.0000
1970:1	0.0030	0.0007	0.0002	0.0000	0.0000
1970:2	0.0116	0.0080	0.0020	0.0007	0.0001
1970:3	0.0032	0.0025	0.0018	0.0004	0.0002
1970:4	0.0019	0.0011	0.0009	0.0007	0.0001
1971:1	0.0014	0.0007	0.0004	0.0003	0.0002
1971:2	0.0044	0.0022	0.0011	0.0007	0.0005
1971:3	0.0211	0.0086	0.0041	0.0020	0.0013
1971:4	0.0063	0.0057	0.0024	0.0011	0.0005
1972:1	0.0037	0.0029	0.0026	0.0011	0.0005
1972:2	0.0017	0.0010	0.0008	0.0007	0.0003
1972:3	0.0012	0.0004	0.0002	0.0002	0.0002
1972:4	0.0012	0.0003	0.0001	0.0001	0.0000

Table 4 (continued)

Date	$P[S_{t-j} = 1 \mid r_t, r_{t-1}, \ldots]$				
(t)	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
1973:1	0.0015	0.0004	0.0001	0.0000	0.0000
1973:2	0.0051	0.0011	0.0003	0.0001	0.0000
1973:3	0.0196	0.0088	0.0018	0.0004	0.0001
1973:4	0.0098	0.0089	0.0041	0.0009	0.0002
1974:1	0.0030	0.0022	0.0020	0.0009	0.0002
1974:2	0.0025	0.0010	0.0008	0.0007	0.0003
1974:3	0.0016	0.0008	0.0003	0.0003	0.0002
1974:4	0.0012	0.0005	0.0003	0.0001	0.0001
1975:1	0.0010	0.0003	0.0001	0.0001	0.0000
1975:2	0.0015	0.0005	0.0002	0.0001	0.0000
1975:3	0.0014	0.0004	0.0001	0.0000	0.0000
1975:4	0.0015	0.0004	0.0001	0.0000	0.0000
1976:1	0.0021	0.0010	0.0003	0.0001	0.0000
1976:2	0.0012	0.0005	0.0003	0.0001	0.0000
1976:3	0.0010	0.0003	0.0001	0.0001	0.0000
1976:4	0.0008	0.0002	0.0001	0.0000	0.0000
1977:1	0.0011	0.0003	0.0001	0.0000	0.0000
1977:2	0.0009	0.0002	0.0001	0.0000	0.0000
1977:3	0.0012	0.0002	0.0001	0.0000	0.0000
1977:4	0.0036	0.0007	0.0001	0.0000	0.0000
1978:1	0.0015	0.0008	0.0002	0.0000	0.0000
1978:2	0.0010	0.0003	0.0002	0.0000	0.0000
1978:3	0.0013	0.0003	0.0001	0.0000	0.0000
1978:4	0.0049	0.0010	0.0002	0.0001	0.0000
1979:1	0.0373	0.0155	0.0029	0.0005	0.0002
1979:2	0.0101	0.0092	0.0038	0.0007	0.0001
1979:3	0.0038	0.0031	0.0029	0.0012	0.0002
1979:4	0.0088	0.0034	0.0028	0.0026	0.0012
1980:1	0.7553	0.3371	0.1273	0.1039	0.0973
1980:2	1.0000	0.9933	0.4131	0.1571	0.1374
1980:3	0.8776	1.0000	0.9962	0.5628	0.0981
1980:4	1.0000	0.9976	1.0000	0.9953	0.6007
1981:1	1.0000	1.0000	0.9974	1.0000	0.9970
1981:2	1.0000	1.0000	1.0000	0.9976	1.0000
1981:3	1.0000	1.0000	1.0000	1.0000	0.9981
1981:4	1.0000	1.0000	1.0000	1.0000	1.0000
1982:1	0.9934	1.0000	1.0000	1.0000	1.0000
1982:2	0.9995	0.9994	1.0000	1.0000	1.0000
1982:3	0.9976	0.9976	0.9976	1.0000	1.0000
1982:4	0.3909	1.0000	1.0000	1.0000	1.0000
1983:1	0.1333	0.1326	1.0000	1.0000	1.0000
1983:2	0.1089	0.1052	0.1046	1.0000	1.0000
1983:3	0.0323	0.0313	0.0301	0.0299	1.0000
1983:4	0.0080	0.0073	0.0070	0.0068	0.0068

Table 4 (continued)

Date (t)	$P[S_{t-j}=1 \mid r_t, r_{t-1}, \ldots]$				
	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
1984:1	0.0029	0.0020	0.0018	0.0018	0.0017
1984:2	0.0045	0.0017	0.0011	0.0010	0.0010
1984:3	0.0021	0.0012	0.0004	0.0003	0.0003
1984:4	0.0027	0.0009	0.0005	0.0002	0.0001
1985:1	0.0994	0.0745	0.0264	0.0151	0.0047
1985:2	0.0282	0.0273	0.0209	0.0072	0.0042
1985:3	0.0238	0.0225	0.0219	0.0187	0.0053
1985:4	0.0102	0.0086	0.0082	0.0079	0.0097
1986:1	0.0030	0.0023	0.0019	0.0018	0.0018
1986:2	0.0014	0.0008	0.0006	0.0005	0.0005
1986:3	0.0010	0.0003	0.0002	0.0001	0.0001
1986:4	0.0012	0.0004	0.0002	0.0001	0.0001
1987:1	0.0013	0.0003	0.0001	0.0000	0.0000
1987:2	0.0010	0.0003	0.0001	0.0000	0.0000
1987:3	0.0010	0.0002	0.0001	0.0000	0.0000

of targeting the Federal funds rate.[13] This change in Fed operating procedures is widely recognized by monetary economists as 'one of the more dramatic events in the recent history of monetary policy' [Spindt and Tarhan (1987, p. 107)] and has been the object of a large number of scholarly studies, including Antoncic (1986), Huizinga and Mishkin (1986), Hardouvelis and Barnhart (1987), Spindt and Tarhan (1987), and Walsh (1987). It is of great interest that the dates of the change in Fed policy correspond precisely with the values imputed by the filter to changes in the regime governing interest rates.

The statistical evidence for a process characterized by such changes in regime in preference to the linear representation (4.1) is overwhelming. For purposes of comparison with the other models in table 2, one can think of combining the nonlinear process for short rates of table 3 with an unrestricted linear representation for long rates [eq. (4.8)]. Note that by fitting four additional parameters relative to a linear representation for r_t (model 1), a striking improvement in the likelihood is achieved. The standard likelihood ratio statistic is $2 \times [482.12 - 450.27] = 63.7$. One usually treats this[14] as $\chi^2(4)$,

[13]Targeting the level of borrowed reserves amounts to pretty much the same thing; see Goodfriend and Whelpley (1986) and Spindt and Tarhan (1987).

[14]A complication arises here in that under the null hypothesis that $\alpha_1 = \omega_1 = 0$, the parameters p and q are unidentified. See the discussion in my 1987 paper. The Lagrange multiplier test reported shortly is immune to these objections and gives equally dramatic results.

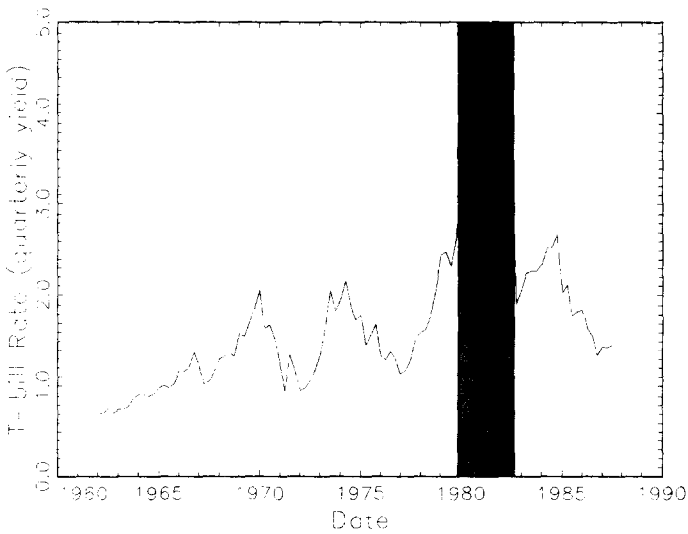

Fig. 1

for which a value only half as large would have led to rejection of this fixed-parameter linear model with a p-value of 10^{-6}.

The strength of this rejection is largely due to the overwhelming evidence of different residual variances associated with the two regimes. It is instructive to compare briefly evidence for heteroskedasticity of the kind predicted by my model with that implied by other popular specifications.

The first panel of table 5 reports Lagrange multiplier tests of the null hypothesis of a constant-parameter linear model with homoskedastic errors against three separate alternatives: (i) a specification in which the variance of residuals depends on the lagged level of interest rates [as in Marsh and Rosenfeld (1983)]; (ii) a specification in which the variance of residuals depends on the lagged squared residual [as in Engle's (1982) ARCH model]; and (iii) a specification in which the variance of residuals depends on the lagged output of my nonlinear filter. The null hypothesis would be overwhelmingly rejected in comparison with any of these three alternatives, though the statistical evidence is strongest when the null is compared against the third specification.

To what extent might the evidence for ARCH and level-dependent heteroskedasticity be due to the change in regime associated with Fed operating procedure? Some preliminary evidence on this issue is presented in the bottom panel of table 5. When we restrict the analysis to the period prior to 1979:4 (when my model would imply homoskedastic errors), we still find evidence of conditional heteroskedasticity from both specifications (i) and (ii). Neverthe-

Table 5

Comparison of alternative specifications of conditional heteroskedasticity.

I. Full sample period

(A) OLS estimate of process for short rate ($t = 1963{:}1\text{–}1987{:}3$)

$$r_t = 0.163 + 0.714 r_{t-1} - 0.004 r_{t-2} + 0.469 r_{t-3} - 0.268 r_{t-4} + \hat{\varepsilon}_t$$

(B) Tests for conditional heteroskedasticity of residuals ($t = 1963{:}2\text{–}1987{:}3$)

(i) *Variance depends on lagged level of interest rate*

$$\underset{}{\hat{\varepsilon}_t^2} = \underset{(0.0617)}{-0.2374} + \underset{(0.0326)}{0.1928} r_{t-1}, \qquad TR^2 = 26.21 \sim \chi^2(1)$$

(ii) *Variance depends on lagged squared residuals (ARCH)*

$$\hat{\varepsilon}_t^2 = \underset{(0.0268)}{0.0519} + \underset{(0.0901)}{0.4697} \hat{\varepsilon}_{t-1}^2, \qquad TR^2 = 21.62 \sim \chi^2(1)$$

(iii) *Variance depends on lagged filter output (regime change)*

$$\hat{\varepsilon}_t^2 = \underset{(0.0262)}{0.0398} + \underset{(0.799)}{0.4845} \mathrm{P}[S_{t-1} = 1 \mid r_{t-1}, \ldots], \qquad TR^2 = 27.15 \sim \chi^2(1)$$

II. Pre-1979 period

(A) OLS estimate of process for short rate ($t = 1963{:}1\text{–}1979{:}3$)

$$r_t = 0.179 + 1.038 r_{t-1} - 0.176 r_{t-2} + 0.314 r_{t-3} - 0.296 r_{t-4} + \hat{\varepsilon}_t$$

(B) Tests for conditional heteroskedasticity of residuals ($t = 1963{:}2\text{–}1979{:}3$)

(i) *Variance depends on lagged level of interest rate*

$$\hat{\varepsilon}_t^2 = \underset{(0.0175)}{-0.0122} + \underset{(0.0122)}{0.0288}\, r_{t-1}, \qquad TR^2 = 5.30 \sim \chi^2(1)$$

(ii) *Variance depends on lagged squared residuals (ARCH)*

$$\hat{\varepsilon}_t^2 = \underset{(0.00599)}{0.02099} + \underset{(0.121)}{0.241} \hat{\varepsilon}_{t-1}^2 \qquad TR^2 = 3.85 \sim \chi^2(1)$$

less, the evidence is far less compelling than the comparable tests that include the entire sample period – the tests for the pre-1979 data reject the null hypothesis of homoskedastic errors at the 5% level but not the 1% level; by contrast, when the full sample period is used, p-values are well below 10^{-6}. Moreover, the parameterization of the conditional heteroskedasticity implied by either process (i) or (ii) seems to change dramatically upon inclusion of the post-1979 data. The implied coefficient relating $\hat{\varepsilon}_t^2$ to r_{t-1} increases by a factor of 7, or a move of 13 standard deviations ($= [0.1928 - 0.0288]/0.0122$) relative to the value implied by the pre-1979 formulation.

I conclude that (a) both regime shifts and ARCH or level effects seem to be present in the data, though the regime shift is by far the most dramatic source of conditional heteroskedasticity over the full sample period; and (b) if we tried to model this conditional heteroskedasticity solely on the basis of specification (i) or (ii) we would have to allow for a discrete shift during

Table 6

Maximum likelihood estimates of nonlinear Markov model as estimated from joint behavior of three-month Treasury bill yields (r_t) and ten-year Treasury bonds (R_t) subject to cross-equation restrictions.

Parameter[a]	Maximum likelihood estimate	Standard error
α_0	1.884	5.127
α_1	1.3156	0.0760
p	0.996819	0.000784
q	0.998106	0.000803
ω_0	0.17429	0.00791
ω_1	0.2245	0.0309
ϕ_1	0.99126	0.00193
ϕ_2	−0.00138	0.00295
ϕ_3	0.00619	0.00274
ϕ_4	0.000141	0.00157
κ_R	0.05058	0.00298
σ_{ε_R}	0.007880	0.000610

[a]Parameters are defined by eqs. (4.4) and (4.5) and were estimated on the basis of observation of both short rates $\{r_t\}$ and long rates $\{R_t\}$ for $t = 1962{:}1$–$1987{:}3$ with the cross-equation restrictions (4.6) imposed.

1979–82 of the coefficients characterizing the dependence of the variance on lagged levels of the interest rate or on the lagged squared residuals.[15]

Model 4: Restricted nonlinear representations for r_t and R_t

The question then is whether this nonlinear process for short rates can also better account for the behavior of long rates as well. The system (4.4)–(4.5) was estimated subject to (4.6). Maximum likelihood estimates of parameters are reported in table 6, and the inferred probabilities for regime shifts given in table 7. The implicit process for the long rate is

$$\begin{aligned} R_t = \underset{(0.00298)}{0.05058} &+ 1.8873\hat{z}_{t|t} + 0.00865\hat{z}_{t-1|t} + 0.0116\hat{z}_{t-2|t} \\ &+ 0.000258\hat{z}_{t-3|t} + 2.4548\,\mathrm{P}[S_t = 1 \mid r_t, \dots] + \varepsilon_{R,t}, \quad \hat{\sigma}_{\varepsilon_R} = \underset{(0.000610)}{0.007880}, \end{aligned} \tag{4.11}$$

$$\hat{z}_{t-j|t} = r_t - \underset{(0.0760)}{1.3156}\,\mathrm{P}[S_{t-j} = 1 \mid r_t, \dots],$$

[15]Failure to model this shift in regime may account for the finding of an explosive ARCH process by Engle, Lilien and Robins (1987), in which the sum of the ARCH coefficients exceeded unity. They tested (in their table 1, p. 402) for a change in the constant term between the specifications labeled I.B.ii and II.B.ii in my table 5, but not for differences in the autoregressive coefficients.

Table 7

Inferred probability, using parameters of process estimated subject to cross-equation restrictions (table 6) and based on information available at date t, that the economy was in the high interest rate, high volatility state at date t (column $j=0$), along with inference using table 6 parameters and based on information available at t about the historical state the economy had been in at dates $t-1$, $t-2$, $t-3$, and $t-4$ (columns $j=1,2,3,4$).

Date	$P[S_{t-j}=1 \mid r_t, r_{t-1}, \ldots]$				
(t)	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
1963:1	0.2090	0.2090	0.2098	0.2106	0.2114
1963:2	0.1034	0.1034	0.1034	0.1047	0.1061
1963:3	0.0482	0.0482	0.0482	0.0482	0.0499
1963:4	0.0242	0.0242	0.0242	0.0242	0.0242
1964:1	0.0109	0.0109	0.0109	0.0109	0.0109
1964:2	0.0048	0.0048	0.0048	0.0048	0.0048
1964:3	0.0021	0.0021	0.0021	0.0021	0.0021
1964:4	0.0009	0.0009	0.0009	0.0009	0.0009
1965:1	0.0004	0.0004	0.0004	0.0004	0.0004
1965:2	0.0002	0.0002	0.0002	0.0002	0.0002
1965:3	0.0001	0.0001	0.0001	0.0001	0.0001
1965:4	0.0000	0.0000	0.0000	0.0000	0.0000
1966:1	0.0000	0.0000	0.0000	0.0000	0.0000
1966:2	0.0000	0.0000	0.0000	0.0000	0.0000
1966:3	0.0000	0.0000	0.0000	0.0000	0.0000
1966:4	0.0000	0.0000	0.0000	0.0000	0.0000
1967:1	0.0000	0.0000	0.0000	0.0000	0.0000
1967:2	0.0000	0.0000	0.0000	0.0000	0.0000
1967:3	0.0000	0.0000	0.0000	0.0000	0.0000
1967:4	0.0000	0.0000	0.0000	0.0000	0.0000
1968:1	0.0000	0.0000	0.0000	0.0000	0.0000
1968:2	0.0000	0.0000	0.0000	0.0000	0.0000
1968:3	0.0000	0.0000	0.0000	0.0000	0.0000
1968:4	0.0000	0.0000	0.0000	0.0000	0.0000
1969:1	0.0001	0.0000	0.0000	0.0000	0.0000
1969:2	0.0000	0.0000	0.0000	0.0000	0.0000
1969:3	0.0000	0.0000	0.0000	0.0000	0.0000
1969:4	0.0000	0.0000	0.0000	0.0000	0.0000
1970:1	0.0001	0.0000	0.0000	0.0000	0.0000
1970:2	0.0003	0.0003	0.0001	0.0001	0.0000
1970:3	0.0001	0.0001	0.0001	0.0001	0.0000
1970:4	0.0001	0.0001	0.0001	0.0001	0.0000
1971:1	0.0001	0.0001	0.0001	0.0001	0.0001
1971:2	0.0001	0.0001	0.0001	0.0001	0.0001
1971:3	0.0013	0.0005	0.0005	0.0005	0.0005
1971:4	0.0008	0.0008	0.0003	0.0003	0.0003
1972:1	0.0007	0.0007	0.0007	0.0003	0.0003
1972:2	0.0003	0.0003	0.0003	0.0003	0.0001
1972:3	0.0002	0.0002	0.0001	0.0001	0.0001
1972:4	0.0001	0.0001	0.0001	0.0001	0.0001

Table 7 (continued)

Date (t)	$P[S_{t-j}=1 \mid r_t, r_{t-1}, \ldots]$				
	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
1973:1	0.0001	0.0001	0.0001	0.0001	0.0001
1973:2	0.0004	0.0002	0.0001	0.0001	0.0001
1973:3	0.0017	0.0011	0.0005	0.0003	0.0003
1973:4	0.0014	0.0014	0.0009	0.0004	0.0003
1974:1	0.0007	0.0007	0.0007	0.0005	0.0002
1974:2	0.0006	0.0006	0.0006	0.0006	0.0004
1974:3	0.0007	0.0007	0.0007	0.0006	0.0006
1974:4	0.0004	0.0004	0.0004	0.0004	0.0004
1975:1	0.0002	0.0002	0.0002	0.0002	0.0002
1975:2	0.0004	0.0004	0.0004	0.0003	0.0003
1975:3	0.0002	0.0002	0.0002	0.0002	0.0002
1975:4	0.0001	0.0001	0.0001	0.0001	0.0001
1976:1	0.0003	0.0002	0.0002	0.0002	0.0002
1976:2	0.0001	0.0001	0.0001	0.0001	0.0001
1976:3	0.0001	0.0001	0.0001	0.0001	0.0000
1976:4	0.0000	0.0000	0.0000	0.0000	0.0000
1977:1	0.0000	0.0000	0.0000	0.0000	0.0000
1977:2	0.0000	0.0000	0.0000	0.0000	0.0000
1977:3	0.0000	0.0000	0.0000	0.0000	0.0000
1977:4	0.0001	0.0000	0.0000	0.0000	0.0000
1978:1	0.0000	0.0000	0.0000	0.0000	0.0000
1978:2	0.0000	0.0000	0.0000	0.0000	0.0000
1978:3	0.0000	0.0000	0.0000	0.0000	0.0000
1978:4	0.0001	0.0000	0.0000	0.0000	0.0000
1979:1	0.0010	0.0004	0.0001	0.0001	0.0000
1979:2	0.0005	0.0005	0.0002	0.0001	0.0000
1979:3	0.0003	0.0003	0.0003	0.0001	0.0000
1979:4	0.0007	0.0005	0.0005	0.0005	0.0002
1980:1	0.0331	0.0164	0.0113	0.0112	0.0111
1980:2	0.9857	0.8539	0.4184	0.2854	0.2839
1980:3	0.2716	1.0000	0.8707	0.4565	0.3116
1980:4	0.9999	0.9628	1.0000	0.8644	0.4448
1981:1	1.0000	1.0000	0.9632	1.0000	0.8643
1981:2	1.0000	1.0000	1.0000	0.9643	1.0000
1981:3	1.0000	1.0000	1.0000	1.0000	0.9643
1981:4	1.0000	1.0000	1.0000	1.0000	1.0000
1982:1	0.9978	1.0000	1.0000	1.0000	1.0000
1982:2	0.9998	0.9998	1.0000	1.0000	1.0000
1982:3	0.9997	0.9997	0.9997	1.0000	1.0000
1982:4	0.2635	1.0000	1.0000	1.0000	1.0000
1983:1	0.1748	0.1748	1.0000	1.0000	1.0000
1983:2	0.1326	0.1326	0.1326	1.0000	1.0000
1983:3	0.0632	0.0632	0.0632	0.0632	1.0000
1983:4	0.0286	0.0286	0.0286	0.0286	0.0286

Table 7 (continued)

Date (t)	$P[S_{t-j}=1 \mid r_t, r_{t-1}, \ldots]$				
	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
1984:1	0.0136	0.0136	0.0136	0.0136	0.0136
1984:2	0.0094	0.0093	0.0093	0.0093	0.0093
1984:3	0.0041	0.0041	0.0041	0.0041	0.0041
1984:4	0.0023	0.0023	0.0023	0.0023	0.0023
1985:1	0.1724	0.1724	0.1714	0.1713	0.1708
1985:2	0.0908	0.0908	0.0908	0.0903	0.0902
1985:3	0.1672	0.1672	0.1672	0.1672	0.1662
1985:4	0.0823	0.0823	0.0823	0.0823	0.0823
1986:1	0.0378	0.0378	0.0378	0.0378	0.0378
1986:2	0.0285	0.0285	0.0285	0.0285	0.0285
1986:3	0.0140	0.0140	0.0140	0.0140	0.0140
1986:4	0.0111	0.0111	0.0111	0.0111	0.0111
1987:1	0.0054	0.0054	0.0054	0.0054	0.0054
1987:2	0.0024	0.0024	0.0024	0.0024	0.0024
1987:3	0.0011	0.0011	0.0011	0.0010	0.0010

where $P[S_{t-j}=1 \mid r_t, \ldots]$ is from column j of table 7. Parameters reported in (4.11) without standard errors were calculated from values in table 6 using eq. (4.6).

Eq. (4.11) is able to account for even more of the variance of long rates than was provided by the unrestricted linear specification (4.8); the R^2 in eq. (4.11) is 0.99997. Indeed, this improved characterization of the long rate is sufficiently important that the restricted nonlinear model is able to achieve a much higher value for the likelihood even than model 3, which used four more parameters than model 4 in order to permit the short rate to follow a completely unrestricted nonlinear process and long rates a completely unrestricted linear process. Of course, there is no contest between the restricted nonlinear model 4 and the unrestricted linear model 1; both use the same number of parameters, but twice the log likelihood ratio is over 100.

We saw that a chief difficulty of the restricted linear model was that the response of the long rate to the short rate is too strong. Investors seemed to be attributing more persistence to the process for short rates than is warranted on the basis of a linear representation such as (4.7). That the success or failure of the expectations hypothesis of the term structure of interest rates depends critically on how one treats the possibility of unit or near unit roots has been recognized by many of the researchers in this field; see, for example, Sargent (1979), Shiller (1981), and Flavin (1983). The persistence of the nonlinear process for interest rates in table 6 results from the high probability of remaining in the current regime (p and q are close to 1) and from the near unit root of $\phi(L)$ in the Gaussian component. These specifications in table 6 are within two standard errors of the unrestricted nonlinear estimates in table 3. Thus, a key feature that allows the restricted nonlinear model to fit the data so well is that the prospect of persistent changes in regime is statistically fairly plausible and would also motivate a strong response of long rates to short rates such as we observe in the data.[16]

One other difference between the restricted and unrestricted parameter estimates for the nonlinear models that is statistically insignificant but nevertheless worth commenting on concerns the comparison between tables 4 and 7. If one uses the restricted parameters of table 6 to draw inference about the historical regimes, the evidence of a return to the original regime ($S_t = 0$) after 1983:1 is much less compelling. This finding might be summarized as follows: given the observed behavior of the long rate, investors were apparently using a model for forecasting changes in regime that still assigned a nonnegligible probability to the continuation of the high interest rate, high volatility regime as late as 1985:3.

Model 5: Unrestricted nonlinear representations for r_t and R_t

Not only does model 4 account for an overwhelming percentage of the variance of long rates, it moreover does so in a way that is consistent with the univariate process assumed for short rates and the rational-expectations hypothesis of the term structure. Here I note that specification 4 would be accepted in preference to other, nonlinear alternatives. The simplest way to construct such an alternative is to estimate eq. (4.5) by OLS with no restrictions on $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$ and with $\mathrm{P}[S_{t-j} = 1 | r_t, \ldots]$ given by table 7. The restrictions (4.6) can then be tested directly. This of course yields the numerically identical result as does an unrestricted regression of the residuals from (4.11) on the set of explanatory variables in (4.5). I report the regression in the latter form for ease of interpretation:

$$\begin{aligned} R_t - \hat{R}_t = \underset{(0.00275)}{0.000482} &+ \underset{(0.00335)}{0.00235} r_t - \underset{(0.00354)}{0.00141} r_{t-1} - \underset{(0.00309)}{0.000306} r_{t-2} \\ &- \underset{(0.00286)}{0.00129} r_{t-3} - \underset{(0.0114)}{0.0168} \hat{P}_{t|t} - \underset{(0.0140)}{0.00588} \hat{P}_{t-1|t} \\ &+ \underset{(0.0137)}{0.0141} \hat{P}_{t-2|t} + \underset{(0.00949)}{0.0111} \hat{P}_{t-3|t} + u_t, \quad \hat{\sigma}_u = 0.00742, \end{aligned} \tag{4.12}$$

[16]The near-unit root complicates the numerical estimation of α_0, which enters the likelihood function in step 2 of the basic filter as $(1 - \phi_1 - \phi_2 - \phi_3 - \phi_4)\alpha_0$. Since moderately sized confidence intervals for $(1 - \phi_1 - \phi_2 - \phi_3 - \phi_4)$ contain zero, virtually any value for α_0 might be consistent with the data, as reflected in the high standard error for $\hat{\alpha}_0$ in table 6. I had the best success in numerically maximizing the likelihood by parameterizing this constant term as $\alpha_0^* = (1 - \phi_1 - \phi_2 - \phi_3 - \phi_4)\alpha_0$, with α_0^* treated as an unrestricted parameter.

where $\hat{P}_{t-j|t}$ denotes $P[S_{t-j}=1|r_t, r_{t-1},\ldots]$ from column j of table 7. The resulting $\chi^2(8)$ test statistic $(2\times[514.18-508.26])=11.84$) leads to ready acceptance of these restrictions ($p=0.16$).

To summarize, the cross-equation restrictions implied by the expectations hypothesis of the term structure of interest rates would be rejected if one assumed a linear process for short rates, but accepted if one used a nonlinear specification. I conclude that once recognition by bond traders of changes in regime is taken into account, the expectations hypothesis of the term structure seems to explain these data quite well.

5. Discussion

This paper presented a model of the term structure of interest rates in which the processes for interest rates were explicitly specified and estimated by full information maximum likelihood. This approach is in contrast to a number of recent tests of the expectations hypothesis of the term structure, which make no attempt to model the process on interest rates directly and may seem to be specification-free. I would like to comment briefly on the implications for the latter approach if the model presented here indeed describes the truth.

One set of results is represented by the observation by Shiller, Campbell and Schoenholtz (1983) that the forward rates implicit in the term structure have essentially no predictive power for future short rates,[17] a result that might be interpreted as offering evidence against the expectations hypothesis. Suppose that short-term interest rates are really driven by the process in tables 3 or 6. Both the Gaussian and the Markov component behave much like a random walk [though we will not see this if we simply fit a constant-parameter linear model as in eq. (4.7)]. Changes in the short rate are very difficult to forecast, and the forward rates implicit in the term structure will be dominated by measurement error. Thus, if the restricted nonlinear model in table 6 (which by construction satisfies the expectations hypothesis) had really generated the data, then the results of Shiller, Campbell and Schoenholtz are precisely what one would have expected.

A second approach that seems to make minimal assumptions about specification regresses the excess one-period holding yields of securities of different maturities on information that was available at the date these bonds were purchased [e.g., Campbell and Shiller (1984), Mankiw and Summers (1984), Fama and Bliss (1987)]. Application of this approach requires careful specification of the conditional heteroskedasticity of the process. For example, Engle, Lilien and Roberds (1987, pp. 399–400) reported that the estimated average excess holding yield falls to a third of its earlier estimated value once the estimation procedure takes into account the ARCH character of the residuals. I have argued above that heteroskedasticity corrections for both

[17] Fama (1984) and Fama and Bliss (1987) offer conflicting evidence.

ARCH effects and regime changes seem necessary for these data. There is further a 'Mexican peso problem' [Krasker (1980)] in evaluating post-1982 data on excess holding yields, as evidenced by the lingering probabilities imputed to investors of a return to the 1979–82 regime (table 7).

More remains to be done. Future research should account more carefully for the role of inflation [as in Engel (1984)], measurement error, and approximation error associated with Shiller's linearization when applied to the recent huge swings in interest rates. But it seems fair to conclude on the basis of this research that the expectations hypothesis of the term structure continues to merit serious consideration by scholars.

Appendix

Here I show that my eq. (3.6) can alternatively be derived directly from Shiller's (1979) eq. (1) by a careful reinterpretation of variables. Let τ index semiannual data in contrast to t in the text which indexed quarterly data. Thus τ is associated with the quarterly observations t and $t+1$; $\tau+1$ with $t+2$ and $t+3$; and so on. Similarly, a bond with n periods to go when the indexation is by quarters has $n^* = n/2$ periods to go when the indexation is semiannual. Following Shiller's convention that superscripts with parentheses denote indexes, while superscripts without parentheses denote exponentiation, (3.4) would be written

$$P_\tau^{(n^*)} = \frac{C}{R_\tau^{(n^*)}} + \frac{1}{\left[1 + R_\tau^{(n^*)}\right]^{n^*}} \left[1 - \frac{C}{R_\tau^{(n^*)}}\right].$$

The six-month holding period yield associated with this bond is

$$H_\tau^{(n^*)} = \frac{P_{\tau+1}^{(n^*-1)} - P_\tau^{(n^*)} + C}{P_\tau^{(n^*)}},$$

which is approximated by Shiller's eq. (7):

$$H_\tau^{(n^*)} \cong \frac{R_\tau^{(n^*)} - \gamma_{n^*} R_{\tau+1}^{(n^*-1)}}{1 - \gamma_{n^*}}, \tag{A.1}$$

where

$$\begin{aligned} \gamma_{n^*} &= \frac{\gamma(1-\gamma^{n^*-1})}{1-\gamma^{n^*}}, \\ \gamma &= \frac{1}{1+\bar{R}} = \frac{1}{1+\bar{r}^2+2\bar{r}}. \end{aligned} \tag{A.2}$$

Compare this with the six-month holding yield from rolling over two three-month Treasury bills:

$$H_\tau^{(1/2)} = (1+r_t)(1+r_{t+1}) - 1$$

$$\cong -\bar{r}^2 + (1+\bar{r})r_{t+1} + (1+\bar{r})r_t. \tag{A.3}$$

The model posits that the expected six-month holding yields differ by a constant risk factor $\phi^{(n^*)}$:

$$\mathrm{E}_\tau H_\tau^{(n^*)} = \mathrm{E}_\tau H_\tau^{(1/2)} + \phi^{(n^*)}, \tag{A.4}$$

or

$$\frac{R_\tau^{(n^*)} - \gamma_{n^*}\mathrm{E}_\tau R_{\tau+1}^{(n^*-1)}}{1-\gamma_{n^*}} = -\bar{r}^2 + (1+\bar{r})r_t + \mathrm{E}_t(1+\bar{r})r_{t+1} + \phi^{(n^*)}. \tag{A.5}$$

Define

$$r_\tau^* \equiv -\bar{r}^2 + (1+\bar{r})r_t + \mathrm{E}_t(1+\bar{r})r_{t+1}, \tag{A.6}$$

and write (A.5) as

$$R_\tau^{(n^*)} = \gamma_{n^*}\mathrm{E}_\tau R_{\tau+1}^{(n^*-1)} + [1-\gamma_{n^*}][r_\tau^* + \phi^{(n^*)}], \tag{A.7}$$

which is the same equation Shiller arrived at on p. 1197, from which his eq. (1) is derived. By the law of iterated projections, his eq. (1) gives the solution to my expression (A.7):

$$R_\tau^{(n^*)} = \frac{(1-\gamma)}{(1-\gamma^{n^*})}\sum_{k=0}^{n^*-1}\gamma^k \mathrm{E}_\tau[r_{\tau+k}^*] + \phi^{(n^*)}. \tag{A.8}$$

Using (A.2), (A.6), and $n^* = n/2$, (A.8) becomes (3.6) as claimed.

References

Antoncic, Madelyn, 1986, High and volatile real interest rates: Where does the Fed fit in?, Journal of Money, Credit and Banking 18, 18–27.

Box, G.E.P. and Gwilym M. Jenkins, 1976, Time series analysis: Forecasting and control, Rev. ed. (Holden-Day, San Francisco, CA).

Campbell, John Y., 1986, A defense of traditional hypotheses about the term structure of interest rates, Journal of Finance 41, 183–193.

Campbell, John Y. and Robert J. Shiller, 1984, A simple account of the behavior of long-term interest rates, American Economic Review, Papers and Proceedings 74, 44–48.
Chiang, Chin Long, 1980, An introduction to stochastic processes and their applications (Krieger, New York).
Cosslett, Stephen R. and Lung-Fei Lee, 1985, Serial correlation in discrete variable models, Journal of Econometrics 27, 79–97.
Cox, John C., Jonathan E. Ingersoll and Stephen A. Ross, 1981, A re-examination of traditional hypotheses about the term structure of interest rates, Journal of Finance 36, 769–799.
Engel, Charles M., 1984, Testing for the absence of expected real profits from forward market speculation, Journal of International Economics 17, 299–308.
Engle, Robert F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987–1007.
Engle, Robert F., David M. Lilien and Russell P. Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, Econometrica 55, 391–407.
Fama, Eugene F., 1984, The information in the term structure, Journal of Financial Economics 13, 509–528.
Fama, Eugene F. and Robert R. Bliss, 1987, The information in long-maturity forward rates, American Economic Review 77, 680–692.
Flavin, Marjorie A., 1983, Excess volatility in the financial markets: A reassessment of the empirical evidence, Journal of Political Economy 91, 929–956.
Flavin, Marjorie A., 1984, Time series evidence on the expectations hypothesis of the term structure, in: Karl Brunner and Allan H. Meltzer, eds., Monetary and fiscal policies and their application, Carnegie–Rochester conference series on public policy, Vol. 20 (North-Holland, Amsterdam).
Goldfeld, Stephen M. and Richard E. Quandt, 1973, A Markov model for switching regressions, Journal of Econometrics 1, 3–16.
Goodfriend, Marvin and William Whelpley, 1986, Federal funds: Instrument of Federal Reserve policy, Federal Reserve Bank of Richmond Economic Review, Sept.; reprinted in 1987: Marvin Goodfriend, ed., Monetary policy in practice (Federal Reserve Bank of Richmond, VA).
Hamilton, James D., 1985, Uncovering financial market expectations of inflation, Journal of Political Economy 93, 1224–1241.
Hamilton, James D., 1987, A new approach to the economic analysis of nonstationary time series and the business cycle, Mimeo. (University of Virginia, Charlottesville, VA).
Hansen, Lars P. and Thomas J. Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, Journal of Economic Dynamics and Control 2, 7–46.
Hardouvelis, Gikas A. and Scott W. Barnhart, 1987, The evolution of Federal Reserve credibility: 1978–84, Mimeo. (Columbia University, New York).
Huizinga, John and Frederic S. Mishkin, 1986, Monetary policy regime shifts and the unusual behavior of real interest rates, in: Karl Brunner and Allan H. Meltzer, eds., The National Bureau method, international capital mobility and other essays, Carnegie–Rochester conference series on public policy, Vol. 24 (North-Holland, Amsterdam).
Krasker, William S., 1980, The 'peso problem' in testing the efficiency of forward exchange markets, Journal of Monetary Economics 6, 269–276.
Liptser, R.S. and A.N. Shiryayev, 1977, Statistics of random processes, Vol. I, General theory (Springer-Verlag, New York).
Mankiw, N. Gregory, Jeffrey A. Miron and David N. Weil, 1987, The adjustment of expectations to a change in regime: A study of the founding of the Federal Reserve, American Economic Review 77, 358–374.
Mankiw, N. Gregory and Lawrence H. Summers, 1984, Do long-term interest rates overreact to short-term interest rates?, Brookings Papers on Economic Activity 1, 223–242.
Marsh, Terry A. and Eric R. Rosenfeld, 1983, Stochastic processes for interest rates and equilibrium bond prices, Journal of Finance 38, 635–646.
Neftci, Salih N., 1982, Optimal prediction of cyclical downturns, Journal of Economic Dynamics and Control 4, 225–241.
Neftci, Salih N., 1984, Are economic time series asymmetric over the business cycle?, Journal of Political Economy 92, 307–328.

Peek, Joe and James A. Wilcox, 1987, Monetary policy regimes and the reduced form for interest rates, Journal of Money, Credit and Banking 19, 273–291.

Roll, Richard, 1971, Investment diversification and bond maturity, Journal of Finance 26, 51–66.

Sargent, Thomas J., 1979, A note on maximum likelihood estimation of the rational expectations model of the term structure, Journal of Monetary Economics 5, 133–143.

Shiller, Robert J., 1979, The volatility of long-term interest rates and expectations models of the term structure, Journal of Political Economy 87, 1190–1219.

Shiller, Robert J., 1981, Alternative tests of rational expectations models: The case of the term structure, Journal of Econometrics 16, 71–87.

Shiller, Robert J., John Y. Campbell and Kermit L. Schoenholtz, 1983, Forward rates and future policy: Interpreting the term structure of interest rates, Brookings Papers on Economic Activity 1, 173–217.

Spindt, Paul A. and Vefa Tarhan, 1987, The Federal Reserve's new operating procedures: A post mortem, Journal of Monetary Economics 19, 107–123.

Walsh, Carl E., 1987, Testing for real effects of monetary policy regime shifts, NBER working paper no. 2116.